DISCRIMINATION OF VOICED PLOSIVES USING TRANSITION PROPERTIES OF THE LPC CEPSTRUM PARAMETERS

Y.Yamashita, M.Yanagida, R.Mizoguchi and O.Kakusho

The Institute of Scientific and Industrial Research, Osaka University, 8-1, Mihogaoka, Ibaraki, 567 JAPAN.

ABSTRACT: In this paper, a discrimination method of voiced plosives is proposed using two sets of parameters; one, that describes the transition of acoustic parameters by fitting their transition loci with regression lines and the other, the acoustic parameters themselves at the beginning point of the transition. The gradient of the regression lines is found to be effective for discriminating among voiced plosives.

1. INTRODUCTION

It is generally assumed that the acoustic properties of voiced plosives lie both near the burst and in the transition part to following vowels.[1] Proposed here is a method for discriminating voiced plosives by employing both instantaneous properties near the burst point and dynamic properties in transient part. The transition properties of a consonant followed by a vowel are especially dependent on the following vowel. In this paper a following-vowel dependent discrimination method is adopted and analysis periods are adjusted for each following vowel. Its performance is evaluated on isolated syllables uttered by 38 male adults.

2. FITTING THE PARAMETER TRANSITION WITH REGRESSION LINES

2.1 Analysis and Discrimination of Voiced Plosives

The acoustic parameters of voiced plosives change drastically at the burst and during the succeeding short period. However, the variations of the parameters in slow transition parts are expected to be sufficiently described with regression lines, i.e. the transition of each acoustic parameter can be approximated with a line. the number of analysis frames to be fitted by regression lines is fixed here to be ten regardless of the frame shift interval.

The LPC analysis of order 12 is performed on 10 successive frames in the transition part starting at the burst point to the following vowel. The analysis start point is defined by time delay Td from the burst point and is determined according to the following vowel together with the frame shift interval Ts. The short-time energy and the LPC cepstrum coefficients are obtained for 10 frames, where the short-time energy is expressed in dB normalized by the short-time power of the following vowel part.

The time series of each parameter is approximated with a regression line. The gradients of the regression lines are employed as the parameters to describe transition properties, and the short-time energy and LPC cepstrum parameters of the first frame are used as those for instantaneous properties. Therefore, 26 parameters in all (13 first frame parameters and another 13 gradient parameters) are employed for mutual discrimination voiced plosives.

Speech samples employed here are 15 isolated CV syllables; /b/, /d/ and /g/ as the leading consonants followed by Japanese vowels /a/, /e/, /i/, /o/ and /u/, uttered by 38 males in a large anechoic chamber. (The total number of utterance is 15x38=570, i.e. 114(=570/5) for each following vowel) The speech samples were quantized at 10 ksamples/sec with 12 bit accuracy after low-pass filtering of 4.5kHz cut-off frequency and -260 dB/oct suppression characteristics.

The discrimination score is evaluated by the leaving-one-out method. In this paper, the Fisher space[2] is employed to reduce the dimension of parameters from 26 down to 2. The discrimination is performed as decision by majorities in 3-nearest neighbors on the 2-dimensional Fisher space.

2.2 Investigation on Analysis Periods

The optimal analysis periods, which are determined by the window length for one frame analysis (Tw), the location of the initial analysis frame (Td) and the frame shift interval (Ts), are investigated for each following vowel assuming that the burst points were detected by visual inspection, and the following vowels, a priori known. The investigation range for each parameter is as follows;

Tw = 10, 15 and 20msec.
Ts = 1, 2, 3, 5 and 7msec.
Td = -12 through 12 by 2msec step.

The discrimination test is performed under 195 (=3x5x13) combinations of analysis conditions.

The discrimination performance was compared among the window length of 10, 15 and 20msec. The discrimination results remain almost the same regardless of the window length. Therefore, the window length is fixed to 20msec in the rest of this paper considering the stability of analysis.

The total length for each CV syllable to be analyzed is determined by the frame shift interval Ts. The analysis start point is identified by Td which is the time delay relative to the burst point. Positive Td means that the analysis is started at Td msec after the burst. The optimal analysis period which yields the best discrimination score for each following vowel is shown in Fig. 1. Table 1(a) shows the best discrimination scores under condition that the burst points were detected by visual inspection, and the following vowel, a priori known.

Table 1 compares the discrimination score with and without employing the gradient parameters of the regression lines. It is recognized from Table 1 that introduction of the gradient parameters improves the discrimination score by 5% on average for the five following vowels. Fig. 2 shows the comparison of the distribution of the phoneme templates on the 2-dimensional Fisher space for both the discrimination schemes with and without parameters for following vowel /u/. The clusters of the with-gradient case have less overlaps and are clearly separated one another compared to the without-gradient case. The Fisher ratio is improved from 3.3 to 12.5.

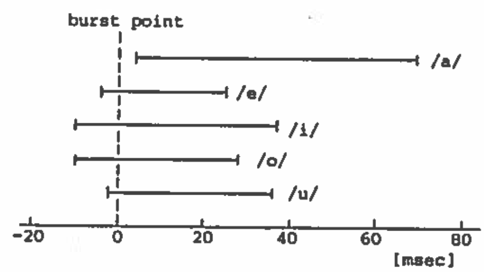
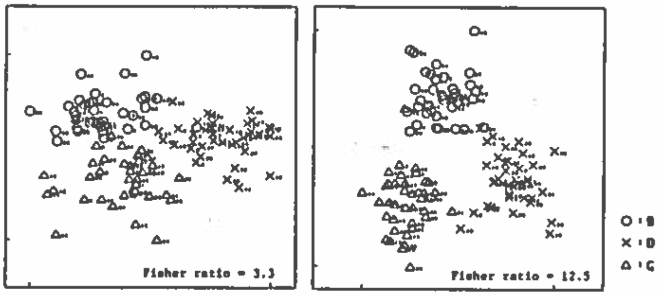


Fig.1 The optimal analysis period for each following vowel.

