Perception of the voiced–voiceless contrast in syllable-final stops

James Hillenbrand
Department of Communicative Disorders, 2299 Sheridan Road, Northwestern University, Evanston, Illinois 60201

Dennis R. Ingrisano
Speech Research Laboratory, Department of Communicative Disorders and Sciences, Wichita State University, Wichita, Kansas 67208

Bruce L. Smith
Department of Communicative Disorders, 2299 Sheridan Road, Northwestern University, Evanston, Illinois 60201

James E. Flege
Department of Biocommunication, UAB Medical Center, University Station, Birmingham, Alabama 35294

(Received 9 June 1983; accepted for publication 6 February 1984)

A computer editing technique was used to remove varying amounts of voicing from the syllable-final closure intervals of naturally produced tokens of /pɛb, pɛd, pɛg, pag, pig, pug/. Vowels for all six syllables were approximately the same duration, and the final release bursts were retained. Identification results showed that voiceless responses tended to occur in relatively large numbers when all of the closure voicing and, in most cases, a portion of the preceding vowel-to-consonant (VC) transition had been removed. A second experiment demonstrated that removal of final release bursts had very little effect on the identification functions. Acoustic measurements were made in an attempt to gain information about the acoustic bases of the listeners' voiced–voiceless judgments. In general, stimuli that subjects tended to identify as voiceless showed higher first-formant offset frequencies and shorter intensity decay times than stimuli that subjects tended to identify as voiced. However, for stops following /i/ and /u/ these acoustic differences were relatively small. We were unable to find a single acoustic measure, or any combination of measures, that clearly explained the listeners' voiced–voiceless decisions.

PACS numbers: 43.70.Dn, 43.70.Ve

INTRODUCTION

The voicing contrast in English stops has been studied rather extensively, but the majority of this work has focused on stops in initial position. The most well-developed approach to describing voicing contrasts in initial stops is the glottal/supraglottal timing view proposed by Lisker and Abramson (1964, 1967, 1970; see Malécot, 1970, for an alternative point of view). Articulatory control of initial stop voicing contrasts is thought to involve variations in the timing of voicing onset relative to articulatory release. The voicing opposition in final stops also involves the timing of a laryngeal gesture relative to articulatory events in the upper airway. For syllables ending in voiceless stops, a laryngeal gesture typically terminates voicing at about the same time that articulatory closure is achieved. For syllables ending in voiced stops, glottal vibration generally continues into at least some portion of the closure interval. In this sense, syllable-final voicing contrasts might be thought of in terms of a "voice offset time" feature somewhat analogous to voice onset time in initial stops.

In articulatory terms, voice offset time would refer to the timing of a phonation-terminating gesture relative to the achievement of articulatory closure. Because of the early termination of glottal vibration in final voiceless stops, these syllables are generally characterized by (1) relatively high first formant ($F1$) terminating frequencies (Wolf, 1978) and (2) silent rather than partially or fully voiced closure intervals (Smith, 1979; Hogan and Roszypal, 1980; Flege and Brown, 1982). Further, measurements that were made as pilot work to the present study show that vowels which precede voiceless stops are generally terminated with a more abrupt drop in intensity as compared to vowels that precede voiced stops (see Derbock, 1977, for similar findings on stops in French and Dutch). The kind of laryngeal timing distinction discussed above, however, does not account for the well-known influence of consonant voicing on the duration of preceding vowels. In phrase-final position in English, vowels preceding voiced consonants are generally about 50 to 100 ms longer than vowels preceding voiceless consonants (e.g., House and Fairbanks, 1953; House, 1961; Peterson and Lehiste, 1960; Chen, 1970).[1]

On the basis of the acoustic properties associated with the syllable-final voicing contrast, perceptual cues to this distinction might involve three kinds of acoustic features: (1) the presence versus absence of a low-amplitude low-frequency "voice bar" during the closure interval, (2) some combination of frequency and intensity characteristics associated with gradual versus abrupt termination of the preceding vowel (i.e., $F1$ offset frequency and intensity decay time), and (3) the duration of the preceding vowel. Although the vowel duration difference is often considered to be of primary perceptual importance to final stop cognate oppositions, experimental evidence supporting this conclusion consists primarily of a series of synthetic speech studies by Raphael and his colleagues (Raphael, 1972; Raphael *et al.*, 1975, 1980).[2] Raphael (1972) used the Pattern Playback to synthesize syllables ending in voiced and voiceless stops and fricatives. In general, final consonants were heard as voiceless when preceded by vowels of short duration and as voiced when preceded by vowels of long duration. Raphael concluded that preceding vowel duration was both a necessary and sufficient cue to syllable-final voicing. Similar findings were reported in two subsequent synthesis studies by Raphael *et al.* (1975, 1980).

The conclusions based on these synthesis studies have not been supported by more recent work involving edited natural speech. For example, Wardrip-Fruin (1982a) showed that vowels preceding final voiced stops could be reduced in duration by one-third without eliciting voiceless responses. Several other investigators using edited natural speech have shown that reducing vowel duration, by itself, does not typically make a naturally produced voiced stop sound voiceless (O'Kane, 1978; Hogan and Rozsypal, 1980; Raphael, 1981; Revoile *et al.*, 1982). It has also been shown that expanding vowel durations of naturally produced syllables ending in voiceless stops does not result in voiced stop judgments (Hogan and Rozsypal, 1980; Revoile *et al.*, 1982).[3]

Most of the evidence from edited natural speech studies suggests that final stop voicing contrasts are cued primarily by acoustic information in the vicinity of articulatory closure. For example, Wolf (1978) made editing cuts at several locations from naturally produced syllables such as /æb/, /æd/, and /æg/. Removal of the entire closure interval produced 16% voiceless responses, while removal of the closure interval and three pitch periods from the vowel-to-consonant (VC) transition produced 70% voiceless responses (see Revoile *et al.*, 1982, for similar findings). These results seem to suggest that final stop voicing contrasts are cued primarily by differences in the way in which the preceding vowel is terminated (Parker, 1974; Walsh and Parker, 1981).

