

SOME EFFECTS OF CONTEXT ON VOICE ONSET TIME IN ENGLISH STOPS*

LEIGH LISKER** and ARTHUR S. ABRAMSON***

Haskins Laboratories, New York

Recent work has led us to the conclusion that the English stop categories /bdg/ and /ptk/ are distinguished by the timing of changes in glottal aperture relative to supra-glottal articulation. In word-initial position, the environment of current interest to us, this is manifested acoustically by voice onset time, that is, the time interval between the burst that marks release of the stop closure and the onset of quasi-periodicity which reflects laryngeal vibration. For citation forms of words this measure of voice onset time completely separates the two phonemic categories. In running speech, however, the separation is less sharp; there is some overlap along the dimension of voice onset time. We have examined running speech in some detail to discover the extent to which certain contextual features are responsible for this overlap. It is clear that the presence of a voiceless stop in a stressed syllable makes for a greater lag in the onset of voicing. In unstressed syllables, an environment of high contextual redundancy as well as low functional yield for the phonemic contrast, there is considerable reduction of the distinction along the dimension. A further increase in voicing lag is noted in syllables bearing the final sentence stress. In addition, the farther such a syllable is from the end of the sentence, the less the effect is likely to be. The importance of voice onset time continues to be apparent, even in running speech, although other effects of context remain to be explored.

VOICING AS A PHONETIC DIMENSION

Among the dimensions which the phonetician finds useful in organizing his description of speech sounds none has a more prominent place than voicing. From what is known of the structure and functioning of the vocal tract it seems reasonable to assert that for any configuration of the supra-glottal cavities the larynx is free either to generate voice or to remain mute¹. Moreover, there is good reason to believe that the quite different acoustic consequences of the two modes of laryngeal operation are identifiable by ear with great ease (Miller and Nicely, 1955; Wickelgren, 1966). But while it may be possible to produce and accurately perceive the voiceless homorganic

* *An early version of this study was presented before the Thirty-Ninth Annual Meeting of the Linguistic Society of America, New York, December 28-30, 1964. A substantial portion of the work was supported by a grant from the National Institute of Child Health and Human Development.*

** *Also, the University of Pennsylvania.*

*** *Also, Queens College, City University of New York.*

¹ *Of course, it does not follow that for all cavity configurations both laryngeal modes will result in speech sufficiently audible for normal communication.*

complement of *any* voiced sound, it is at the same time true that voicing as a *distinctive feature* is more often found associated with stops than with other kinds of speech sounds. It is, as a matter of fact, not at all easy to find a language in which voicing is distinctive for sounds other than stops and fricatives, that is, sounds characterized by closure or severe constriction and the build-up of considerable air pressure in the oral cavity.² It is, in particular, most uncommon to find voice playing an unambiguously distinctive role in the case of sounds produced with minimal constriction. (Vowels, it has been said more than once, are "normally voiced".) Whatever the reasons may be for this apparent affinity between stops and distinctive voicing, it has led us to focus our attention on that class of speech sounds when looking for acoustic correlates of the voicing dimension.

Auditorily the presence of voice is perceived as buzz emanating from the larynx. The acoustic feature regularly associated with this buzz consists of pulses occurring at fairly regular intervals, giving rise to a sensation of pitch. Our concern is to show how well a simple measure of these pulses serves to separate the English stops into a more voiced set /bdg/ and a less voiced set /ptk/.

Physically speech sounds may differ continuously in two different ways with respect to voicing: in the *intensity* of the voicing pulses and in the *duration* of pulsing relative to the supraglottal events which characterize the sounds⁴. For the stops, if not for other speech sounds, it is relatively easy to define measures of voicing that are primarily durational and can be precisely fixed in time relative to other articulatory features. The reason for this is that acoustically the interval of closure for a stop is fairly well marked, particularly as to when the occlusion is broken. On spectrograms this point is very clearly shown by the presence of a "burst" or brief interval of high-intensity noise. The onset of this burst can be fixed with considerable certainty, and thus provides a convenient reference point for making measurements of voicing duration. However, stops in different contexts show bursts of varying prominence: in initial position they are very easily detected, but in final position there may well be no burst at all (in the case of stops without oral release). Particularly in initial position, then, is it feasible to relate the degree of voicing of a stop to the time relation between the burst and the onset of pulsing. This measure of voice onset time (hereafter simply VOT) has been applied to a broad sampling of prevocalic word-initial stops to provide the data for the present study.

² Languages in which laterals, nasals or vowels are said to utilize voicing distinctively do exist, e.g. Cheyenne (Davis, 1962).

³ E.g., Jakobson, Fant and Halle (1952, 2.333) and Laziczius (1961, p. 70). We might hazard the guess that for languages in which stop voicing is not distinctive, the stops are "normally voiceless." Considerations of this sort might be thought relevant to the question of "marked" versus "unmarked" features in phonology.

⁴ This point has been made before. See, e.g., Schmitt (1947, p. 166) and Fischer-Jørgensen (1963).

In choosing the time dimension as the basis for describing the degree of voicing of a stop we cannot entirely ignore the intensity dimension. From spectrographic examination it is not always easy to decide whether or not pulses of glottal origin are present in a given stretch of speech, and this fact makes it necessary to set some intensity threshold value as a criterion for identification. Such a threshold may be set to reject only the pulses present in the background noise that is bound to intrude in any recording ("hum"), or it may be set high enough to reject pulses originating at the glottis that are below the threshold of audibility at some reasonable distance from the signal source. While each kind of threshold setting is of course appropriate to a particular class of questions within the general field of experimental phonetics, we have chosen the one which matches the psycho-physical threshold⁵.

Our choice of the VOT measure as an appropriate index of degree of voicing does not mean that the initial stops of English differ only in respect to the feature of pulsing. There are other acoustic features associated with degree of voicing that we might have selected for primary attention. Indeed, we need not have considered voicing at all, but instead have chosen to base our search for acoustic cues on either of the impressionistic phonetic dimensions of aspiration and articulatory force as the primary basis for distinguishing the two categories of initial stops. One compelling reason for our selection of voicing and its primary acoustic concomitant, pulsing, is that, of the other two phonetic dimensions, aspiration has a contrastive value limited to particular contexts, while "articulatory force" has no agreed-upon physical meaning. Differences in VOT, on the other hand, are readily measured and appear to distinguish the two sets of English stops in all positions of contrast. Moreover, we can plausibly consider *all* the acoustic features, and the auditory phonetic ones as well, to be interrelated effects of a single underlying mechanism⁶. Although much is still unknown about the physiology of phonation, it is possible to make certain inferences as to how the action of the larynx in conjunction with supra-glottal events generates the various acoustic features which characterize the initial stops of English. We may suppose, to begin with, that the initiation of speech involves the development of a sub-glottal overpressure and a subsequent movement of air through the glottis into the supra-glottal cavities. If the glottis is initially open, as it is when resting, then there will be little hindrance to the flow of air. At some point during normally phonated speech, however, the speaker closes down the glottis and, given a suitable balance of aerodynamic forces and muscular tensions, the vocal folds will begin to vibrate. This shift in degree of glottal opening and mode of laryngeal operation occurs more or less in step with a change in supra-glottal articulation, from the closure phase of the stop

⁵ *This choice was dictated by our interest in the perceptual properties of acoustic features in the speech signal, as well as in the relation of such features to articulation.*

⁶ *This matter has been discussed at some length in Lisker and Abramson (1964). Rothenberg (1966) presents a comprehensive account of stop production.*

to the progressively more open tract of the following vowel. The acoustic consequences of this combination of laryngeal and oral gestures depend very much on their relative timing. Laryngeal vibration provides the near-periodic carrier that we call voicing. Voicing yields harmonic excitation of a low-frequency band during the interval of oral closure, and of the full formant pattern after release of the stop occlusion. Should the gesture of glottal closure and the onset of pulsing be delayed until some time after the release, however, then there will be no signal preceding the release, but instead there will be an interval between the release and voicing onset when the relatively unimpeded air rushing through the open glottis provides the turbulent excitation of the voiceless carrier that we know as aspiration. This aspiration phase is accompanied by considerable attenuation of the first formant⁷, an effect presumably due to the presence of the tracheal tube below the open glottis.

In a recently published study (Lisker and Abramson, 1964) we reported extensive data on the voicing of word-initial stops in a variety of languages that differ considerably both in number of stop categories and in the phonetic features said to distinguish them. These data indicated that the VOT measure can serve as a very effective basis for distinguishing physically between homorganic categories, but they also suggested the need for a closer look at the individual languages than our cross-language survey afforded. In the present paper we shall begin with a review of the previously presented data for American English, and then go on to new data that give a more detailed picture of the relations between stop voicing and certain other features of that language.

