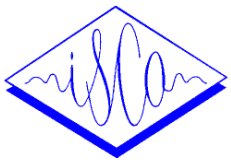

3rd International Conference on
Spoken Language Processing (ICSLP 94)
Yokohama, Japan
September 18-22, 1994

SPOKEN LANGUAGE DISCRIMINATION USING SPEECH FUNDAMENTAL FREQUENCY

ITAHASHI Shuichi, ZHOU Jian Xiong and TANAKA Kimihito
Institute of Information Sciences and Electronics
University of Tsukuba
1-1-1 Tennodai, Tsukuba, Ibaraki 305, Japan

ABSTRACT

This paper describes classification methods of spoken languages based on fundamental frequency (F_0) contours of speech. Speech data were taken from language learning tapes. Six languages including Japanese, Korean, Chinese, English, French and German were used. First, F_0 contour was approximated by a set of polygonal lines so that the mean square error between the lines and F_0 values was minimized; the optimum boundaries of the lines were determined using a dynamic programming procedure. The starting frequency, slope and duration of each line were calculated. Seventeen parameters including mean values and standard deviations of the above parameters were used for the analysis. Then parameters derived from F_0 pattern were analyzed by using principal component analysis. We also tried discriminant analysis of these parameters. Results show that the six languages can be classified based on these parameters.
Keywords: Slope of F_0 contour, Principal Component analysis, Discriminant analysis.

1. INTRODUCTION

Many languages are used in the world. Most human languages have dialects and some of the dialects differ significantly within the same language. It could happen that a speaker of one dialect cannot understand some other dialect. On the other hand, it could also be the case that people of a certain language are able to communicate with people of some other language. Most speech recognizers are designed to accept a single language today. However, it will be necessary for future speech recognizers to accept a variety of languages. Discrimination of spoken languages will play an important role as preprocessing for multi-language speech understanding [1-4].

Characteristics of languages are represented by either segmental or prosodic features of speech. The fact that speech intonation plays an important role in spoken language understanding suggests that prosodic features could also be the basis of spoken language identification. In addition, fundamental frequency is supposed to be more robust than segmental parameters. In noisy environment like a subway train, we can hear accent or intonation of a conductor's announcement, even when it is difficult to recognize what he/she said. Moreover, long term average of such parameters is more stabe.

This paper describes a method of spoken language discrimination, based on various parameters derived from fundamental frequency contours of speech [5-7]. The fundamental frequency (F_0) contour was approximated by a set of polygonal lines to minimize the mean square error between the lines and F_0 values; the optimum boundaries of the lines were determined using a dynamic programming procedure. The starting frequency, slope, and duration of each component line were calculated. Seventeen parameters including mean values and standard deviations of the above parameters were used for the analysis. Then a F_0 pattern was analyzed by using principal component analysis and discriminant analysis.

2. SPEECH MATERIAL AND F_0 EXTRACTION

Speech material was taken from language learning tapes of six languages including Japanese, Korean, Chinese, English, French and German. Speakers were five males for each language. The duration of each material was about 40 seconds for each language.

The fundamental frequency contour of speech was extracted by the AMDF method with a 16 kHz sampling frequency, 16 bit quantization, 30 ms analysis window and 10 ms frame interval. Speech intervals with speech powers greater than a certain threshold were detected automatically. Most F_0 extraction methods, including the AMDF method, sometimes make errors. Most of them are double-pitch or half-pitch errors. We have devised a simple error correction method. It calculates the mean F_0 for frames whose power is above a certain threshold. Raw F_0, $F_0 * 2$ and $F_0/2$ were compared with the mean F_0 and the one which is closest to the mean F_0 is adopted as a corrected F_0 value. Actually, $log_2 F_0$ was used in the following analysis.

3. VARIOUS PARAMETERS DERIVED FROM F_0 CONTOURS

The F_0 pattern was approximated by a set of polygonal lines which minimize the square error between each line and F_0 values. The optimum boundary of each line component of a polygonal line was determined by a dynamic programming procedure [8].

The approximation error decreases as the number of lines becomes large. It is necessary to approximate an F_0 contour with necessary and sufficient number of approximation lines to show typical characteristics of the F_0 contour of each language. One of the parameters we can use to determine the suitable number of component lines is the rate of decrease in the approximation errors. It does not necessarily give a globally optimal approximation. The number of lines should be determined according to the following conditions:

1) Keep the number of lines as small as possible.
2) Try to describe global characteristics rather than local features.

Suitable number of lines was determined so that the approximation error becomes smaller than a certain threshold. The starting frequency, slope and duration of each component line were obtained, then their mean values, standard deviations (SD) and mean values of duration-weighted slope were calculated.

3.1 Slope parameters

Table 1 shows the 17 parameters derived from F_0 contours which were used for principal component analysis.