The present study was designed to examine cues to the perception of final stop voicing in greater detail by using a fine-grained editing technique on naturally produced syllable-final stops in several vowel environments. Wolf (1978) and Revoile *et al.* (1982) studied cues to final stop voicing using stimuli in the environment of /æ/, a vowel that might tend to accentuate differences in both $F1$ offset frequency and decay time. The open vocal tract configuration of vowels such as /æ/ results in relatively high intensities and high-frequency first formants. As the vocal tract constricts for the final closure, substantial decreases are seen in overall intensity and in the frequency of the first formant. Therefore, editing cuts that truncate the VC transition would result in relatively high $F1$ offset frequencies and short decay times. On the other hand, more constricted vowels would be expected to show less dramatic intensity and frequency changes in the VC transition. For this reason it is unclear whether the pattern of results obtained with /æ/ would also be seen when editing cuts are made from stimuli in the context of more constricted vowels, such as /i/ and /u/.

In the present study voicing cues were examined using syllable-final voiced stops varying in place of articulation and vowel environment. Editing techniques were used to determine how much of the final voiced segment had to be removed before subjects heard final voiceless stops. The study involved two listening experiments and a series of *post hoc* acoustic analyses of the edited stimuli. The stimuli for experiment 1 were edited in such a way as to retain the final release bursts. Experiment 2 examined the role of release bursts by asking subjects to identify edited stimuli both with and without final bursts. The purpose of the acoustic measurements was to provide preliminary tests of several hypotheses regarding acoustic cues to final voicing contrasts.

I. EXPERIMENT 1

A. Stimuli

1. Recording and measurement

The stimuli consisted of a combination of naturally produced and edited tokens of /pεb, pεd, pεg, pag, pig, pug, pεp, pεt, pεk, pak, pik, puk/. A male talker produced several repetitions of each syllable. Audio recordings were made in a sound-treated booth with a headset microphone (Shure SM11) and a reel-to-reel tape deck (Akai GX – 4000 DB). The tape-recorded stimuli were then low-pass filtered at 4 kHz and digitized at a 10-kHz sample frequency. Oscillographic representations of 100-ms segments of the stimuli were displayed on a high-resolution graphics terminal (Tektronix 4010). The oscillograms were used to segment each stimulus into (1) voice onset time (VOT) of initial /p/, (2) vowel, (3) final stop closure, and (4) final stop release burst.[4] For the stimuli ending in voiced stops, these measurements were used to select which of the multiple repetitions would be used in the perception tasks. Tokens were required to meet two criteria: (1) a fully voiced closure interval measuring about 75 ms (± 5ms) and (2) an audible final burst measuring 5–15 ms in duration. Measurements of the six stimuli ending in voiced stops are shown in Table I. For comparison the table also gives duration measurements from the six syllables ending in voiceless stops. These stimuli were included as a reliability check on listener responses and were not required to meet any durational criteria.

TABLE I. Time measurements (in ms) for (1) VOT of the initial stop, (2) vowel duration, (3) closure duration of the final stop, (4) burst duration, and (5) total syllable duration. The values in parentheses indicate the vowel and syllable durations after application of the editing technique that equalized vowel durations for stimuli ending in voiced stops.

	VOT	Vowel	Closure	Burst	Syllable
pεb	37	122 (112)	77	7	243 (232)
pεd	47	131 (110)	71	4	253 (232)
pεg	67	106 (106)	76	8	257 (257)
pag	61	174 (107)	73	16	324 (257)
pig	49	149 (113)	73	13	284 (248)
pug	65	142 (108)	75	5	288 (254)
pεp	30	82	...[a]	...[a]	140
pεt	58	111	90	11	270
pεk	41	105	93	4	243
pak	95	116	108	18	337
pik	51	86	88	23	248
puk	88	85	94	12	279

[a] This token was unreleased.

2. Control of vowel duration

To hold any potential influence of vowel duration constant across the six continua, a computer editing procedure was used to modify the vowel durations of the six stimuli ending in voiced stops. Beginning with the third pitch period of the vowel, every other pitch period was removed until the vowel was within one-half pitch period of 110 ms. To avoid introducing clicks, editing cuts were made at zero crossings. By removing every other pitch period, abrupt discontinuities in fundamental frequency and formant frequencies were avoided. Since the editing program was able to shorten but not lengthen vowels, all of the stimuli were adjusted to accomodate the shortest vowel. The vowel in /pɛg/ was the shortest at 106 ms and was left unmodified. Vowel and syllable durations after application of this editing technique are shown in parentheses in Table I.

3. Editing of closure voicing

Varying amounts of glottal pulsing were removed from the closure intervals of the six syllables ending in voiced stops. Figure 1 shows a continuum based on the natural token of /pɛg/. The continuum consisted of 13 stimuli ranging from 0–120 ms of signal removed, in 10-ms steps. Because the closure interval for the original /pɛg/ was 76 ms in duration (see Table I), the stimuli from 80 to 120 ms in the continuum involve the removal of all of the voicing from the closure interval and a portion of the preceding vowel. Editing of the stimuli was accomplished by multiplying the original digitized waveforms by a series of weighting functions of the type shown in Fig. 2. This figure displays a weighting function superimposed on the stimulus from the /pɛg/ series that is labeled "20" in Fig. 1. Values in the weighting function ranged from zero (complete attenuation of the signal) to one (no signal attenuation). The nominal value of 20 ms refers to the amount of time that the weighting function remained at zero. The zero-amplitude interval was preceded by a 15-ms linear decay function which was intended to simulate the gradual amplitude reduction that generally occurs in natural speech when voicing terminates prior to release. The decay function also reduced the possibility that the editing cuts would introduce transients into the stimuli. It should be emphasized that the 20-ms nominal value refers *only* to the zero-amplitude portion of the weighting function and does not include that portion of the signal attenuated by the 15-ms decay function. Following the zero-amplitude interval, the weighting function returned instantaneously to a value of one in order to retain the release burst. A series of weighting functions was calculated individually for each stimulus so that this final full-amplitude portion of the weighting function exactly fit the release burst.

