Role of formant transitions in the voiced-voiceless distinction for stops

Kenneth N. Stevens and Dennis H. Klatt

Research Laboratory of Electronics and Department of Electrical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts
(Received 2 August 1971; revised 1 November 1973)

Previous research on acoustic cues responsible for the voiced-voiceless distinction in prestressed English plosives has emphasized the importance of voicing onset time with respect to plosive release (VOT). Voiced plosives in English normally have a short VOT (less than 20–30 msec) and a significant formant transition is present following voice onset. Voiceless plosives in prestressed position, on the other hand, have relatively long VOT's (greater than about 50 msec) and the formant transitions are essentially completed prior to voice onset. Our experiments with synthetic speech compare the role of VOT and the presence or absence of a significant formant transition following voicing onset as cues for the voiced-voiceless distinction. The data indicate that there is a significant trading relationship between these two cues. The presence or absence of a rapid spectral change following voice onset produces up to 15-msec change in the location of the perceived phoneme boundary as measured in terms of absolute VOT. One can speculate that the auditory system may be predisposed to detect the presence or absence of a rapid spectrum change as a general property of acoustic inputs. If this is the case, then the acquisition of the voiced-voiceless distinction in infants may be conditioned initially by the presence or absence of this property at the onset of voicing rather than by absolute VOT.

Subject Classification: 70.30.

INTRODUCTION

The acoustic cues for the distinction between voiced and voiceless[1] stop consonants in initial prestressed position in English have been examined in detail through measurements on spoken consonant-vowel syllables (Lisker and Abramson, 1964), and through experiments in which the acoustic parameters of synthetic stimuli are manipulated (Liberman, Delattre, and Cooper, 1958; Lisker and Abramson, 1970). The data from the studies with synthetic speech have suggested that the acoustic characteristic providing the simplest and most direct indication of whether a stop consonant is voiced or voiceless is the time from the release of the closure to the onset of vocal-cord vibration, called the voice-onset time (VOT). If this time is greater than about 25–40 msec (depending on consonantal place of articulation), the consonant is voiceless; for smaller voice-onset times, the consonant is identified as voiced. The data from natural utterances show that the voice-onset times for the two classes of consonants in prestressed position are clustered around 0–20 msec and about 50 msec or more,[2] and intermediate voice-onset times are rarely found. Furthermore, identification and discrimination judgments for synthetic stimuli indicate that there is a sharp perceptual boundary between voiced and voiceless responses. Discrimination of pairs of stimuli within a given class is poor, while discrimination for stimuli near the phoneme boundary is good (Liberman, Harris, Kinney, and Lane, 1961).

Recent experiments on the discrimination of synthetic syllables by infants (Eimas *et al.*, 1971) have drawn further attention to the importance of the voice-onset time as a property which the auditory mechanism uses to categorize speech sounds into classes. The Eimas experiments suggest that infants cannot discriminate changes in VOT within a region that adults would classify as one phoneme, but can readily discriminate changes of similar magnitude that span a phoneme boundary.

A conclusion that one might make from all of these findings is that the auditory system is, as it were, "wired" to place a boundary at some fixed absolute noise duration. Study of a variety of phonetic features indicates, however, that the acoustic differences between minimal pairs of phonemes in a given phonetic environment appear to be characterized by distinctively different *properties* rather than by differences in the magnitude of a given acoustic parameter that takes on a range of values (Jakobson, Fant, and Halle, 1963; Stevens, 1972). As far as we know, there is no evidence to suggest that a 20- to 40-msec noise duration preceding a vowel-like sound represents a natural perceptual boundary.[3] Further experiments are needed to resolve this question.

Examples of spectrograms of the syllables /*d*ɑ/ and /*t*ɑ/ are shown in Fig. 1. The difference in VOT is clearly seen in the spectrograms. Another difference between the sounds, however, is that the voiced stop has a well-defined transition in the first formant (as well as the second formant) after the onset of voicing, whereas the F_1 transition for the voiceless stop is essentially nonexistent after the onset of voicing. The lack of formant transitions after voicing onset for the aspirated consonants indicates that the rapid movements of the supraglottal articulators (the tongue tip in the case of Fig. 1) are essentially complete before the vocal cords are in a configuration appropriate for the onset of voicing. Based on synthesis experiments (Liberman, Delattre, Gerstman, and Cooper, 1956), it is known, in fact, that the duration of these transitions is of the order of 40 msec or less.

