STUDIA PHONOLOGICA XI (1977)

Voicing Features in the Perception and Production of Stop Consonants by Japanese Speakers

Katsumasa SHIMIZU

I. Introduction

Since it has become possible to control acoustic dimensions in speech synthesis, various experiments of speech perception have been carried out on the identification and discrimination of vowels and consonants. Among such experiments, the ones for consonants are mainly concerned with voicing features such as voiced and voiceless and place features such as labials, alveolars, and velars. The present study is to examine voicing features in the perception and production of the Japanese speakers.

It is generally known that the distinction between voiced and voiceless stop consonants is made not only by the existence or absence of laryngeal pulsing but also by voice onset time (VOT). VOT is generally defined as the interval between the release of the articulators and the onset of laryngeal pulsing. Lisker and Abramson (1964) extensively studied the VOT values of the initial stop consonants in 11 languages and demonstrated three major modes of voicing in terms of the VOT values; voiced unaspirated stops, voiceless unaspirated stops, and voiceless aspirated stops[1]. Languages such as Japanese and English use two modes of voicing, while languages such as Thai and Eastern Armenian use three modes of voicing, as mentioned above. Although VOT is recognized to be one of the significant features in stop consonants, it is not applicable to distinguish voiced aspirated stops from voiced unaspirated stops in such languages as Marathi and Hindi.

In the synthesis of speech sounds, the acoustic correlate to VOT is manifested by the cutback of the onset of the first formant (F-1) relative to the higher formants in the spectrograms[2]. Liberman et al. (1958) showed that the successive F-1 cutback effectively converted voiced into voiceless stop consonants. By using synthetic speech sounds, some experiments have been made to study the perceptual relevance of major categories of stop consonants. Lisker and Abramson (1970) studied the identification modes of English, Spanish, and Thai speakers by using synthetic speech sounds with continuously varied VOT.

Among experiments of speech perception involving voicing features, some

Katsumasa SHIMIZU (清水克正): Instructor, Dept. of Linguistics, Kyoto Univ., and also a faculty member of Nagoya Gakuin Univ. The author is doing research in linguistics under the direction of Prof. Tatsuo Nishida.

recent issues are concerned with the nature of voicing feature detectors. Some neurological examinations are made on the feature detectors by Abbs and Sussman (1971). The detectors are neurological systems which are selectively responsive to specific features of stimuli and are supposed to clarify the decoding system in speech perception. It is assumed that perception of voicing features is mainly accomplished by the feature detectors which extract the VOT information from speech sounds. The existence of such feature detectors is usually attested by the selective adaptation test in which repetitive presentation of stimulus should fatigue the detectors and reduce the sensitivity, by which the boundary of phonetic categories in identification would be altered. Although there is no disagreement on the existence of such detectors among investigators, the problem on the detectors is at what stage in speech processing such detectors operate; that is, linguistic or auditory level. Eimas & Corbit (1973) assumed that such detectors are linguistic in nature, while Bailey (1973) and Pisoni & Tash (1975) assumed that they operate at low level acoustic information and therefore are auditory in nature. No convincing evidence has so far been presented to support either linguistic or auditory level. Another problem is whether the detectors operate at a mediating level of perception and production, and several experiments have been made to examine the articulatory effects on speech perception (Cooper, et al., 1975 & 1976).

Under such research trends of voicing features, the purposes of the present study are to examine the identification of voiced and voiceless consonants by the Japanese speakers by using the synthetic speech sounds of VOT continuum and to examine the adaptation effect of the voicing feature detectors; specifically to examine how repetitive articulation affects the speech perception.

II. EXPERIMENTAL METHOD

The experiment can be divided into three parts: 1) the experiment of identification, 2) the experiment by repetitive listening and 3) the experiment by repetitive articulation.

1) Subjects: Twelve Japanese speakers (2 female and 10 male subjects) took part in the experiment as a subject, ages 18~22. All subjects were undergraduate students of Nagoya Gakuin Univ. They had normal hearing ability and had no known neurological defects. For the experiment by repetitive listening, six speakers out of the above served as a subject. They were paid for their participation.

