STUDIA PHONOLOGICA XIX (1985)

Discrimination of Stop Consonant in Monosyllabic Speech Including Glottal Stop

Shigeyoshi KITAZAWA and Shuji DOSHITA

INTRODUCTION

Phonologically, in Japanese, the glottal stop is not recognized as a phoneme. Acoustically, it lacks the initial burst noise that characterizes stop consonants. For these reasons, the glottal stop is not usually inlcuded in studies of stop consonants.

However, there are several reasons why it is desirable to do so. For one, generative phonology describes the glottal stop as a phoneme which precedes an initial vowel.[1] Further, both the articulation and the acoustic characteristics are similar to those of consonants. Finally, in recognition tasks glottal stops are easily confused with consonants.

The articulation of a "glottal stop" is similar to that of a stop consonant. At the beginning of phonation of a vowel in the morpheme initial position, the vocal cords are much closer than in the normal breathing position, and usually close the glottis.

The acoustic characteristics of a glottal stop vary within a range that overlaps the range of variation of stop consonants:

a) at high utterance intensity, a burst like waveform can occur at the onset of an isolated vowel;

b) conversely, in some cases the burst waveform of a glottal stop cannot be seen. Glottal stops lack the formant transitions of stop consonants, but voiceless stops may lack them also;

c) in some cases of voiced stops, the murmur that normally precedes the burst does not appear, resulting in a waveform similar either to a voiceless stop or a glottal stop.

In speech recognition tasks, detecting the presence of an initial stop consonant (i.e. discriminating it from a glottal stop) is particularly difficult for stop consonants.

For all these reasons, we decided to incorporate a glottal stop /ʔ/ among the stop consonants in our experiments.

We have already shown that the burst spectrum and the transitive spectra

Shigeyoshi KITAZAWA (北澤茂良): Associate professor, Department of Computer Science, Faculty of Engineering, Shizuoka University.
Shuji DOSHITA (堂下修司): Professor, Department of Information Science, Faculty of Engineering, Kyoto University.

that follow the burst in stop consonants are effective for discrimination concerning the place of articulation (given voiceless/voiced distinction).[2] The discrimination performance remains good even if the speaker and the following vowel are unspecified.

In this paper, we try to show that there are similar features in the onset of vowels and those features are effective for discrimination from plosive consonants.

