LEIGH LISKER
and *ARTHUR S. ABRAMSON*

A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements*

I. Introduction

Theoretical background. In diagnosing the phonetic basis for our ability to distinguish between phonemic categories, linguists often invoke the dimension of voicing to call some categories "voiced" and others "voiceless". In the case of stop consonants it is usual to label as voiced, categories characterized by the presence of glottal buzz during the interval of articulatory closure, while absence of buzz during this interval is a mark of voiceless stops. Acoustically the two kinds of stops are in most cases easily distinguished by reference to their spectrographic patterns; for voiced stops the formantless segment corresponding to the closure interval is traversed by a small number of low-frequency harmonic components, while in the case of voiceless stops the closure interval is essentially blank. But while this difference is an adequate basis for the physical separation of stop categories in many languages, there are some, like English, for which it works only in part. Although in medial position English /b d g/ are voiced and /p t k/ voiceless, in initial position both sets are commonly produced with silent closure intervals and should therefore be classed as voiceless according to the definitions cited. While phoneticians rarely call initial /b d g/ out-and-out voiceless stops, they regularly cite at least one other phonetic attribute, that of aspiration, which reliably distinguishes /p t k/ from /b d g/ both in initial position and medially before a stressed syllabic.

*A preliminary report, "Voicing Lag and English Initial Stops," was read before the Sixth Annual Conference on Linguistics of the Linguistic Circle of New York, held on May 6, 1961. Earlier versions of the present cross-language study were presented at the Sixty-Sixth Meeting of the Acoustical Society of America, Ann Arbor, Michigan, November 6–9, 1963 and the Thirty-Eighth Annual Meeting of the Linguistic Society of America, Chicago, Illinois, December 28–30, 1963. This work was supported in part by the National Science Foundation under Grant G-23633.

Reprinted from WORD
Volume 20, No. 3, December 1964

In many non-final positions,[1] then, /b d g/ are voiced and /p t k/ voiceless, while in some, /p t k/ are released with an audible explosion and an interlude of noise and /b d g/ are not. Thus, differences of voicing and aspiration, either singly or in conjunction, are said to separate the two sets of English stop phonemes, although neither is alone sufficient to distinguish them over the entire range of contexts in which both are found.

In addition to voicing and aspiration a third phonetic dimension, one of articulatory force, is widely cited as still another basis for separating the stop categories of English and many other languages. Although the assessment of articulatory force appears ultimately to be a matter of proprioceptive judgment, this judgment is said to depend directly on the audible features of closure duration and the loudness of the stop explosion.[2] Thus it is said that English /p t k/ are in general more forcefully articulated than /b d g/. In fact it is often asserted that this fortis/lenis difference is the primary mark of the /p t k/:/b d g/ set of contrasts, on the ground that it alone is operative in every position in which these contrasts are observed. In the current terminology, the fortis/lenis difference is the distinctive feature separating the two categories, while any concomitant differences of voicing and aspiration are systemically redundant, quite aside from whatever importance they might have as cues for perception.

In seeking to determine experimentally the acoustic cues by which listeners distinguish between English /b d g/ and /p t k/, we have of course been interested in discovering some single best measure by which to separate the two phoneme categories. In trying to use accepted phonetic descriptions of English as a guide to our research, we have been hampered by certain ambiguities in the treatment of articulatory force. Many phonetic statements imply that the three dimensions of voicing, aspiration and force of articulation are taken to be mutually independent coordinate dimensions of description; and yet there are reasons for wondering whether this is so. No one of the physical measures, whether physiological or acoustic, that have been proposed as correlates of the fortis/lenis

[1] As for final position, our impression of the literature on English phonemics is that three *observable* phonetic features, voicing, release and vocalic length, are attested for the contrast between /b d g/ and /p t k/. The occlusion of /b d g/, it is said, may be partially voiced or not voiced at all. If the final stops are audibly released, as they sometimes are, /p t k/ will have a stronger, more aspirated release. The single differentiating feature that all descriptions agree is regularly present is the greater length of vowels followed by /b d g/, but this is usually treated as a matter of vowel allophonics. (See, for example, the references cited in footnote 22.)

[2] See K. L. Pike, *Phonetics* (Ann Arbor, 1951), pp. 128–129; R-M. S. Heffner, *General Phonetics* (Madison, 1949), p. 120; R. Jakobson, C. G. M. Fant and M. Halle, *Preliminaries to Speech Analysis* (M.I.T. Technical Report No. 13, 1952), p. 36.

dimension, has been shown *not* to be significantly connected with voicing or aspiration. And in fact an examination of the phonetic literature generally fails to turn up any language which is said to possess stop categories that differ only in force of articulation.[3] For languages in which the fortis/lenis difference is invoked, it is too often the case to be accidental that voiceless and aspirated stops are discovered to be fortis, while voiced and unaspirated ones are at the same time lenis. In languages whose stop categories are said to differ on all three dimensions, the total number of such categories seems never to exceed four. The ambiguous status of the terms "fortis" and "lenis" (or "tense" and "lax") is also reflected in statements by several writers[4] to the effect that a number of phonetic features, *among them voicing and aspiration*, may be taken as manifestations of an underlying division of stops on the basis of a fortis/lenis opposition. So far as we are aware, only one recent study, Gunnar Fant's *Acoustic Theory of Speech Production*,[5] suggests that the ensemble of acoustic features that are used as evidence for a dimension of articulatory force may be plausibly grouped together without any need for positing an independent fortis/lenis difference; in fact Fant associates all these features instead with differences in the position and activity of the glottis during the various phases of stop production, and our own work convinces us that Fant's views are entirely correct.

The acoustic features that we may suppose to be useful physical correlates of the manner contrasts between stop categories are in part readily visible in spectrograms,[6] although in general there is all too little solid evidence for asserting that any given feature is to be connected exclusively with some one of the phonetic dimensions of voicing, aspiration and articulatory force. At the same time it is reasonable to claim as a salient acoustic correlate of voicing the periodic character of a speech signal, which shows up in narrow-band spectrograms as a set of one or more harmonic traces and in wide-band spectrograms as a series of regularly spaced vertical striations. Aspiration too is spectrographically unambiguous; it registers as noise (i.e., random stippling), mostly at the frequencies of the second and third formants of contiguous pattern segments. In the case of the fortis/lenis relation, as has been said, the dimension of articulatory force seems

[3] In the occasional instance that appears in the literature, the description never clearly excludes the involvement of aspiration, presumably because any distinguishing aspiration present is taken to be a sign of fortisness.

[4] See, e.g., R. Jakobson and M. Halle, "Tenseness and Laxness," in *Roman Jakobson: Selected Writings*, I. *Phonological Studies* (The Hague, 1962), pp. 554–555.

[5] The Hague, 1960, pp. 224–225.

[6] These are listed, for example, in L. Lisker, "Closure Duration and the Intervocalic Voiced-Voiceless Distinction in English," *Language* XXXIII (1957), pp. 43, 45.

to us to be of doubtful status; certainly none of the acoustic features which have been suggested as correlates of a fortis/lenis dimension[7] is demonstrably independent of voicing.

The two features correlated with voicing and aspiration—periodic pulsing at the frequency of the voice pitch and noise in the frequency range of the higher formants—have an interesting relation to one another, at least in the case of the stops in English; each feature tends to be prominent in spectrograms only where the other is absent. Thus if a portion of a spectrographic pattern indicates the presence of voicing, then the noise feature is absent or much obscured, while if noise is strongly marked then periodic pulsing is usually not discernible. Now if we locate a pattern segment with reference to the instant of release of the stop closure,—and this event is marked by an abrupt increase in the amplitude and frequency spread of the signal—then we may define the amount or degree of voicing of a stop as the duration of the time interval by which the onset of periodic pulsing either precedes or follows release. In thus giving up the absolute definition of the term "voiced stop" with which we began, we are free to say that a difference of voicing not only separates voiced from voiceless stops,[8] but that it equally well distinguishes aspirated from unaspirated stops, where the latter are both commonly called voiceless. The noise feature of aspiration, instead of being considered coordinate with voicing, is then regarded simply as the automatic concomitant of a large delay in voice onset. In English, at least, this seems reasonable: /b d g/ and /p t k/ probably differ everywhere in the time of voice onset relative to release, but in certain positions the presence of aspiration noise tells us something about the absolute magnitude of delay in the onset time following /p t k/ releases.

On the basis of the considerations just presented it seems reasonable to begin a study aimed at finding the acoustic features which serve as cues for the manner differentiation of stops by fixing attention on the timing relation between voice onset and the release of occlusion. This measure is both easy to make and at the same time most promising as providing the single best basis for the physical discrimination of stop manner categories. Moreover, while this timing relation is to be connected immediately to the phonetic dimension of voicing, the underlying glottal mechanism which controls it is also responsible, presumably, for generating some of the features that have been taken to be acoustic manifestations of aspiration

[7] See especially Jakobson, Fant and Halle, 1952, pp. 36, 38. Additional references and discussion are to be found in L. Lisker, "On Hultzén's 'Voiceless Lenis Stops in Prevocalic Clusters'," *Word* XIX (1963), 376–387.

[8] As defined absolutely.

and articulatory force.[9] Finally, in choosing this feature, we took into account the fact that any findings based on spectrographic analysis would ultimately have to be corroborated by experiments involving the use of speech synthesis. The synthesizers available at the Haskins Laboratories permit us to manipulate most precisely and conveniently the relative timing of the acoustic features marking release and the onset of voicing.

It may be objected that much of the foregoing discussion applies only to English, where the features of voicing and aspiration are distributed within the stop categories according to a particular pattern. The present study was undertaken to gather the data needed in order to determine, with some degree of precision, just how important our measure of the relative timing of voice onset is for the physical specification of stop categories in languages generally.

Purpose of the study. The purpose of the present study is to see how well a single dimension, voice onset time, serves to separate the stop categories of a number of languages in which both the number and phonetic characteristics of such categories are said to differ. Attention will be limited to word-initial position before vowels. For each of the languages chosen the phonetic features for which differentiating functions are usually claimed were such that we could expect to find initial stop categories marked by differences in voice onset time. For two of the languages, however, Hindi and Marathi, we did not expect this feature to mark off the so-called voiced aspirates. The eleven languages that were examined fall into three groups according to the number of stop categories: (1) two-category languages: American English, Cantonese, Dutch, Hungarian, Puerto Rican Spanish, and Tamil; (2) three-category languages: Korean, Eastern Armenian, and Thai; (3) four-category languages: Hindi and Marathi.

Procedure. The general procedure involved the spectrographic analysis of high-quality tape recordings made in an acoustically treated room. Each of our informants, seventeen in all,[10] produced a set of words chosen to include a sampling of all the initial prevocalic stops[11] found in his language.

[9] See G. Fant, *Acoustic Theory of Speech Production* (The Hague, 1960), p. 279.

[10] The number of informants used for each language depended upon availability. They were all educated speakers of standard varieties of their languages. For Tamil some further specification of dialect is given.

[11] The linguist may wish to include affricates among the "stops" or "plosives" in his phonemic analysis of a language. We have excluded affricates from the present study, e.g. in Hungarian and Cantonese, limiting ourselves to plain prevocalic stops, namely occlusives that have an abrupt rather than a fricative release.