2) Stimuli

The stimuli consist of three sets of synthetic speech sounds prepared by a parallel resonance synthesizer at the Haskins Laboratories. The first set consists of 24 variants of /ba-pa/ series with the VOT continuum from −30 to +85 msec. The second set consists of 24 variants of /da-ta/ series with the VOT continuum

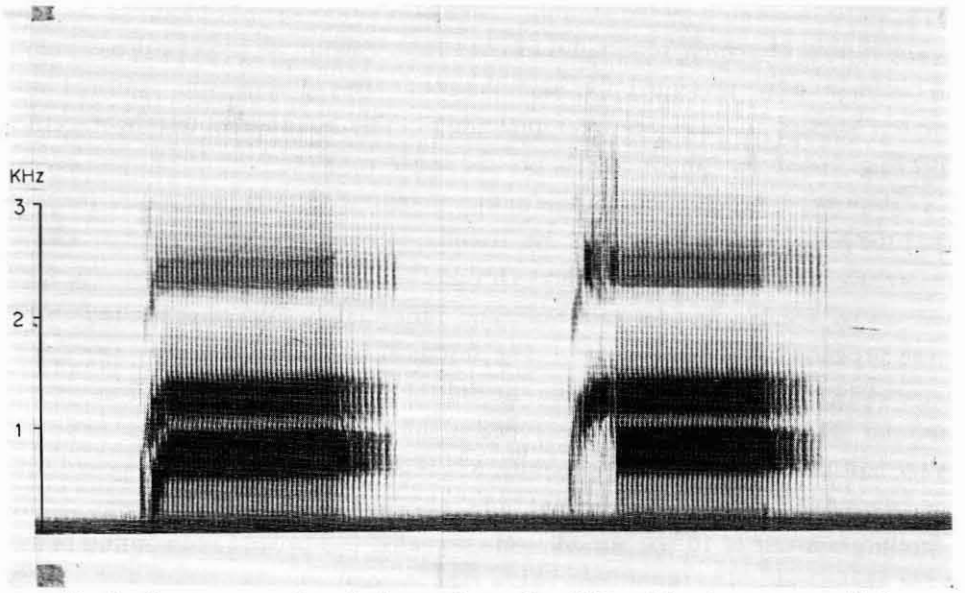

Fig. 1. Spectrograms of synthetic speech sounds: slight voicing lag represents [ba] and a long voicing lag represents [pʰa].

from −40 to +80 msec. The third set consists of 25 variants of /ga-ka/ series with the VOT continuum from −50 to +70 msec. The variations in VOT were made by varying the onset of the first formant relative to the onset of higher formants. The stimuli in each set vary in 5 msec steps. The stimuli were each 450 msec long.

3) Procedure

Identification: Subjects were asked about their hearing and neurological defects. They were tested at sound-proofed language laboratory booths of Nagoya Gakuin Univ. in two separate sessions. In these sessions, three sets of stimuli were presented in the order of /da-ta/, /ba-pa/, and /ga-ka/ series. The first and second sets consisted of 96 stimuli with 4 presentations of each syllable and the third set consisted of 100 stimuli with 4 presentations of each syllable. Stimuli were presented binaurally in the random order by a Sony taperecorder at a comfortable listening level. The interval between stimuli were about 3 seconds. The subjects were asked to identify the stimuli by circling either /ba/ or /pa/ in labial set, either /da/ or /ta/ in alveolar set, and either /ga/ or /ka/ in velar set in the response sheet. Some trials were given for practice to each subject. The identification tests lasted for about one hour in each session.

Repetitive Listening: The adapting stimuli in repetitive listening were taken from test stimuli and were the syllables which were clearly identified as /ba/, /da/, and /ga/. The subjects were tested one at a time in the sound-proofed IAC booth. In each set, the adapting stimulus was first presented repeatedly for two min. (about 100 presentations of the syllable) and immediately after this, subjects

were asked to identify 10 test stimuli. After this initial identification, the adapting stimulus was presented repeatedly for one min. (about 50 presentations) and 10 test stimuli were presented for identification. This process was repeated until all test stimuli in each set were presented. The adaptation test by repetitive listening lasted for about 1.5 hours for each subject.

Repetitive Articulation: Subjects were required to utter repeatedly [ba] and [pa] for labial set, [da] and [ta] for alveolar set, and [ga] and [ka] for velar set before identifying the test stimuli. In each set, each subject uttered repeatedly the voiced syllable first for one min. and immediately after this, 10 test stimuli were presented for identification. For instance, in /ga-ka/ set, after the subject repeatedly uttered [ga] for one min., he did the identification test of 10 test stimuli in velar set and this process was repeated until 50 test stimuli were presented. After half an hour, the same procedure was repeated to the voiceless /ka/ syllable; i.e., the subject uttered [ka] for one min. and immediately after this, he did the identification test of 10 test stimuli. He repeated this process until 50 test stimuli were presented. The same procedure was carried out for labial and alveolar sets.