ACOUSTIC CHARACTERISTICS

The onset of a vowel in the initial position is variable, depending on the in-

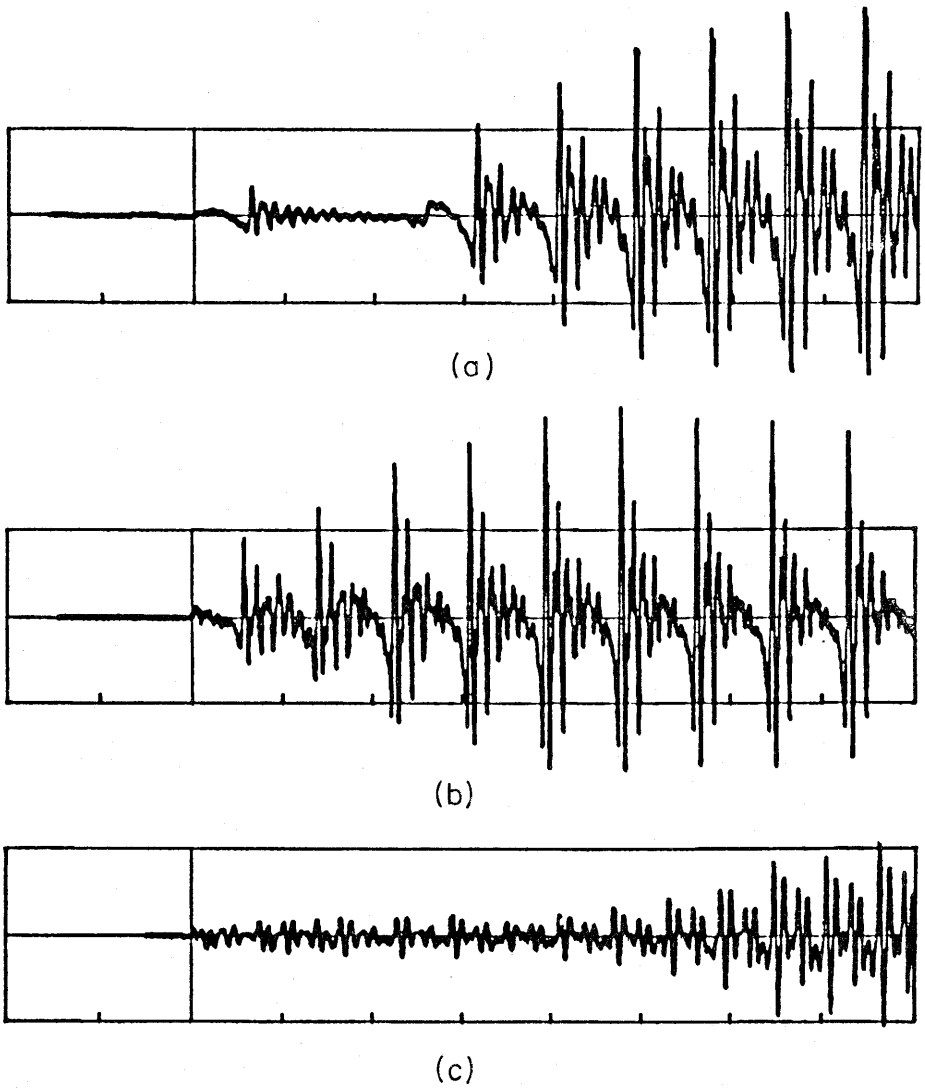

Fig. 1. Various onsets of a vowel /a/ of each different speaker. (a) a stop like vowel onset, (b) an immediate and regular vowel onset, and (c) a frication like onset.

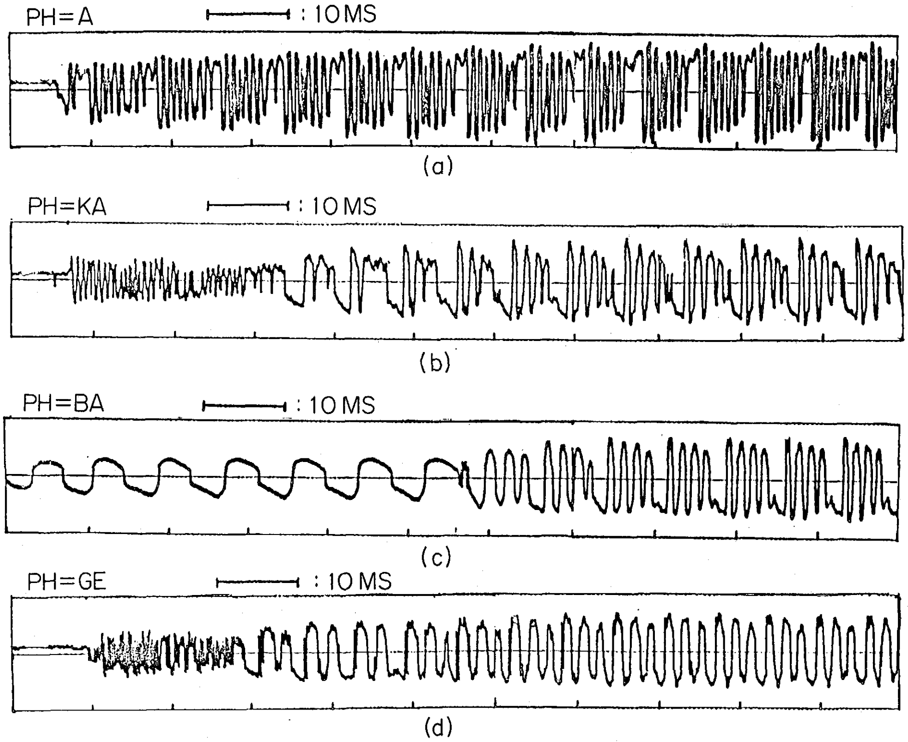

Fig. 2. Waveforms of initial part of CV syllable. (a) onset of a vowel /a/, (b) burst with aspiration noise for /ka/, (c) voiced stop consonant with pre-burst murmur, and (d) voiced stop consonant without murmur.

tensity of utterance. Fig. 1 shows examples of typical onsets. In the case of a strong phonation, the glottis is tightly closed. A stop like explosion will appear if the air pressure under the glottises is high enough, otherwise a stable oscillating waveform starts from the beginning. In the case of a weak phonation, the glottis is loosely closed, and there appears a short aspirative noise [^{h}] or a transitive irregular waveform. Differences in phonation of glottal stops are related to situations and idiosyncrasies of individual people.

The presence of a burst noise is the most evident contrast between stop and initial vowel onset as shown in Fig. 2. However, the distinction between isolated vowel and voiceless stop is difficult in terms of formant changes. At the onset of an initial vowel, there are very slight formant changes. In voiceless stops, a short burst comes first and then follows a short transition. Formant changes of a voiceless stop are slight, and are often difficult to observe, especially in bilabials.