These examples suggest that one of the possible cues

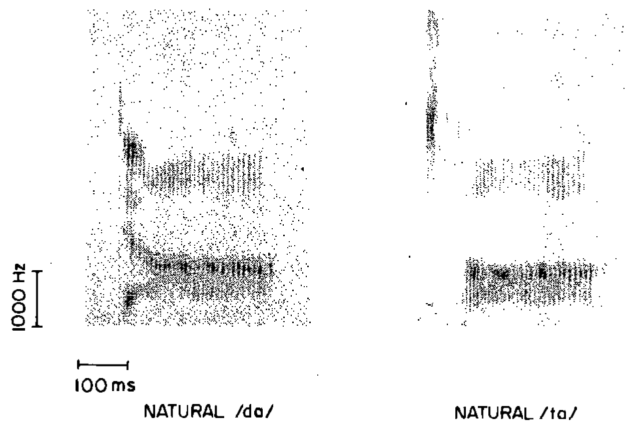

FIG. 1. Spectrograms of the syllables /$d\alpha$/ and /$t\alpha$/ as spoken in isolation by an English talker.

for the voiced-voiceless distinction for stops in pre-stressed position is the presence or absence of a significant and rapid spectrum change at the onset of voicing. According to this hypothesis, absence of such a spectrum change would be a requirement for perception of a well-formed voiceless stop. It is known, of course, that a rapid spectrum change always occurs at the *release* of a stop consonant. We are postulating here that, in the case of a voiceless stop consonant, another rapid spectrum change at the onset of voicing is a negative cue for voicelessness.

I. DESCRIPTION OF STIMULI AND EXPERIMENTS

In order to examine these questions, we have performed two experiments with synthetic consonant-vowel stimuli in which voice-onset times and transition durations were manipulated, and we have obtained listener responses to these stimuli. The stimuli were produced on a digital terminal-analog speech synthesizer (Klatt, 1972). The synthesizer contains sources for frication, aspiration, and voicing. The spectra of frication and aspiration are flat to 5 kHz and the spectral envelope of the periodic voicing source falls off at −12 dB/octave above 300 Hz. The synthesizer configuration includes five digital resonators connected in cascade to simulate the vocal-tract transfer function for vowels, another set of digital resonators for the fricative transfer function, and a radiation characteristic. The synthesizer produces 5-kHz bandwidth speech or speechlike sounds.

A. Experiment 1

This experiment was designed to determine whether the boundary between unaspirated and aspirated stop consonants at a VOT in the range 20–40 msec represents a characteristic of the auditory processing of acoustic stimuli independent of whether the stimuli are speech or nonspeech.

Stimuli for this experiment are schematized in Fig. 2. Each stimulus consisted of a brief 5-msec burst of broad-band noise followed by an interval of silence, and this was followed in turn by the onset of a synthetic vowel with fixed formants. The formants corresponded roughly to values appropriate for the vowel [ɛ]. The fundamental frequency of the vowel had a falling inflection from 135 to 100 Hz over its 265-msec duration. The level of the noise burst was adjusted to be a few decibels above the level of the vowel in the frequency region above 2000 Hz.

There were ten different experimental stimuli, corresponding to silent intervals T_G in 5-msec steps from 0 to 40 msec. None of the stimuli could be readily interpreted as speech events.

Several replications of the stimuli were arranged in random order and were presented to a group of listeners. The stimuli were described to listeners as consisting of a brief burst of noise and a buzz-like sound. The task of the listeners was to judge whether or not there was a silent interval between the burst of noise and the onset of the buzz.

Average data from four subjects are shown in Fig. 3. The abscissa represents the time from onset of noise burst to onset of buzz. (Since the duration of the noise burst is 5 msec, this time is 5 msec longer than the silent interval.) We shall call this the voice-onset time, even though the stimuli were not interpreted as speech by the listeners.

The data indicate that up to a VOT of about 15 msec, no silent interval could be heard. The listeners apparently regarded the onset of the noise burst and the onset of the buzz as being essentially simultaneous (although this question was not specifically asked of the listeners). For a VOT greater than about 25 msec, listeners usually reported the presence of a silent interval. The VOT at the 50% point is about 20 msec.

