STUDIA PHONOLOGICA XX (1986)

Discrimination on Japanese Stop Consonants Using Pair-Wise Discrimination Method.

Shuji DOSIHTA, Shigeyoshi KITAZAWA, Masa-aki ISHIKAWA,
Hiro-aki KOJIMA, Yoshihiro NISHIMURA

Introduction

This paper describes the method to discriminate stop consonants in Japanese mono-syllabic speech independently of speakers, using running spectra near the burst point. The method is based on the statistical feature selection and the statistical discrimination for multiple groups.

We first tried the conventional, multiple group discriminant method using linear discriminant functions, and showed it was possible to discriminate stops with high accuracy provided that voiceless/voiced distinction was beforehand given. But we could not get satisfiable discriminant score when we tried to discriminate all the stops including both the voiceless and the voiced.

The reason will be that, although in multiple group discrimination much amount of information are needed to discriminate a larger number of groups, the features are selected only as an averaged result for all the groups and the best features separating one phoneme from others can not always be selected. This problem will become more crucial as the increase of the number of groups to be discriminated.

To improve this probrem we propose an alternative method, a multiple group discriminant method using pair-wise discrimination.

This method consists of two steps. At first, two group linear discriminant analysis is performed for each pair of groups. Next, by combining the results of these analyses, we obtain discriminated results for the whole groups.

We show where the probrem lays in the former method, and also show the later method we propose in this report is effective for the discrimination of voiced and voiceless stops.

Analysis of Speech Signal

Samples we use are mono-syllables; stop consonants /? p, t, k, b, d, g/ followed

Shuji DOSHITA (堂下修司): Professor, Department of Information Science, Faculty of Engineering, Kyoto University.
Shigeyoshi KITAZAWA (北澤茂良): Associate professor, Department of Computer Science, Faculty of Engineering, Shizuoka University.
Masa-aki ISHIKAWA (石川雅朗): RICOH Corporation.
Hiro-aki KOJIMA (児島宏明): Master cource student, Department of Information Science, Faculty of Engineering, Kyoto University.
Yoshihiro NISHIMURA (西村嘉洋): Master course student, Department of Information Science, Faculty of Engineering, Kyoto University.

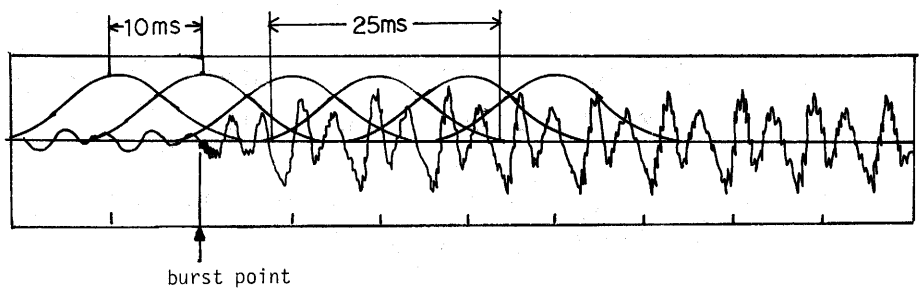


Fig. 1. Example of time windows and the waveform (mono-syllable /bi/).

by one of the five vowels /a, i, u, e, o/ uttered by 89 male speakers (3402 sampls in all) articulated without training or regulation. /?/ means the void of consonants, i.e., pure vowel.

A speech signal is low-pass filtered and digitized at 18.5 kHz with an accuracy of 12 bits.

We analize each sample by the following procedure.

1) Detect the burst point of a consonant by the human observation of the change of the amplitude and the zero-crossing of the waveform.
2) Cut out six frames from the speech signal with a time window, where each frame is 25 ms wide and delayed 10 ms from preceeding frame. And the center of second frame is located at the burst point. (See Fig. 1)
3) For each frame, calculate the smoothed spectrum and the mean squared prediction error by the 26th-order linear prediction algorithm.
4) Transform or merge each spectrum into 28 variables corresponding to critical band filter outputs.
5) Thus, each frame is analyzed to generate 29 variables (28 variables calculated in 4), and the mean squared prediction error in 3)). So, we get 174 input variables (29 variables $\times$ 6 frames) as the measurements of a sample.

MULTIPLE GROUP DISCRIMINANT METHOD USING LINEAR DISCRIMINANT FUNCTION

At first, we tried to classify samples using a conventional linear discriminant function.

In order to obtain an effective discriminant function (that is, one with a low error rate), variables, i.e., the feature variables which have discriminating power must be selected before construction of discriminant function, excluding the noisy or irrelevant input variables.

