VOICE TIMING FOR STOP CLASSIFICATION IN CONVERSATIONAL ENGLISH

Arthur S. Abramson and Leigh Lisker
Haskins Laboratories, New Haven, Connecticut, U.S.A.

ABSTRACT

Acoustic research demonstrating the role of the temporal control of the glottis in separating English "voiced" and "voiceless" stops has mostly used citation forms. We have examined stops in two conversations. For each stop we see whether the voice pulsing is interrupted. If there is a break, we measure the duration of the break with reference to the articulatory release. The results show that temporal control is quite robust even in running speech.

BACKGROUND

Considerable earlier acoustic [e.g., 1], perceptual [e.g., 2] and physiological or articulatory [e.g., 3] work by us and many others [e.g., 4] has demonstrated the significance of the temporal control of the valvular action of the larynx for the division of English stop consonants into the traditional "voiced" and "voiceless" categories. These studies have mostly examined rather deliberate speech: citation forms and short expressions and sentences read aloud.

Although the term "voice onset time" (VOT) has come to be widely used, it was meant by us in the first place to refer to utterance-initial position. Thus, laryngeal pulsing might begin at the moment of closure-release (zero time), before it (voicing lead with time in negative units), or after it (voicing lag in positive units). For widespread varieties of English, with special reference to American English in our work, initial /bdg/ normally have zero-onset of voicing or very short lags of 10 ms or so, although some speakers show voicing lead. Utterance-initial /ptk/ commonly have a rather long voicing lag of some 30 to 40 ms.

The temporal dimension is, of course, not linear in its acoustic manifestations. Voicing lead appears as excitation of the first one or two harmonics during the articulatory closure. Voicing lag appears as noise-excitation of both the release-burst and as much of the formant-pattern as emerges until the onset of pulsing. If the lag is long enough, the turbulence and somewhat attenuated first formant will be heard as aspiration. Experiments with speech synthesis and manipulated natural speech have shown that some several acoustic consequences of voice timing can serve as perceptual cues to the phonological distinction in context-free experiments.

In running speech, with stops occurring immediately after vowels or other consonants, as well as after pauses, the concept of VOT should be broadened to that of "laryngeal timing" or maybe "voice timing" [5]. Tokens of /bdg/ after other voiced consonants or vowels are very likely to have unbroken glottal pulsing in their closures, while /p/ and /k/ before unstressed syllables often have such short voicing breaks as to be heard as unaspirated. (In the latter context "underlying" /t/, as well as /d/, commonly appears in American English as a voiced flap.).

Limiting ourselves for now to the acoustic signal, we wish to assess the stability of the temporal factor in spontaneous fluent English. This is part of our larger interest in the robustness in casual running speech of the differentiating properties and perceptual cues that have long been known for citation forms and careful speech.

PROCEDURE

We recorded about ten minutes of spontaneous conversation held in separate sessions by each of two couples, All four people were native speakers of American English whose minor differences in regional dialects in no way impeded communication. In each couple the man and woman knew each other well and were quite used to talking into microphones; moreover, they were quite at ease with us. Each couple chatted in a relaxed way about personal and professional topics of their own choice without knowing anything of our research goals. Listening to the recordings, we found the conversations intelligible, spontaneous, fluent, and friendly.

After digitizing the recorded speech at 22Kh, we used the Signalyze™ computer program to obtain waveforms and FFT spectrograms. For each conversation, omitting all instances of overlap between speakers and distortions caused by coughing, laughter, and the like, we picked out for analysis all acoustically measurable tokens of the six stops in stressed position, as well as all measurable tokens of /bpgk/ in unstressed position. That is, we excluded the voiced flaps so typical of American English, because any residual contrast in this context between /d/ and /t/ seems to depend upon properties other than voice timing, such as vowel length and quality. We did not include the few instances in our corpus of stops under emphatic stress. We also excluded stops in consonant clusters with /s/ as the first member; here there is clearly no voicing contrast. As for stress, anything not unstressed was taken to be stressed without any attempt at finer gradations.