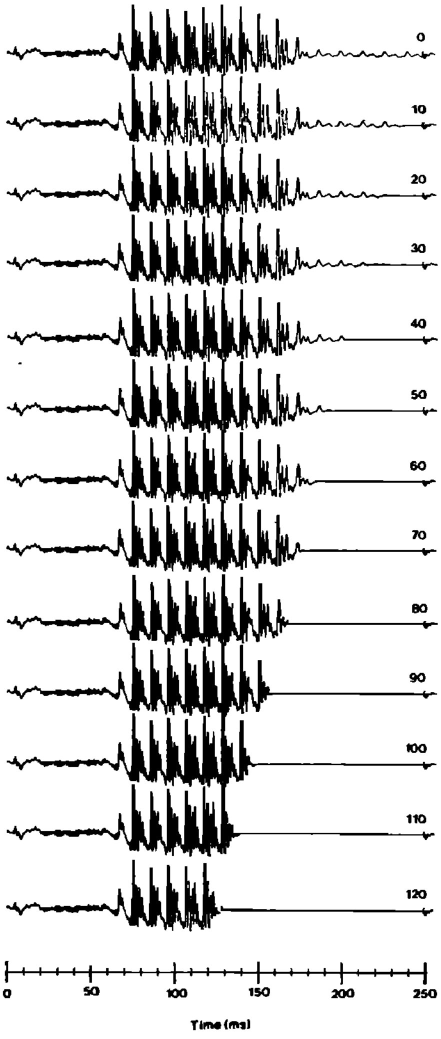

FIG. 1. Stimulus continuum constructed by making increasingly large editing cuts form the voiced closure interval of a naturally produced /pɛg/. The numbers to the right of each oscillogram are nominal values that reflect the amount of signal that was removed by the editing procedure (see text).

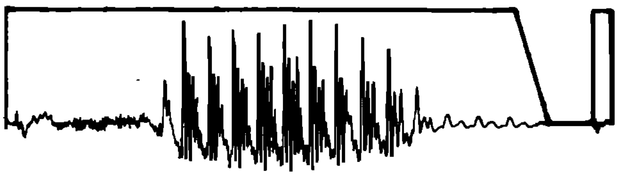

FIG. 2. Oscillogram of the 20-ms stimulus from the /pɛg/ continuum shown with the associated weighting function. Starting 35 ms prior to the final burst, the weighting function decays linearly over 15 ms from a gain of one to a gain of zero, remains at zero for 20 ms, then returns instantaneously to a gain of one for the duration of the burst. Note that the nominal value of 20 ms refers only to the amount of time that the weighting function remains at zero.

B. Subjects and procedures

Subjects were 23 Northwestern University students with no reported history of hearing or speech problems. Presentation of stimuli and collection of subjects' responses were under the control of a laboratory computer equipped with a high-speed disk drive and a 12-bit D/A converter. At the output of the D/A converter the stimuli were low-pass filtered at 4 kHz, amplified, attenuated, and delivered binaurally over matched TDH-49 headphones. The output attenuator was adjusted so that signals peaked at 78 dBA.

Each stimulus continuum consisted of 14 tokens: one exemplar of the natural voiced stop, 12 edited versions, and one natural production of the voiceless cognate.The identification tests consisted of ten randomly ordered presentations of the 14 stimuli. On each trial subjects were asked to press a button labeled B,D,G or a button labeled P,T,K. Each sub-

TABLE II. Mean voiced–voiceless category boundaries, standard deviations, and ranges for each stimulus continuum. The values (in ms) represent the amount of signal removed from the final voiced segments of syllables ending in voiced stops.

Continuum	Mean	Standard deviation	Range
/pɛb/	68	11	48– 88
/pɛd/	77	14	58– 93
/pɛg/	77	13	62–122
/pag/	70	7	55– 80
/pig/	61	6	54– 80
/pug/	63	7	50– 78

ject was tested on all six continua and each received a different ordering of the conditions.

C. Results and discussion

Table II shows voiced–voiceless category boundaries and measures of dispersion for each continuum. Category boundaries were calculated by linear interpolation of the 50% crossover from voiced to voiceless. In general, category boundaries tended to occur in the 60- to 80-ms range. Figures 3 and 4 show the percentage of voiced responses as a function of the size of the editing cut. Figure 3 shows that there was a slightly earlier crossover for the labial continuum as compared to the alveolar and velar continua. Figure 4 shows earlier crossovers for /pig/ and /pug/, the two vowels with low-frequency first formants. Although the place and vowel effects seem to be fairly small in terms of absolute magnitude, two separate repeated measures ANOVAs showed that both effects were significant [place: $F(2,44) = 4.9, p < 0.05$; vowel: $F(3,66) = 20.2, p < 0.01$].

In general, the listening tests showed that voiceless responses were not elicited in large numbers until the closure interval and, in most cases, a portion of the VC transition had been removed. This finding is consistent with the idea that information in the VC transitions is important to the perception of final voicing contrast. A more detailed discussion of these findings will await the results of a series of acoustic analyses of the stimuli, described in Sec. III.

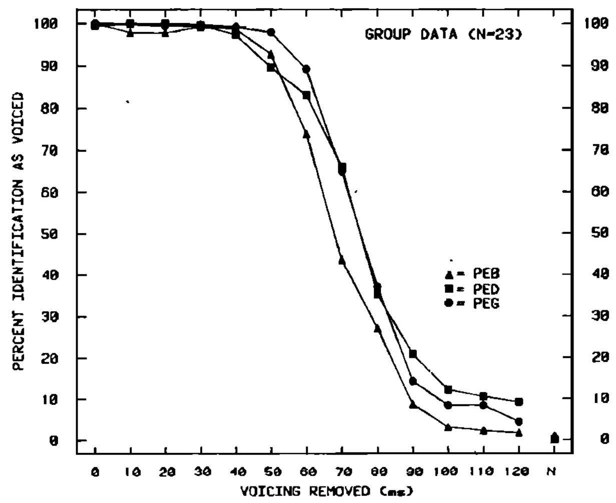

FIG. 3. Identification results showing the effect of varying the place of production of the final stop. Each function represents the pooled results from 23 listeners. The "N" on the abscissa represents responses to the unedited voiceless stops.