ISOLATED WORDS

Measurement procedure

The great bulk of our VOT measurements, both those previously presented and those made more recently, are derived from spectrograms prepared on a modified Kay Sonagraph⁸ set for wide-band analysis. VOT determinations were made with a precision that permitted rounding to the nearest five milliseconds, and values thus determined will be presented as frequency distributions in which the observed VOT range is divided into ten-millisecond intervals. In stating the VOT value of a stop we have also

⁷ This feature of first-formant "cutback" has been extensively studied as a cue to the voicing contrast in English initial stops (Liberman, Delattre and Cooper, 1958).

⁸ The standard drum was replaced by one of larger diameter, which provided a time scale of 7.5 in./sec. instead of the usual 5 in./sec.

adopted the convention of assigning a value of zero to the instant of burst onset, while *negative* numbers are used to indicate the number of milliseconds by which voice onset precedes or *leads* the stop release, and *positive* numbers represent the time interval by which voice onset *lags* behind release. Thus by our convention VOT values vary directly with degree of *voicelessness*⁹.

Findings

Our first body of data (Fig. 1)¹⁰ consists of VOT values for the initial stops of some five hundred words uttered as isolated items by four speakers. These words, part of a much larger set that had been chosen so as to minimize the chance that the speakers would give particular attention to the stops, consisted almost entirely of monosyllables¹¹.

For the population investigated a number of generalizations seem warranted. First of all, the members of each homorganic pair of phonemes occupy distinctly different ranges along the VOT dimension; hence the measure provides a sufficient basis for identifying any stop of a *specified place of occlusion* as a member of either the /bdg/ or the /ptk/ set. Secondly, we note that the position of the boundary between homorganic categories is partially dependent on the place of stop occlusion, and this dependence is such that for all stops in which voicing does not lead release the VOT value tends to increase as the place of closure moves back in the mouth¹². Thirdly, it appears that while /ptk/ have distributions that are essentially unimodal, /bdg/ show values that fall into two discontinuous ranges, with modes at about -100 msec. and near

⁹ This convention does not of itself imply that we assign some priority to one voicing state or the other, as is done by those linguists, e.g., Hockett (1955, pp. 166 - 167), who distinguish between "marked" and "unmarked" features on certain phono-distributional grounds. At the same time, it seems to us that a case can be made for regarding the glottis as normally closed in speech, from which it would follow that a positive opening gesture is executed in the production of a voiceless stop or fricative.

¹⁰ These data are adapted from Lisker and Abramson (1964, Figs. 2 - 4).

¹¹ This selection of word types was not deliberate. Of the thousand most frequent words listed in Thorndike and Lorge (1952), 66% of those with initial stops are monosyllables and 18% are dissyllables with stressed first syllables.

¹² This relation is not confined to English (Lisker and Abramson, 1964, p. 399), so its basis is presumably to be sought in acoustico-physiological considerations rather than linguistic ones. One possibly relevant factor is the tendency for labial stops to be released more rapidly than the others (Fant, 1960, p. 199).

Voice Onset Time in English Stops

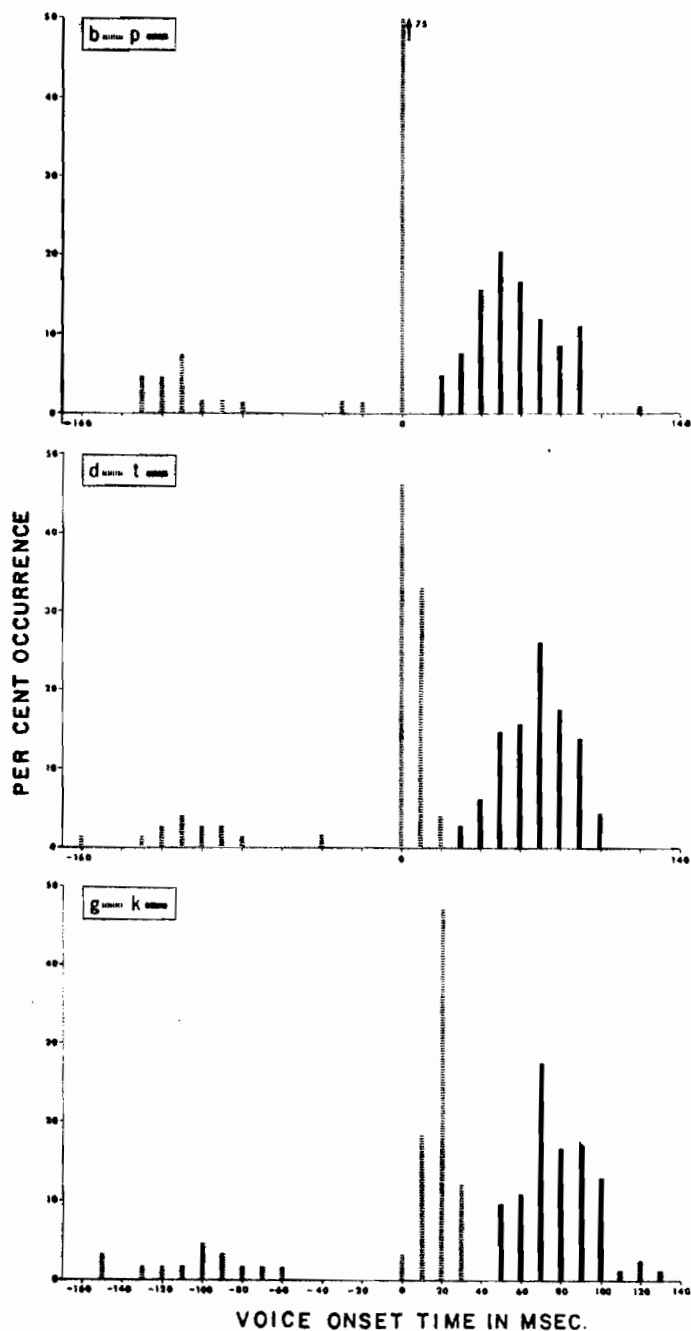


Fig. 1. Isolated words (four speakers)

	/b/	/p/	/d/	/t/	/g/	/k/
Av.	-101 1*	58	-102 5	70	-88 21	80
N.	17 51*	102	13 63	116	13 53	84

* Two sets of figures for average VOT and the number of tokens are given for /bdg/ to avoid grouping positive and negative values as items of a single continuous population.

zero¹³. These findings all have a certain theoretical interest. The first two bear on the general question of the mutual independence of phonetic dimensions; the last illustrates the point that a physical measure whose linguistic relevance is established to the extent that a phonemic contrast is reflected in the distribution of its observed values, cannot in turn serve as the basis for inferring the existence of a phonemic contrast.

The data of Fig. 1 may be taken as conclusive evidence that the VOT measure provides by itself a sufficient basis for the physical separation of /b/ and /p/, /d/ and /t/, and /g/ and /k/, when these stops occur in the specific contexts we have been considering. But these data can only be taken to demonstrate the relevance of our measure for *analysis*, that is, for the physical characterization of the stops as linguistic elements. In no sense can they be equated with data from experiments in *synthesis*, for only data of the latter kind could show the perceptual significance of the VOT dimension. Since both aspects of speech are of interest to our research group, such experiments have in fact been conducted (Liberman, Delattre and Cooper, 1958; Abramson and Lisker, 1965), and they provide complete assurance that the observed variability along the VOT dimension is directly related to listeners' judgments of initial stops as /bdg/ or /ptk/.

Useful as the VOT measure may be for identifying stops that begin isolated words, we have still to determine its value as a means of identifying stops in running speech. The burden of the present paper is concerned with the extension of our measurement procedure to word-initial stops in utterances longer than a single word, and with our search for factors other than membership in particular phonemes that help determine the timing of voice onset.

SENTENCES

Measurement problems

When word-initial stops are at the same time initial in running speech, it is in general no more difficult to determine their VOT values than in the case of isolated words. When, however, such stops occur in other positions within running speech, the measurement of VOT may present certain problems. Even in rather short sentences not embodied in a longer discourse, isolating the stop sounds in spectrograms must be done with considerable care, since wide amplitude variations may obscure some of the acoustic details.

¹³ In other words we have two phonetic categories of /bdg/ rather than a single one, as we might have expected to find. In fact the relation between the two distribution modes for each member of the /bdg/ set is not overtly different from the relation between phonemically distinct categories in certain other languages (Lisker and Abramson, 1964, pp. 400 - 402). We may speculate that the two varieties of English /bdg/ are alike characterized by early onset of glottal narrowing, and that pulsing is a secondary feature which may or may not begin before release. We may further suppose that more detailed information about the operation of the larynx will make clear why /bdg/ show two preferred ranges of VOT values separated by a region for which very few cases are recorded.