For each instance of a stop we recorded data on glottal pulsing in the vicinity of the closure and release. With no interruption in pulsing, the item was called "unbroken." An interruption before the release was called a "negative break," and one after the release was a "positive break." The durations of these breaks were measured in the waveforms with reference to the spectrograms. Negative breaks were measured only if there were clear spectral signs of an acoustic discontinuity before the closure with no indication, acoustic or auditory, of a pause. Thus, a stop in utterance-initial position or preceded by a pause could never have a negative break. Also, if a negative break included the closure of a preceding stop, it was not measured. For each stop a "full break" was also entered in our data, whether this was the sum of negative and positive breaks or just the duration of the only one of them available in the utterance.

It is not surprising that in our randomly produced corpus of speech, the stop consonants were unevenly represented across the categories. As a result, we used unpaired two-tailed *t*-tests for assessment of statistical significance.

There were not enough tokens for us to focus on narrower segmental and prosodic contexts. We have postponed any statistical treatment of our two levels of stress.

RESULTS

The means and standard deviations of the full voicing breaks in ms for all four speakers are shown in Figure 1. The average voicing break of the voiceless stops is indeed higher, but there is much overlap of the standard deviations. To this we must add the observation that 84 voiced stops, 62% of that category, had unbroken voicing. They do not appear in the figure.

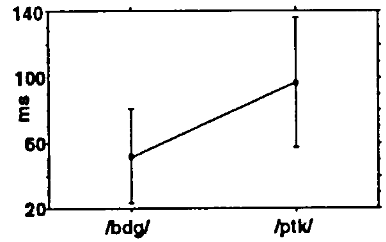

Figure 1. Full voicing breaks: Means and standard deviations for the pooled data of all four speakers. Voiced n= 51; voiceless n = 276.

The data of Figure 1 are broken down into the four speakers in Table 1. Here we see that the difference is significant for all four speakers, although the level is lower for DS and JH. It is interesting to note that the voiceless stops outnumbered the voiced ones by far.

The means and standard deviations of the negative voicing breaks for all four speakers are shown in Figure 2. The data are given separately for the speakers in Table 2, where we can see that while the differences are highly significant for MC and JH, they are not significant for the other two, DS and DL. As for the latter two, however, it must be borne in mind that they do show significant differences in Table 1, so it will be important to see how they fare with positive breaks.

Table 1. Full voicing breaks in ms: Means, standard deviations, and significance levels for the four speakers' unpaired t-tests.

Spkr:	DS	DL	MC	JH
/bdg/				
M	72	62	30	45
SD	27	25	23	18
n	8	22	15	6
/ptk/				
M	106	91	109	82
SD	34	31	49	34
n	41	112	79	44
df	47	132	92	48
t	-2.6	-4.1	-6.1	-2.6
p <	.02	.001	.001	.02

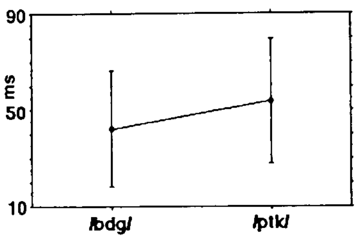

Figure 2. Negative voicing breaks: Means and standard deviations for the pooled data of all four speakers. Voiced n=45; voiceless n=242.

Table 2. Negative voicing breaks in ms for four speakers.

Spkr:	DS	DL	MC	JH
/bdg/				
M	61	48	27	25
SD	15	24	19	19
n	7	22	10	6
/ptk/				
M	51	45	67	54
SD	21	18	33	21
n	40	112	79	44
df	45	132	87	48
t	1.1	-.7	-3.7	-3.2
p <	.3, ns	.5, ns	.001	.003

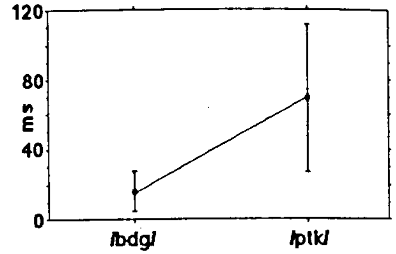

Figure 3. Positive voicing breaks in ms for four speakers. Voiced n=62; voiceless n=293.