For each word the informant was then told to make up two sentences to show its use in initial and non-initial positions. He was urged to utter these sentences with the fluency and naturalness of normal conversation. The informant recorded each word and each sentence twice. Departures from this elicitation procedure are noted for English and Thai sentences in Part III.

Wide-band spectrograms[12] of the recordings were made, and from these, voice onset times were measured by marking off the interval between the release of the stop and the onset of glottal vibration, that is, voicing. The point of voicing onset was determined by locating the first of the regularly spaced vertical striations which indicate glottal pulsing, while the instant of release was found by fixing the point where the pattern shows an abrupt change in overall spectrum. Oral closure is marked spectrographically by the total or almost total absence of acoustic energy in the formant frequency range; oral release is marked by the abrupt onset of energy in the formant frequency range.[13] In Figure 1, which shows wide-band spectrograms with displays of relative amplitude, we have three stop+vowel syllables illustrating three common conditions of voice onset time. In the first, voicing begins before the release of the stop; in the second, just after the release; in the third, voice onset lags considerably behind the release. According to their usual phonetic descriptions, the first stop is voiced and unaspirated, the second is voiceless and unaspirated, and the third is voiceless and aspirated. We have adopted the convention of assigning zero-time to our reference point, the instant of release; thus, measurements of voice onset time before the release are stated as negative numbers and called voicing lead, while measurements of voice onset time after the release are stated as positive numbers and called voicing lag. Each measurement was rounded to the nearest five milliseconds as a reasonable estimate of attainable precision.

In some few cases where word-initial stops were medial in sentences, it was not possible to equate the presence of vertical striations with audible glottal vibration. In such cases the striations[14] were very faint and confined

[12] The sound spectrograph used was the Sona-Graph of the Kay Electric Company, Pine Brook, New Jersey. The wide-band setting rather than the narrow one was used for its better time resolution; this effect was enhanced by substituting a large drum for the standard one to provide an expanded time scale on the paper of 7.5 in./sec. as against 5 in./sec.

[13] This acoustic energy may be concentrated just at the formant frequencies, or it may take the form of a "burst" or very brief interval of noise having a somewhat broader frequency spread.

[14] These inaudible pulses are undoubtedly the result of glottal activity, for they occur at the same frequencies as the audible variety. For further discussion of this phenomenon, see **Inferences as to glottal mechanisms** in IV.

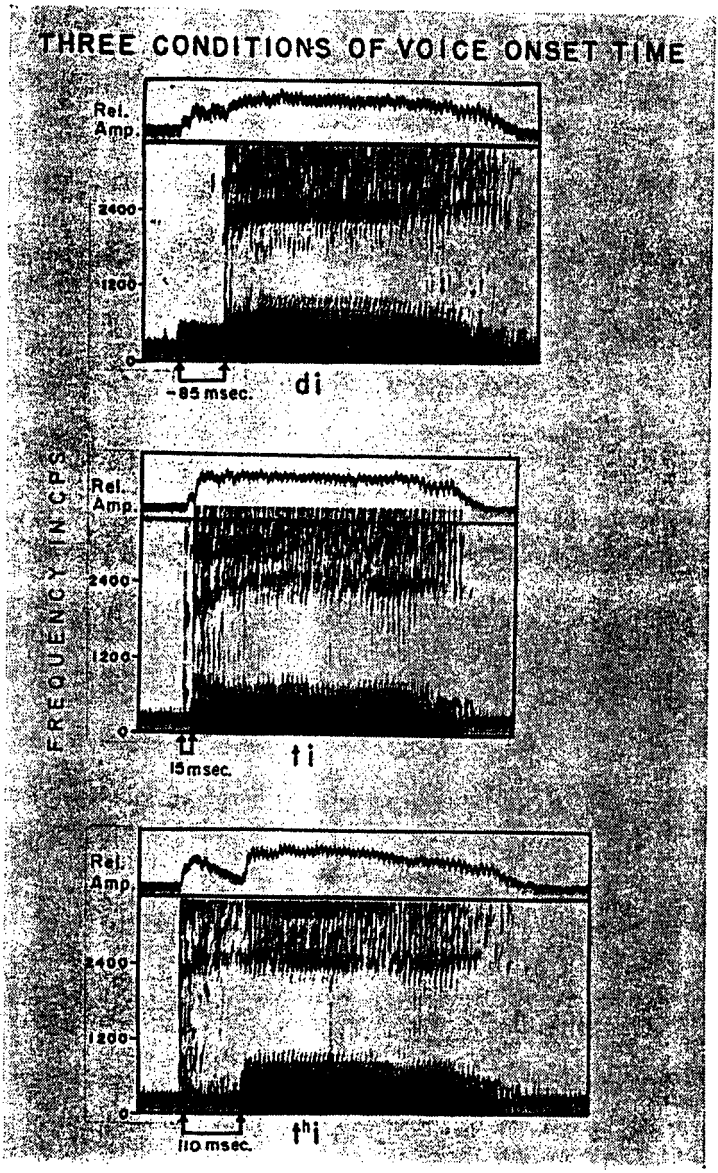

FIGURE 1. Wide-band spectrograms showing three conditions of voice onset time: voicing lead, short voicing lag, and long voicing lag. (Examples from Thai.)

to the bottom of the pattern. Moreover, careful listening tests showed that no pulses could be detected by ear. In carrying out our measurements, we ignored these striations.

II. Stops in Isolated Words

In this section we present our findings for the initial stops of isolated words in each of the eleven languages studied. These languages fall into three groups, depending on whether they have a maximum of two, three, or four categories of stops at each place of articulation. Each language is introduced by a chart of its stop phonemes with a statement of the phonetic features commonly recognized as distinctive for the stops. This is followed by a table of averages and ranges of voice onset time for all the stop phonemes of the language. The frequency distributions underlying these numbers are then shown graphically in Figures 2–4 for the two-category languages and Figures 5–7 for the three- and four-category languages. Thus, for example, the bilabial stops of the first language, Dutch, are to be found in Figure 2, the dental stops in Figure 3, and the velar stops in Figure 4.

Two-category languages.

1. Dutch

LABIAL	b	p
DENTAL	d	t
VELAR		k

Dutch is said to have a contrast of voiced and voiceless unaspirated stops,[15] except in the velar position where only /k/ is found. Moulton, however, simply calls them lenis and fortis.[16] In Table 1 we show data based on measurements of initial stops uttered by a single native speaker of standard Dutch. The first line gives average values of voice onset time for each category; the second line shows the ranges of values observed; and the third line records the number of tokens of each stop phoneme measured.

[15] See A. Cohen, C. L. Ebeling, K. Fokkema and A. G. F. van Holk, *Fonologie van het Nederlands en het Fries* (s'-Gravenhage, 1961), pp. 33–34, in which the two stop series are called voiced and voiceless. At the same time, however, the authors note that "in most parts of the world voicing goes together with another articulatory feature, one that is referred to by the term *lenis* or *soft*. The voiceless sounds are characterized as *fortis* or *sharp*." (Our translation.)

[16] William G. Moulton, "The Vowels of Dutch: Phonetic and Distributional Classes," *Lingua* XI (1962), p. 308.

TABLE 1. Voice Onset Time in Msec: Dutch
(1 speaker)

	/b/	/p/	/d/	/t/	/k/
Av.	−85	10	−80	15	25
R.	−145: −50	0:30	−115: −45	5:35	10:35
N.	22	46	32	56	60

2. *Puerto Rican Spanish*

LABIAL	b	p
DENTAL	d	t
VELAR	g	k

Puerto Rican Spanish is said to have a contrast of voiced and voiceless unaspirated stops.[17] The allophonic complexity of these stops will be discussed in Table 13, where it is relevant.

TABLE 2. Voice Onset Time in Msec: Spanish
(2 speakers)

	/b/	/p/	/d/	/t/	/g/	/k/
Av.	−138	4	−110	9	−108	29
R.	−235: −60	0:15	−170: −75	0:15	−165: −45	15:55
N.	17	20	16	16	14	20

3. *Hungarian*

LABIAL	b	p
DENTAL	d	t
VELAR	g	k

The two sets of Hungarian stops are described as voiced and voiceless unaspirated respectively.[18] A palatal set /ɟ c/, which is often classed with the stops, will not be included here. (See footnote 11).

[17] In this respect it is like other dialects of Latin American Spanish as well as Iberian Spanish. See Tomás Navarro, *El Español en Puerto Rico* (Río Piedras, P. R., 1948), p. 58. But the two sets of Spanish stops are said to be distinguished by the tense/lax feature in Sol Saporta and Heles Contreras, *A Phonological Grammar of Spanish* (Seattle, 1962), p. 39.

[18] R. A. Hall, Jr., *An Analytical Grammar of the Hungarian Language* (Baltimore, 1938), pp. 16–17; J. Lotz, *Das ungarische Sprachsystem* (Stockholm, 1939), p. 27; József Tompa, ed., *A mai magyar nyelv rendszere. Leíró nyelvtan.* I. (Budapest, 1961), p. 81.

TABLE 3. Voice Onset Time in Msec: Hungarian
(2 speakers)

	/b/	/p/	/d/	/t/	/g/	/k/
Av.	−90	2	−87	16	−58	29
R.	−125: −65	0:10	−130: −65	10:25	−70: −35	20:35
N.	8	12	7	12	6	7

4. *Tamil*

LABIAL	b	p
DENTAL	d	t
VELAR	g	k

Our data for the Tamil initial stops are taken from the speech of an educated Brahman of South Arcot District in Madras State; they may be considered fairly representative of the speech of well educated Tamilians in the area that includes Madras City and the country to the east and south. Of the more recently published statements on the phonology of the language of this region some[19] describe the initial stops as voiced and voiceless, but at least one study[20] states categorically that "in stop consonants paired by similarity of articulatory position, tenseness is the significant feature in opposition to laxness, rather than voicelessness in opposition to voice." In addition "tense stops, unless affricated, are aspirated" (p. 361).

TABLE 4. Voice Onset Time in Msec: Tamil
(1 speaker)

	/b/	/p/	/d/	/t/	/g/	/k/
Av.	−74	12	−78	8	−62	24
R.	−100: −55	0:45	−105: −35	0:30	−110: −35	15:35
N.	8	42	16	8	13	13

5. *Cantonese*

LABIAL	p	p^h
DENTAL	t	t^h
VELAR	k	k^h

The two sets of stops are described[21] as being voiceless unaspirated and

19 L. Lisker, "The Tamil Occlusives: Short vs. Long or Voiced vs. Voiceless?" *Indian Linguistics*, Turner Jubilee Vol., I (1958), 294–301.

20 Murray Fowler, "The Segmental Phonemes of Sanskritized Tamil," *Language* XXX (1954), 360–367.

21 Yuen Ren Chao, *Cantonese Primer* (Cambridge, Mass., 1947), p. 20; Diana Kao, *The Structure of the Cantonese Syllable* (Ph.D. dissertation in preparation, Columbia University, 1964), chap. 2.

voiceless aspirated respectively. In addition to the three places of articulation given in our chart an alveolar series is often included among the stops; these are phonetically affricates, however, and are therefore excluded from consideration (see footnote 11). Sometimes a fourth series of labialized velars is posited, but these too will not be treated here.