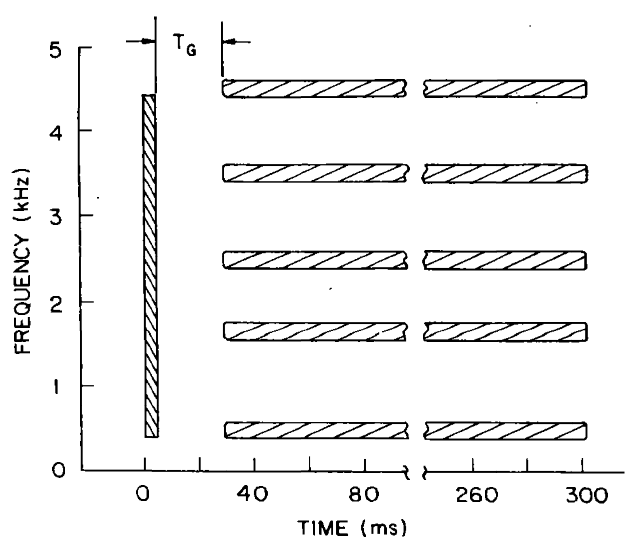

FIG. 2. Schematized representation of nonspeech stimuli used in Experiment 1. The initial vertical bar is a noise burst, and the horizontal bars represent formants in a vowellike sound. The silent interval T_G is varied in 5 msec steps from 0 to 40 msec.

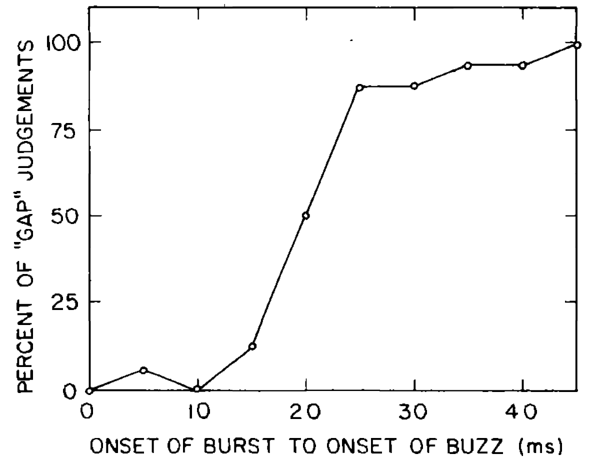

FIG. 3. Average percentage of times a silent interval was heard as a function of the time from burst onset to buzz onset for the stimulus ensemble shown in Fig. 2.

B. Experiment 2

A series of synthetic consonant-vowel stimuli were generated in which onset times and transition durations were manipulated independently, and listener responses to these stimuli were obtained. Sixteen stimuli similar to either /*tɑ*/ or /*dɑ*/ were produced by selecting a number of combinations of voice-onset time (VOT) and rate of formant motion. Each stimulus consisted of a brief 5-msec burst of frication, a variable interval of aspiration, and then voicing for the remainder of the utterance, as shown at the top of Fig. 4. Formant frequencies follow a time course shown at the bottom of Fig. 4. Formant bandwidths were set at constant values of 50, 80, 120, 180, and 300 Hz for the first through fifth formants, respectively.

Four formant trajectories are defined in Fig. 4, corresponding to moderate (a) through rapid (d) consonant-vowel transition times. Four values of voice-onset time were selected for each trajectory. All the transition times are within the range that yields a reasonable /*dɑ*/ stimulus for the shortest VOT's. Preliminary tests were performed in order to place the /*d-t*/ phoneme boundary within the range of voice-onset times for each trajectory. VOT values for the stimuli ranged from 15 to 55 msec.

Control parameters were updated at discrete 5-msec intervals. The fundamental frequency was held fixed at 130 Hz from voicing onset to 80 msec and fell linearly to 100 Hz from 80 msec to the end of the stimulus at 350 msec. The first voicing pulse in an utterance occurred precisely at the desired onset of voicing. The voicing amplitude was set to a constant value during the voicing interval and was off otherwise.

The frication burst spectrum was produced by exciting a single resonator set to the formant frequency and bandwidth values of the fifth formant, as is to be expected from analysis of natural speech (Stevens, 1968). Frication source amplitude was set to produce a fifth-formant level in the burst about 6 dB greater than the fifth-formant level during the following steady vowel.

The amplitude control for the aspiration source was set to a constant level during the aspiration interval and was off otherwise. The amplitude was set so that the excitation in the region of F3 and F4 remained at the same level in going from aspiration to voicing (Stevens, 1971), i.e., there was no discontinuity in level in this frequency region in passing from noise burst to buzz excitation.