Finally, the means and standard deviations of the positive voicing breaks for all four speakers are shown in Figure 3. The data are given separately for the speakers in Table 3. For one of the speakers, JH, the difference is barely significant; for the other three, however, it is highly significant. This also accounts for the significant difference between the full breaks of DS and DL found in Table 1.

Table 3. Positive voicing breaks: Means and standard deviations of the pooled data for the four speakers. Voiced n=62; voiceless n=293.

Spkr:	DS	DL	MC	JH
/bdg/				
M	20	16	12	20
SD	13	12	6	19
n	9	32	15	6
/ptk/				
M	55	45	43	39
SD	19	22	28	23
n	46	125	78	44
df	53	155	91	48
t	-5.4	-7.2	-4.3	-1.9
p <	.001	.001	.001	.06,?

DISCUSSION

In the history of speech research certain acoustic properties have been found to have the power to differentiate the phonemes of languages in the production of citation forms or other short utterances. Our research was motivated by a desire to investigate the stability of one of those properties, voice timing, for the distinction between voiced and voiceless stop consonants in American English spontaneous speech.

What with all the contextual redundancy and top-down information present in running speech, one might expect the phonetic rendition of many phonemic distinctions to be somewhat less precise than in more deliberate speech. That is, with so much other information in the discourse, clarity of expression moment by moment ought to be less important. Indeed, just the great temporal variation often observed might blur some distinctions, especially, perhaps, those that include temporal control as an important mechanism.

Despite all the pressures to which such a distinction as consonantal voicing might be vulnerable in running speech, our findings support the general robustness of temporal control of the larynx as an important factor in voicing distinctions in spontaneous conversation. Some generalizations emerge from our sampling of four speakers.

Once the flaps, with their allegedly underlying /d/ and /t/, are eliminated from consideration, it is only the voiced stops that show unbroken pulsing in non-initial position. Thus it is that in our corpus just over 60% of the instances of /bdg/ are distinguished from /ptk/ by this factor alone. As for the rest, relative duration of the voicing break in the region of the closure and release does a rather good job of separating the categories. Even without taking our two levels of stress into account, we find that the voiceless stops have longer voicing breaks than the voiced stops. In addition, it appears that breaks after the articulatory release (positive breaks) bear more of the burden than breaks before the release (negative breaks). Our data are insufficient for examination of narrower phonetic contexts, such as particular vowels.

A preliminary look suggests that a quantitative treatment of the differences linked to stress will remove some of the overlap remaining between the two voicing categories. We plan to do this. Furthermore, we are planning perceptual tests of the validity of our findings.

ACKNOWLEDGMENT

This research was supported by NIH Grant HD-01994 to the Haskins Laboratories. In addition, facilities and help were given to the authors by their respective universities, The University of Connecticut and The University of Pennsylvania.

REFERENCES

[1] Lisker, L. & Abramson, A.S. (1964), "A cross-language study of voicing in initial stops: Acoustical measurements," *Word*, vol. 20, pp. 384–422.
[2] Abramson, A.S. & Lisker, L.(1970), "Discriminability along the voicing continuum: Cross-language tests," In *Proceedings of the 6th International Congress of Phonetic Sciences, Prague* (pp. 569–573). Prague: Academia.
[3] Sawashima, M., Abramson, A.S., Cooper, F.S., & Lisker, L. (1970), "Observing laryngeal adjustments during running speech," *Phonetica*, vol. 22, pp. 193–201.
[4] Löfqvist, A. (1980. "Interarticulator programming in stop production," *Journal of Phonetics*, vol. 8, pp. 475–490.
[5] Abramson, A.S. (1977). "Laryngeal timing in consonant distinctions," *Phonetica*, vol. 34, pp. 295–303.