TABLE 5. Voice Onset Time in Msec: Cantonese
(1 speaker)

	/p/	/pʰ/	/t/	/tʰ/	/k/	/kʰ/
Av.	9	77	14	75	34	87
R.	0:20	35:110	5:25	45:95	25:55	70:115
N.	15	15	12	15	15	15

6. English

LABIAL	b	p
ALVEOLAR	d	t
VELAR	g	k

Recent descriptions of English are in general agreement that initial /p t k/ are aspirated and /b d g/ are not, but there is considerable diversity as to the relative emphasis put on differences in voicing and differences in force of articulation. For certain writers[22] the two series differ primarily in voicing, but the voiced set may be additionally characterized as produced with weaker articulatory force than the voiceless set. Other writers, however, prefer to place primary stress on the difference in articulatory force, while the voicing difference is said to be secondary, on occasion even absent.[23]

TABLE 6. Voice Onset Time in Msec: English
(4 speakers)

	/b/	/p/	/d/	/t/	/g/	/k/
Av.	1/−101	58	5/−102	70	21/−88	80
R.	0:5/−130: −20	20:120	0:25/−155: −40	30:105	0:35/−150: −60	50:135
N.	51/17	102	63/13	116	53/13	84

[22] J. S. Kenyon, *American Pronunciation*, 10th ed. (Ann Arbor, 1951), pp. 41, 45, 121–131; C. F. Hockett, *A Manual of Phonology* (Baltimore, 1955), pp. 115–116; A. A. Hill, *Introduction to Linguistic Structures* (New York, 1958), p. 32; A. J. Bronstein, *The Pronunciation of American English: an Introduction to Phonetics* (New York, 1960), pp. 59–60; H. A. Gleason, *An Introduction to Descriptive Linguistics*, 2nd ed. (New York, 1961), pp. 22, 24, 247–248; M. W. Bloomfield and L. Newmark, *A Linguistic Introduction to the History of English* (New York, 1963), p. 67.

[23] G. L. Trager and H. L. Smith, Jr., *An Outline of English Structure* (Norman, Okla., 1951), p. 29; A. C. Gimson, *An Introduction to the Pronunciation of English* (London, 1962), pp. 146–147.

It should be noted that we give two sets of values for /b d g/. To have given a single set of values would have meant lumping positive and negative values of voice onset time as items of a single population, and it appeared rather that these are distributed within two discontinuous ranges. In such a situation it would be misleading to determine single average values of onset time for the members of the /b d g/ set. Moreover, it is relevant that instances of positive and negative values do not occur randomly in our material. In the following table we see how the two kinds of /b d g/ were distributed for each of our four speakers of American English.

TABLE 6a. Number of Occurrences of Lag and Lead

Speaker	/b/	/d/	/g/
AA	+ 6	+30	+18
	− 0	− 0	− 0
LL	+22	+16	+18
	− 0	− 0	− 0
CC	+22	+17	+17
	− 0	− 1	− 1
TR	+ 1	+ 0	+ 0
	−17	−12	−12

A single speaker TR was responsible for 95% of all the stops produced with voicing lead, while one other speaker, CC, produced the remainder of such stops. Conversely, speaker TR produced 41 of his 42 /b d g/ tokens with voicing lead; CC, for his part, had positive values (lag) in 56 of the 58 /b d g/ stops he produced. Thus it appears that our speakers do not randomly produce such stops with positive and negative values of relative onset time; rather, each speaker, in isolated words at least, always or nearly always produces a single kind of /b d g/.

Three-category languages.

1. Eastern Armenian

LABIAL	b	p	p^h
DENTAL	d	t	t^h
VELAR	g	k	k^h

The three series are called mediae, tenues and aspiratae by the author of one recently published textbook;[24] these terms are presumably equivalent to voiced, voiceless unaspirated and voiceless aspirated. Another source[25] characterizes these stops as voiced, voiceless glottalized

[24] P. L. Movsessian, *Armenische Grammatik* (Vienna, 1959), p. 17.

[25] G. H. Fairbanks and E. W. Stevick, *Spoken East Armenian* (New York, 1958), pp. xv–xviii.

unaspirated and voiceless aspirated. The very much earlier description by Rousselot[26] referred to them as lenis, fortis and aspirated.

TABLE 7. Voice Onset Time in Msec: Eastern Armenian
(1 speaker)

	/b/	/p/	/pʰ/	/d/	/t/	/tʰ/	/g/	/k/	/kʰ/
Av.	−96	3	78	−102	15	59	−115	30	98
R.	−195:−35	0:15	60:105	−195: −90	10:20	35:80	−190: −150	15:50	60:135
N.	23	12	10	13	14	8	14	14	16

2. *Thai*

LABIAL	b	p	pʰ
DENTAL	d	t	tʰ
VELAR		k	kʰ

The stops of Thai are most often described as voiced, voiceless unaspirated and voiceless aspirated.[27] Hockett adds the comment that the voiceless unaspirated stops "do not tend to be non-distinctively glottalized."[28]

TABLE 8. Voice Onset Time in Msec: Thai
(3 speakers)

	/b/	/p/	/pʰ/	/d/	/t/	/tʰ/	/k/	/kʰ/
Av.	−97	6	64	−78	9	65	25	100
R.	−165: −40	0:20	25:100	−165: −40	0:25	25:125	0:40	50:155
N.	31	32	33	33	33	33	32	38

3. *Korean*

LABIAL	p	pᶜ	pʰ
DENTAL	t	tᶜ	tʰ
VELAR	k	kᶜ	kʰ

Our choice of symbols for representing the stops of Korean is not based on the practices of linguists specializing in the language; it represents rather a simple broad transcription based on our own phonetic judgments of the productions of several speakers of the Seoul dialect. Published phonetic descriptions of the stops show considerable variation.[29] The first

[26] P.-J. Rousselot, *Principes de phonétique expérimentale*, Vol. I (Paris, 1924; 1st pub., 1901–1908), pp. 596–599.

[27] M. R. Haas, G. V. Grekoff, R. C. Mendiones, W. Buddhari, J. R. Cooke and S. C. Egerod, *Thai-English Student's Dictionary* (Stanford, 1964), p. xi. This analysis has been presented by M. R. Haas in numerous earlier publications. See also A. S. Abramson, *The Vowels and Tones of Standard Thai: Acoustical Measurements and Experiments* (Bloomington, 1962), p. 4.

[28] Hockett, *Manual*, p. 115.

[29] S. E. Martin, "Korean Phonemics," *Language* XXVII (1951), 519–533; F. Lukoff, *A Grammar of Korean* (Ph.D. dissertation, University of Pennsylvania, 1951), pp. 5–8; Mieko S. Han, *Acoustic Phonetics of Korean*, Technical Report No. 1 (University of California, 1963), pp. 20–21.

series, /p t k/, have been called voiceless, tense, long, and glottalized, though not all these terms have been used by everyone. The series /pᶜ tᶜ kᶜ/ are said to be, in initial position, voiceless, lax, and slightly aspirated. /pʰ tʰ kʰ/ are described as voiceless and heavily aspirated, but lax. Although for our present purpose it is efficient to regard the stops of each series as unit phonemes, two well known analyses establish only the second type as unit phonemes, while those of the first and third types are represented as sequences of two phonemes each; the first type is taken either as a case of gemination or as a sequence of the simple stop phoneme and a phoneme of "glottal tension," while the third type is a sequence of the simple stop followed by a phoneme /h/. For our purpose the choice of one of these analyses over the others is a matter of no consequence.

TABLE 9. Voice Onset Time in Msec: Korean
(1 speaker)

	/p/	/pᶜ/	/pʰ/	/t/	/tᶜ/	/tʰ/	/k/	/kᶜ/	/kʰ/
Av.	7	18	91	11	25	94	19	47	126
R.	0:15	10:35	65:115	0:25	15:40	75:105	0:35	30:65	85:200
N.	15	30	21	16	24	12	16	34	12

Four-category languages.

1. Hindi

LABIAL	b	bʰ	p	pʰ
DENTAL	d	dʰ	t	tʰ
DOMAL	ḍ	ḍʰ	ṭ	ṭʰ
VELAR	g	gʰ	k	kʰ

The Hindi stops may be arranged, with respect to the place and manner dimensions of phonetic description, in a 4×4 array. The manner dimension itself is, in turn, analyzed into the two independent components of voicing and aspiration, so that the sixteen categories may also be located in a 2×2 manner array:

	voiced	voiceless
unaspirated	b d ḍ g	p t ṭ k
aspirated	bʰ dʰ ḍʰ gʰ	pʰ tʰ ṭʰ kʰ

It is of course possible to argue that the aspirated categories are better analyzed into phoneme sequences consisting of either voiced or voiceless stop and a phoneme /h/.[30] For our purpose, however, it makes no difference whether the aspiration is taken as a phonetic component of half the stop categories or as the manifestation of a phoneme /h/.

[30] See John J. Gumperz, "Phonological Differences in Three Hindi Dialects," *Language* XXXIV (1958), 212–224; also C. F. Hockett, *Manual*, esp. p. 107.

TABLE 10. Voice Onset Time in Msec: Hindi
(1 speaker)

	/b/	/bʰ/[31]	/p/	/pʰ/	/d/	/dʰ/	/t/	/tʰ
Av.	−85	−61	13	70	−87	−87	15	67
R.	−120: −40	−105:0	0:25	60:80	−140: −60	−150: −60	5:25	35:100
N.	16	15	18	18	18	18	16	16

	/ɖ/	/ɖʰ/[31]	/ʈ/	/ʈʰ/	/g/	/gʰ/	/k/	/kʰ/
Av.	−76	−77	9	60	−63	−75	18	92
R.	−115: −30	−110:0	0:20	45:80	−95: −30	−160: −40	10:35	75:100
N.	18	15	18	18	17	16	16	18

2. *Marathi*

LABIAL	b	bʰ	p	pʰ
DENTAL	d	dʰ	t	tʰ
DOMAL	ɖ	ɖʰ	ʈ	ʈʰ
VELAR	g	gʰ	k	kʰ

In general the Marathi stops are phonetically similar to the Hindi; distributionally they differ in that Marathi /ɖ/ and /ɖʰ/ do not occur in word-initial position (and thus are not represented in the body of data given below). The four categories differ phonetically with respect to the two intersecting dimensions of voicing and aspiration, although at least one recent statement[32] characterizes the voiceless stops as also fortis, the voiced as lenis, and the aspirated voiced stops as "rather fortis."

TABLE 11. Voice Onset Time in Msec: Marathi
(1 speaker)

	/b/	/bʰ/	/p/	/pʰ/	/d/	/dʰ/	/t/	/tʰ/
Av.	−117	−95	11	76	−111	−87	10	65
R.	−160: −80	−160: −65	0:25	40:110	−175: −65	−110: −40	0:20	40:85
N.	14	16	11	14	20	19	17	14

	/ʈ/	/ʈʰ/	/g/	/gʰ/	/k/	/kʰ/
Av.	0	63	−116	−89	24	87
R.	0:10	35:75	−160: −75	−120: −45	10:40	60:105
N.	10	14	18	18	14	14

Overall relations. From a comparison of the mean values and ranges given in the tables just presented it is quite clear that, on the whole, differences in

[31] In each of the categories /bʰ/ and /ɖʰ/ there was a single aberrant production with voicing lag: +20 for one word said to begin with /bʰ/ and +25 for one of the words recorded for /ɖʰ/. Both were excluded from our table, since careful listening led us to believe that they might well be regarded by the speaker as faulty productions, although it seemed very unlikely that they would be identified as any stops other than /bʰ/ and /ɖʰ/.