In this report we used stepwize variable selection method in the application program package BMDP7M (Biomedical Computer Program-P) to select effective variables out of 174 input variables.

This method starts by calculating the between-group F value for each variable. A variable is selected as a feature if

1) it maximizes the between-group F value, and
2) the between-group F value exceeds a threshold value (we give it beforehand).

And a variable is removed as a feature if

1) the between-group F value fails to exceed a threshold value, and
2) it minimizes the between-group F value.

Applying this procedure stepwise we can select a set of feature variables.

Linear discriminant functions which includes meaningless variables will not be the effective one. So variable selection is very important for this method.

Using these selected variables, we build-up linear discriminant function.

Now suppose that there are n groups G_1, G_2,...,G_n, to which samples are discriminated.

A common covariance matrix $\sum$, is given by

$$\sum = \frac{1}{s-n} \sum_{k=1}^{n} W_k$$

now, $W_k = (s_k - 1)\sum_k$

$$s = \sum_{k=1}^{n} s_k$$

where, $\sum_k$ is the covariance matrix in group G_k, and

s_k is the size of samples from group G_k.

Here we assume that $\sum_1 = \sum_2 = \sum_3 = \cdots\cdots = \sum_n (\equiv \sum)$. If this assumption is not true, discriminant functions will be quadratic. By a quadratic discriminant function, we may get more satisfiable results than by linear discriminant functions. But quadratic discriminant function may give us an unreliable result if we have only small number of samples, compared with the number of variables. So we take this assumption and build-up linear discriminant functions.

We use x (vector expression of feature variables) as a measurements of a samples to be discriminated.

Maharanobis D^2 distance between x and the mean vector in G_k is given by

$$D_k^2 = (x - \mu^k)' \sum^{-1} (x - \mu^k)$$

where, μ^k is the mean vector in G_k

We assign x to group $G_\natural$ which minimizes Maharanobis D^2 distance, that is,

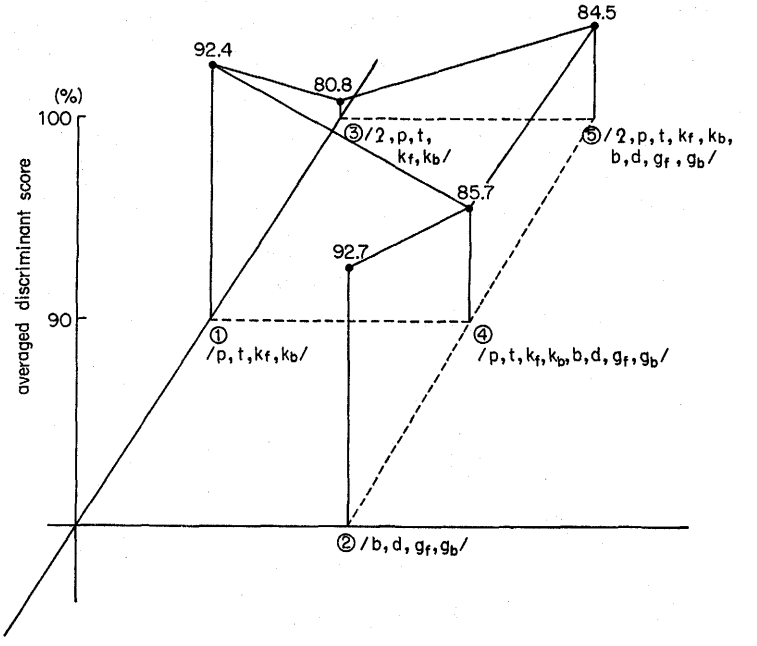
$$D_\natural^2 = \min_{i=1}^{n} D_i^2$$

Experimental Results 1

Samples examined are stops /?, p, t, k, d, b, g/ followed by one of the five vowels /a, i, u, e, o/ uttered by 89 male speakers (3402 samples in all).

We made the following experiments, using these samples.

a) discrimination of $/p, t, k_f, k_b/$
b) discrimination of $/b, d, g_f, g_b/$
c) discrimination of $/?, p, t, k_f, k_b/$



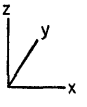


Fig. 2. Averaged discriminant scores for discrimination of various set of groups, obtained by multiple group discriminant method using linear discriminant functions.
x axis indicates whether groups include the voiced or not.
y axis indicates whether groups include the voiceless or not.
z axis indicates averaged discriminant scores for each groups.

d) discrimination of /*p, t, kf, kb, b, d, gf, gb*/

e) discrimination of /?, *p, t, kf, kb, b, d, gf, gb*/

In each experiment, we classified samples independently of the following vowels.