Occurrences of prevocalic /bdg/ in non-initial position in sentences can usually not be measured for VOT when voicing is present in segments preceding the stop closure. The voicing of such a preceding environment tends to continue unbroken into the closure interval. Moreover, in measuring VOT for /ptk/ in the same environment, there may be certain complications. Most of the time, especially when the stops initiate stressed syllables, voicing is interrupted and the usual voicing lag is observed. But sometimes, during the lag period, faint vertical striations at glottal rates are seen near the baseline of the spectrogram below the clearly aperiodic high-frequency noise of aspiration. In an earlier discussion of this problem (Lisker and Abramson, 1964, pp. 416-418) we told how we were unable to amplify these vibrations enough to make them audible and suggested that "the laryngeal oscillations of the 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 name 'edge vibrations' for this hypothesized behaviour." In that study we adduced various kinds of evidence, direct and indirect, to buttress this reasoning. Recent investigations based on trans-illumination of the larynx lend further support (Abramson, Lisker and Cooper, 1965, p. 611; Lisker, Abramson, Cooper and Schvey, 1966).

In connection with the present study, we have tried to learn how close to auditory threshold typical specimens of edge vibration come. Since these vibrations normally appear to be rather sinusoidal, i.e., they consist essentially of a single harmonic, our approach was to compare the intensity of a specimen of edge vibration with the intensity of a just audible pure tone of the same frequency¹⁴. Representative instances of edge vibration were found in the /k/ of the word *Carl* in the expression "Couldn't Carl . . ." and the word *catching* in ". . . from catching . . .", both taken from sentences uttered by one of the male speakers used in this study. For three listeners tested with random presentations of pure tones of varying intensity, edge vibrations during the stop closures were from 1 to 8.5 db. above auditory threshold. This finding is not in itself damaging to our earlier assessment of edge vibrations, since it is only maintained that a break in audible pulsation is crucial and not that it need coincide with all aspects of the stop. This expectation is confirmed by the finding that edge vibrations observed during the *voicing lag phase* of these productions were from 2 to 20 db. below auditory threshold.

It should be pointed out that certain shortcomings in this little experiment that were tolerated for the sake of methodological simplicity almost certainly biased the results

¹⁴ Oscillograms of the speech samples and comparison tones were made on a Honeywell Regulator Co. Visicorder, 1508. The pure tones were made with a Hewlett-Packard 200 CD oscillator, and intensity values were read from a Hewlett-Packard 400 D vacuum-tube voltmeter. The speech signals and the tones were fed through Allison variable filters, Model 2AB, to either the Visicorder or a pair of Permoflux PDR-1 headphones. For the threshold determinations the intensities of the tones were regulated by means of a Hewlett-Packard 5-watt 600-ohm attenuator set, Model 350B. We are indebted to Mr. David Speaker who designed the set-up for us.

toward assigning more audibility to edge vibrations than they really have. (1) The test tones were sustained, while edge vibrations occur in short spurts. (2) The test tones were presented at a constant amplitude, whereas edge vibrations represent a sudden drop in level between two rather high levels. (3) The test tones were presented in the clear, while edge vibrations, at least those observed after the release of a stop, are accompanied by the turbulence of aspiration. In all, the results of this simple experiment support our earlier conclusion that the weak glottal pulses sometimes observed instrumentally in non-initial occurrences of English /ptk/ are inaudible at least for the period from the release of the stop to the onset of normal voicing, i.e., the voicing-lag phase.

Comparison of isolated words and sentences

The four speakers who provided our sample of stops in isolated words also produced a number of sentences (of between eight and fifteen syllables each) containing a total of 800 word-initial stops in a variety of contexts. The VOT distributions for these stops are given in Fig. 2, from which it is evident that the relations between homorganic categories are not quite as they were found to be for stops in isolated words. To be sure, there is still a basis for saying that a significant correlation exists between magnitude of voicing lag, for both /ptk/ and /bdg/, and closure place along the front-to-back dimension. Moreover, /bdg/ still show divided distributions; the relative paucity of instances of voicing lead is only apparent, since fully 50% of the recorded /bdg/ are cases where voicing was continuously present over an interval that included segments preceding the stops. Such cases of "unbroken voicing", for which no VOT values can be specified by our measurement procedure, are undoubtedly to be classed together with the /bdg/ showing voicing lead that are reported in Fig. 2. (These latter are about equally divided between items in absolute sentence-initial position and non-initial cases characterized by a break in voicing somewhere before the release.) What is very clearly different about the distributions of Fig. 2 is that it is no longer possible to completely separate contrasting homorganic categories solely on the basis of the VOT measure; their values along this dimension show ranges of overlap. If we compare the average VOT values for each stop category (see legends to Figs. 1 and 2), we find that the time intervals by which voice onset either leads or lags behind release are almost always significantly shorter in sentences than in isolated words; in fact, the magnitude of the differences between homorganic stops without voicing lead is of the order of 60 msec. for the case of isolated words, while in sentences these differences are only half as large. Nor can the overlap between contrasting categories that we see in Fig. 2 be attributed to a greater variability in the production of stops in sentences, since these show, if anything, *less* variability in the timing of voice onset than do the stops of isolated words.

The finding that, with respect at least to VOT, the stops of isolated words are better separated than those in longer utterances, is not very surprising, for it confirms the general conviction as to a distinction between "spoon-fed" and "normal" speech. It is not immediately obvious, however, as to how this finding should be interpreted.

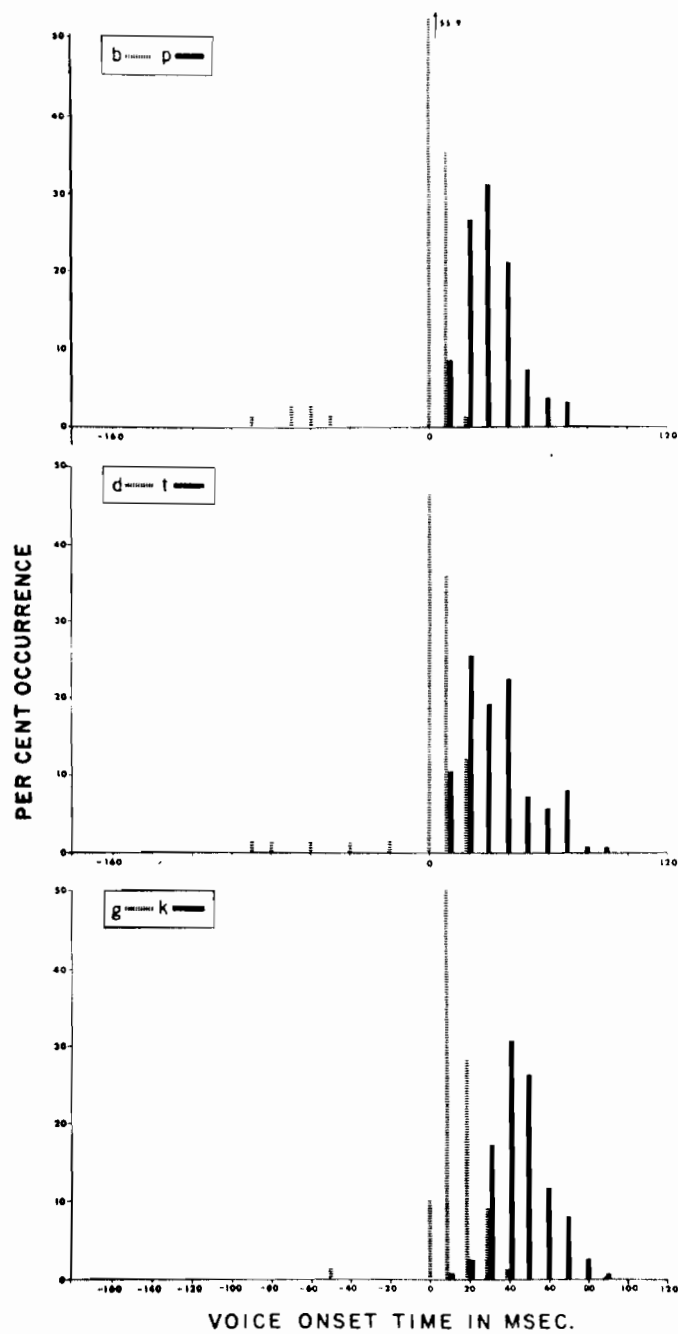
Voice Onset Time in English Stops

Fig. 2. Sentences (four speakers)

	/b/	/p/	/d/	/t/	/g/	/k/
Av.	-64 5	33	-56 8	38	-45 16	48
N.	6 71	110	5 79	125	1 77	163

The apparent reduction in the usefulness of the VOT measure as a sufficient basis for identifying stops with respect to their voicing state can be understood in two rather different ways: (1) as an indication of some general "blurring" of distinctiveness that may be inevitable at the higher articulation rates of the longer utterances, where linguistic constraints will take up any "slack" in intelligibility (Lieberman, 1963; Pickett and Pollack, 1963; Fry, 1964); or (2) as evidence for the existence of other acoustic features, whose contribution to intelligibility may be redundant for deliberate speech but indispensable at higher rates of transmission. Concerning the first possibility we can say very little, for there is as yet very little information available on the subject of phoneme as distinct from word intelligibility in utterances longer than monosyllables. It is our personal impression¹⁵, however, that stops whose measured VOT values fall in the zone of overlap may be ambiguous with respect to the /bdg/-/ptk/ contrast. If, on the other hand, it is the second possibility that is generally realized in speech, i.e., that a stop indeterminate with respect to the VOT dimension shows no loss in intelligibility, then the acoustic features which provide the substitute cues by which such stops are identified remain to be determined. The question as to which of the two possibilities mentioned corresponds to the truth is not, however, one that will be considered in the present discussion. Instead, our concern here is to look for correlations between certain contextual factors and VOT values, in order to determine whether the English stops are as categorically distinct for any specified context as we have found them to be in isolated words.