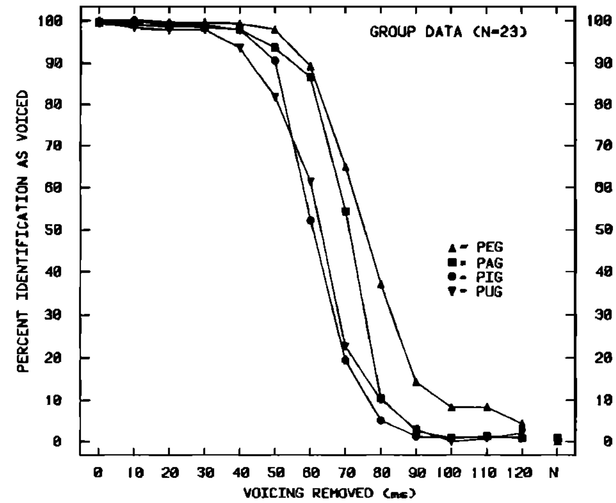

FIG. 4. Identification results showing the effect of varying the vowel environment. Each function represents the pooled results from 23 listeners. The "N" on the abscissa represents responses to the unedited voiceless stops.

II. EXPERIMENT 2

The main finding of experiment 1 was that voiceless responses did not predominate until relatively large amounts of voicing had been removed from the stimuli. The possibility exists, however, that the syllable-final release bursts contained cues which biased subjects toward voiced responses. This possibility motivated a second experiment with a new group of 11 subjects which examined the role of release bursts. The stimuli for this experiment consisted of the /pɛb/, /pɛd/, and /pɛg/ continua from experiment 1. Stimuli from each continuum were presented to the subjects both with and without release bursts. The release bursts were eliminated simply by aligning a cursor to redefine the end of the stimulus. Subject-selection criteria, instrumentation, and procedures were identical to those described for experiment 1.

A. Results and discussion

The results from experiment 2 are shown separately for each place of articulation in Fig. 5. Although the burst conditions show earlier crossovers in each comparison, the differences in the category boundaries are only 3–4 ms. A two-way repeated measures analysis of variance fell just short of significance for the burst versus no burst comparison [$F(1,10) = 4.6, p < 0.07$]. In contrast to the results of experiment 1, phonetic boundaries were not significantly affected by differences in place of articulation [$F(2,20) = 3.0, p$ NS]. The place-by-burst interaction did not approach significance.

The relatively minor role played by bursts in this experiment should be interpreted in light of the fact that the release bursts were relatively brief (15 ms or less). Longer release bursts for voiced stops sometimes show evidence of periodicity (e.g., Wolf, 1978) and might, therefore, influence voicing

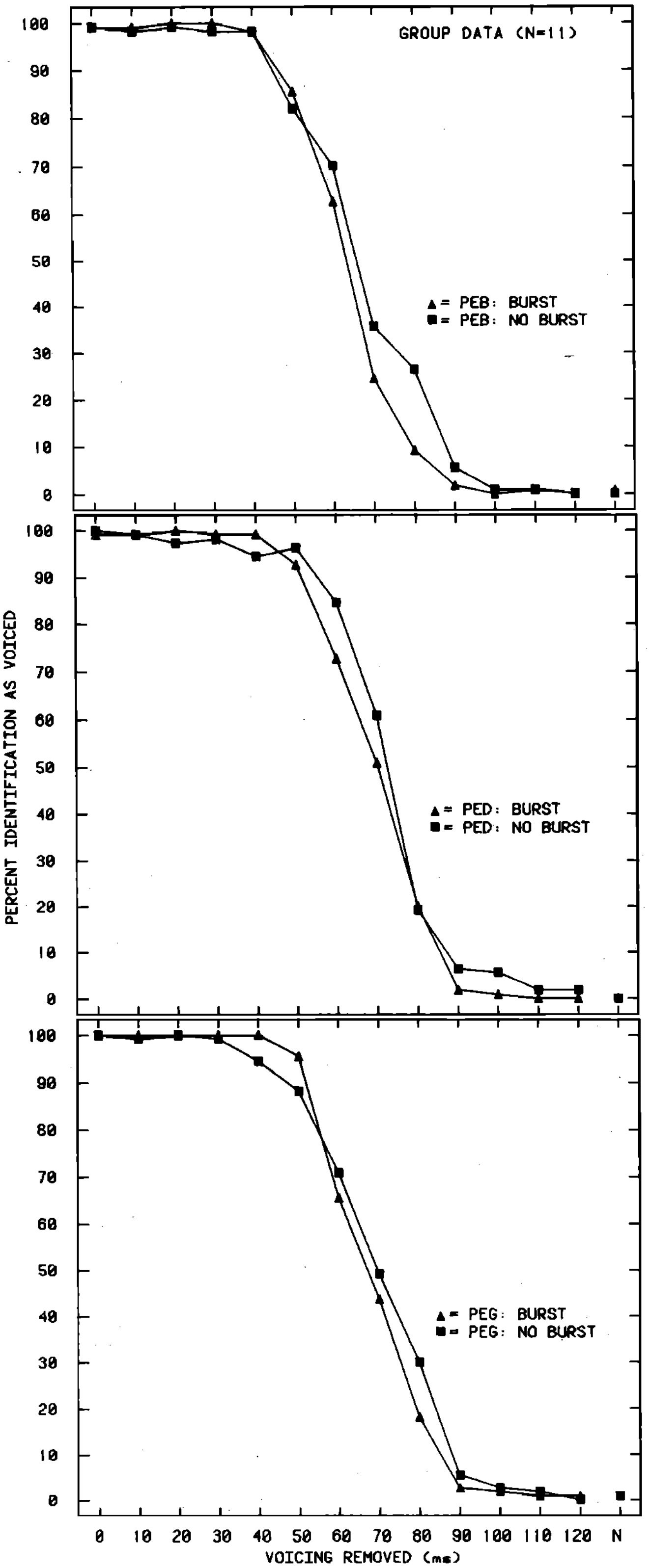

FIG.5. Identification results showing the effect of removing the release burst. Each function represents the pooled results from 11 listeners. The "N" on the abcissa represents responses to the unedited voiceless stops.

judgments. There is also evidence that the removal of final release bursts from voiceless stops has a greater effect on voicing judgments than the removal of bursts from final voiced stops (Malécot, 1958; Revoile *et al.*, 1982). However, the general conclusion that final bursts from voiced stops do not strongly influence voicing judgments is supported by several other studies (Malécot, 1958; Wolf, 1978; Raphael, 1981; Wardrip-Fruin, 1982a; Revoile *et al.*, 1982). This result is also sensible in light of production data showing that a large proportion of final stops are produced without release bursts (Rositzke, 1943; Crystal and House, 1982).