Table 1: Parameters derived from F_0 contour

1: SD of F_0 of analyzed segments
2: SD of P_0 of analyzed segments
3: Correlation coefficient of F_0 and P_0
4, 5: Number of P and N
6, 7: Mean slope of P and N
8, 9: SD of P and N
10, 11: Duration-weighted mean of P and N
12, 13: Mean starting frequency of P and N
14, 15: SD of starting frequency of P and N
16, 17: Relative starting frequency of P and N

Abbreviations utilized in the above table.
F_0: Fundamental frequency
P_0: Speech power
P: Approximation line of positive slope
N: Approximation line of negative slope
SD: Standard deviation

3.2 Slope occurrence frequency ratio parameters

It was observed that Chinese has more polygonal line segments with steeper slope than other languages [7]. Therefore, following parameters were calculated.

1. $R_n = (Z_n)/(-Z_n)$,
2. $R_{n+10} = (Z_n)/(all)$,
3. $R_{n+20} = (-Z_n)/(all), \qquad n = 1, 2, \cdots, 10$

In the above notations, (Z_n) denotes number of lines whose slopes are in the range $[n-1, n)$; $(-Z_n)$ denotes number of lines whose slopes are in the range $[-n, -n+1)$. (all) designates number of all lines.

4. ANALYSIS RESULTS AND DISCUSSION

Figure 1 shows an example of analyzed fundamental frequency values (dots) and approximated polygonal lines.

4.1 Principal component analysis of slope parameters

The parameter set was used for principal component analysis to extract effective parameters of fewer numbers. The first principal component has 49 % contribution rate and the cumulative contribution rate of lower four components reaches 84 %.

Figures 2 and 3 show the results of principal component analysis of 17 parameters. Each point in the figures corresponds to each speaker. The first principal component (P1) and the second component (P2) plane shows that Korean is separted from the other languages; Japanese and Chinese are also separable from the others. Separation of three Asian languages from three European languages seems quite good. However, three European languages are difficult to discriminate on P1-P2 plane. Figure 2 indicates that six languages are divided into five groups, i.e., Japanese, Chinese, Korean, German and (English, French) on P1-P2 plane. English and French are mostly separated on P4-P5 plane as shown in Fig. 3.

4.2 Discriminant analysis of slope parameters

Discriminant analysis was conducted with Mahalanobis' distance measure. The parameters were selected by using the step-wise method. Table 2 shows discrimination rates and selected parameters for closed and open experiments. In the closed experiment all samples were used when calculating the discriminant function, while in the open experiment the sample to be discriminated was excluded when calculating the discriminant function. In Table 2, "6 groups" shows the results of all six languages discrimination from the

other languages. Similarly, "2 groups" shows the results of discrimination between European and Asian languages. In the closed experiments we obtained 100% discrimination rate for six languages and 87% for discriminating two groups of European and Asian languages. In the open experiments we got 80% for six languages and 83% for two language groups.

Table 2: Results of discriminant analysis (%)
(Slope parameters)

	6 groups	*2 groups*
Closed experiment	100.0	86.7
Open experiment	80.0	83.3
Selected parameters	4,5,12,13,14	4,11

4.3 Principal component analysis of slope occurrence frequency ratio parameters

The first principal component showed a contribution rate of 35 % and the cumulative contribution rate of lower eight components reached 82 %. Figures 4 and 5 show the results of principal component analysis of 30 parameters. These figures show some what similar tendency as Figures 2 and 3 but with a little better separation among 6 languages.

4.4 Discriminant analysis of slope occurrence frequency ratio parameters

Table 3 shows discrimination rates and selected parameters for closed and open experiments. The notations are the same as those in Table 2. In the closed experiments we obtained 100 % for six languages discrimination and also 100 % for the discrimination into two groups of European and Asian languages. In the open experiments we obtained 77 % for six languages and 96 % for the discrimination into two groups.

Table 3: Results of discriminant analysis (%)
(Slope occurrence frequency ratio parameters)

	6 groups	*2 groups*
Closed experiment	100.0	100.0
Open experiment	76.7	96.0
Selected parameters	$R_1, R_{13}, R_{15}, R_{25}$	$R_2, R_3, R_5, R_{20}, R_{25}$

5. CONCLUSION

Fundamental frequency contours of six languages were approximated by a set of polygonal lines, and two sets of parameters were derived for the analyses. They were analyzed using principal component analysis and discriminant analysis. The results showed that the proposed method is effective for spoken language discrimination. Further experiments are necessary with more speakers, more languages and female speech material.

ACKNOWLEDGEMENTS

This research was supported in part by Grant-in-Aid for Scientific research from Ministry of Education Science and Culture No. 03452158 and in part by Real World Computing Partnership subcontract from Ministry of International Trade and Industry.

REFERENCES

[1] Y. Ueda and S. Nakagawa, "Prediction for Phoneme/Syllable/Word-Category and Identification of Language Using HMM", Proc. ICSLP90, Paper 27-6, pp.1209-1212 (1990).

[2] M. Sugiyama, "Automatic Language Recognition Using Acoustic Features", Proc. ICASSP91, Paper 16.S12.8, pp.813-816 (1991).

[3] Michael Savic, Elena Acosta and Sunil K. Gupta, "An Automatic Language Identification System", Proc. ICASSP91, Paper 16.S12.9, pp.817-820 (1991).