CONTEXTUAL VARIABLES

Further sentence data

As a necessary preliminary to studying the factors most likely to be responsible for the differences between the data of Figs. 1 and 2, we greatly expanded the set of sentences originally used. The next set provided a much more comprehensive variety of contexts in which to measure stop voicing. Furthermore, in order to check on the reliability of our four speakers of American English, who, as it happened, were all students of linguistics, we collected recordings of our new sentences from ten additional speakers, most of them linguistically naive persons. The new data thus obtained are represented in the graphs of Figs. 3.1 - 3.3. They are essentially in agreement with the data of Fig. 2. Perhaps the only difference, and it is a very small one, is that the new group of speakers has rather longer intervals of voicing lag for both /ptk/ and the instances of /bdg/ with positive VOT values. A close comparison of Fig. 2 with

¹⁵ *Listening carefully to our recordings, we found that certain stops, although they sounded natural enough, could not be judged unequivocally as /bdg/ or /ptk/. We have since collected samples of spontaneous speech (radio interviews and round-table discussions) with a view to excising portions with which to see how well naive listeners can identify the stops when there is insufficient context to identify the host word or morpheme.*

Voice Onset Time in English Stops

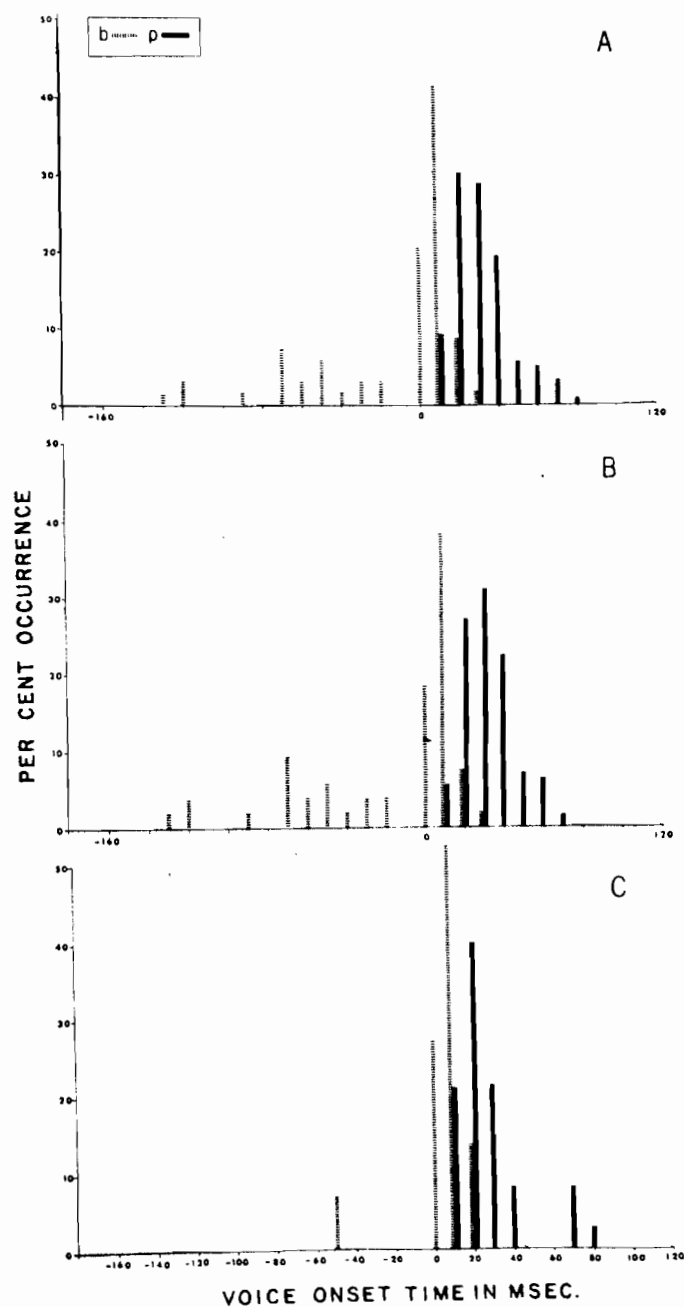


Fig. 3.1 Labial stops in sentences (ten speakers)

	/b/	/p/
(A) General :	Av. -61 10	34
	N. 20 50	168
(B) Stressed :	Av. -62 10	35
	N. 19 36	130
(C) Unstressed :	Av. -50 10	34
	N. 1 14	168

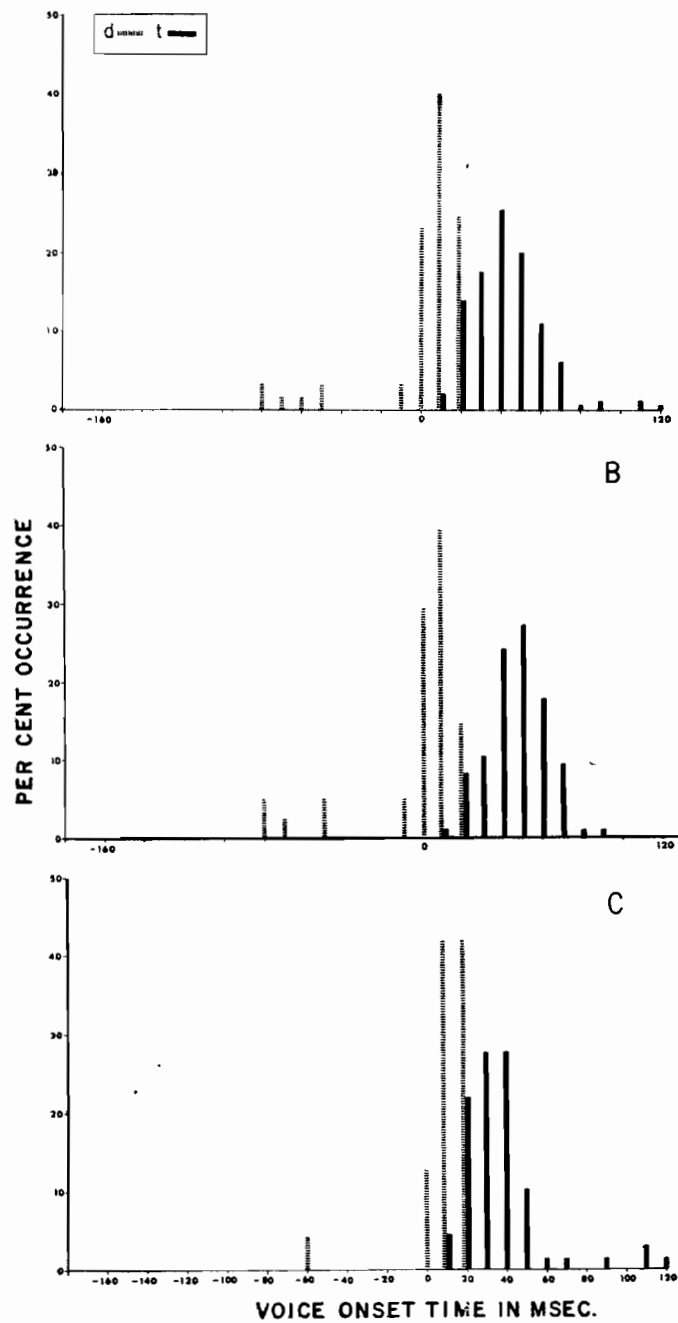


Fig. 3.2 Alveolar stops in sentences (ten speakers)

	/d/	/t/
(A) General :		
Av.	-50 12	45
N.	8 57	165
(B) Stressed :		
Av.	-49 10	48
N.	7 34	96
(C) Unstressed :		
Av.	-55 15	40
N.	1 23	69