Tape recordings of a number of replications of these 16 stimuli in random order were prepared, and were presented over monaural headphones to five listeners, who were asked to identify each stimulus as /*dɑ*/ or /*tɑ*/. Spectrograms of stimuli unanimously judged to be /*dɑ*/ and /*tɑ*/ are shown in Fig. 5.

Each listener made 12 responses to each of the stimuli. Results for two of the listeners are shown in Fig. 6. These data represent, in a sense, two extremes of performance of the listeners. For one listener (AWFH), the boundary between /*dɑ*/ and /*tɑ*/ responses occurred at a fixed time relative to completion of the formant transitions, while for other listeners (e.g., VZ) the boundary tended to be at a nearly fixed time relative to plosive release. However, the latter listener showed at least some tendency for the absolute VOT at the phoneme boundary to be longer for stimuli with longer transitions. When expressed in terms of voice-onset time,

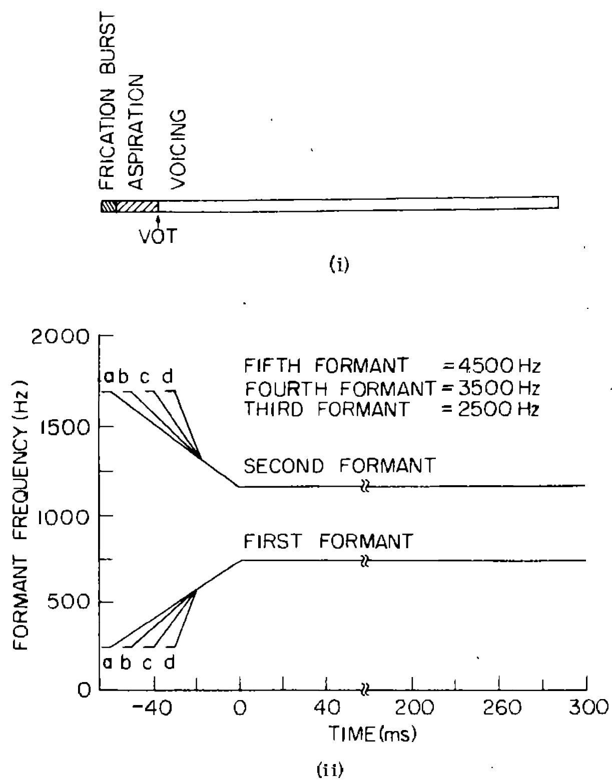

FIG. 4. A synthetic utterance consists of a frication burst and a variable interval of aspiration, followed by onset of voicing, as shown in part (i). The lowest two formant frequencies follow one of four possible trajectories that are labeled (a) through (d) in part (ii) of the figure. Zero on the time scale corresponds to the time when the formant transitions are completed.

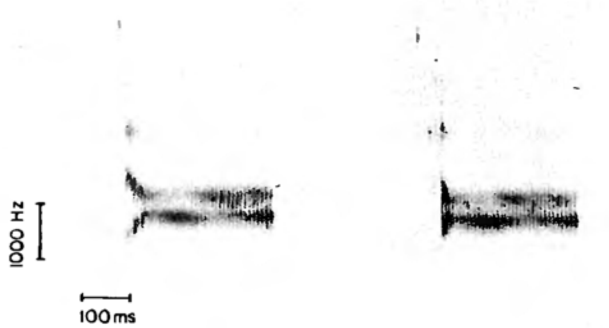

FIG. 5. Spectrograms of two of the synthetic utterances. The stimulus at the left had a VOT of 15 msec and was unanimously identified as /dɑ/. The stimulus at the right had a VOT of 55 msec and was unanimously identified as /tɑ/.

the phoneme boundary moved about 25 msec for a 30-msec change in the duration of the formant transition in the data of listener AWFH, but the phoneme boundary moved only about 5 msec for VZ.

Average data for the five listeners are shown in Fig. 7. Each point represents the average VOT at the 50% response point for the five listeners. For this figure, the abscissa is the voice-onset time relative to the consonantal release. The average VOT at the phoneme boundary moved from 26 to 39 msec for a 30-msec change in formant transition duration. The horizontal line in Fig. 7 indicates a constant voice-onset time; the data points would lie along this line if the absolute VOT provided the only cue for the voiced-voiceless distinction. The line sloping up to the left represents the locus of constant duration of the transition from voicing onset to termination of the transition. Data points would lie along a line with this slope if the cue for the voiced-voiceless distinction were the presence or absence of substantial formant transitions after voicing onset. The line in Fig. 7 is in fact drawn for a formant transition duration during voicing of 23 msec, as indicated by the arrow in the inset. The amount of spectrum change occurring as a result of the rising transition of the first formant during this interval is relatively small compared with the spectrum change resulting from the entire transition from 250–750 Hz. The average rise in intensity of the higher formants due to the F1 transition is about 5 dB in the former case and about 20 dB in the latter (Fant, 1956; Stevens and House, 1961).