[32] A. R. Kelkar, *The Phonology and Morphology of Marathi* (Ph.D. dissertation, Cornell University, 1958), pp. 4, 16.

voice onset time may serve as a basis for separating the various manner categories in each of the languages examined. At the same time, however, there is some indication that the measure of voice onset time is also, to a certain extent, sensitive to the *place* of stop closure, for the velars seem to have consistently higher values than the other stops. Because this may well have the effect of producing apparently overlapping distributions if stops of the same manner but different places of articulation are taken together, we shall keep separate the data for the three general positions of stop closure. An example of this effect is the relation between English /b d g/ and /p t k/; the two classes of stops overlap in the +20 to +30 range, but this is in fact almost entirely a matter of overlap between /g/ and /p/, for our data show no range of values common to /b/ and /p/ or to /d/ and /t/ or to /g/ and /k/. With our present data we can not rule out the possibility that some of the remaining overlap is a function of the uncontrolled variable of vocalic environment.

In order to see in detail just how the measure of voice onset time serves as a device for separating stop categories it is perhaps easier to display our data in the form of graphs of frequency distribution. These graphs enable us to see at a glance whether the values determined for items belonging to any particular category tend to cluster near some favored or "modal" value. In Figs. 2–7 each category is represented by a family of vertical lines, where the height of any one line indicates the percentage of items belonging to the category whose values of voice onset time fall within a ten millisecond interval at the value shown by the location of the line along the horizontal axis of the graph. It is assumed that as the number of measured items is increased the family of lines representing them will tend more and more to conform to one or another type of "smooth" distribution, that is, that the line connecting the ends of those lines will increasingly approximate a smooth curve. A number of categories represented in our figures show frequency distributions that are far from smooth, but in general our data provide a reasonably good indication of the range of values that a larger sample of stops in the various languages might occupy.

If we begin by looking at the frequency distributions for the two-category languages (Figs. 2, 3, 4) we see, first of all, that four of the languages (Dutch, Spanish, Hungarian, Tamil) are grossly similar in having one set of stops with negative values and the other with zero or small positive values of voice onset time. Cantonese is similar to these in having a category whose values lie in the region at and just above zero, but its second category shows considerably higher positive values. English resembles Cantonese, except for the lone speaker (of the four who served as informants) who produced /b d g/'s with negative values of voice onset

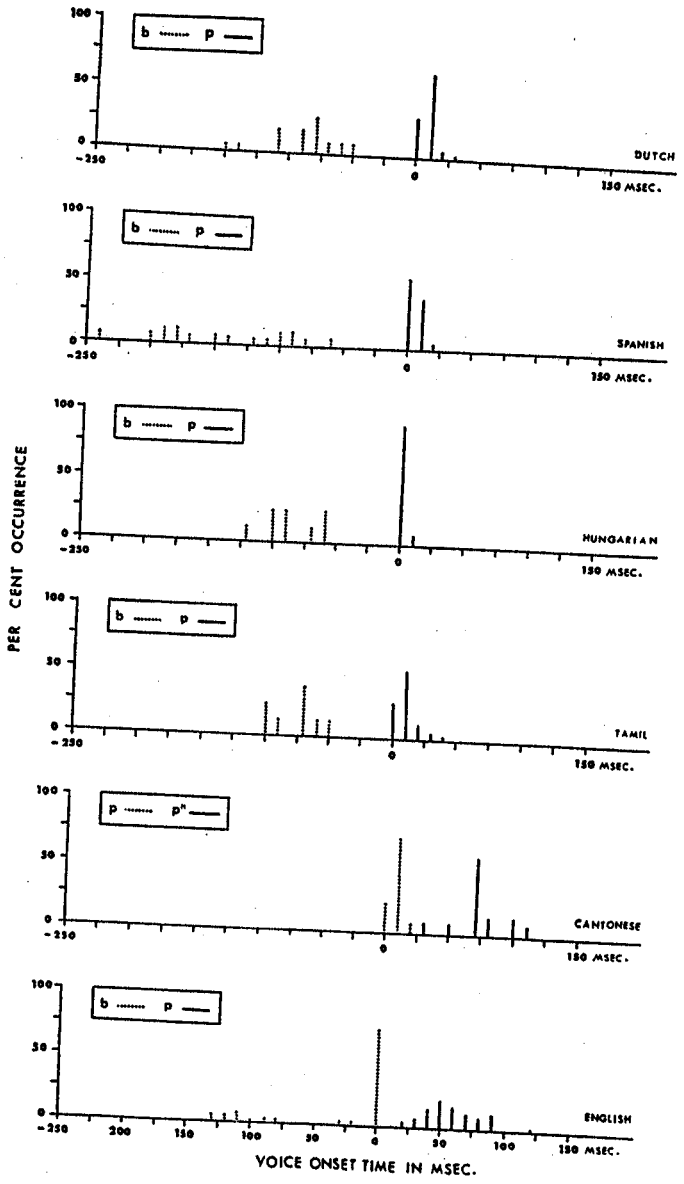

FIGURE 2. Voice onset-time distributions: labial stops of two-category languages.

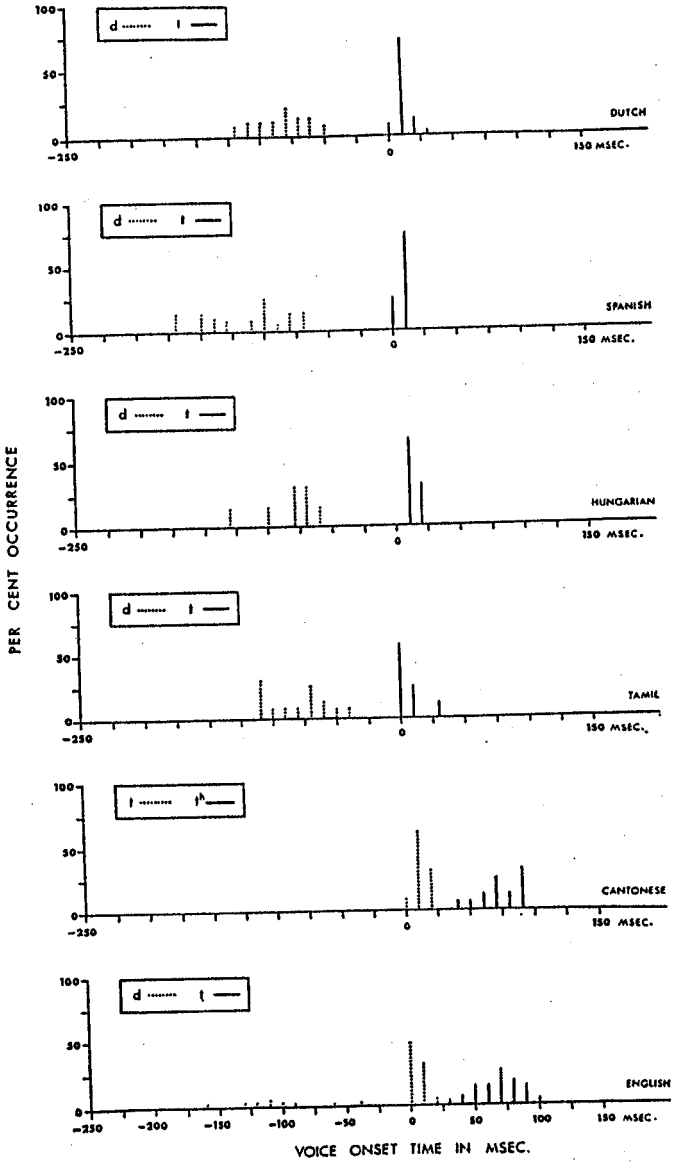

FIGURE 3. Voice onset-time distributions: apical (dental and alveolar) stops of two-category languages.

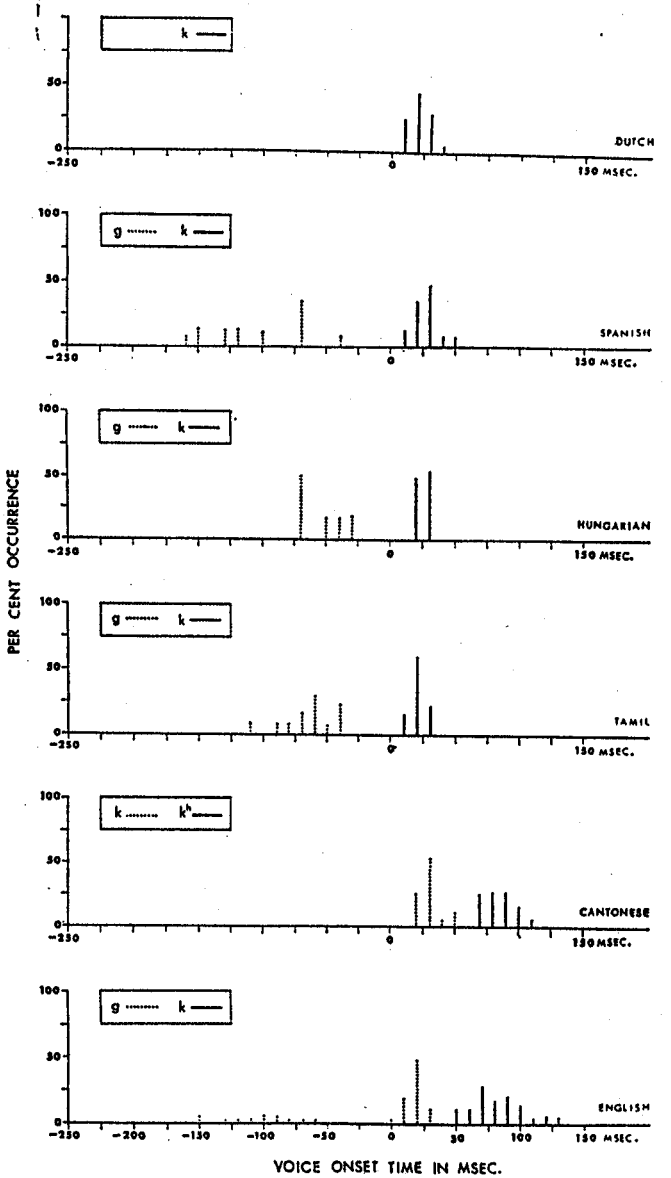

FIGURE 4. Voice onset-time distributions: velar stops of two-category languages.

time. Despite some differences in the magnitudes of averages and ranges of values from language to language the overall similarity among these six languages is striking, for it appears that the stop categories overall fall generally into three ranges—one from about −125 to −75 msec, one from zero to +25 msec, and a third from about +60 to +100 msec. The median values for these ranges are −100, +10 and +75 msec respectively. Thus we may say that Dutch, Spanish, Hungarian and Tamil show essentially the same distribution of values, occupying the ranges about −100 and +10 msec, that Cantonese locates its stop categories in the +10 and +75 msec ranges, and that English shows values distributed like those of Cantonese, except for a scattering of items in the −100 msec range.[33]

The situation in the case of the three- and four-category languages is represented in Figs. 5, 6 and 7. Values measured for Eastern Armenian and Thai line up very precisely with the ranges which the two-category languages occupy jointly; their three categories are distributed over the ranges centering at −100, +10 and +75 msec. The remaining three-category language, Korean, is peculiar in that all of its stops are located in the positive half of the voice onset time continuum; the resolution between the two lower-valued categories is not very good, while the third category shows average values that are rather greater than those found in any of the other languages. But while the distribution of values is thus somewhat anomalous, we cannot say with reasonable assurance that our measure of voice onset time fails to separate the three categories of Korean stops; it will certainly suffice to distinguish the aspirated set from the other two and it may still well be the single most important measure for separating the latter.