Voice Onset Time in English Stops

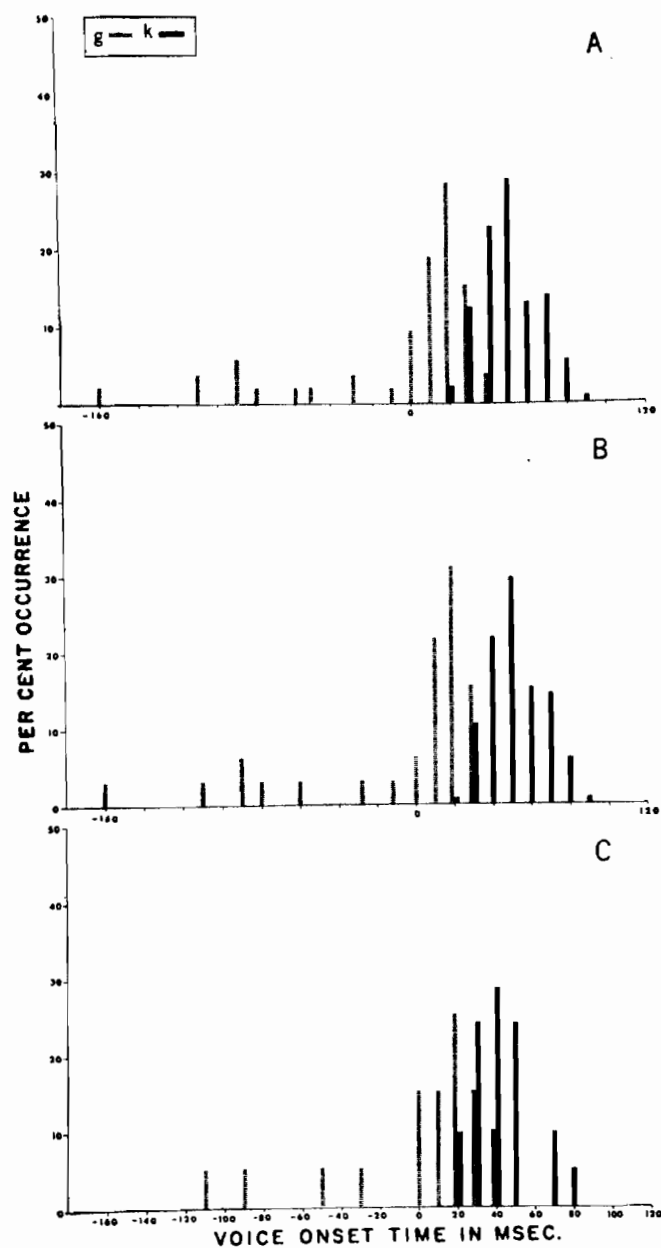


Fig. 3.3 Velar stops in sentences (10 speakers)

	/g/	/k/
(A) General :	Av. -73 20	53
	N. 12 40	144
(B) Stressed :	Av. -77 20	55
	N. 8 24	123
(C) Unstressed :	Av. -66 21	45
	N. 4 16	21

Figs. 3.1A, 3.2A and 3.3A strongly suggests that the data derived from our linguistically trained speakers and the other speakers form a single homogeneous population. If in fact there are differences between the two groups that should be taken into account, they are not gross enough to be apparent here.

Non-factors

Three variables may be mentioned in order to be eliminated from further consideration as factors controlling the time of voice onset. They are (1) initial *vs.* non-initial position, (2) utterance tempo, and (3) vocalic environment. The first variable was eliminated by inspection of the data for the fourteen speakers used in the present study, for these data show that it makes no difference whether a stop is in absolute initial position or elsewhere within a sentence, except insofar as non-initial /bdg/ may be characterized by unbroken voicing. The second variable was ruled out by the finding that, for a fairly broad sampling of the utterances of two of our speakers, no correlation could be established between VOT values and rate of syllable production. To test for the third variable, the same two speakers recorded a total of 108 tokens of /bdg/ and the same number of /ptk/ before twelve syllabic nuclei. Despite our feeling that any connection between VOT and vocalic environment would be most easily perceived in the citation forms measured¹⁶, no correlation between the two emerged from the data.

The three variables which have just been ruled out as significant factors governing the time of voice onset in stop production may indeed turn out, upon more exhaustive examination, to be not entirely irrelevant. What is already most certain, however, is that they are negligible in comparison with the variables to which we now turn attention.

Stress

One feature that we have good reason to believe significant for the timing of voice onset is stress. Phonetic descriptions of English all point out that the voiceless stops are invariably aspirated when they begin stressed syllables and unaspirated when they begin unstressed syllables that are non-initial within a word. Moreover, some descriptions add that in word-initial position the amount of aspiration in /ptk/ may vary considerably with degree of stress (Gimson, 1962, p. 146). Since we have reason to believe that the feature of aspiration is directly related to the timing of voice onset, it follows that we should expect differences in stress to be reflected in our VOT measurements.

¹⁶ *It must be admitted that this result does not rule out the possibility that a large sampling of running speech, well balanced for vowels following stops, would reveal that this "non-factor" is indeed present as a redundant feature. Conceivably the rather deliberate speech likely to be used in citation forms washes out any such effect by enhancing the status of VOT as the primary cue and does not, contrary to our expectation, provide maximum opportunity for an effect of vocalic environments to appear.*

In order to determine whether VOT values can be correlated with stress differences we divided the stops represented in the A displays of Figs. 3.1 - 3.3 into a stressed and unstressed class whose distributions are shown in the B and C displays respectively of the same figures. For the most part, our "stressed" class comprises syllables said to bear "primary" or "secondary" stress in the widely used Trager-Smith analysis (Trager and Smith, 1951), together with a few instances of "tertiary" stress; all others are "unstressed". Indeed, although many levels of loudness (or any other perceptual attribute of stress) are discriminable, it seems to us that the linguistic relevance of anything more elaborate than a two-level system is at best marginal (Lieberman, 1965, pp. 50 - 51, 53). Although there is, unfortunately, a considerable imbalance in the numbers of each category for the two stress conditions, our data are sufficient to show that stress and VOT are not strictly independent of one another. The nature of the relation is, however, rather different from the one which holds between VOT and place of articulation for stops beginning isolated words, since the division into stressed and unstressed varieties does not yield non-overlapping distributions for contrasting categories under either condition of stress. Close comparison of the B and C displays suggests a number of tentative generalizations: (1) stressed /ptk/ tend to be produced with longer delays in voice onset than do unstressed /ptk/; (2) those /bdg/ not characterized by voicing lead show a small contrary tendency, that is, they tend to have greater VOT values in unstressed than in stressed position; (3) there is a smaller incidence of /bdg/ with unbroken voicing in unstressed position. The third generalization is not evident in the figures themselves but in the underlying data. As a consequence of the just-mentioned tendencies, the separation between contrasting categories is less clearly defined for the unstressed than for the stressed varieties of stops.

Although the frequency distributions of Figs. 3.1 - 3.3 indicate a relation between VOT and stress that seems reasonably unambiguous, we must add that the effects of stress on VOT are rather limited; they involve chiefly /ptk/ and are moreover small as compared, for instance, with the differences between stops in isolated words (Fig. 1) and those in sentences generally (Fig. 2). Where the difference between mean values for /ptk/ in isolated words as against sentences is about 25 msec., the comparable difference between stressed and unstressed /ptk/ is only about 6 msec. Thus, it seems that stress differences alone are not sufficient to explain why the VOT measure is an adequate basis for completely separating /ptk/ and /bdg/ categories in isolated words but not in the longer utterances. It is, moreover, difficult to believe that matters would be greatly helped by replacing our simple binary classification of syllables as either stressed or unstressed with a more complicated one.

A full assessment of stress as a factor governing VOT is hampered by our failure to include items with unstressed initial stops in the word list which provided the data for Fig. 1. In order to repair this deficiency three of the speakers who provided the material for the sentence data were asked to record a list of isolated words containing both stressed and unstressed initial stops. For /ptk/ the mean values of VOT were those given in Table 1.

TABLE 1

Mean voice onset time in msec.: isolated words (3 speakers)

	STRESSED	UNSTRESSED
/p/	59	38
/t/	67	45
/k/	84	55

These values are noteworthy in several respects. First of all, they are in conformity with our earlier observation on the relation between place of stop occlusion and VOT. Secondly, the values for the stressed stops are in very close agreement with those given in the legend to Fig. 1, while the values for the unstressed stops fall generally within the range occupied by both varieties when found in sentences. Moreover, the differences between stressed and unstressed VOT values are quite close, being 21, 22 and 29 msec. for /p/, /t/ and /k/ respectively. Thus it appears that the VOT difference between stressed and unstressed /ptk/ is about four times greater in isolated words than in sentences (a figure of 24 msec. *vs.* 6 msec.). The relation between the stress difference and the word-sentence difference is perhaps made clearer if mean VOT values for /ptk/ are compared, as in Table 2.