Voiced stops are distinguished from voiceless stops or isolated vowels by the vocalization prior to the burst, i.e. weak prevoicing vibration (murmur) during the closure period, but sometimes mainly owing to the individual difference, there is no murmur. Fig. 3 shows typical voiced stops preceded by murmur (left) and

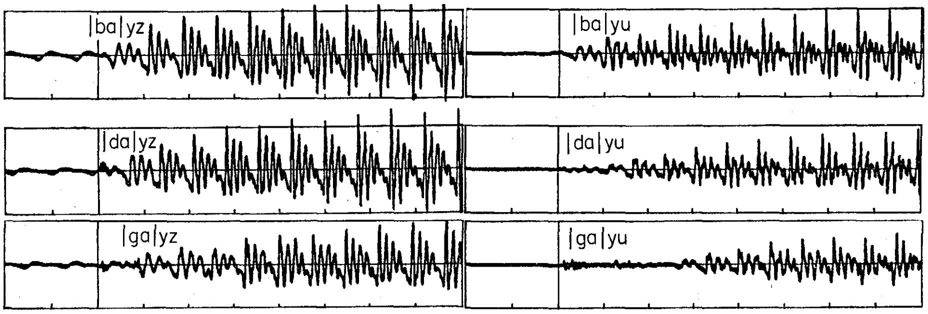

Fig. 3. Typical voiced stop consonants preceded by murmur (left) and voiced stop consonants without murmur (right). The left half utterances are of the same speaker and the right half utterances are of the other same speaker.

voiced stops without murmur (right). The utterances come from two different speakers and show the individuality. The waveforms of voiced stops lacking murmur are difficult to distinguish from those of a voiceless consonant, or of a glottal stop.

EXPERIMENTAL METHOD

Phonemes examined are voiceless stops /p,t,k/ and voiced stops /b,d,g/ followed by one of five vowels, i.e. 30 kinds of monosyllabic /CV/s, plus five monosyllabic vowels, i.e. /ʔV/, where /ʔ/ is a glottal stop and /V/ is a vowel. In Table 1 all syllables are shown. Syllables /ti/, /tu/, /di/, and /du/ are phonated as [ti], [tu], [di], and [du]. Speakers consists of 89 male students, 5 of them contributed three repetitions of each syllable, and the remaining 84 men one each.

According to our method for stop consonant analysis, the feature for place of articulation is extracted from the spectra after the burst. The burst point is visually determined. The corresponding point for /ʔ/ is the onset of voicing. Spectrum is analyzed every 10ms up to 5 frames (with 25ms window duration). Details of analysis method are the same as described in reference 2. Classification is based on a set of linear discriminant functions. We consider the recognition rates under various combinations of phonetic features, and use them to evaluate the discrimi-

Table 1. Stop consonant vowel syllables examined.

/ʔa/	/ʔi/	/ʔu/	/ʔe/	/ʔo/
/pa/ /ta/ /ka/	/pi/ /ti/[ti] /ki/[qi]	/pu/ /tu/[tu] /ku/	/pe/ /te/ /ke/[qe]	/po/ /to/ /ko/
/ba/ /da/ /ga/[Ga]	/bi/ /di/[di] /gi/[gi]	/bu/ /du/[du] /gu/[Gu]	/be/ /de/ /ge/[ge]	/bo/ /do/ /go/[Go]

nant capabilities of individual features.

Here, assuming that the coarticulatory factor is negligible as a first order approximation, we will discriminate consonants independent of the vowel followed. And also assuming that the deformations due to individuals are negligible, we will classify each consonant phonated by a different speaker in one category, that is, speaker independent recognition.

Experimental Results

1. Discrimination in the place of articulation of stop consonants

Consonant discrimination was experimented for the voiceless stops and the voiced stops separately. Table 2 shows results of consonant recognition. For consonants with the same place of articulation followed by one of five vowels pooled with equal weight (vowel independent consonant recognition), 92.0% correct recognition rate was achieved for voiceless stops and 90.3% for voiced stops. These results show that phonetic features of stop consonants are expressed in spectrum patterns near the burst point and that the vowel followed have an insignificant influence on those features. Vowel dependent recognition do improve performance, however, the number of sample was insufficient to declare the results confidently.