III. RESULTS

The phonetic boundaries of the voiced and voiceless consonants for labials, alveolars, and velars were extrapolated for each subject from the responses in the experiment of identification. The boundary was defined as that point on the stimulus scale which would receive 50% responses from either category involved[3]. The results are shown in Table 1.

From Table 1, it can be found that some subjects such as YS and MI show

Table 1. The identification value for the distinction of voiced and voiceless stop consonants by the Japanese speakers (in milliseconds).

Subjects	ba–pa boundary	da–ta boundary	ga–ka boundary
SK	15 msec	32 msec	24 msec
NS	17	35	28
HT	14	32	30
HA	13	18	23
YS	23	20	24
KY	−20	12	22
FU	35	27	32
MI	27	27	25
TH	35	25	33
TY	15	25	28
HK	13	25	20
NSa	25	33	23
Mean	18	26	26

Table 2. Shift of the phonetic boundary in milliseconds of VOT for the adaptation experiment of the Japanese speakers.

Subjects	Base boundary /b−p/	Adaptation with [ba]	Base boundary /d−t/	Adaptation with [da]	Base boundary /g−k/	Adaptation with [ga]
SK	15 msec	00	32 msec	− 7	24 msec	−1
NS	17	+3	35	− 8	28	+4
HT	14	−2	32	−12	30	−8
HA	13	+7	18	− 3	23	−3
YS	23	−8	20	00	24	00
KY	−20	−5	12	+ 5	22	−5

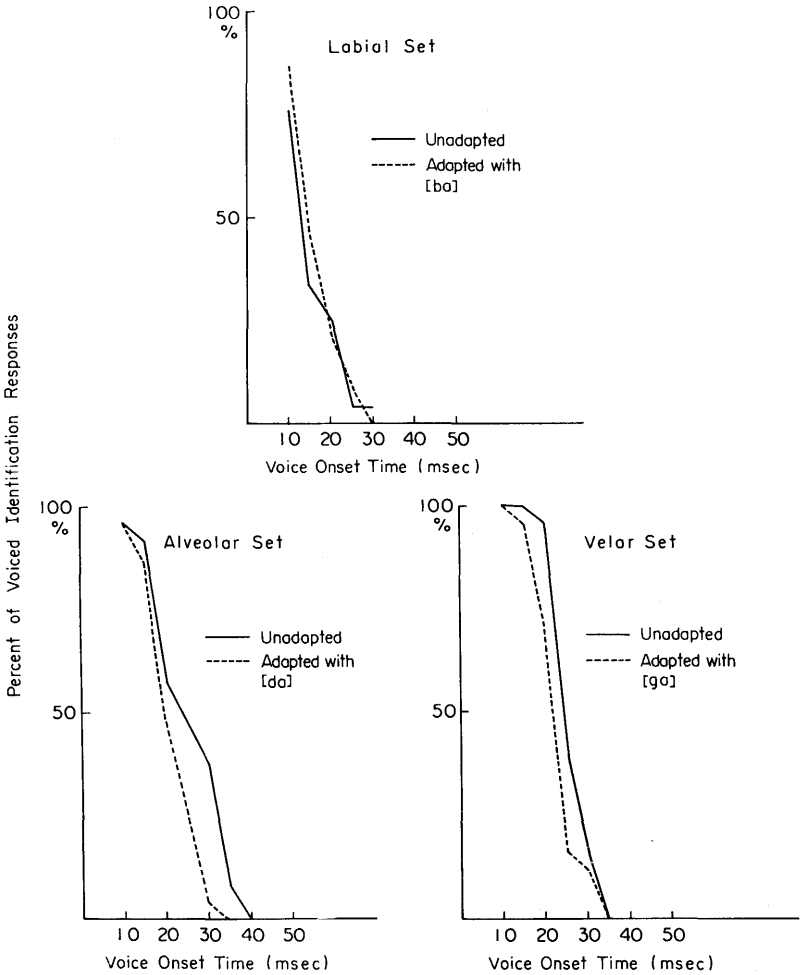

Fig. 2. Percentage of voiced identification responses obtained with and without adaptation for six Japanese subjects. The solid lines indicate the unadapted identification functions, while the dotted lines indicate the identification after adaptation with voiced stops.

similar values for the distinction of voiced and voiceless consonants in three sets, while other subjects such as HT, TY, and HK show similar ones for alveolar and velar sets. Furthermore, the identification of voiced and voiceless consonants in three sets was categorical, as generally predicted.