An additional experiment was performed to determine the minimal first-formant transition that can be detected at the onset of the synthetic vowel /ɑ/. Five stimuli were prepared with formant transition durations of (a) 0, (b) 5, (c) 10, (d) 15, and (e) 20 msec. The rate of the formant transition was held constant in all cases at 500 Hz/60 msec. This rate is typical of a first-formant transition in a stop release and it is the same as the slowest rate of transition used in the previous experiments.

Three subjects participated in ABX tests comparing the stimulus without a formant transition with each of the four stimuli containing various amounts of formant transition. The results were plotted on a graph of percent correct versus transition duration, and a smooth curve was fitted to the data points. This curve crossed the 75% correct detection threshold at a transition duration of about 13 msec.

This just-detectable transition duration at the stimulus onset is about 10 msec less than the maximum transition duration that triggers a /t/ response in Fig. 7. The agreement between these two threshold values of transition duration is not unreasonable, if account is taken of the fact that the burst and aspiration in the plosive-vowel stimuli may inhibit detection of the transition at the onset of voicing.

II. DISCUSSION

The results of Experiment 1 are in reasonable agreement with the experimental data of Hirsh (1959) and of Hirsh and Sherrick (1961), who carried out a series of experiments to determine the difference in onset time that is required for a listener to judge which of two stimuli with diverse characteristics came first. These investigators found a time difference of about 20 msec for a variety of stimulus conditions, including cross-modal presentation of the stimuli. Various components of a complex stimulus that are packaged in the first 20–

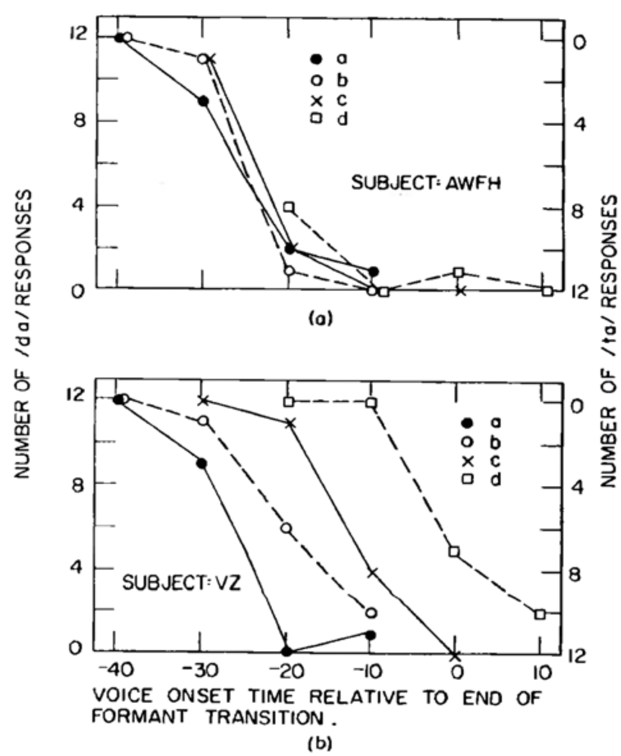

FIG. 6. Identification data for two subjects. The number of /dɑ/ responses is plotted as a function of time of onset of voicing, where zero on the scale corresponds to the completion of the formant transition shown in Fig. 4. A family of four curves is drawn to indicate how the responses change as a function of rate of formant motion; the labels *a*, *b*, *c*, *d* refer to the transition types shown in Fig. 4.