The minimal importance of final bursts would seem to imply that the duration of the closure interval was not a strong voicing cue for these stimuli. Since the final release burst marks the end of the closure interval, it seems logical to assume that removal of this marker would alter the listeners' impression of the duration of the interval. Therefore, the failure to find large differences between the burst and no burst conditions suggests that closure duration did not play a significant role in the subjects' voicing judgments. This is consistent with Raphael (1981) who showed that extending final /g/ closures to durations appropriate for /k/ did not elicit /k/ responses from listeners.

III. ACOUSTIC ANALYSIS

One of the difficulties in interpreting the labeling results is that category boundaries occurred very close to the beginning of the closure interval. Location of the boundary in this region makes it difficult to decide between two very different (and possibly nonmutually exclusive) alternatives concerning cues to syllable-final voicing. One possibility is that subjects' decisions were based on the presence versus absence of audible voicing during the closure intervals. Another possibility is that decisions were based on gradual versus abrupt termination of the preceding vowels. Our strategy in this section was to try to make some speculations about the acoustic bases of our listeners' voiced–voiceless judgments based on acoustic analysis of the stimuli. In particular, we tried to find systematic acoustic differences between stimuli that straddled the voiced–voiceless boundary. Three parameters were examined: (1) presence versus absence of closure voicing, (2) the offset frequency of the first formant, and (3) rate of intensity decay at stimulus offset. For all of these analyses, we compared measurements from the *most edited* stimulus that produced 75% or more voiced responses with the *least edited* stimulus that produced 75% or more voiceless responses. Since these represent the most physically similar stimuli that produced qualitatively different labeling responses, we reasoned that comparing measurements from these pairs might reveal acoustic differences that listeners used to make voiced–voiceless judgments.

The first possibility that was considered was that listeners' judgments might be related to the presence versus absence of closure voicing. Closure voicing can be measured either from time- or frequency-domain representations. On a spectrogram, the disappearance of formants above $F1$ is typically used to mark the start of the closure inteval (e.g., Peterson and Lehiste, 1960).[5] If a closure interval is unvoiced, $F1$ and the higher formants will generally terminate at approximately the same time. This is simply another way of saying that the first formant or "voice bar" is not usually evident during the closure interval for unvoiced stops. However, if a closure interval is voiced, a low-intensity, low-frequency first formant or "voice bar" will usually extend beyond the termination of the higher formants. For the present study closure voicing was defined as the interval between the offset of the second formant and the offset of the first formant. This spectrum-based measurement technique was preferred to a time-domain method because it allowed us to evaluate the possibility that the resolution of temporal order might be involved in the perception of final stop voicing contrasts. For example, it is possible that a final stop will tend to sound voiced if the termination of $F1$ is delayed by a significant amount relative to the offset of the upper formants. Conversely, a final stop might sound unvoiced if $F1$ and the higher formants terminate at about the same time. If this hypothesis is correct, then stimuli that listeners tend to hear as voiced should show asynchronous offsets (i.e., the termination of $F1$ is delayed relative to the upper formants) while stimuli that listeners tend to hear as unvoiced should show synchronous offsets (i.e., $F1$ and the higher formants terminate at roughly the same time).

Measurements were based on extraction of formant frequency, amplitude, and bandwidth by linear predictive coding (LPC) analysis (Markel and Gray, 1976). These parameters were extracted every 6.4 ms over 20-ms Hamming-windowed segments. The digital spectrograms produced with this technique were used to measure the offset time of the first formant relative to the offset time of the second formant. This can be thought of either as a measurement of the amount of voicing present during the closure interval or as a measurement of relative offset time. The decision regarding the termination of a formant was based on a combination of an increase in formant bandwidth, a decrease in formant level, and the presence of abrupt discontinuities with previously extracted formant frequencies. Gross errors in formant extraction were avoided by checking the LPC derived measurements against conventional spectrograms (Kay Sonagraph, model 6061B).

The results of these measurements are shown in Table III.[6] The results do not support the idea that subjects' voicing decisions were based primarily on relative offset time or, stated differently, on the presence versus absence of closure voicing. In most cases $F1$ and $F2$ terminated simultaneously for stimuli on both sides of the voiced–voiceless boundary. Pastore (1983) has shown that the difference limen for offset asynchronies is about 7 ms for nonspeech stimuli 300 ms in duration. If this is a reasonable estimate for speech sounds, comparisons from only two of the six continua would show audible differences in relative offset time.

The second parameter that was measured was the terminating frequency of the first formant. Measurements of $F1$ offset frequency were made using the LPC analyses described above. Table III gives $F1$ offset frequencies for stimuli on opposite sides of the voicing boundary for each of the six continua. In each comparison $F1$ offset frequency is higher for the stimulus on the voiceless side of the boundary. However, the effect is much larger for the vowels with high-frequency first formants. $F1$ offset frequency differences ranged from 76 to 258 Hz for stimuli in the environment of /ɛ/ and /a/ but were only 17 Hz for /i/ and 23 Hz for /u/. In fact, the first-formant contours for the /pig/ and /pug/ stimuli were very nearly flat; consequently, little variation in $F1$ offset frequency would be expected regardless of the location of the editing cuts. Although the appropriate psychophysical data are not available, it seems unlikely that $F1$ offset differences of 17 and 23 Hz could be responsible for the large differences in labeling responses.[7]

TABLE III. Acoustic measurements from stimuli on opposite sides of the voiced–voiceless boundary for the six continua. The first column gives the percentage of voiced responses that were elicited by the stimulus. The values in parentheses are estimated difference limens for decay time (see text). Stimulus labels indicate the continuum from which the token was drawn and the size of the editing cut. For all of the analyses, comparisons were made between the most edited stimulus that was heard as voiced at least 75% of the time and the least edited stimulus that was heard as voiceless at least 75% of the time.