As mentioned in the experimental procedure, six subjects out of the above took part in the experiment by repetitive listening. The response data in the adaptation with voiced stop consonants were analyzed in the same manner as in the data of identification. The boundary shifts caused by adaptation are the differences between the base boundary (unadapted) and the adapted boundary. The results of six subjects are shown in Table 2.

The adapting stimuli were taken from test stimuli. They were −30 msec

Table 3. Shift of the phonetic boundary in milliseconds of VOT for the experiment by repetitive articulation.

a) b−p phonetic boundary

Subjects	Base boundary	Shift after [p] articulation	Shift after [b] articulation
SK	15 msec	− 2	− 5
NS	17	− 4	−12
HT	14	+11	− 9
HA	13	+ 7	+ 7
YS	23	+ 7	− 3
KY	−20	+20	+15
FU	35	−10	−15
MI	27	− 4	−14
TH	35	0	−12
TY	15	− 2	−10
HK	13	+ 2	+ 2
NSa	25	− 2	−10

b) d−t phonetic boundary

Subjects	Base boundary	Shift after [t] articulation	Shift after [d] articulation
SK	32	+ 1	−12
NS	35	0	0
HT	32	− 7	− 2
HA	18	+ 7	+ 2
YS	20	+10	+10
KY	12	+13	+ 6
FU	27	+ 8	− 2
MI	27	− 2	− 2
TH	25	+10	+ 8
TY	25	+ 5	0
HK	25	− 5	− 5
NSa	33	−15	−15

c) g–k phonetic boundary

Subjects	Base boundary	Shift after [k] articulation	Shift after [g] articulation
FU	32	+ 8	+ 3
MI	25	+ 3	0
TH	33	0	0
TY	28	− 5	− 8
HK	20	+ 8	− 5
NSa	23	+ 5	0