TABLE 2

Mean voice onset time for /ptk/ (3 speakers)

	STRESSED	UNSTRESSED
Words	70	46
Sentences	41	35

From these figures it appears that both stressed and unstressed stops suffer a reduction in VOT as we go from words to sentences, but that, because the stressed stops are reduced by very much greater amounts than are the unstressed, the stress effect on VOT is itself very much reduced in the case of stops in sentences. It is also apparent from Table 2 that the large stress effect in the case of isolated words is almost entirely a matter of the overlong delay of voice onset in the production of the stressed stops.

The mean values for /ptk/ in isolated words that are given in Table 1 do not of themselves tell us anything about their relations with /bdg/ in similar contexts. We have already found, in our first body of data (Fig. 1), that stressed /ptk/ and

/bdg/ in isolated words occupy distinct ranges along the VOT dimension. From the sentence data of Figs. 3.1 - 3.3 we know that the two categories have overlapping distributions even in the case of the stressed varieties. It should be called to mind, however, that these displays indicate a much greater percentage of overlap than in fact there is, since data are shown only for those occurrences of /bdg/ that could be measured for VOT; for nearly all instances of non-initial /bdg/ laryngeal pulsing is unbroken and there is complete separation from /ptk/ along the VOT dimension. In addition, if our convention of measuring VOT from the point of stop release obscures the possible perceptual relevance of pulsing breaks that start during stop closure in non-initial /ptk/, the percentage of overlap is further exaggerated. It is also of interest to note that the unstressed /ptk/ of isolated words, although their mean values are nearly the same as those of the stressed /ptk/ of sentences, nevertheless are unlike the latter in that they occupy ranges of VOT values which are quite distinct from the ranges occupied by /bdg/.

In assessing the data just presented, we felt some uncertainty as to their being a representative sample of casual relaxed speech, for in listening to the recordings of our ten naive speakers we were struck by the lack of spontaneity of a number of them. In fact, it seemed to us that linguistically more sophisticated speakers might well do a better job of representing the normal productions of naive speakers than our naive speakers themselves had done. As a check on our findings so far, therefore, a new set of sentences was recorded by two of the linguists who had provided the data of Figs. 1 and 2. Because the measurements so far indicate quite clearly that /ptk/ are rather more context-sensitive than /bdg/, the new sentences were heavily loaded with /ptk/ in both stressed and unstressed positions. Their VOT values (Fig. 4), although somewhat smaller than those recorded in Figs. 3.1 - 3.3, are not dramatically different from the latter; with respect to the relation between stressed and unstressed varieties the two sets of data closely match one another. Their VOT values (Fig. 4) are in general somewhat smaller than those of Figs. 3.1 - 3.3, as can be seen by comparing ranges and means of the distributions for corresponding categories, and this difference may reflect the observed difference in speaking style between the two groups of speakers¹⁷. However, the differences between the two sets of data are not dramatic; in particular there is a very close match with respect to the relation between stress and VOT.

Position in the sentence

So far, then, it appears that stress has an effect on VOT that is most clearly observable in the /ptk/ stops, and that this effect makes for a somewhat greater separation between /ptk/ and /bdg/ categories in stressed positions than in unstressed.

¹⁷ This inference depends, of course, upon the validity of our impression that the two linguists did indeed speak in a more relaxed, "natural" manner than the ten speakers.

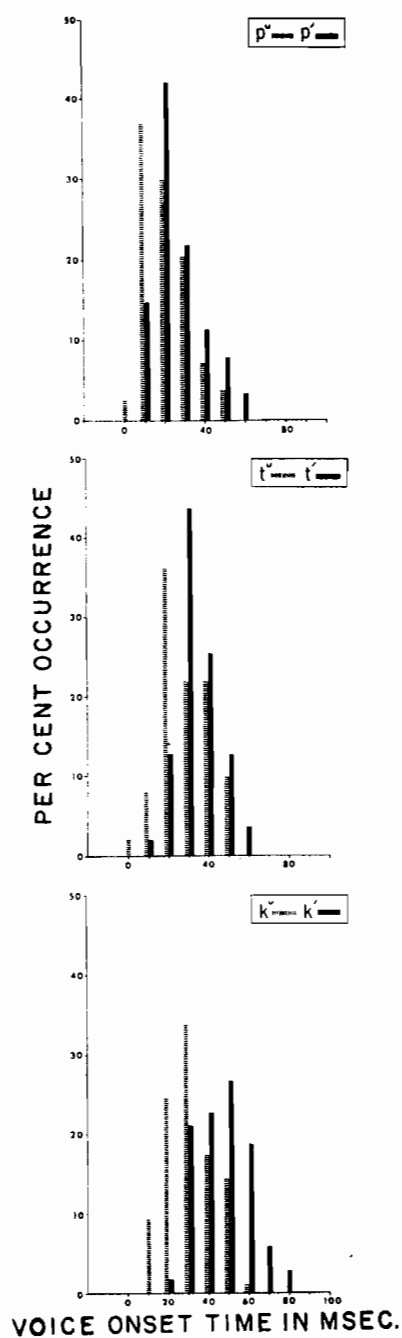


Fig. 4. Stressed (^l) and unstressed (°) /ptk/ in sentences (two speakers)

	/p°/	/p ^l /	/t°/	/t ^l /	/k°/	/k ^l /
Av.	23	30	32	38	33	50
N.	84	124	50	55	98	123

There still remains, even in the case of stressed position, a residual overlap between homorganic categories. In order to discover whether there are not other contextual factors important for VOT we had our two non-naïve speakers record still another set of sentences. This time we restricted attention to the single category /k/, using a limited number of words beginning with this phoneme and combining them in various orders to give reasonable English sentences. The measurements obtained from this new sample were analyzed in several ways. First of all, mean VOT values were calculated for stressed and unstressed varieties of /k/ as a check on the generality of our finding from the sentence set previously measured. These values were in close agreement with the figures for /k/ given in the legend to Fig. 4. Next we examined our new /k/ measurements to see whether VOT depended in any way on position within the sentence. With a single notable exception, there appeared to be no relation between our measure and the position of the stop, whether position is defined in relation to syntactic constituents or merely by reference to the number of preceding or following words or syllables in the sentence. This generally negative finding is not necessarily significant, since the variety of sentence types studied was limited. The exception is presented by /k/ initiating words in sentence-final position, which happened to be where the main sentence stress fell, in all the sentences of the set recorded. The mean VOT value for stressed syllables in that position is significantly greater than the mean for /k/ in stressed positions in the set overall: 59 msec. as against 48. For unstressed /k/, on the other hand, the mean value in sentence-final words is the same as the overall mean for unstressed position.

It is revealing to look at the data for initial /k/ in sentence-final words in yet another way. Fig. 5 shows the VOT distributions for (A) stressed monosyllables (e.g., *cold*), (B) stressed dissyllables (e.g., *curtain*) and (C) unstressed polysyllables (e.g., *canary*). (C) is no more different from (B) than (B) is from (A). That is, not only is the placement of the main sentence stress significant in controlling VOT values, but also the number of syllables following this stress appears to have an effect. This effect remains to be explored further with polysyllabic words of varying length with the stress on the first syllable.

Enhancement of the voicing difference

The matter of the bi-modal distribution of VOT values for /bdg/ has been discussed (see also Lisker and Abramson, 1964, pp. 394 - 395). Our sampling of speakers is not broad enough to enable us to decide whether a tendency on the part of a speaker to produce either of the two modes, lead or short lag, is to be explained as an idiolectal or dialectal feature of English. The phonetic literature sheds little light on the question. One aspect of the problem that seemed solvable with the available speakers was the question of whether speakers were likely to use one mode rather than the other under certain conditions. More specifically, if a speaker does not habitually show voicing lead in running speech, is he likely to use lead somewhat more often in more deliberate speech and thus enhance the /bdg/ : /ptk/ distinction? To test for this enhancement effect, we examined the utterances of the ten informants

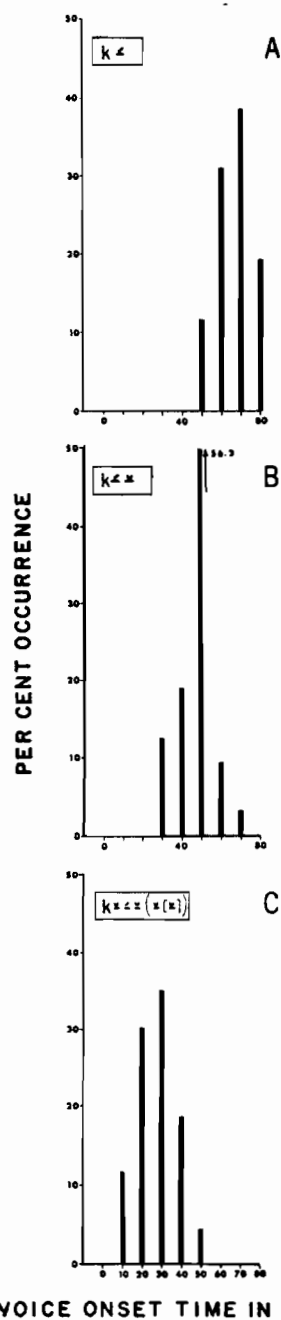


Fig. 5. Initial /k/ in sentence-final words (two speakers)

(A) Stressed monosyllables :	Av.	68
	N.	26
(B) Stressed dissyllables :	Av.	49
	N.	32
(C) Unstressed polysyllables :	Av.	31
	N.	43

who had supplied the data for Fig. 3 and chose two speakers who sometimes used voicing lead in /bdg/ : RH, 24% of the tokens, and LS, 5%. Although all the instances of lead were found in stressed syllables, these percentages take both stressed and unstressed voiced stops into account. This occasional use of lead suggested a potential for the enhancement of the distinction along the VOT dimension in more deliberate speech.