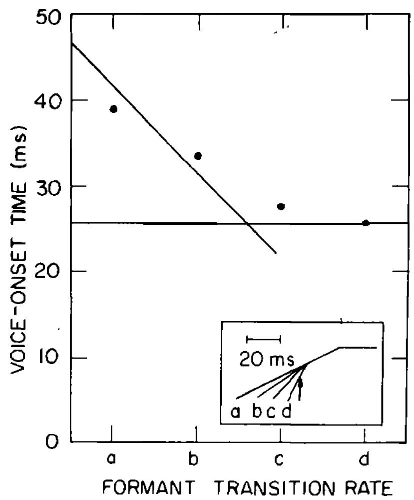

FIG. 7. The identification data of five subjects are summarized by plotting the average VOT at the phoneme boundary (see text) for each rate of formant motion. The VOT in this case is the time from consonantal release to onset of voicing. The letters on the abscissa are identified with different rates of transition, as indicated in the inset. A horizontal line could be fitted to the data if the phoneme boundaries were based only on the delay of voicing with respect to plosive release. A sloping line could be fitted to the data if the phoneme boundaries were based on the presence or absence of a detectable formant transition (or of a fixed duration of residual transition) during the initial portion of the voicing. The sloping line in this figure corresponds to a time of voicing onset indicated by the arrow in the inset.

25 msec are somehow integrated and processed as a unit. If two components with different acoustic characteristics have onset times more than 20–25 msec apart, they are perceived as being successive events.

The findings of Experiment 2 are consistent with the data of Liberman, Delattre, and Cooper (1958) who examined the responses of listeners to a series of consonant-vowel stimuli which differed only in the amount of cutback of the first formant relative to the onset time. In other words, the stimuli were produced by starting with a sound like that of /*da*/ in Fig. 1, and by "blanking out" the initial part of the F1 transition in varying amounts. Their data show that the stimuli with larger F1 cutback were identified as beginning with /*t*/. The boundary between /*d*/ and /*t*/ responses occurred when the cutback was at a point where most of the F1 transition was completed. In a recently reported experiment, Summerfield and Haggard (1972) have also noted that the presence of an F1 transition after voicing onset is a cue for voicing.

Based on the results of Experiments 1 and 2, we can suggest a possible strategy that a listener uses to discriminate between voiced and voiceless stop consonants in prestressed position. We postulate first that the cue for the presence of a consonantal segment is a rapid change in the acoustic spectrum occurring at a point where there is an abrupt or discontinuous increase in intensity in some frequency range (Stevens, 1971).[4] This rapid change in the acoustic spectrum is in the frequency range above about 1000 Hz, and occurs over a brief time interval of 20–30 msec. The characteristics of this transient spectrum shift provide some of the cues for place of articulation for the consonant.

This hypothesis can be applied to the stimuli used in Experiment 2, which have an abrupt initial intensity increase, coupled with a rapid change in the spectrum in the next few tens of msec. When the onset of voicing is delayed relative to the beginning of the stimulus, there is a second discontinuity due to an abrupt increase in energy in the first formant range. If the VOT is less than 20 msec, the two onsets are perceived as simultaneous (from Experiment 1 and from Hirsh's results), and the stimulus is identified as a voiced consonant.

For a VOT that exceeds 20-odd msec, the initial onset and the onset of voicing are perceived as successive events. If the rapid change in spectrum resulting from the formant transitions is essentially completed before the onset of voicing, the consonant is identified as voiceless. If, however, there are substantial formant transitions following the onset of voicing, then there are two conflicting cues. The long VOT tends to indicate a voiceless consonant, whereas the rapid spectrum change at the onset of voicing is characteristic of a voiced consonant. There are, in fact, two successive onsets that contain rapid spectrum changes. Some listeners (e.g., VZ) assign more weight to the absolute VOT, whereas others (e.g., AWFH) will not accept a stimulus with an appreciable transition at the onset of voicing as a voiceless consonant regardless of the VOT. The most acceptable voiceless consonant is, of course, one for which both criteria are satisfied: the VOT exceeds 20 msec and the formant transitions are essentially completed before voicing onset occurs. These were in fact judged by listeners to be the most "natural" voiceless consonants.

Indirect evidence supporting the importance of the presence or absence of rapid spectrum change at voicing onset is found in data on segmental durations in prestressed plosive-sonorant constant clusters (Klatt, 1973b). In a word such as "breed," VOT is typically about 10 msec and the [*r*] segment is rather short. For a word beginning with [*pr*], the VOT is increased to about 60 msec. Measurements indicate that in this cluster the consonant [*r*] is increased in duration by about 30 msec. If the sonorant had not been lengthened, voicing onset would have occurred during the rapid formant motions of the sonorant-vowel transition, leading to a false cue for voicing. The increased sonorant duration results in an onset of voicing that is well before the sonorant-vowel transition, thus preserving the cue for voicelessness.