Table 1 Discrimination score of voiced plosives(I).
Burst point detection : by visual inspection
Following Vowels : a priori known

Gradient Parameters	/a/	/e/	/i/	/o/	/u/	average
	Discrimination Score(%) Following Vowel					
(a)used	96	89	92	95	93	93
(b)not used	94	83	84	91	89	88



(a)without the gradient parameters   (b)with the gradient parameters

Fig.2 A comparison of sample distribution on the Fisher space (following vowel /u/).

3. DETECTION OF BURST POINTS AND RECOGNITION OF FOLLOWING VOWELS

In the previous section, it is assumed that the burst points are detected by visual inspection and that vowels, a priori known. This section describes a recognition system of voiced plosives with automatic detection of burst points and automatic recognition of following vowels.

3.1 Automatic Detection of Burst Points

Automatic detection of burst points is realized by using the distance measure of the LPC cepstrum parameters between a lame pair of frames, where two frames start at the same point, and end at different points. The lengths of the long and the short frames are 20 and 17msec, respectively. The correct detection rate of the proposed method for burst point detection is evaluated under the following criterion. If the difference of the detected point and the real burst point is less than 3msec, the detection is presumed to be correct. Under the criterion above, the score is 92% in average.

3.2 Recognition of Following Vowels

Recognition of following vowels is realized also employing decision by majorities on the Fisher space projected from a 16-dimensional space scaned by the LPC cepstrum parameters. For test samples, 5 frames near the center of vowel part are analyzed and assigned to one of the five Japanese vowels according to decision by majorities in 5-nearest neighbors in the Fisher space. Then, the final decision is made again by majorities in the result of the successive five frames. The recognition score of the following vowels by this algorithm is 99% using leaving-one-out method.

4. AUTOMATIC DISCRIMINATION OF VOICED PLOSIVES

In this section, voiced plosives are automatically discriminated. The following vowel is first recognized, and the Fisher space is automatically chosen among those prepared for the five Japanese vowels separately. The analysis periods are determined referring the burst point detected automatically.

Table 2(a) shows the recognition scores based on the automatic identification of the following vowels. The notation of the recognition unit C in Table 2 indicates that the recognition score is that concerning the consonants only, regardless of recognition error concerning the vowels. In case of recognition unit CV, the score is recognition including the vowel identification, i.e. the score means recognition rate of CV syllables. Although the score is a little bit worse than Table 1, however, the score is over 90% in average. Table 2 shows the comparison of the score obtained by using the gradient parameters and that without using them. From this Table, it can be said that introduction of the gradient parameters improves the score by 5% for automatic CV syllable recognition.

Table 2 Discrimination score of voiced plosives(II).
Burst point detection : automatic
Following vowel recognition : automatic

Gradient Parameters	Rec. Unit	Discrimination Score(%) Following Vowel /a/	/e/	/i/	/o/	/u/	average
(a)used	C	93	90	91	95	89	92
	CV	93	88	91	95	89	91
(b)not used	C	91	82	82	89	87	86
	CV	90	82	82	89	85	86

Table 3 Discrimination score of voiced plosives by the following vowel independent discrimination system.
Burst point detection : automatic

	Following Vowel /a/	/e/	/i/	/o/	/u/	average
Score(%)	85	80	84	84	88	84

5. FOLLOWING-VOWEL INDEPENDENT DISCRIMINATION

As described above, a Fisher space is prepared for each following vowel for vowel-dependent discrimination aiming at improvement in discrimination ability. In order to justify the vowel-dependent discrimination, a following-vowel independent discrimination is tried on the speech data. Table 3 shows the scores of following vowel independent discrimination. The test samples are classified into three categories with decision by majorities in 5-nearest neighbors in a 4-dimensional Fisher space. The vowel-dependent scheme is proved to be very effective for discrimination of voiced plosives.

6. CONCLUSION

A following-vowel dependent discrimination system for voiced plosives is proposed employing additional parameters for describing the dynamic properties. The dynamic properties are extracted as the gradients of regression lines which approximate the transition of the acoustic parameters The short-time energy and the LPC cepstrum parameters are employed as the acoustic parameters here. The speech samples are CV syllables uttered by 38 male adults with voiced plosives for C and Japanese vowels for V.

The analysis periods are adjusted according to following vowels. In case that the burst points are automatically detected and following vowels are recognized by the system, discrimination score is 91%. It is proved to be effective for discrimination of voiced plosives

(1) to introduce gradient parameters of regression lines to describe the dynamic property, and

(2) to adjust the analysis condition for each following vowel and adopt a following-vowel dependent algorithm.

REFERENCE

[1] Tanaka,K : "A parametric representation and a clustering method for phoneme recognition --Application to stops in a CV environment", ASSP-29, 6, pp.1117-1127 (1981).
[2] Duda,R.O. et al. : "Pattern Classification and Scene Analysis", John Wiley, NY, pp.114-121 (1973).