Stimulus	Percent voiced responses	Closure voicing	$F1$ offset frequency	Decay time
pɛb50	93	26	226	48 (16)
pɛb80	25	0	395	20
pɛd60	83	6	227	34 (13)
pɛd90	21	0	485	16
pɛg60	89	0	305	29 (12)
pɛg90	14	0	486	16
pag60	87	0	334	49 (16)
pag80	10	0	410	27
pig50	90	0	232	59 (18)
pig70	20	0	249	40
pug50	82	13	159	73 (21)
pug70	23	0	182	54

The last parameter considered was intensity decay time. A computer program was used to measure overall rms intensity every 6.4 ms over 20-ms time windows. Following the procedure recommended by van den Broecke and van Heuven (1983), amplitude decay time was expressed as the time needed for signal intensity, expressed in decibels, to drop from 90% of its peak value to 10% of its peak value. Table III shows amplitude decay times for stimuli on opposite sides of the voicing boundary for the six continua. In all cases the stimulus that subjects tended to hear as voiceless showed a more abrupt drop in intensity. On the average, the decay times for stimuli on the voiced side of the boundary were about 1.7 times longer than those on the voiceless side. Although this effect was quite consistent, it is not immediately clear whether these decay time differences are audible. Van den Broecke and van Heuven (1983) have shown that the difference limen (DL) for decay time can be predicted quite accurately ($r = 0.99$) from reference decay time by

$$\mathrm{DL} = 0.2t + 6.07,$$

where t is reference decay time. The decay time differences measured from our stimuli exceed difference limens estimated from this formula for all comparisons except for the stimuli drawn from the/pug/ series. The difference in decay time for this pair was 19 ms, compared to an estimated difference limen of 21 ms. For the /pig/ continuum, however, the measured decay time difference exceeded the estimated DL by only 1 ms.

IV. GENERAL DISCUSSION

To summarize briefly, subjects were asked to identify computer edited CVCs ending in stops. In general, subjects tended to change from hearing voiced stops to hearing voiceless stops when the closure interval and a portion of the vowel-to-consonant transition had been removed. Removal of the final bursts did not significantly alter the identification functions. Experiment 1 showed a small effect for place of articulation. However, since this effect was not replicated in experiment 2, it is not clear whether the place of articulation of the final stop is an important factor. The effect for vowel environment was statistically significant, but it is not clear how this finding should be interpreted. The two continua that were based on vowels with relatively low first-formant frequencies —/i/ and /u/— changed from voiced to voiceless 10–15 ms earlier than stimuli in the environment of the two high *F* 1 vowels. Given the possibility that *F* 1 offset frequency might be a voicing cue, we had anticipated a vowel context effect in the opposite direction. We reasoned that if a low *F* 1 offset frequency was a cue to voicing, an inherently low first formant might favor voiced responses. The results did not confirm this prediction: stimuli in the context of low *F* 1 vowels showed slightly earlier crossovers from voiced to voiceless. It should be kept in mind, however, that each continuum was based on a single naturally produced token. For that reason, what appears to be a systematic vowel context effect might simply reflect idiosyncratic characteristics of the particular tokens. We are currently designing a synthetic speech study to examine vowel effects in more detail.

Acoustic analysis of stimuli straddling the voiced–voiceless boundary provided few definitive answers regarding cues to final stop voicing contrasts. However, the results of these measurements do allow some tentative conclusions. Based on the measurements of closure voicing, it does not appear as though a voiced closure interval is required for subjects to report hearing a voiced stop. This finding may provide at least a partial explanation for the tendency of speakers to devoice relatively large amounts of closure intervals of syllables ending in phonologically voiced stops (Smith, 1979; Hogan and Rozsypal, 1980; Flege and Brown, 1982). It appears as though the tendency of speakers to devoice is supported by perceptual strategies on the part of listeners that do not rely on the presence versus absence of closure voicing.

Acoustic analyses were also used to investigate the possibility that subjects' voicing decisions were based on the manner of termination of the preceding vowel. The two manifestations of gradual versus abrupt vowel termination that we investigated were the offset frequency of the first formant and amplitude decay time. When stimuli on opposite sides of the voicing boundary were compared, large differences in *F* 1 offset frequency were seen only for the stimuli with relatively high-frequency first formants. This suggests that a relatively high *F* 1 offset frequency is not a necessary cue to the perception of a voiceless stop. It is possible, however, that *F* 1 offset frequency plays a role in some vowel environments but not others. A similar pattern was seen in the decay time measurements. For all six continua, the stimulus on the voiceless side of the boundary showed a more abrupt drop in amplitude than the corresponding stimulus on the voiced side of the boundary. However, in some vowel environments the differences were quite small in relation to estimated DLs for decay time. Again, the /i/ and /u/ environments showed relatively small differences.

Taken as a whole, the acoustic analyses were not as definitive as we had hoped: no single acoustic dimension separated stimuli on opposite sides of the voicing boundary for all six continua. By itself this kind of outcome should not come as a surprise. Many investigators in the phonetic perception area are accustomed to thinking less in terms of unitary cues to a contrast and more in terms of trading relations among a number of cues (see Repp, 1982 for a review). The results of the present study, however, do not appear to lend themselves well to this kind of description. For example, it is possible to view abrupt vowel termination as a perceptual dimension that can be cued by a high *F* 1 offset frequency, or by a short decay time, or by some appropriate combination of these two properties. Recall that the /pig/ and /pug/ continua showed very small differences in *F* 1 offset frequency. On the assumption that *F* 1 offset frequency and decay time trade against one another, one would expect to find that the lack of contrast in *F* 1 offset frequency was "compensated" by a relatively large contrast in decay time. This was clearly not the case since the differences in decay time for the /pig/ and /pug/ series were at best barely audible. The trade-off approach could, of course, be extended one step further by suggesting that gradual versus abrupt vowel termination trades against another dimension, such as relative offset time. A trading relation approach would suggest that the relatively small contrast in *F* 1 offset frequency and decay time would be compensated by relatively large differences in relative offset time. The data do not support this hypothesis since there was no difference in voice offset time for the /pig/ series and only a 13-ms difference of the /pug/ series.