The proposed mechanism for discriminating prestressed voiced from voiceless consonants suggests a way of explaining the results of Eimas *et al.* (1971) on the response of infants to the voiced-voiceless distinction for stop consonants.[5] It is postulated that the auditory system provides one kind of response when there is a significant transition in F1 after onset of voicing, creating a rapid spectrum change, and another kind of response where there is no such rapid spectrum change. Stimuli which differ in this property yield auditory responses that are qualitatively different. The particular VOT's for the stimulus pair that was discriminated by the children in the experiment of Eimas *et al.*, were 20 and 40 msec. The stimulus with a VOT of 20 msec has a rapid spectrum change after onset of voicing whereas the stimulus with a VOT of 40 msec has negligible formant transitions after voicing onset. Stimuli with voice-onset times up to 20 msec have the common attribute that there is a rapid spectrum change at the onset of voicing, and these stimuli are judged to be similar by infants. On the other hand, stimuli with voice-onset times greater than 40 msec have the common property that the formant transitions are completed before onset of voicing, and thus are not discriminated from one another by infants.

If there is a requirement that the formant transitions be essentially completed before onset of voicing, then a simple explanation exists for the longer VOT's that are observed for velars than for dentals, and for dentals than for labials in prestressed position (Lisker and Abramson, 1964). It is known that the duration of the movement of the articulator that forms the closure is greatest for the tongue body, less for the tongue tip, and least for the lips. Measurements of rates of formant transitions, and speech synthesis experiments with various rates of formant transitions also show the slowest rates for velars. When the rate of transition is slower, the onset of voicing must be increased if the transition is to be essentially completed before voicing begins.

This change in VOT for voiceless aspirated stops from one place of articulation to another raises a question as to what strategy is used by a speaker to actualize these different VOT's. The mechanism suggested by Klatt (1973a), in which glottal closure is initiated reflexly by the rapid pressure drop that occurs in the mouth cavity (and in the region of the glottis) could account for the differences not only for stressed consonant-vowel syllables but also for the increased VOT in initial voiceless stop consonants in consonant clusters. Klatt observes that the time from completion of the initial burst of frication noise (when the mouth pressure presumably drops rapidly) to voicing onset is roughly independent of consonant place of articulation or of the identity of the following segment. The frication noise burst is longer for the velars, for which the rate of opening of the constriction is slower than for dentals and labials.

The strategy proposed here for discriminating voiceless from voiced stop consonants applies specifically to single consonants preceding stressed vowels and possibly for initial consonants in clusters. When the consonants occur in other environments, different strategies may have to be invoked in order to distinguish voiced from voiceless stops. Thus, for example, if the consonant follows a stressed vowel and is either in final position or preceding an unstressed vowel, the length of the vowel is often enough to signal the voicing feature—the vowel is shortened preceding a voiceless consonant (Denes, 1955; House, 1961). In the intervocalic environment, the length of the stop gap may also provide a cue, the gap being shorter for the voiced consonant (Lisker, 1957). In intervocalic pre-unstressed position, the voice-onset time for voiceless stops is usually shorter than in the prestressed position, and in this phonetic environment VOT measurements alone do not reliably separate voiced from voiceless stops (Lisker and Abramson, 1967(. All of these comments indicate that the voiced-voiceless distinction for a stop is not triggered by the same acoustic property (or properties) independent of the phonetic environment in which it appears. The properties considered in this study apply only to consonants in prestressed position. It is, of course, a common observation that the acoustic manifestation of a particular phonetic feature depends on the environment of other features in which the segment occurs.

ACKNOWLEDGMENT

This research was supported in part by the National Institutes of Health (Grant No. NS-04332).

[1]The term voiceless as applied to stop consonants in this paper refers to voiceless *aspirated* stops of the type that occurs in English. Voiceless unaspirated stop consonants also occur in many languages.

[2]These numbers apply to labials and apicals; they are somewhat greater for velars.

[3]A possible exception is the finding that the temporal order of the onsets of two different acoustic stimuli with appropriate characteristics can be judged only if the time between onsets is 20 msec or more (Hirsh, 1959; Hirsh and Sherrick, 1961). The relevance of these results to the present problem are discussed later.

[4]Note that neither the /*h*/ nor the glottal stop /ʔ/ have such a rapid spectrum change at the onset of voicing, where there is abrupt intensity increase (at low frequencies in the case of /*h*/). Linguists characterize these segments as nonconsonantal.