We use an expression /*kf*/ for /*k*/ followed by /*i,e*/, and /*kb*/ for /*k*/ followed by /*a,u,o*/. /*gf*/ and /*gb*/ are same as the case of /*k*/.

Fig. 2 shows discriminant scores of each experiment. (In this report, we obtain discriminant scores by Jack-knife method to estimate the performance in the classification of future observations.)

Discriminant scores of experiments a) and b) (both about 92%) show that our method is effective for discrimination of a smaller number of groups.

But results of experiments d) and e) show that our method isn't effective enough for discrimination of a lager number of groups.

Problems in a Conventional Multiple Group Discriminant Method

In the method we described in the preceeding section, the features which contribute to the separation of all the groups on the average are selected. But, in general, the best feature separating one group from others may be different from

those separating other one. Consequently, the variables which can best separate one group from others may not be selected, if they have no discriminating power for other groups. This will cause the loss of information in feature variables.

On the other hand, the discrimination among multiple groups can be realized by the combination of discrimination between two groups constituting the multiple groups. By doing so, the merit is attained to select the optimal feature variables to separate each pair of groups best.

Multiple Group Discriminant Method Using Pair-Wise Discrimination

The method we describe in this section consists of two steps. In the first step (pair-wise diacrimination step), two group linear discriminant analysis is performed for each pair of groups. Next, in the second step (multiple group discrimination step), by combining the results of the first step, a multiple group discrimination is performed. Fig 3 shows a diagram of the multiple group discriminant method using pair-wise discrimination method.

1) Pair-wise discrimination step

We have groups /ʔ, p, t, k_f, k_b, b, d, g_f, g_b/ to which samples are to be discriminated. At first, we perform two group linear discriminant analysis for every pair of groups.

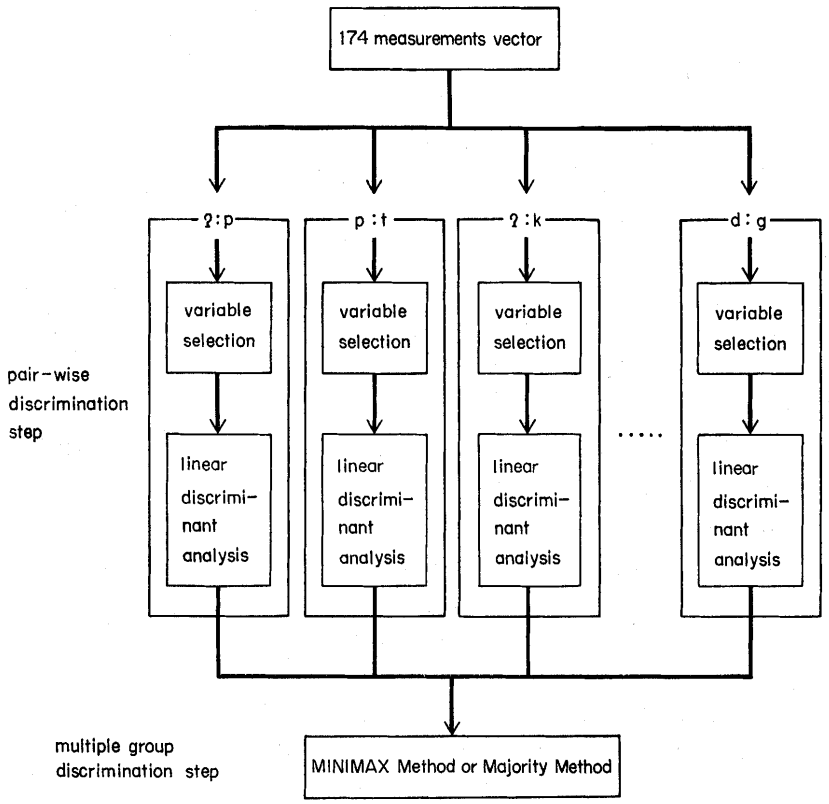


Fig. 3. Diagram of multiple group discriminant method using pair-wise discrimination.

At this case, selection of variables is done for every pair. For each pair of groups, using stepwise variable selection method, as stated in the previous section, we select feature variables which are effective for the discrimination of that pair of groups.