The two four-category languages, Hindi and Marathi, present us with our only clearcut cases in which the measure of voice onset time is insufficient for distinguishing among all the stop categories of a language. To be sure, the voiced unaspirated and voiced aspirated stops show differences in average values that are almost systematic; nevertheless they occupy ranges that are nearly coextensive. It seems very likely that the voiced aspirates are distinguished from the other voiced category by the presence of low amplitude buzz mixed with noise in the interval following release of the stop.

Two final observations with regard to the data of Figs. 2–7 can be made. The first is the rather curious one that not a single one of the two-category languages locates its categories where we might expect to find them, that

[33] Very recently Lawrence Raphael of Queens College of the City University of New York examined initial stops in isolated Persian words and found them to be much like English stops in the way they lie along the dimension.

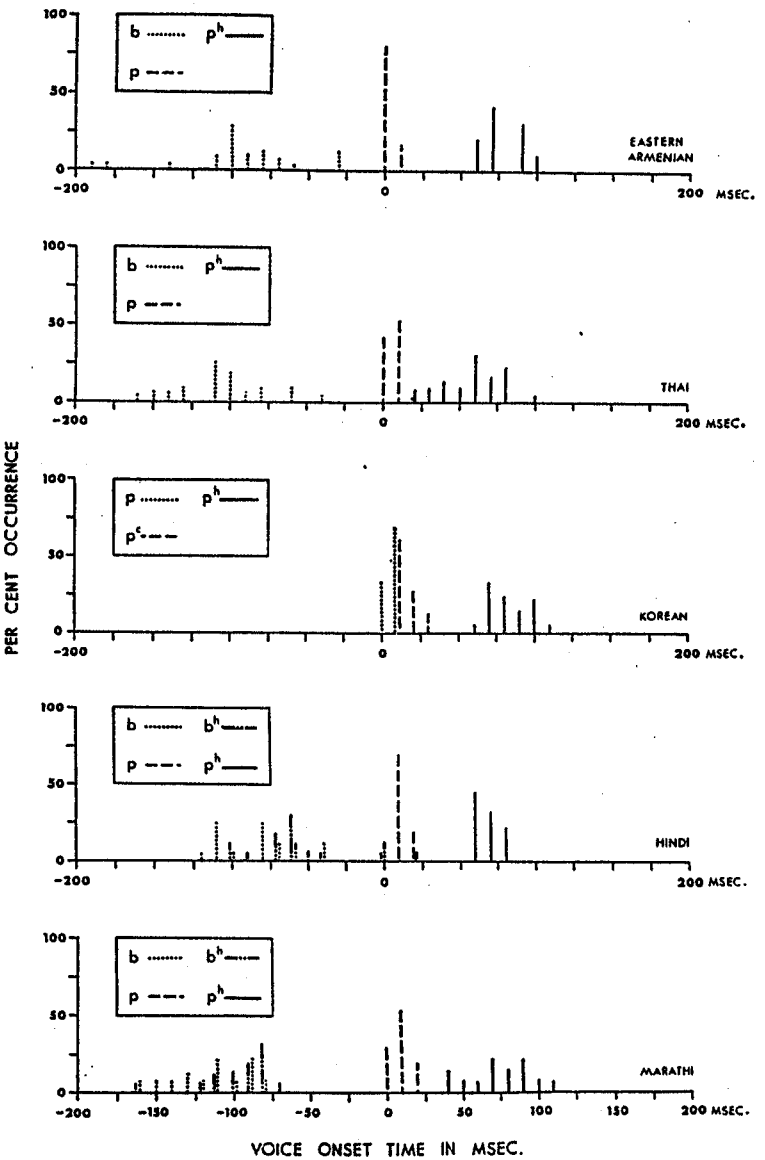

FIGURE 5. Voice onset-time distributions: labial stops of three- and four-category languages.

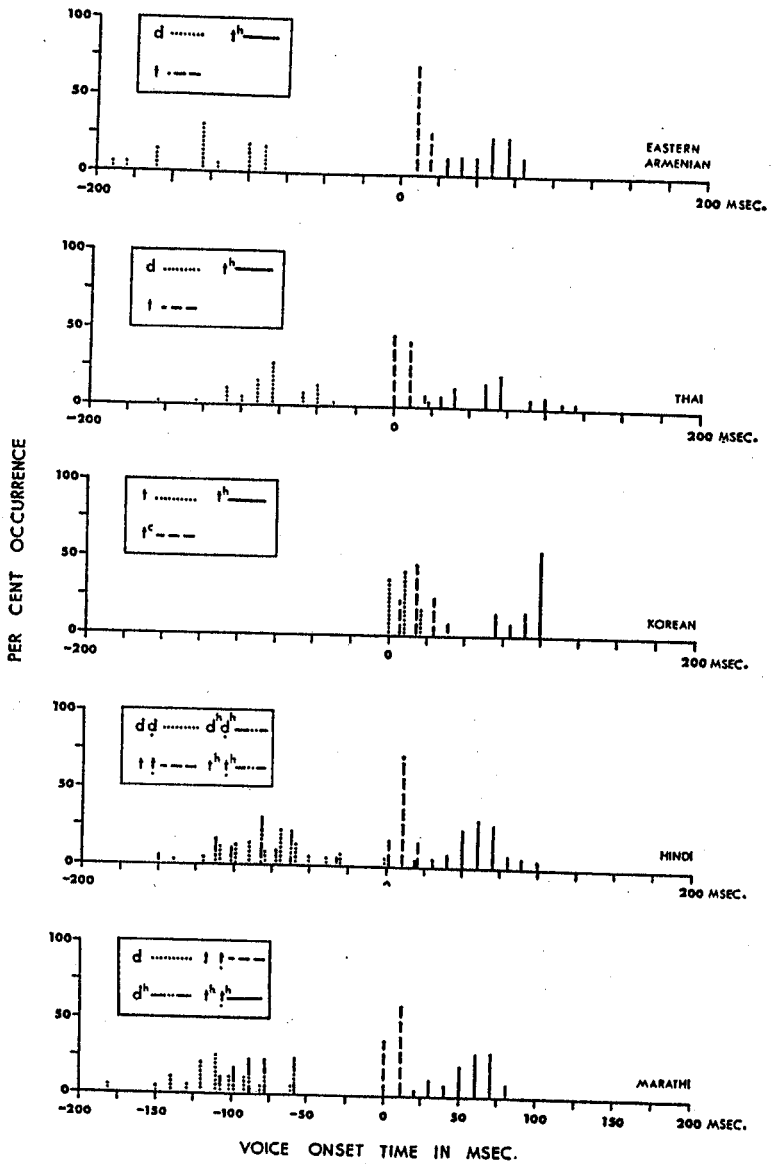

FIGURE 6. Voice onset-time distributions: apical (dental and domal) stops of three- and four-category languages.

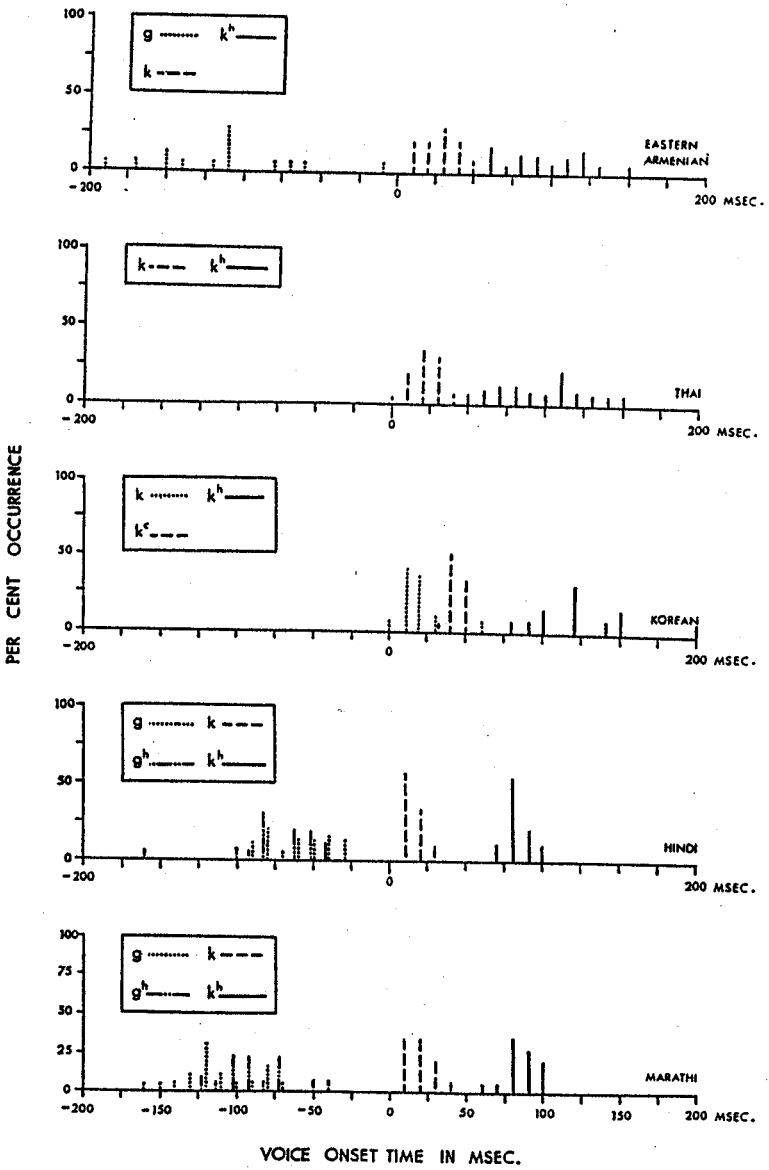

FIGURE 7. Voice onset-time distributions: velar stops of three- and four-category languages.

is, at opposite ends of the continuum of voice onset time. This fact, if it is a reflection of the situation in languages generally, is evidence for the view that in the phonetic "realization" of phonemic contrasts human beings fall considerably short of utilizing all the phonetic space available to them. The second observation concerns the relation between frequency distribution and the existence of "holes" in the phonemic inventory of a language. In both Dutch and Thai the velar series of stops is deficient as compared with the labial and apical sets, and we might plausibly expect to find that the velars found in those languages have rather broader distributions than the other stops. This is not the case, however; instead the distributions for Dutch /k/ and Thai /k/ and /k^h/ are exactly what they would be if their velar series conformed in number of phonemes to the other series.