Experiments on discrimination between /ʔ/ and stop consonants are harder. Consonant /ʔ/ was paired with the minimal contrasting set of phonemes /p, t, k/. Under the vowel independent experiment also, the average recognition rate of consonants was reduced 3.5% to 88.5% due to incertitude introduced by /ʔ/. Nevertheless, this score shows that the spectrum at the burst or near the vowel onset is sufficient to discriminate between /ʔ, p, t, k/. The features used in discrimination were those supposed to be effective to contrast the plosiveness, that is, the absolute

Table 2. Recognition results of stop consonants of 89 speakers'. The difference of the succeeding vowel was not taken into account, i. e. the vowel independent recognition. Results marked with * are vowel dependent recognition, i. e. a consonant and vowel pair (a syllable) was recognized as an individual group.

consonants discriminated	percent correct
p, t, k, q	92.0%
b, d, G, g	90.3
ʔ, p, t, k, q	88.5
ʔ, p, t, k, q	96.0*
ʔ, p, t, k, q, b, d, G, g	69.7
ʔ, p, t, k, q, b, d, G, g	86.3*
ʔ, {p, t, k, q}, {b, d, G, g}	82.4

energy levels of higher frequency components and lower frequency components and the time varying transitions of these energies.

In similar experiments in which CVs with different vowels are recognized separately, i.e. vowel dependent recognition, the average recognition rate over five vowels was 96.0%. This fact indicates that vowel-dependant glide information also contributes to stop consonant recognition. Furthermore, in a discrimination experiment between all 7 consonants /ʔ, p, t, k, b, d, g/, where the features of voicing, place of articulation and presence of burst must be discriminated simultaneously, an 86.3% correct recognition rate was obtained. Most of the errors were due to errors concerning voicing, that is, voiced consonants were misclassified into voiceless and *vice versa*. In the next section the experiments on voicing are further considered.

2. Discrimination in the voicing feature

There are voicing contrasts in stop consonants, such as /p-b/, /t-d/, and /k-g/. The glottal stop, classified as unvoiced, has no voiced counterpart. Acoustically, the presence of a murmur during closure of the vocal tract is a characteristic of voiced stop consonants. Experiment on the voiced/voiceless distiction among all stop consonants resulted in 82.4% correct recognition as shown at the bottom in Table 2. The most effective feature was the presence of low frequency energy just before the burst, and the next was the vowel onset spectrum pattern.

In experiments on the voiced/voiceless distiction within each place of articulation, i.e. discrimination between minimal pairs, velars obtained the best result, next came dentals, and then bilabials. However, in all cases more than 90% correct recognition rates were achieved (see Table 3 in the right column). The most effective feature was of course the presence of a murmur preceding the burst, and the next was the ratio of high frequency components over the total burst energy. Except for bilabials, the vowel onset spectrum was also effective.

Although the presence of the preceding murmur is the main characteristic of voiced consonants, 10.8% of all the analyzed voiced stop consonants were lacking the murmur, i.e. they started immediately with the burst. These voiced stops could be discriminated from voiceless stops, but the decrease in recognition rate was as much as 12.5% for bilabials, although there was no difference for velars.

Table 3. Discrimination experiments concerning the voiced/voiceless distinction. $_o$b, $_o$d, $_o$G, and $_o$g, correspond to voiced stop consonants without murmur and $_\smile$b, $_\smile$d, $_\smile$G, and $_\smile$g correspond to voiced stop consonants with murmur.

consonants contrasted	percent correct	consonants contrasted	percent correct
p, $_o$b	78.5%	p, $_\smile$b	91.0%
t, $_o$d	81.4	t, $_\smile$d	92.6
k, $_o$G	94.9	k, $_\smile$G	96.0
q, $_o$g	95.9	q, $_\smile$g	95.8

Table 3 shows results for individual pairs of consonants without murmur (left) and with murmur (right). The recognizable features were the increase and rise-time of high frequency energy at the burst onset, for bilabials and dentals. For velars, the absence of a significant decrease in recognition accuracy suggested that the sustained noise sound after the burst contains sufficient features to discriminate from voiceless.

Conclusion

From the discriminant analysis of the spectra of the stop burst and between the burst and the onset of the followed vowel, the following conclusions are drawn:

1) /ʔ/ was well discriminated from its minimal contrasting voiceless stops;
2) the difference between voiced and voiceless was well discriminated by the presence or absence of the murmur preceding the burst;
3) the discrimination of voicing of voiced stops without preceding murmur was possible but slightly difficult, bilabials were much affected, but velars were not.

References

1) J. D. McCawley, The Phonological Component of a Grammer of Japanese (The Hague, 1968).
2) Shigeyoshi Kitazawa and Shuji Doshita, "Speaker and vowel independent recognition of CV initial plosives by reduction and integration of running spectra," the Seventh International Conference on Pattern Recognition, Montreal, Canada, July 30–Aug. 2, 1984.

(Aug. 31, 1985, received)