A set of variables which are effective for the discrimination of a pair will be defferent from a set of variables effective for discrimination of other pairs. For example, variables selected for the pair of /p/ and /k/ will be different from those selected for /p/ and /b/. Therefore, different feature space will be formed for each pair.

A sample to be discriminated is fed into every two group discriminant analysis in parallel.

Now suppose that a sample x is fed into two group linear discriminant analysis for pair of groups G_1 and G_2. A posteriori probability to the group pair G_1, G_2, for this sample is given by,

$$P_{G_i}(G_1: G_2)(x)=\pi_i \exp(-D_i^2/2)/\sum_{m=1}^{2}\pi_m \cdot \exp(-D_m^2/2) \quad (i=1, 2)$$

where, π_i is a priori probability of this sample belonging to group G_i.

D_i^2 is Maharanobis D^2 distance between x and the mean vector in group G_i.

$$P_{G_1}(G_1: G_2)(x)+P_{G_2}(G_1: G_2)(x)=1.$$

In this case, we use only the variables selected for pair G_1 and G_2.

EXAMPLE 1

Suppose that there are three groups, /p/,/t/ and /k/. For a sample x to be discriminated, we will get $P_p(p: t)(x)$; $P_t(p: t)(x)$, $P_p(p: k)(x)$; $P_k(p: k)(x)$ and $P_t(t: k)(x)$; $P_k(t: k)(x)$ for three pairs.

Thus for a sample to be discriminated, we obtain $2\cdot {}_nC_2$ a posteriori probabilities, which are the outputs of two group linear discriminant analysis for each pair of groups.

We will introduce two different method for multiple group discrimination, combining these a posteriori probabilities.

11) Multiple group discrimination step

Suppose that there are n groups to which samples are discriminated that is, G_1, G_2, G_3,...,G_n

1. Majority method

i) calculate $DP_{G_i}(G_j: G_k)(x)$ for each pair G_j, G_k, $1 \leq j < k \leq n$

$$DP_{G_i}(G_j: G_k)(x)=\begin{cases} 0 \text{ if } P_{G_i}(G_j: G_k)(x)<0.5 \\ 1 \text{ if } P_{G_i}(G_j: G_k)(x)\geq 0.5 \end{cases}$$
$$i=j \text{ or } k$$

ii) calculate $S_{G_i}(x)$ for group $G_i (i=1, n)$

$$S_{G_i}(x)=\sum_{k\neq i} DP_{G_i}(G_i: G_k)(x)$$

iii) sample x will be d scriminated to group $G_{\hat{s}}$, where

$$S_{G\sharp}(x)=\max_{*=1}^{n}(S_{G*}(x))$$

EXAMPLE 2

In the case of EXAMPLE 1, if we get

$$P_p(p:t)(x)=0.7,\quad P_t(p:t)(x)=0.3,$$
$$P_p(p:k)(x)=0.6,\quad P_k(p:k)(x)=0.4,$$
$$P_t(t:k)(x)=0.8,\quad P_k(p:k)(x)=0.2,$$

then

$$DP_p(p:t)(x)=1,\quad DP_t(p:kt)(x)=0,$$
$$DP_p(p:k)(x)=1,\quad DP_k(p:k)(x)=0,$$
$$DP_t(t:k)(x)=1,\quad DP_k(t:k)(x)=0,$$

and

$$S_p(x)=2,\ S_t(x)=1,\ S_k(x)=0.$$

At this case we will discriminate x to /p/.

$DP_{Gi}(G_i:G_j)(x)$ takes value 1 when two group discriminant analysis for a pair G_i, G_j judges that a sample x might be discriminated to G_i, not to G_j. (Notice that sample x may belong to the other group G_k, $k\neq i$ nor j. Even in this case, x is forced to be judged as G_i or G_j.)

$S_{Gi}(x)$ indicates how many times the judgements that x might belong to G_i are done. In this method, the majority of these judgements decides the group to which the sample is finally discriminated. So we call this method 'majority method'.

2. MINMAX method

In the case of EXAMPLE 2, one may suppose that the sample x belongs to /t/ observing the highest probability $P_t(t:k)(x)=0.8$. But notice that $P_t(t:k)(x)$ says nothing about /p/, since $P_t(t:k)(x)$ is calculated regardless of the distribution of samples from /p/. So the sample x may belong to /p/. On the other hand, we can suppose that the sample x doesn't belong to /k/ observing $Pk(t:k)(x)=0.2$.