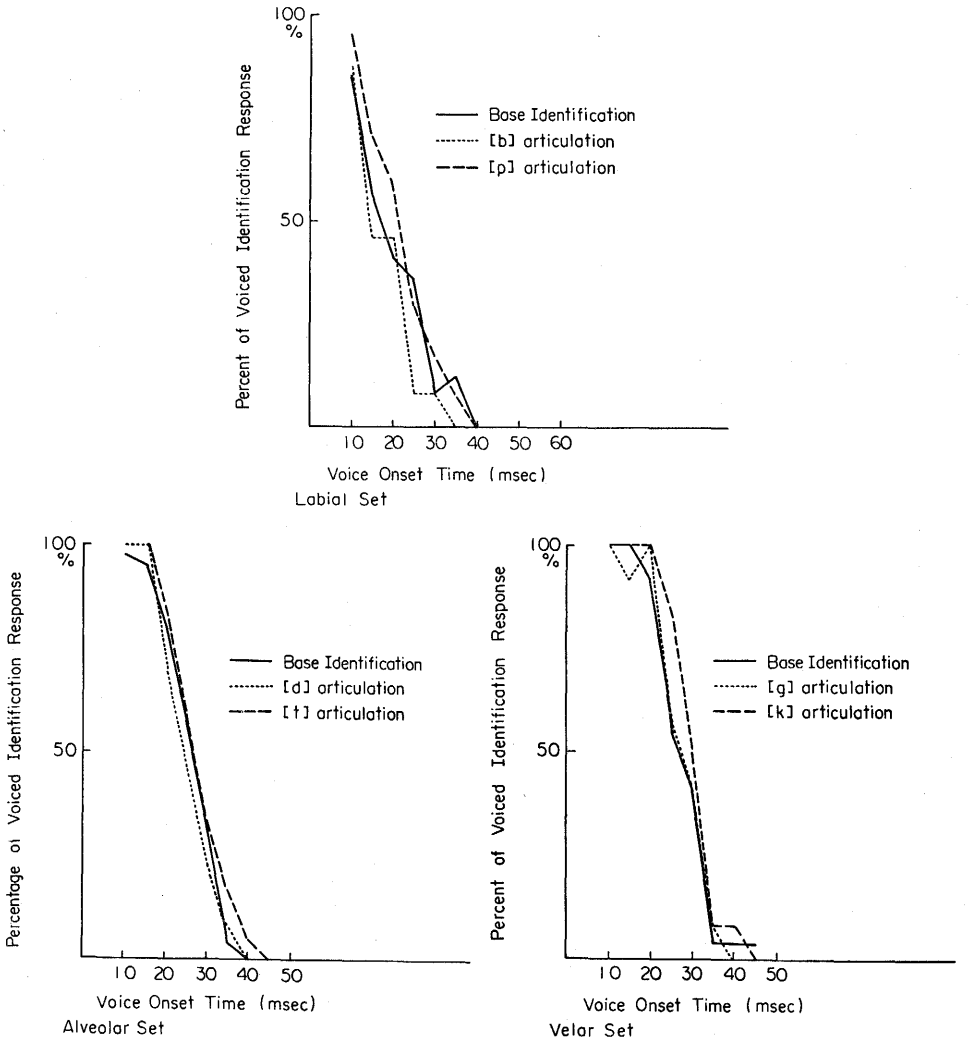

Fig. 3. Percentage of voiced identification responses [b, d and g] obtained with and without articulatory effects. The solid lines indicate the base identification functions and the dotted and dashed lines, the identification functions after repetitive articulations of voiced and voiceless consonants.

for labial set, −40 msec for alveolar set, and −50 msec for velar set. From Table 2, it can be pointed out that the adaptation effect with [da] caused a notable shift of phonetic boundary in a predicted direction and the effect with [ga] also caused a shift of the boundary. In these two sets, subjects gave more responses belonging to voiceless category. These are in general agreement with the experimental data so far published (Eimas & Corbit, 1973). In case of labial set, however, the adaptation effect with [ba] was not so consistent and some subjects did not show any adaptation effect. In Figure 2, the adaptation effects are shown in curves for three sets.

The articulatory effects in speech perception were examined for twelve subjects in labial and alveolar sets and for six subjects in velar set. The response data in the experiment by repetitive articulation were also analyzed in the same way as was followed in the experiment by repetitive listening. The shifts of phonetic boundaries are shown in Table 3 and the effects by repetitive articulation are shown in curves in Figure 3.

In examining the articulatory effects in speech perception, it is expected that there is a common mechanism of perception and production and that the perceptual boundaries would be shifted to the directions as in repetitive listening. From Table 3 and Figure 3, it is apparent that articulatory effects in perception are different in three sets of test stimuli. In the overall examination, the effects were weak and not so significant in three sets, especially in alveolar set. In labial set, however, some boundary shifts were found in the articulation of [ba] and [pa], though the effects were less in [ba]-articulation. In the case of velar set, the boundary was shifted to the predicted direction in [ka]-articulation, while not in [ga]-articulation.

IV. DISCUSSION

It is well known that VOT is an index to laryngeal timing for three modes of homorganic stop consonants. Lisker and Abramson (1970) studied the identification of voiced and voiceless stop consonants for the speakers of English, Spanish, and Thai. With the addition of their data, the distinction between voiced and voiceless stop consonants by the Japanese speakers can be shown as follows[4]:

	English	Spanish	Japanese
/b−p/	+20∼+30	+10∼+20	+15∼+20 ($\bar{X}$=18)
/d− t/	+30∼+40	+20∼+30	+20∼+30 ($\bar{X}$=26)
/g−k/	+30∼+40	+20∼+30	+25∼+30 ($\bar{X}$=26)

(VOT in milliseconds)

The perceptual values for distinction of voiced and voiceless stop consonants by the Japanese speakers are similar to those by the Spanish speakers. It can be said that Japanese and Spanish speakers have similar characteristics in the perception of voicing features. In examining the above data, it seems that if languages

have binary distinction of voicing features, the perceptual boundaries will lie in the VOT value from +20 to +40 msec.

Adaptation effects by repetitive listening to voiced syllables were found in alveolar and velar sets, and the effects were in general agreement with the research results so far published, but no significant effect was found in labial set. If we are free from the concerns in stimuli preparation and experimental procedure, the results in labial set were not predictable and some idiosyncratic properties may be relevant for labial stop consonants.