One additional issue that deserves attention is the relation between voicing cues for initial stops and those for final stops. It has been suggested that the perception of voicing contrasts in initial stops is based on the resolution of relative onset time (Pisoni, 1977; see also Miller *et al.*, 1976; Summerfield, 1982; Hillenbrand, 1984). Based on the present results, it does not seem likely that relative offset time is strongly involved in the perception of final stop voicing contrasts. There were many instances in which stimuli were labeled differently without showing differences in relative offset time. It is clear, however, that relative onset time is not the only cue to initial stop voicing contrasts. Differences in the onset frequency of the first formant also exert a strong influence on initial stop voicing judgments. In general, low *F* 1 onset frequencies tend to favor the perception of voiced stops and high *F* 1 onset frequencies tend to favor the perception of voiceless stops (Lisker, 1975; Summerfield and Haggard, 1977). The obvious final position analog of *F* 1 onset frequency would be *F* 1 offset frequency. Wolf has speculated that, "the amount of low-frequency energy in the vicinity of onset or offset may serve as a common basis for the perception of the voicing distinction in initial and final position" (1978, p. 299). Wolf's conclusions, however, were based largely on experiments using stops following /ae/, a vowel with a high-frequency first formant. The present results suggest that $F1$ offset frequency differences may play little or no role when a stop follows a vowel with a low-frequency first formant.

ACKNOWLEDGMENT

This work was supported by research funds from NIH grants T32NS07100-04 and 5S05RR07028.

[1] Although voicing-dependent vowel duration differences have received a good deal of attention in the literature, it is important to emphasize that large effects are seen primarily in phrase-final position and in isolated words. These differences are much smaller in nonphrase-final position (Umeda, 1975; Klatt, 1973, 1975, 1976). Under certain conditions these durational differences disappear altogether in connected speech (Crystal and House, 1982).

[2] To simplify this review, we are restricting attention to voicing contrasts involving postvocalic stop consonants. Other studies have examined the role of vowel duration in final fricative voicing contrasts (e.g., Denes, 1955; Derr and Massaro, 1980; Hogan and Rozsypal, 1980; Port and Dalby, 1982; Massaro and Cohen, 1983) and of stops preceded by nonvocalic segments (Raphael *et al.*, 1975).

[3] It is very likely that vowel duration plays a role in the perception of final stop voicing when other voicing cues are ambiguous. A recent study by Wardrip-Fruin (1982b) showed that vowel duration becomes an important voicing cue when stimuli are presented in noise.

[4] No attempt was made to separate the steady-state vowel from the vowel-to-consonant transitions. The most difficult segmentation decision was the boundary between the vowel offset and the start of the stop closure interval. This decision was based on a drop in signal intensity and a smoothing of the waveform, indicating that most of the radiated energy is at or near the voice fundamental.

[5] Attenuation is perhaps a better word than disappearance. Sweep-tone measurements by Fujimura and Lindqvist (1971) have shown that a second formant peak is clearly present in the vocal-tract transfer function during stop occlusion (see also Fant, 1962, p. 13). However, the level of $F2$ is some 10 dB lower than $F1$. This fact, in combination with the low-frequency bias of the glottal source spectrum, means that only the first formant is visible on a spectrogram during the period of time corresponding to articulatory occlusion.

[6] The measurements of closure voicing given in Table III will not agree with values obtained simply by subtracting the nominal amount of "voicing removed" from the unedited closure durations reported in Table I. There are two reasons for this discrepancy. First, as indicated in Sec. I A the nominal stimulus values refer only to the zero-amplitude portion of the weighting functions that were used in the editing procedure. Since the nominal stimulus values do not take into account the 15-ms linear decay function, these values will underestimate the amount of closure voicing that was attenuated. Second, the closure voicing measurements given in Table I were based on oscillograms while those in Table III were based on LPC derived digital spectrograms. (Reasons for switching to the frequency-domain technique are explained in Sec. III.) These two methods should produce similar but not necessarily identical results.

[7] Since data are not avilable on the discrimination of $F1$ offset frequency differences, speculations about audibility must rely on difference limen data using stimuli with stationary formants. The smallest estimate of the difference limen for first-formant frequency is 15 Hz, based on a reference signal with a 300-Hz $F1$ (Flanagan, 1955). Mermelstein (1978), however, used similar stimuli and found a 50-Hz difference limen for $F1$. Even using the smaller estimate provided by Flanagan, the 17-and 23-Hz differences measured from our stimuli would be just detectable in a same/different task.

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

Crystal, T. H., and House, A. S. (1982). "Segment durations in connected speech signals: Preliminary results," J. Acoust. Soc. Am. 72, 705–716.

Denes, P. (1955). "Effect of duration on the perception of voicing," J. Acoust. Soc. Am. 27, 761–764.

Derbock, M. (1977). "An acoustic correlate of the force of articulation," J. Phon. 5, 61–80.

Derr, M. A., and Massaro, D. M. (1980). "The contribution of vowel duration, $F0$ contour, and fricative duration as cues to the /juz/-jus/ distinction," Percept. Psychophys. 27, 51–59.

Fant, C. G. M. (1962). "Descriptive analysis of the acoustics of speech," LOGOS 5, 3–17.

Flanagan, J. L. (1955). "A difference limen for vowel formant frequency," J. Acoust. Soc. Am. 27, 613–617.

Flege, J., and Brown, W. S. (1982). "The voicing contrast between English /p/ and /b/ as a function of stress and position-in-utterance," J. Phon. 10, 335–345.