Thus, we can get a negative information for assignment by observing a lower value of a posteriori probabilities. So, at first, for the group G_i $(i=1, n)$ we calculate the minimum value of the a posteriori probabilities. If such a minimum value for G_i is small enough, we can safely judge that the sample doesn't belong to G_i. So we discriminate the sample to group $G_\sharp$ which has the maximum value among minimum value of a posterior probabilities of each group G_i. Such the maximum value indicates a risk of judgements that the sample does not belong to that group like an elimination method of negative results or like a refutation method. The assignment rule is given by,

i) calculate $M_{Gi}(x)$ for group $G_i(i=1, n)$, where

$$M_{Gi}(x)=\min_{*\neq i}(P_{Gi}(G_i:G_*)(x))$$

ii) discriminate a sample x to $G_\sharp$, where

$$M_{G\sharp}(x)=\max_{*=1}^{n}(M_{G*}(x))$$

EXAMPLE 3

In the case of EXAMPLE 2, the minimum value of a posteriori probability for each group is given by $M_p(x)=0.6$, $M_t(x)=0.3$, $M_k(x)=0.2$. In this case, x is discriminated to /p/, the same result with majority method. We call this method 'MINMAX method'.

EXPERIMENTAL RESULTS 2

Samples are mono-syllabic speech /?, *p*, *t*, *k*, *b*, *d*, *g*/ followed by /*a*, *i*, *u*, *e*, *o*/ uttered

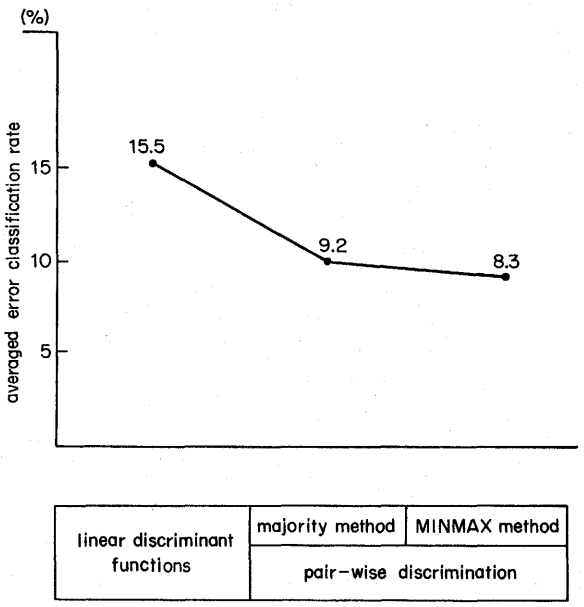


Fig. 4. Averaged error classification rate by each method for /?, p, t, k_f, k_b, b, d, g_f, g_b/.

Table 1. Classification result (Confusion matrix) by MINMAX method

consonants	percent correct	?	p	t	k_b	k_f	b	d	g_b	g_f
?	94	466	25	1	2	0	0	0	0	1
p	90	30	444	16	1	2	2	0	0	0
t	93	3	14	431	4	12	0	1	0	0
k_b	95	0	2	9	281	0	0	0	4	0
k_f	99	0	0	2	0	195	0	0	0	1
b	87	5	42	1	1	0	432	12	3	0
d	87	0	3	37	0	0	13	405	1	6
g_b	91	2	0	2	15	1	3	4	266	1
g_f	93	1	0	1	0	8	1	3	0	184
Total	91									

by 89 male speakers (3402 samples in all).

We discriminate stops /ʔ, *p*, *t*, k_f, k_b, *b*, *d*, g_f, g_b/ followed by vowels independently of speakers. Fig. 4 shows the averaged error classification rate by each discriminant method. 90.8% of all the samples are correctly classified by 'majority method' and 91.2% by 'MINMAX method', while 84.5% by linear discriminant function for multiple groups stated before. Table 1 shows confusion matrix by 'MINMAX method'.

Conclusion

At first we tried to discriminate stops, using multiple group discriminant function. But we faild to get satisfactly discriminant score especially for the discrimination for a larger number of groups.

Next, we proposed multiple group discriminant method using pair-wise discrimination, assuming that differences among all the groups are well described by a set of features extracted for each pair of groups.

We used two group linear discriminant analysis for each pair of groups using variables that optimize the separation of each pair. Then we discriminated stops to multiple group using results of these two group linear discriminant analysis. And we could get satisfiable improvement, that is, 91.2% of all the samples were correctly classified.

The results lead to the conclusion that multiple group discriminant method using pair-wise discrimination is effective for discrimination of Japanese voiceless and voiced stops.

(Aug. 31, 1986, received)