To provide citation forms for comparison with the sentence data, we had our two speakers record a randomized list of 70 words beginning with the six stops. These were mixed with fifteen words beginning with other consonants so as to keep the speakers from being overly aware of the stops.

Finally, we presented the two speakers with a list of pairs of words minimally distinguished by voicing in stops and gave them the following instructions : "This is a list of pairs of rhyming words. We wish to examine the differences between the members of each pair. Please speak clearly and pause after each word." Each pair of stops was placed before four vowels, making twelve minimal pairs of words. The twelve pairs were recorded with the voiced stops first and then with the voiceless stops first. This was done twice, yielding a total of sixteen tokens of each stop spoken by each speaker. It was thought that calling the speakers' attention to the stop contrasts in this way would provide an opportunity for maximum enhancement of the distinction.

For each of the three conditions, Table 3 shows the percentage of stops characterized by voicing lead. Since we could not find minimal pairs distinguished by initial unstressed stops, we eliminated observations for unstressed stops altogether in order to make the three conditions comparable.

TABLE 3
Voicing lead in stressed /bdg/ (2 speakers)

	FREQUENCY	NO. OF STOPS EXAMINED
Sentences	23%	30
Randomized Words	46%	97
Minimal Pairs	68%	96

The incidence of voicing lead clearly reveals an enhancement effect. Voicing lead is used considerably more often in isolated words than in sentences, and most of all in minimal pairs. The two speakers, however, differ in where they show the major enhancement effect. The data of Table 3 are broken down in Table 4 to show this difference.

TABLE 4
Voicing lead in stressed /bdg/

Speaker :	RH		LS	
	Freq.	N.*	Freq.	N.*
Sentences	38%	16	7%	14
Randomized Words	83%	48	10%	49
Minimal Pairs	88%	48	48%	48

* Number of stops examined

For RH the major break is between sentences and words, while LS mainly enhances the minimal pairs.

Another sign of enhancement would be a widening of the separation between the two categories along the VOT dimension. To test for this, we obtained VOT measurements for all the remaining stressed stops in the three sets, i.e., all the tokens of stressed /ptk/ as well as all the tokens of /bdg/ that did not have voicing lead. The average values are given in Table 5.

TABLE 5
Voice onset time in msec. (2 speakers)

	STRESSED /bdg/	STRESSED /ptk/
Sentences		
Av.	12	45
N.	23	71
Randomized Words		
Av.	17	74
N.	55	106
Minimal Pairs		
Av.	16	74
N.	31	96

It is clear that here too we have an enhancement effect. It distinguishes the two groups of words from the sentences. The average voicing lag of /ptk/ increases 1.6 times, so that, in spite of a slight increase of the voicing lag of /bdg/, there is considerable widening of the separation between the two phonemic categories in the isolated words as compared with the sentences. When taken separately, the two speakers show the same effect.

This explanation of enhancement effect suggests a need for caution in evaluating statements to the effect that voicing lead ("full voicing") is a normal mode in English word-initial /bdg/. It is very likely that a speaker who exhibits voicing lead in rather deliberate productions will not do so in running speech. Conversely, the speaker who seldom uses voicing lead in relaxed colloquial speech may start phonating before the release of the stop if he is made conscious of his speech behaviour or is asked to enhance the distinction between voicing categories. The ease with which these stops take on unbroken voicing in certain environments, as well as their lower burst intensity as compared with /ptk/, leads one to infer that even in productions of /bdg/ without voicing lead the glottis must be in a state that is nearly suitable for phonation.¹⁸ If this inference is reasonable, it is not surprising that the speaker can assert enough control over his larynx to bring phonation on before the release of the stop.

DISCUSSION OF RESULTS

In our study of word-initial stops in English the finding which deserves primary emphasis is that the two categories /ptk/ and /bdg/ are characterized by significantly different distributions of VOT values. This relation holds true quite independently of several contextual factors investigated. At the same time, however, the timing of voice onset relative to stop release is not entirely independent of certain of those factors, for we find regular differences between stops in isolated words and in sentences, as well as between stops which initiate stressed syllables and those beginning unstressed ones. This variability of voicing behaviour is, in the case of the /bdg/ category, chiefly a matter of whether or not the stop is immediately preceded by a voiced interval, for in such a context there is usually no break in pulsing and hence no VOT value to be associated with the stop. For those stops of both categories to which the VOT measure is applicable, contextual variability is mainly of differences within the /ptk/ category. These stops show maximum VOT values in stressed position in isolated words, minimum values in unstressed position in sentences, and approximately the same values in both unstressed position in words and stressed position in sentences. These differences, together with relatively minor ones noted for those /bdg/ with measurable VOT values, are reflected as differences in the distance between categories along the VOT dimension for the four contexts distinguished. Those contexts may be arranged, in descending order of magnitude of category separation, as follows: stressed in words, unstressed in words, stressed in sentences, and unstressed in sentences. For the first two contexts the extent of separation is such that /ptk/ and /bdg/ occupy non-overlapping ranges; for the third, and even more for the fourth, there is appreciable overlap between their VOT distributions. Despite

¹⁸ Our current work on transillumination of the larynx indicates so far that in productions of /bdg/ of this type the glottis tends to close during the occlusion, as if in anticipation of phonation.

these context differences it remains true, that not only are the mean VOT values greater for /ptk/ than for /bdg/ in the same context, but that no contextual variant of /ptk/ has a mean value as small as that of any variant of /bdg/.

The observation that a stop cannot always be categorized unambiguously on the basis of its VOT value can be interpreted in several different ways, and these need not be taken to be mutually exclusive. Thus it might be supposed that the insufficiency of our measure so far as running speech is concerned means that :

- (1) our analysis of stress, which admits of only two degrees, is too coarse to show that under precisely similar stress conditions /ptk/ and /bdg/ are adequately distinguished on the basis of VOT alone ;
- (2) the contrast between /ptk/ and /bdg/ is not perfectly maintained in running speech, particularly in unstressed position ;
- (3) the particular measure applied was not well chosen ;
- (4) no single acoustic feature can be expected to serve as a sufficient basis for the physical separation of linguistically distinct sound categories.

Of these four points the first two are questions as to the adequacy of our phonological description of English, while the latter two are more directly addressed to the question of interpreting the measurement data. It is worthwhile to take each point separately and, assuming it to be valid, to see just how it affects our estimate of the usefulness of the VOT measure as a correlate of the /ptk/ - /bdg/ contrast.

To suppose that our division of word-initial syllables into two classes with respect to stress implies that there are only two levels of prominence in English is to suppose what in fact was far from our intention ; the relation between phonetic prominence and linguistic stress level in English is certainly more complex. It is however hard to see how a finer stress analysis will entirely eliminate VOT overlap between contrasting members of the two stop phoneme categories. The class of unstressed stops represented by the data of the (C) displays in Fig. 3.1 - 3.3 includes only stops which initiate syllables that would be judged unstressed according to the Trager-Smith system, and it is just this class of stops for which the category separation along the VOT dimension is most poorly defined. On the other hand, if our stressed class were more finely divided, it might well turn out that for each of the new classes of stops the VOT overlap is less than that shown in the (B) displays of Fig. 3.1 - 3.3. Now in fact our data, specifically those given in Fig. 5, do provide a certain basis for this expectation, for it appears that stops which begin monosyllabic words bearing the primary stress in the sentence fall into two sharply separated ranges along the VOT dimension.

The lack of experimental data on the intelligibility of individual phonemes in running speech makes it difficult to exclude the possibility that the overlap between categories is matched by a corresponding perceptual overlap. This, in effect, raises questions as to the validity of our classification of the stops that were measured. Would in fact any carefully designed labelling test corroborate our phonemic assignments of stops whose VOT values falls in the overlap region ? If not, would such items be labelled in haphazard fashion, or would they be identified in a way consistent with their VOT values but independent of their morphophonemic (or perhaps merely orthographic)

status? If either of the latter questions were answerable in the affirmative, we should have to say that the /ptk/ - /bdg/ contrast is not fully maintained in running speech, but that this state of affairs is not inconsistent with the view that a VOT difference is essential to that contrast¹⁹.