Finally, it is of interest to see how values for the voice onset time measure are distributed for all eleven of our languages taken together. In Figure 8 the overall frequency distributions for each of our three general places of articulation are given after normalizing so that all stop categories in each language are in effect represented by the same number of measurements. From this figure it appears that the distribution of values is essentially tri-modal, corresponding generally to the ranges centering at -100, $+10$ and $+75$ msec. The three modes differ considerably in their degree of peakedness and in the sharpness of their separation from the other modes of the distribution. It is furthermore apparent from Figure 8 that the positive modes for the velar place of articulation have somewhat higher values than either the labial or the apical stops. It should also be noted that each of the three frequency distributions shows a "hole" in the region just below zero, although at the moment there is no evidence to support speculation on the meaning of the observation.

III. Stops in Sentences

The ultimate usefulness of measuring voice onset time depends on how effectively it enables us to identify the stops in running speech. We have not as yet had a really serious look at long stretches of speech, but we have compared values for words spoken in isolation with those for the same words embedded in longer sentences, both in initial and non-initial positions (see **Procedure** in I). The sentence data for our eleven languages are presented in Tables 12–22. The average value and range of voice onset time in milliseconds, as well as the number of items measured for each phoneme, are given in the same format as in Tables 1–11, except that a separate set of items for non-initial position appears in each table.

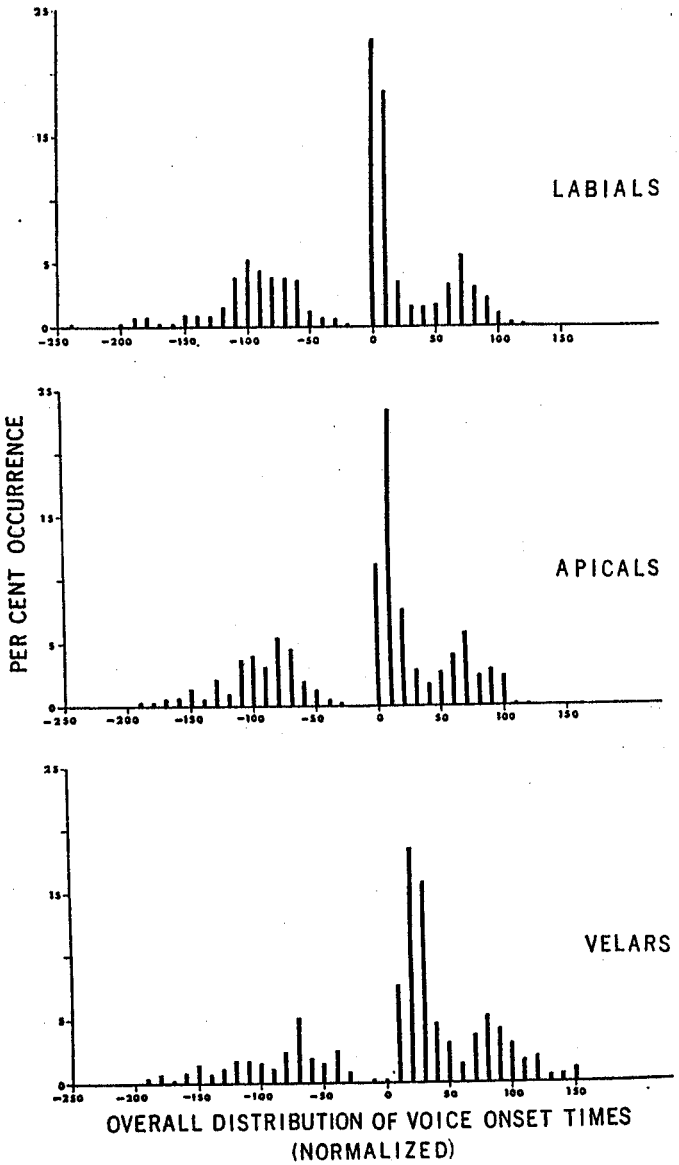

FIGURE 8. Overall frequency distributions of voice onset-time values, normalized so that all stop categories are equally represented.

Measurement data.

TABLE 12

Dutch
(1 speaker)

	/b/	/p/	/d/[34]	/t/	/k/
			Initial		
Av.	−41	11	−51	16	34
R.	−65:0	0:25	−70: −40	10:20	25:60
N.	10	10	11	9	10
			Non-initial		
Av.	*	9	−40*	20	33
R.	—	0:25	−70: −10	0:55	25:40
N.	—	10	2	12	10

* An asterisk next to an average means that in addition to the data recorded, cases of unbroken voicing were observed, i.e., voicing that proceeds unbroken from a preceding voiced environment into the stop-closure interval. If an asterisk is the sole entry, it means that all occurrences of the stop showed unbroken voicing.

TABLE 13

Puerto Rican Spanish
(2 speakers)

	/b/	/p	/d/	/t/	/g/	/k/
			Initial			
Av.	−110	4	−109	7	−92	25
R.	−175: −35	0:15	−170: −55	0:15	−145: −50	15:55
N.	15	19	16	14	16	19
			Non-initial			
Av.	−90*	4	*	8	*	20
R.	−90	0:15	—	0:15	—	10:30
N.	1	21	—	16	—	18

*In addition to unbroken voicing, Spanish stop phonemes in non-initial position have fricative allophones in some environments. The latter were not measured.

TABLE 14

Hungarian
(1 speaker)

	/b/	/p/	/d/	/t/	/g/	/k/
			Initial			
Av.	−55	0	−70	20	−61	28
R.	−70: −35	0	−90: −35	10:35	−95: −40	20:45
N.	6	6	6	6	6	6

[34] One occurrence of initial /d/ showed a voicing lag, not lead, of 10 msec.

Non-initial

Av.	*	4	*	24	*	34
R.	—	0:15	—	20:30	—	10:45
N.	—	6	—	6	—	6

TABLE 15
Tamil
(1 speaker)

	/b/	/p/	/d/	/t/	/g/[35]	/k/
			Initial			
Av.	−61	12	−64	10	−56	27
R.	−75: −40	0:45	−100: −45	0:25	−80:10	15:40
N.	10	37	13	8	13	11

[35] Two occurrences of initial /g/ showed voicing lag. One was 10 msec and the other, 5 msec.

Non-initial

Av.	*	6	*	6	*	10
R.	—	0:20	—	0:15	—	10:30
N.	—	43	—	8	—	10

TABLE 16
Cantonese
(1 speaker)

	/p/	/p^h/	/t/	/t^h/	/k/	/k^h/
			Initial			
Av.	11	58	15	62	34	68
R.	10:15	50:75	10:25	45:85	25:50	55:80
N.	6	6	4	5	6	6
			Non-initial			
Av.	9	39	15	66	23	67
R.	0:15	20:55	10:25	45:110	10:45	60:75
N.	5	5	6	6	5	6

TABLE 17[36]
English
(4 speakers)

	/b/	/p/	/d/	/t/	/g/	/k/
			Initial			
Av.	7/−65	28	9/−56	39	17/−45	43
R.	0:15/−65	10:45	0:25/−90: −20	15:70	0:13/−45	30:85
N.	24/2	24	45/5	26	24/1	25

[36] The investigation of voice onset time in English was conducted before the format for the present study had been set, and the method of obtaining stops embedded in

Non-initial

Av.	4/−63*	34	7*	37	16*	49
R.	0:20/−90: −50	15:70	0:20	15:95	0:40	15:90
N.	47/4	86	34	99	53	138

TABLE 18

Eastern Armenian

(1 speaker)

	/b/	/p/	/pʰ/	/d/	/t/	/tʰ/	/g/	/k/	/kʰ/
				Initial					
Av.	−72	5	51	−107	13	35	−66	23	83
R.	−140: −55	0:10	35:65	−150: −70	10:15	30:40	−130:0	15:30	55:100
N.	14	6	6	3	2	2	7	7	8
				Non-initial					
Av.	−47*	7	53	*	10	47	−21*	27	76
R.	−55: −30	0:10	35:65	—	10	45:50	−35:0	15:55	45:95
N.	3	6	6	—	2	2	4	8	8

TABLE 19[37]

Thai

(1 speaker)[38]

	/b/	/p/	/pʰ/	/d/	/t/	/tʰ/	/k/	/kʰ/
				Initial				
Av.	−35	8	37	−53	15	63	15	69
R.	−50: −20	0:15	25:45	−60: −45	15	55:70	15	55:90
N.	2	4	3	2	2	2	2	4
				Non-initial				
Av.	−66*	11	50	−38*	8	43	16	74
R.	−90: −20	0:20	25:80	−60: −20	0:20	20:95	0:25	30:135
N.	8	5	12	10	6	18	6	29

sentences was different. We simply composed several lists of sentences containing not only the words of Part II but also many other words with initial stops. The informants were told to read them as naturally as they could. For the double entries under /b d g/, see the discussion accompanying Tables 6 and 6a.

[37] As in the case of English, voice onset time in Thai was investigated before the format for this study had been set. To obtain stops embedded in sentences we listened to high-quality tape recordings of about half an hour's worth of unrehearsed conversational and narrative material and picked out passages containing prevocalic stops. The only phones accepted as utterance-initial stops were those that appeared after juncture, which was taken to be any break in discourse such that what followed was impressionistically equivalent to the initiation of speech. This procedure yielded somewhat fewer initial stops than non-initial ones.

[38] This speaker is one of the three entered in Table 8.

TABLE 20
Korean
(1 speaker)

	/p/	/pᶜ/	/pʰ/	/t/	/tᶜ/	/tʰ/	/k/	/kᶜ/	/kʰ/
					Initial				
Av.	7	22	89	11	30	100	20	48	125
R.	0:15	15:40	55:115	0:20	15:40	75:130	0:30	35:75	80:175
N.	14	28	24	15	21	12	14	35	10
					Non-initial				
Av.	5	13*	75	12	22*	78	21	44*	93
R.	0:10	10:20	40:130	0:25	10:45	50:120	10:35	30:65	55:175
N.	14	10	23	16	12	10	14	11	10

* The noteworthy thing about unbroken voicing (see Table 12) in Korean is that it occurs in the middle category rather than the leftmost one. The matter will be brought up again in the general discussion of the sentence data.