One of the purposes of the present study is to examine articulatory effects on speech perception. Some experiments have been reported on the articulatory effects, but most of them are mainly concerned with place features, not with voicing features. Through such experiments, it is suggested that there exists a common mechanism for perception and production of speech sounds. (Cooper & Lauritsen, 1974) In the present study, repetitive articulation of [ba], [pa] and [ka] caused a shift of phonetic boundaries to the predicted directions. The reasoning for this shift can be as follows: repetitive articulation of CV syllables should fatigue the voicing feature detectors and would reduce the sensitivity of the detectors in perception. It can be pointed out, however, that there are some differences in the strength of articulatory effects; that is, the feature detector for voiceless feature is more sensitive than the detector for voiced feature and the detectors for labials are more sensitive than other place features. This may indicate that there exists a mediating mechanism for perception and production for some features, but it is not clear why the articulatory effects did not come out for other features. These results partially conform to the ones by Cooper and Lauritsen (1974) in which they examined the perceptuo-motor effect and such effects were found for the voiceless stop consonants, but not for the voiced stop consonants[5]. From these studies, it can be pointed out that some features can be independently processed from articulatory factors and the detectors for each feature do not necessarily function at a mediating level for perception and production. Some detectors function as a common mechanism, while others may not. Further experiment will be needed to confirm separate processing of some feature detectors from articulation and to examine how neural commands in articulation affect the processing of auditory and linguistic information of speech sounds.

ACKNOWLEDGEMENTS

I wish to thank Dr. Tatsuo Nishida of Kyoto Univ. who has helped my study of linguistics in various ways and Dr. Michael Studdert-Kennedy of City University of New York for valuable suggestions on the experiments of speech perception. The present study was in part supported by the grant of the Japanese Ministry of Education.

NOTES:

1. It is pointed out by Abramson(1976) that the VOT for voiced unaspirated stops centers at −100 msec, the one for voiceless unaspirated stops centers at +10 msec and the one for voiceless aspirated stops centers at +75 msec.
2. The F-1 cutback is not a sole relevant acoustic cue to distinguish the voiced from voiceless stop consonants. For further discussions and arguments, see Liberman, et al. (1958)
3. This extrapolation method on defining the boundary was taken in Donald (1976).
4. The figures of English and Spanish were taken from Lisker and Abramson (1970) and indicate the range in which phonetic boundaries lie in the identification curves of voiced and voiceless stop consonants.
5. Some supports for the separate processing of voiced and voiceless features can be found in the distributional pattern of the VOT values in production. The VOT values for a voiceless segment vary in a wide range from +30 to +120 msec, while the ones for a voiced segment vary in a narrow range from 0 to +30 msec. This may indicate that two features of voiced and voiceless receive independent neural commands in production.

REFERENCES

Abbs, J. H. and Sussman, H. M. (1971), "Neurophysiological feature detectors and speech perception: A discussion of theoretical implications", *Journal of Speech and Hearing Research,* 14, 23–36.

Abramson, A. S. (1976), "Laryngeal timing in consonant distinctions", *Status Report on Speech Research,* 47, 105–112.

Bailey, P. (1973), "Perceptual adaptation for acoustical features in speech", *Speech Perception,* The Queen's University of Belfast, Series 2, 29–34.

Cooper, W. E., Blumstein, S. E. and Nigro, G. (1975), "Articulatory effects on speech perception: a preliminary report", *Journal of Phonetics,* 3, 87–98.

Cooper, W. E., Billings, D. and Cole, R. E. (1976), "Articulatory effects on speech perception: a second report", *Journal of Phonetics,* 4, 219–232.

Cooper, W. E. and Lauritsen, M. R. (1974), "Feature processing in the perception and production of speech", *Nature,* 252, 121–123.

Donald, L. (1976), "The effects of selective adaptation on voicing in Thai and English", *Status Report on Speech Research,* 47, 129–136.

Eimas, P. D. and Corbit, J. D. (1973), "Selective adaptation of linguistic feature detectors", *Cognitive Psychology,* 4, 99–109.

Liberman, A. M., Delattre, P. G. and Cooper, F. S. (1958), "Some cues for the distinction between voiced and voiceless stops in initial position", *Language and Speech,* 1, 153–167.

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. (1970), "The voicing dimension: some experiments in comparative phonetics", *Proceedings of the 6th International Congress of Phonetic Sciences,* 563–567.

Pisoni, D. B. and Tash, J. (1975), "Auditory property detectors and processing place features in stop consonants", *Perception and Psychophysics,* 18(6), 401–408.

(Aug. 31, 1977, received)