Fujimura, O., and Lindqvist, J. (1971). "Sweep-tone measurements of vocal-tract characteristics," J. Acoust. Soc. Am. 49, 541–558.

Hillenbrand, J. (1984). "Perception of sine-wave analogs of voice onset time stimuli," J. Acoust. Soc. Am. 75, 231–240.

Hogan, J., and Rozsypal, A. (1980). "Evaluation of vowel duration as a cue for the voicing distinction in the following word-final consonant," J. Acoust, Soc. Am. 67, 1764–1771.

House, A. (1961). "On vowel duration in English," J. Acoust. Soc. Am. 33, 1174–1178.

House, A. S., and Fairbanks, G. (1953). "The influence of consonant environment upon the secondary acoustical characteristics of vowels," J. Acoust. Soc. Am. 25, 105–113.

Klatt, D. H. (1973). "Interaction between two factors that influence vowel duration," J. Acoust. Soc. Am. 54, 1102–1104.

Klatt, D. H. (1975). "Vowel lengthening is syntactically determined in a connected discourse," J. Phon. 3, 129–140.

Klatt, D. H. (1976). "Linguistic uses of segmental duration in English: Acoustic and perceptual evidence," J. Acoust. Soc. Am. 59, 1208–1221.

Lisker, L. (1975). "Is it VOT or a first-formant transition detector?, "J. Acoust. Soc. Am. 57, 1547–1561.

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," Lang. Speech 10, 1–28.

Lisker, L., and Abramson, A. S. (1970). "The voicing dimension: Some experiments in comparative phonetics," in *Proceedings of the 6th International Congress of Phonetic Sciences, Prague, 1967* (Academia, Prague), pp. 563–567.

Malécot, A. (1958). "The role of releases in the identification of released final stops," Lang. 34, 370–880.

Malécot, A. (1970). "The lenis-fortis oppostion: Its physiological parameters," J. Acoust. Soc. Am. 47, 1588–1592.

Markel, J., and Gray, A. (1976). *Linear Prediction of Speech* (Springer-Verlag, New York).

Massaro, D. M., and Cogen, M. M. (1983). "Consonant–vowel ratio: An improbable cue in speech," Percept. Psychophys. 33, 501–505.

Mermelstein, P. (1978). "Difference limens for formant frequencies of steady-state and consonant-bound vowels," J. Acoust. Soc. Am. 63, 572–580.

Miller, J. D., Wier, C. C., Pastore, R. E., Kelly, W. J., and Dooling, R. J. (1976). "Discrimination and labeling of noise-buzz sequences with varying noise-lead times: An example of categorical perception," J. Acoust. Soc. Am. 60, 410–417.

O'Kane, D. (1978). "Manner of vowel termination as a perceptual cue to the voicing status of post-vocalic stop consonants," J. Phon. 6, 311–318.

Parker, F. (1974). "The coarticulation of vowels and stop consonants," J. Phon. 2, 211–221.

Pastore, R. E. (1983). "Temporal order judgment of auditory stimulus offset," Percept. Psychophys. 33, 54–62.

Peterson, G., and Lehiste, I. (1960). "Duration of syllable nuclei in English," J. Acoust. Soc. Am. 32, 693–703.

Pisoni, D. B. (1977). "Identification and discrimination of the relative onset time of two-component tones: Implications for voicing perception in stops," J. Acoust. Soc. Am. 61, 1352–1361.

Port, R. F., and Dalby, F. (1982). "Consonant/vowel ratio as a cue for voicing in English," Percept. Psychophys. 32, 141–152.

Raphael, L. J. (1972). "Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English," J. Acoust. Soc. Am. 51, 1296–1303.

Raphael, L. J. (1981). "Durations and contexts as cues to word-final cognate opposition in English," Phonetica 38, 126–147.

Raphael, L. J., Dorman, M., Freeman, F., and Tobin, C. (1975). "Vowel and nasal duration as cues to voicing in word-final stop consonants: Spectrographic and perceptual studies," J. Speech Hear. Res. **18**, 389–400.

Raphael, L. J., Dorman, M. F., and Liberman, A. M. (1980). "On defining the vowel duration that cues voicing in final position," Lang. Speech **23**, 297–307.

Repp, B. H. (1982). "Phonetic trading relations and context effects: New evidence for a phonetic mode of perception," Psychol. Bull. **92**, 81–110.

Revoile, S., Pickett, J. M., Holden, L. D., and Talkin, D. (1982). "Acoustic cues to final stop voicing for impaired- and normal-hearing listeners," J. Acoust. Soc. Am. **72**, 1145–1154.

Rositzke, H. A. (1943). "The articulation of final stops in general American speech," Am. Speech **18**, 32–42.

Smith, B. L. (1979). "A phonetic analysis of consonant devoicing in children's speech, " J. Child Lang. **6**, 19–28.

Summerfield, Q. (1982). "Differences between spectral dependencies in auditory and phonetic temporal processing: Relevance to the perception of voicing in initial stops," J. Acoust. Soc. Am. **72**, 51–61.

Summerfield, Q., and Haggard, M. P. (1977). "On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants," J. Acoust. Soc. Am. **62**, 435–448.

Umeda, N. (1975). "Vowel duration in American English," J. Acoust. Soc. Am. **58**, 434–445.

van den Broecke, M. P. R., and van Heuven, V. J. (1983). "Effect and artifact in the auditory discrimination of rise and decay time," Percept. Psychophys. **33**, 305–313.

Walsh, T., and Parker, F. (1981). "Vowel termination as a cue to voicing in post-vocalic stops," J. Phon. **9**, 105–108.

Wardrip-Fruin, C. (1982a). "On the status of temporal cues to phonetic categories: Preceding vowel duration as a cue to voicing in final stop consonants," J. Acoust. Soc. Am. **71**, 187–195.

Wardrip-Fruin, C. (1982b). "Final stop voicing: Vowel duration the primary cue in noise," paper presented at the American Speech and Hearing Association Convention, Toronto, November, 1982.

Wolf, C. (1978). "Voicing cues in English final stops," J. Phon. **6**, 299–309.