TABLE 21
Hindi
(1 speaker)
Initial

	/b/	/bʰ/	/p/	/pʰ/	/d/	/dʰ/	/t/	/tʰ/
Av.	−89	−65	12	63	−88	−78	11	63
R.	−115: −75	−105:0	5:20	50:75	−120: −50	−105: −65	5:20	35:80
N.	8	8	8	8	8	8	8	8

	/ḍ/	/ḍʰ/	/ṭ/	/ṭʰ/	/g/	/gʰ/	/k/	/kʰ/
Av.	−74	−74	11	57	−47	−59	16	84
R.	−140: −30	−90: −35	5:15	50:65	−70: −35	−85: −30	10:25	65:105
N.	8	8	8	8	8	8	8	8

Non-initial

	/b/	/bʰ/	/p/	/pʰ/	/d/	/dʰ/	/t/	/tʰ/
Av.	*	*	9	46	0*	*	14	45
R.	—	—	0:20	35:55	0	—	10:20	30:70
N.	—	—	8	8	1	—	8	8

	/ḍ/	/ḍʰ/	/ṭ/	/ṭʰ/	/g/	/gʰ/	/k/	/kʰ/
Av.	0*	*	8	48	10*	*	21	63
R.	0	—	0:15	15:60	10	—	0:35	50:80
N.	1	—	8	8	1	—	8	8

TABLE 22
Marathi
(1 speaker)

Initial

	/b/	/bʰ/	/p/	/pʰ/	/d/	/dʰ/	/t/	/tʰ/
Av.	−106	−73	0	35	−93	−71	11	54
R.	−145: −85	−90: −60	0	25:55	−110: −75	−85: −60	10:15	35:80
N.	4	4	4	4	8	8	4	4

	/t/	/tʰ/	/g/	/gʰ/	/k/	/kʰ/
Av.	11	35	−81	−67	21	73
R.	10:15	30:40	−90: −70	−85: −40	20:25	60:80
N.	4	4	6	6	4	4

Non-initial

	/b/	/bʰ/	/p/	/pʰ/	/d/	/dʰ/	/t/	/tʰ/
Av.	*	*	0	45	*	0*	15	60
R.	—	—	0	40:50	—	0	15	50:70
N.	—	—	4	2	—	1	3	4

	/t/	/tʰ/	/g/	/gʰ/	/k/	/kʰ/
Av.	3	53	*	*	13	66
R.	0:10	40:65	—	—	0:20	60:75
N.	4	4	—	—	4	4

Comparison with words. Tables 12 to 22 reveal that by and large the sentence data are congruent with the word data. Although the voice onset time dimension effectively separates stop categories in sentences, there is, nevertheless, some effect of embedding the stops in running speech. First of all, for categories with voicing lead, we find that in non-initial position voicing usually proceeds unbroken from a preceding voiced environment into the closure interval; any interruption of glottal buzz depends on what sounds occur before the stop. This effect is also observed in categories with short voicing lag in two languages, English /b d g/ (Table 17) and the Korean middle category /pᶜ tᶜ kᶜ/ (Table 20).[39] It is difficult to see, off-hand, why it should be the Korean middle category and not the lowest one that

[39] It must be understood that while the negative as well as positive marking of the categories named affects the absolute magnitude of voice onset time, it does not cause a weakening of the power of the dimension; indeed, it enhances it. Where Korean and English stops have unbroken voicing (i.e., voicing that has started earlier in the utterance and not stopped at the time of the stop closure), they are better separated from the categories with interrupted voicing and lag than they are in other environments.

takes on unbroken voicing;[40] however, it should be noted that in all the other languages but English, categories with zero onset time or very short lag are like Korean /p t k/ in not showing unbroken voicing. Another effect of embedding stops in sentences is that voice onset time values, both lead and lag, tend to be a bit compressed in comparison with the values measured in the citation forms of words. If this tendency, admittedly not a strong one, persists in a larger sampling of utterances, it suggests that not only vowel duration but also voice onset time is likely to be affected by temporal compression in rapid speech. In the several cases where this compression occurs, there is, of course, a reduction of the gap separating those categories from each other along the voice onset time dimension. There is no question, however, of a reduction to the point of serious overlap between otherwise distinct categories.

Several uncontrolled variables in the sentence data have to be mentioned. In our desire to get normal connected speech, we made no attempt to control the rate of utterance, but simply told the informants to say the sentences naturally. Stress variation was considerable, especially in English where we did not hesitate to examine stops that occurred in initial position in words other than the key words. (See footnote 36.) No control over vocalic environments was exercised. When feasible, we had the informant, or some other speaker of the language, check his recordings, but we are quite sure that occasional losses of contrast have remained in our data because of informants' slips. Aside from accidental losses of contrast, there is also the question of unstable contrasts. If, in a given language system, two phonemes tend to alternate with some freedom, e.g. English /ð/ and /θ/, one might wonder whether that contrast is not in a state of flux. The phonemic descriptions available to us indicated nothing as to the relative stability of the stop contrasts, yet we suspect that this may be a factor in some of our data, for example, the Korean. In spite of these variables, which may in part be uncontrollable, the resolving power of the voice onset time dimension is good.

IV. Discussion and Plans

Inferences as to glottal mechanisms. In phonetic investigations one would like to show what physiological mechanisms underlie acoustic features that differentiate phonemes. We believe that certain inferences about glottal behavior will explain the findings of this study, as well as other studies, of

[40] See, however, S. Martin (1951) for the positing of a glottal component for the lowest-valued category in Korean. Glottalization has been attributed to the stops of other languages too. All such assertions, based on auditory evaluation, clearly require physiological confirmation.

the voicing distinction in stops. It is to be understood that this is not merely the obvious question of whether or not there is laryngeal vibration. Instead, it seems evident that a fairly complicated acoustic output is dependent upon the relatively simple matter of varying the area of the glottis. If the speaker closes the glottis down enough for phonation, he does not directly "command" the vocal folds to vibrate;[41] rather, he makes the necessary muscular adjustments that set the conditions for vibration when sufficient airflow is supplied.[42]

At this point it may be useful to go through a brief review of the salient features of voice production.[43] The myoelastic-aerodynamic theory, involving a counterbalancing of aerodynamic forces and muscular tensions, is the most widely accepted explanation of phonation. When air pressure beneath the closed glottis is sufficiently high, the vocal folds are blown apart and a puff of air is released into the supralaryngeal cavities. With the resulting drop in subglottal air pressure, the folds snap together again[44] under the impact of two restoring forces: (1) tension of the intrinsic muscles of the larynx, and (2) a drop in air pressure along the margins of the folds caused by air rushing through the glottis (the Bernoulli effect). Voice onset can occur either with the folds completely adducted or somewhat abducted. The laryngeal repetition rate (fundamental frequency) is regulated by the length, mass and tension of the vocal folds in conjunction with varying amounts of air pressure from the lungs.[45] The puffs of air thus repeatedly released by

[41] Unless one accepts Husson's generally discredited theory of direct neural triggering of laryngeal pulses. See R. Husson, "Etude des phénomènes physiologiques et acoustiques fondamentaux de la voix chantée," Thèse. *Revue Scientifique* (Paris, 1950). For criticism, see the works cited below in footnotes 43 and 45.

[42] A theory of the perceptual relevance of motor commands is presented in A. M. Liberman, F. S. Cooper, K. S. Harris and P. F. MacNeilage, "A Motor Theory of Speech Perception," *Proceedings of the Speech Communication Seminar*, Vol. II, Royal Institute of Technology (Stockholm, 1962).

[43] For good anatomical drawings of various views of the larynx, see, e.g., Raymond C. Truex and Carl E. Kellner, *Detailed Atlas of the Head and Neck* (New York, 1948), Figures 66–74.

[44] This is a commonly observed mode of vibration. The description does not take into account the way the mode varies with different voice registers. For example, the system can be kept in oscillation with little or no contact between the folds at the end of each cycle.

[45] For a good exposition in greater detail, see, e.g., Giles W. Gray and Claude M. Wise, *The Bases of Speech*, 3rd ed. (New York, 1959), pp. 163–171. For a critical survey of work on phonation, see Janwillem van den Berg, "Myoelastic-Aerodynamic Theory of Voice Production," *Journal of Speech and Hearing Research*, I (1958), 227–244. For work on subglottal air pressure, see Peter Ladefoged, "Sub-Glottal Activity during Speech" in *Proceedings of the Fourth International Congress of Phonetic Sciences*, ed., A Sovijärvi and P. Aalto (The Hague, 1962), pp. 73–91.

the glottis excite the resonant frequencies of the vocal tract, producing voice.

The glottis can vary considerably in size and shape. This is effected principally by movements of the arytenoid cartilages under the control of the intrinsic muscles of the larynx. The vocal folds extend from the thyroid cartilage in front to the vocal processes (anterior angles) of the arytenoid cartilages in back. These pyramidal cartilages, set on the upper back edge of the cricoid cartilage of the trachea, can tilt backwards, stretching and possibly tensing the folds, and forward, relaxing them. They can glide toward and away from each other, and rotate to bring the vocal processes together or apart;[46] thus the front three-fifths of the laryngeal opening, the membranous glottis, can be closed while the cartilaginous glottis in back is open, or the whole glottis can be open or shut. Consequently, wide variations in the area and shape of opening are possible.[47]

By and large, the concept of voice onset time offers no physiological difficulty. Phonation simply starts at some point in time relative to the release of the stop closure. (During a voicing lag, aspiration will be heard if the vocal tract resonates to turbulent air passing through the open glottis.) It must be acknowledged that in making this physical measure one may be misled into including a short span of pulsation, perhaps just one cycle, that is too weak to be audible. We are certain that occasional errors of this sort can make no significant difference in the way the stop categories have been shown to lie along the dimension.

The question of audibility of glottal pulses is a crucial one. There is no point, it would seem, in speaking of the distinctive relevance of a phonetic feature that is not perceptible. In the **Procedure** we stated that from time to time, in an environment of preceding voicing, non-initial stops with apparent voicing lag showed faint vertical striations at glottal rates near the baseline of the spectrogram below the clearly aperiodic high-frequency noise of aspiration. It seems that some investigators[48] have observed such cases and seized upon them to cast doubt on the primacy of the voiced/voiceless distinction. This may reflect a failure to discriminate between acoustic features and their preceptual correlates; that is, there is a danger of giving primary emphasis to an instrumentally detectable acoustic disturbance that, in the situation, can have no auditory consequences. Despite the almost negligible number of our recorded utterances that showed this

46 Harold M. Kaplan, *Anatomy and Physiology of Speech* (New York, 1960), p. 120.

47 Kaplan, pp. 128–129.

48 For example, Fred W. Householder in his review of Ernst Pulgram, *Introduction to the Spectrography of Speech*, *International Journal of American Linguistics* XXVII (1961), p. 178.

phenomenon, we feel we must discuss the matter, if only to satisfy the observers who have made so much of it. Donning our phoneticians' caps, we listened carefully and repeatedly to the utterances in question without detecting voicing in the hold or release of the stops. In addition, in passing the tapes over the playback head of the tape recorder at extremely slow speeds and very high amplification, a procedure which yields a distinct impression of periodic pulsing in the case of normal voicing, we heard nothing but the noise of release and aspiration. We believe that inferences can be made from the available knowledge of glottal mechanisms to account for these anomalous cases.

To date it has not been possible to look directly at the action of the vocal folds in running speech; therefore we can only make inferences from other kinds of evidence. Spectrograms suggest that the laryngeal oscillations of a preceding voiced environment may simply continue for a while even after the glottis has begun to open for a voiceless stop; these vibrations are so low in intensity that any auditory effect they might have by themselves seems to be masked out by the stop burst and the noise of turbulent air rushing through the glottis. We propose the term "edge vibrations" for this hypothesized behavior.[49] That edge vibrations occur at all in speech is evident in the high-speed motion pictures of the larynx now available. These films are limited so far to speech with the mouth wide open. In her film "Vocal Cord Action in Speech: a High-Speed Study,"[50] Elizabeth Uldall shows vowel productions with laryngeal vibrations occurring before the glottis has closed at the beginning of the vowel and after the glottis has opened at the end of the vowel. Vibration occurs with an even wider glottal aperture for an initial [h].