[5]The experiments of Eimas *et al.* were performed with stimuli on a continuum between /*po*/ and /*bo*/ rather than /*tɑ*/ and /*dɑ*/, but the acoustic cues for the voiced-voiceless distinction are substantially the same in the two cases.

Denes, P. (**1955**). "Effect of Duration on the Perception of Voicing," J. Acoust. Soc. Am. **27**, 761–764.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., and Vigorito, J. (**1971**). "Speech Perception in Infants," Science **171**, 303–306.

Fant, C. G. M. (**1956**). "On the Predicability of Formant Levels and Spectrum Envelope from Formant Frequencies," in *For Roman Jakobson*, edited by M. Halle *et al.* (Mouton, The Hague), pp. 109–120.

Hirsh, I. J. (**1959**). "Auditory Perception of Temporal Order," J. Acoust. Soc. Am. **31**, 759–767.

Hirsh, I. J. and Sherrick, C. E. Jr. (**1961**). "Perceived Order in Different Sense Modalities," J. Exp. Psychol. **62**, 423–432.

House, A. S. (**1961**). "On Vowel Duration in English," J. Acoust. Soc. Am. **33**, 1174–1178.

Jakobson, R., Fant, C. G. M., and Halle, M. (**1963**). *Preliminaries to Speech Analysis* (MIT Press, Cambridge, Mass.).

Klatt, D. H. (**1972**). "Acoustic Theory of Terminal Analog Speech Synthesis," *Proc.* 1972 *Int. Conf. on Speech Communication and Processing*, No. 72 CHO 596-7 AE (IEEE, New York), pp. 131–135.

Klatt, D. H. (**1973a**). "Voice Onset Time, Frication and Aspiration in Word-Initial Consonant Clusters," Res. Lab. Electron. Quart. Progr. Rept. No. 109, M. I. T., pp. 124–136.

Klatt, D. H. (**1973b**). "Durational Characteristics of Prestressed Word-Initial Consonant Clusters in English," Res. Lab. Electron. Quart. Progr. Rept. No. 108, M. I. T., pp.

253–260.

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**, 153–167.

Liberman, A. M., Delattre, P. C., Gerstman, L., and Cooper, F. S. (1956). "Tempo of Frequency Change as a Cue for Distinguishing Classes of Speech Sounds," J. Exp. Psychol. **52**, 127–137.

Liberman, A. M., Harris, K. S., Kinney, J. A., and Lane, H. (1961). "The Discrimination of Relative Onset-Time of the Components of Certain Speech and Non-Speech Patterns," J. Exp. Psychol. **61**, 379–388.

Lisker, L. (1957). "Closure Duration and the Intervocalic Voiced-Voiceless Distinction in English," Language **33**, 42–49.

Lisker, L. and Abramson, A. S. (1964). "A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements," Word **20**, 384–422.

Lisker, L. and Abramson, A. S. (1967). "Some Effects of Context on Voice Onset Time in English Stops," Language and Speech **10**, 1–28.

Lisker, L. and Abramson, A. S. (1970). "The Voicing Dimension: Some Experiments in Comparative Phonetics," in *Proc. 6th Int. Cong. Phonetic Sciences, Prague* 1967 (Academia Publ. House of Czechoslovak Acad. of Sci., Prague), pp. 563–567.

Stevens, K. N. (1968). "Acoustic Correlates of Place of Articulation for Stop and Fricative Consonants," Res. Lab. Electron. Quart. Progr. Rept. No. 89, M. I. T., pp. 199–205.

Stevens, K. N. (1971). "The Role of Rapid Spectrum Changes in the Production and Perception of Speech," in *Form and Substance* (Festschrift for Eli Fischer-Jørgensen). (Akademisk Forlag, Copenhagen), pp. 95–101.

Stevens, K. N. (1972). "The Quantal Nature of Speech: Evidence from Articularoty-Acoustic Data," in *Human Communication: A Unified View*, edited by E. E. David Jr. and P. B. Denes (McGraw–Hill, New York), pp. 51–66.

Stevens, K. N. and House, A. S. (1961). "An Acoustical Theory of Vowel Production and Some of its Implications," J. Speech and Hearing Res. **4**, 303–320.

Summerfield, A. and Haggard, M. (1972). "Perception of Stop-Voicing—A Rate-Specific Transition Detector?" Speech Perception: Report of Research in Progress Series 2 (No. 1), Dept. Psychol, Queen's Univ. of Belfast, pp. 1–14.