The selection of the VOT measure was dictated in part by the fact that it is easily determined from spectrograms. Needless to say, high visibility in spectrograms is no guarantee of corresponding prominence in auditory perception. The possibility of finding a more useful correlate of the /ptk/ - /bdg/ contrast than the timing of voice onset cannot be excluded *a priori*. In fact, however, we have good evidence, from experiments in speech synthesis, to show that VOT differences are well perceived auditorily, and that they bear directly on the division of the English stops into two categories. The only other clearly relevant acoustic feature that has thus far been identified is the so-called "cut-back" of the first formant, but this feature should not be considered independent of VOT²⁰. As for other acoustic features, there is nothing in the literature to suggest that any one of them is likely to be more useful than the one we have chosen. Nor is the situation helped appreciably by asserting that the root of the difficulty is a faulty diagnosis of the contrast as voiced/voiceless rather than fortis/lenis, since no one has proposed an acoustic measure (or any other kind of measure, for that matter), correlative with the latter phonetic dimension, which is more useful than the one of VOT.

If we suppose that there is no acoustic measure more effective than the one we have chosen, and if that one is not sufficient for stops in running speech, then it seems appropriate to ask what the basis is for our expectation that the two stop categories should be distinguishable with respect to a single simple acoustic dimension. The answer to this quite possibly may be found in the failure to make a distinction

¹⁹ It is certainly of interest, and possibly significant as well, that the context in which overlap along the VOT dimension is most extensive, i.e., at the beginning of word-initial unstressed syllables, is also one in which the /ptk/ - /bdg/ contrast has an almost negligible functional yield. If the contrast is in fact vestigial in this position, we have a situation similar to that of the vowels in unstressed syllables: instability of contrast relationships and reduction of the physical distances between categories. This situation is very different from one in which there is a "hole in the pattern", whether in the inventory of phonemes generally or in their distributions. In such a case we find little tendency for categories to encroach on the physical space appropriate to the missing element. An example is provided by Thai, whose stop phonemes may be arrayed as follows:

b	p	p ^h
d	t	t ^h
	k	k ^h

The VOT values for /k/ are distributed precisely as we should expect them to be if Thai possessed a phoneme /g/. (Lisker and Abramson, 1964, p. 411.)

²⁰ We have made exploratory experiments in synthesis in which VOT and first-formant cutback were independently varied, and it seems very clear that in the simulation of normally voiced speech neither feature is alone sufficient to cue the contrast between the two stop categories. The situation is somewhat different in the case of synthesized whispered speech, where the feature of first-formant cutback is a sufficient cue in the absence of pulsing throughout the utterance.

between two rather different kinds of phonetic descriptions: the linguistic variety, which is essentially only precise and detailed enough to justify some phonological analysis, and one which treats exhaustively of all the features characterizing the utterance types of a language. The fact that many linguists, apparently on principle, reject the second kind of description *even as a goal* means that we are left with simplistic accounts of the phonetic basis for phonemic contrasts. These accounts, however adequate to the linguist's aim of justifying a maximally simple orthography that is not inconsistent with his auditory impressions, imply an economy in the number and variety of features signalling contrasts that is spurious. The acoustic phonetic literature shows very few cases where linguistically distinct categories are distinguished by no more than a single acoustic measure, even where analysis is limited to "citation speech" (i.e., nonsense syllables, isolated words, and the like), and certainly none for running speech. If the VOT measure fails to separate /ptk/ and /bdg/ even for the limited range of contexts examined within running speech, then it seems very possible that there exists in fact no single acoustic feature by which the category membership of any given stop may be determined independently of context. Instead, there is reason to believe that /ptk/ are invariably characterized by a gesture of laryngeal opening, and that such a gesture does not accompany the articulations of /bdg/²¹. This articulatory difference, which we take to be the distinctive feature underlying the contrast, may of course be associated in specific contexts with other features that can be called redundant. To say that the difference between the open and closed states of the glottis is the basis for the /ptk/ - /bdg/ contrast is not to imply, however, that the acoustic differences between the two categories also comprise one that is distinctive and a remainder that is redundant. All that is implied is that some one or another sub-set of the acoustic features which reflect opening of the larynx serves to distinguish /ptk/ from the other stops in each context in which this contrast is utilized in the language.²²

By way of conclusion we may return to the supposition expressed early in this paper, namely, that the phonetic differences among contextual variants of the English stop phonemes reflect differences in the relative timing of laryngeal and supra-glottal gestures. If it is granted that such timing differences are to some extent linguistically determined, it seems at the same time reasonable to ask whether there is not some physiological basis for certain of the variations in VOT that have been observed. Thus the greater values of VOT recorded for /ptk/ initiating stressed syllables may perhaps be due to the larger opening of the larynx in that context. If further investigations such as the transillumination study already referred to should establish a relation between size and duration of glottal opening and stress, then it seems reasonable to think that the greater sub-glottal air pressure which characterizes stressed syllables (Ladefoged, 1962) may play some purely mechanical role in determining the time of voice onset.

²¹ This dependence is discussed under "Voicing as a Phonetic Dimension" at the beginning of this article. See also fn. 7.

²² This is not to say that the size of the opening nor its shape need be the same in all environments.

REFERENCES

- ABRAMSON, A. S. and LISKER, L. (1965). Voice onset time in stop consonants : acoustic analysis and synthesis. 5e Congrès International d'Acoustique : Rapports (Liège), Ia, A51.
- ABRAMSON, A. S., LISKER, L. and COOPER, F. S. (1965). Laryngeal activity in stop consonants. *Haskins Laboratories Status Report*, 4, 6.1.
- DAVIS, I. (1962). Phonological function in Cheyenne. *IJAL*, 28, 36.
- FANT, G. (1960). Acoustic Theory of Speech Production (s'Gravenhage).
- FISCHER-JØRGENSEN, E. (1954). Acoustic analysis of stop consonants. *Miscellanea Phonetica II*, 42.
- FISCHER-JØRGENSEN, E. (1963). Beobachtungen über den Zusammenhang zwischen Stimmhaftigkeit und intraoralem Luftdruck. *Zeitschr. f. Phonetik*, 16, 19.
- FRY, D. B. (1964). The correction of errors in the reception of speech. *Phonetica*, 11, 164.
- GIMSON, A. C. (1962). Introduction to the Pronunciation of English (London).
- HOCKETT, C. F. (1955). A Manual of Phonology (Baltimore).
- JAKOBSON, R., FANT, C. G. M. and HALLE, M. (1952). Preliminaries to Speech Analysis (Cambridge, Mass.).
- LADEFOGED, P. (1962). Sub-glottal activity during speech. Proc. of the Fourth International Congress of Phonetic Sciences. *Helsinki*, 1961, 73 - 91.
- LAZICZIUS, J. (1961). Lehrbuch der Phonetik (Berlin).
- LIBERMAN, A. M., DELATTRE, P. C. and COOPER, F. S. (1958). Some cues for the distinction between voiced and voiceless stops in initial position. *Language and Speech*, 1, 153.
- LIEBERMAN, P. (1963). On the redundancy of speech and articulation and perception. *J. acoust. Soc. Amer.*, 35, 807 (A).
- LIEBERMAN, P. (1965). On the acoustic basis of the perception of intonation by linguists. *Word*, 21, 40.
- LISKER, L. and ABRAMSON, A. S. (1964). A cross-language study of voicing in initial stops : acoustical measurements. *Word*, 20, 384.
- LISKER, L., ABRAMSON, A. S., COOPER, F. S. and SCHVEY, M. H. (1966). Transillumination of the larynx in running speech. *J. acoust. Soc. Amer.*, 39, 1218 (A).
- MILLER, G. A. and NICELY, P. E. (1955). An analysis of perceptual confusions among some English consonants. *J. acoust. Soc. Amer.*, 27, 338.
- PICKETT, J. M. and POLLACK, I. (1963). Adjacent context and the intelligibility of words excised from fluent speech. *J. acoust. Soc. Amer.*, 35, 807 (A).
- ROTHENBERG, M. (1966). The Breath-Stream Dynamics of Simple-Released-Plosive Production (University of Michigan, Ph.D. diss.).
- SCHMITT, A. (1947). Die neuhochdeutschen Verschlusslaute. *Zeitschr. f. Phonetik*, 1, 149.
- THORNDIKE, E. L. and LORGE, I. (1944). The Teacher's Word Book of 30,000 Words (New York).
- TRAGER, G. L. and SMITH, JR., H. L. (1951). An Outline of English Structure (Norman, Oklahoma).
- WICKELGREN, W. A. (1966). Distinctive features and errors in short-term memory for English consonants. *J. acoust. Soc. Amer.*, 39, 388.