A way of directly photographing edge vibrations, or indeed any laryngeal vibrations, during consonant articulations has not yet been worked out. The method of photo-electric glottography, recently developed by B. Sonesson,[51] affords an indirect way of recording the area of the glottis as it varies in time. The procedure used by us was designed independently by Franklin S. Cooper, who followed suggestions made by Paul Moore of the University of Florida. The system uses a powerful light source, a light guide, and a photoelectric cell. The light guide is a light-conducting rod

[49] The spectrum of the edge vibrations seems to consist only of one or two weak harmonics at the low-frequency end. This can be seen better in narrow-band spectrograms.

[50] University of Edinburgh and Swiss Federal Institute of Technology, 1957.

[51] B. Sonesson, *On the Anatomy and Vibratory Pattern of the Human Vocal Folds* (Lund, 1960).

shaped to fit the contour of the roof of the subject's mouth[52] and shower an intense beam of light down onto the larynx. In a darkened room, light passing through the open glottis can be seen coming through the skin of the neck below. The photocell is held against the neck at this point. The amount of light entering the cell varies in proportion to the changing area of the glottis. An oscillographic record of the photocell output reveals whether the glottis is open or closed and whether the folds are vibrating or not.[53]

Through transillumination of the larynx we obtained a number of glottal traces for English voiceless stops in medial position after voiced phones and before both stressed and unstressed vowels. In every instance the glottis opened abruptly, but the vocal folds did not always stop vibrating at the same moment. They sometimes continued to vibrate until the glottis began closing to reach normal phonatory position again. In one production, before the stressed vowel of *depict*, the folds vibrated through the whole open phase, with some weakening of the pulses after the glottis reached its peak opening. These findings are supported by high-speed oscillograms of such utterances recorded simultaneously by means of an air microphone, which picks up the signal as it arrives at the ear, and a throat microphone, which is more directly responsive to laryngeal vibrations. The throat trace will sometimes show very weak glottal modulation (edge vibration) on the wave form of the aspiration noise of English /p t k/, while the air trace shows only noise. Of course, if the edge vibrations are a bit higher in amplitude, although still inaudible, they will appear in the air trace too, just as they do in spectrograms. Further work with improved instrumentation is needed, but the findings of this little side study clearly lend credence to our thoughts on edge vibrations.

Another anomaly that requires explanation is the failure of the dimension of voice onset time to separate the voiced aspirate stops from the voiced inaspirates of Hindi and Marathi. Voicing lead of much the same duration is found in both categories, and indeed it distinguishes the pair from the other stop categories. Auditory impressions suggest that the

52 The two subjects to try it so far find that they can tolerate the light guide only for short periods of time, but while it is in place, they can articulate pretty well. F. S. Cooper is now designing a modification of the system with a light guide made of a flexible fiber optics bundle that can be passed through the nose to interfere even less with speech. The next step will be to try to supplement transillumination with direct photography by bringing up an image through a separate light bundle to a motion picture camera.

53 Another indirect method of obtaining the glottal waveform is that of inverse filtering, in which the effects of the resonances of the vocal tract are removed from the speech wave. So far, this method can best be used for sustained sounds and not in running speech.

voiced aspirates are released with breathy voice or murmur.[54] These impressions are supported by spectrograms in which, upon release of the stop, the voicing is seen to take on a special character. There is a period of glottal periodicity, sometimes intermittent, mingled with random noise in the formant regions, all at relatively low amplitude. This voiced aspiration sounds about as long as voiceless aspiration, but it is difficult or even impossible to make physical measurements of its duration, because it merges more or less imperceptibly with the following normal voicing.

Instead of the time of onset of voicing it is the kind of voicing that distinguishes the voiced aspirates; thus, here too we should be able to find a laryngeal mechanism to account for the feature. It is physiologically reasonable to infer that the murmur is characterized by glottal vibration of a sort that allows a steady leakage of turbulent air either through the incompletely closing folds or between the open arytenoid cartilages.[55] In his film "Voice Production: the Vibrating Larynx,"[56] Jw. van den Berg uses a specimen larynx to demonstrate breathy voice. Under contraction of the lateral cricoarytenoid muscles air escapes between the arytenoid cartilages while the membranous glottis is vibrating.[57] We hope to test these inferences by transilluminating the larynx of a native speaker of Hindi or Marathi. It is also possible that there will turn out to be differences in subglottal activity associated with the production of aspirated and unaspirated voiced stops.

Coming back to the more general cases, those in which voice onset time does distinguish stop categories, we often find other features in association with distinctions along the dimension. They are: delay in the onset of the first formant (F1 cutback),[58] differences in buccal air pressure and burst intensity, and the maintenance of contrasts in whispered speech. We believe that these too are susceptible of explanation in terms of glottal mechanisms, but it will be more convenient to discuss these matters in a sequel to this article.[59]

[54] In the occasional instance of /b^h d^h g^h/ without voicing lead, the murmured release is nevertheless present.

[55] Eli Fischer-Jørgensen, *Almen Fonetik*, 3rd ed. (Copenhagen, 1960), p. 63.

[56] University of Groningen, the Netherlands, 1960.

[57] Murmur, voiced [fi], etc., are discussed in J. Lazicius, *Lehrbuch der Phonetik* (Berlin, 1961), pp. 45, 60–62.

[58] A. M. Liberman, P. C. Delattre, and F. S. Cooper, "Some Cues for the Distinction between Voiced and Voiceless Stops in Initial Position," *Language and Speech* I (1958), 153–166.

[59] "A Cross-Language Study of Voicing in Initial Stops: Experiments in Perception," (in preparation).

The fortis/lenis feature in English. Let us return briefly to a topic presented in the Introduction, the positing of a fortis/lenis distinction for English stop consonants. English /b d g/ used to be regularly labelled "voiced" and /p t k/, "voiceless". At some point however, certain linguists decided that a definition of the terms "voiced" and "voiceless" in strictly phonetic terms made them not completely accurate as descriptors, for we can read, in some accounts at least (see footnotes 22 and 23) that initial allophones of /b d g/ may be more or less voiceless. Such linguists have pretty well adopted the practice of labelling the two sets of stop phonemes "lenis" and "fortis," asserting that the two sets are more generally distinguished on the basis of differences in force of articulation than in voicing. Now, if all these terms are assumed to have phonetic meanings, we must ask whether this change nets us any gain in precision of description, for we have exchanged a phonetic dimension, voicing, which has a clear articulatory and acoustic meaning, for one which is considerably less well defined both articulatorily and acoustically; furthermore, attempts at a purely physiological definition of fortisness in stops have thus far yielded nothing reliable.[60] It seems to us that the greater generality claimed for the fortis/lenis dimension of description depends upon the vagueness of its phonetic reference, a vagueness which makes it difficult to decide whether the new terms really refer to a physical difference which serves to separate the /p t k/ and /b d g/ sets or whether they are merely phoneme-class labels masquerading as phonetic categories. A closer look at the fortis/lenis feature will be taken in the sequel to this article.

Experiments in perception. Although our data generally support the view that the dimension of voice onset time is a sufficient acoustic differentiator of stop categories across a wide variety of languages, it remains to be demonstrated that this feature is perceptually relevant and sufficient.[61] We are preparing series of synthetic-speech syllables of the form stop + vowel in which the voice onset time varies in small steps along a continuum that matches the ranges observed in the present study. That is, each series will encompass the ranges of all eleven languages, as well as many others, although for a given language there will no doubt be variants that are in-

[60] Gloria Lysaught, Robert J. Rosov, and Katherine S. Harris, "Electromyography as a Speech Research Technique with an Application to Labial Stops," *Journal of the Acoustical Society of America* XXXIII (1961), 842 (Abstract); K. S. Harris, G. Lysaught, and M. H. Schvey, "Experimental Studies of the Production of Oral and Nasal Labial Stops," (in preparation).

[61] This has in fact been demonstrated for English (Liberman *et al.*, 1958) but not for the other languages.

appropriate in the sense of both lead and lag. A series will be prepared for each of the three major places of articulation.

Randomized sequences of stimuli of each syllable-type will be presented for phoneme labelling to speakers of several of the languages under consideration. We intend these experiments to test three hypotheses: (1) Perceptual boundaries between phoneme categories will fall along the voice onset time dimension; (2) the phoneme boundaries will vary from language to language in general accord with the measurements obtained from real speech; (3) best synthesis will require that voicing onset be represented acoustically by the simultaneous starting of both periodic pulsation (buzz) and the first formant.[62] Preliminary experimental data, obtained from trials of one such test series, support these hypotheses. Where category boundaries along the dimension are not very sharp, one must look for possible perceptual instability of the phonemic distinction in question. The contrast between the two Korean lower-valued categories seems a likely prospect. A necessary precaution, then, will be to run tests of the intelligibility of the distinction in real speech before turning to the experiments with synthetic speech.

Experiments will also be conducted to determine acuity of perceptual discrimination of stop variants across phoneme boundaries in comparison with discrimination of variants within phoneme categories. The cross-language situation promises to be of much interest in relation to earlier findings of the Haskins group,[63] as well as in connection with more recent thought and work on a possible link between perception and articulation.[64] This line of inquiry is made all the more interesting because the categorical kind of production suggested by the three cross-language modes of Figure 8; attempts at mimicry of the synthetic variants should throw further light on this.

Work on the various experiments in perception is now in progress, and it is hoped that an article presenting the results together with the implications of both studies will soon be ready for publication.

Summary. Linguists often find it useful to divide the phonemes of a language into "voiced" and "voiceless" categories. For stops, some languages are said to utilize aspiration in conjunction with voicing to yield two, three or

[62] The third hypothesis is based on the work of Liberman, Delattre and Cooper (1958), p. 64.

[63] A. M. Liberman, K. S. Harris, H. S. Hoffman, and B. C. Griffith, "The Discrimination of Speech Sounds within and across Phoneme Boundaries," *Journal of Experimental Psychology* LIV (1957), 358–368.

[64] A. M. Liberman *et al.* (1962).

four categories, while in other languages categories are said to be distinguished solely by differences in aspiration. Some linguists, moreover, speak of fortis and lenis categories. Despite the fact that these features of voicing, aspiration and force of articulation are usually treated as independent dimensions of phonetic description, there are some grounds for considering them to be plausible consequences of a single underlying variable. In the search for acoustic features which serve as cues for the perception of stop consonants in initial position we have focused our attention on spectrographic measurements of the time interval between the burst that marks release and the onset of periodicity that reflects laryngeal vibration. This measure of voice onset time has been applied to word-initial stops in eleven languages and has been found to be highly effective as a means of separating phonemic categories, although these languages differ both in the number of those categories and in the phonetic features usually ascribed to them. The boundaries between contrasting categories along the continuum of voice onset time vary from language to language, but this variation is so far from random in nature that we may speak of three general phonetic types from which the categories of a particular language are selected. It would seem that such features as voicing, aspiration and force of articulation are predictable consequences of differences in the relative timing of events at the glottis and at the place of oral occlusion.

Department of Linguistics
University of Pennsylvania
Philadelphia, Pa. 19104

Haskins Laboratories
305 East 43rd Street
New York, N.Y. 10017