[4] Yeshwant K. Muthusamy and Ronald A. Cole, "Automatic Segmentation and Identification of Ten Languages Using Telephone Speech", Proc. ICSLP92, Paper pp.1007-1010 (Oct. 1992).

[5] S. Itahashi and T. Yamashita, "A method of dialect discrimination in Japanese", Proc. 14th ICA, Paper G6-6, Beijing (Sep. 1992).

[6] S. Itahashi and T. Yamashita, "A discrimination method between Japanese dialects", Proc. ICSLP92, pp.1015-1018, Banff (Oct. 1992).

[7] J. Zhou and S. Itahashi, "Feature Extraction for Spoken Language Discrimination Using Fundamental Frequency", Technical Rep. IEICE. Paper SP93-99, pp.61-66 (Nov. 1993).

[8] S. Itahashi, "Description of Speech Data Patterns by Several Functions with Applications to Formant and Fundamental Frequency Trajectories", STL-QPSR 2-3/1978, pp.1-22 (1978).

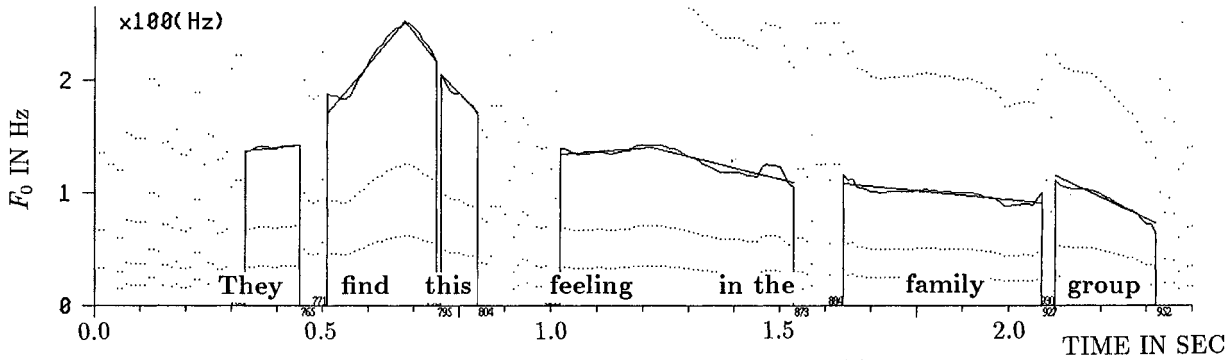

Fig. 1 F_0 contour and approximated lines.

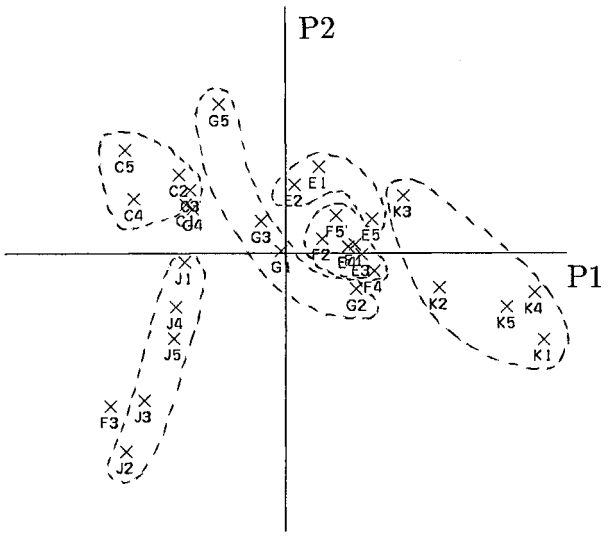

Fig. 2 Six languages on P1 - P2 plane. (Slope parameters)

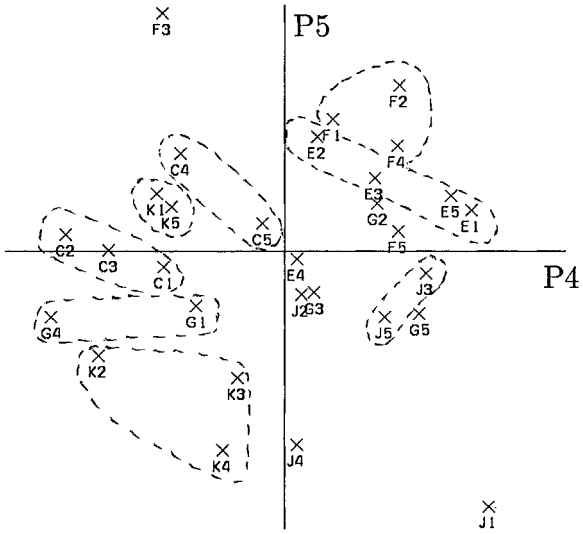

Fig. 3 Six languages on P4 - P5 plane. (Slope parameters)

Symbols	*Languages*
J1 - J5	Japanese
K1 - K5	Korean
C1 - C5	Chinese
E1 - E5	English
F1 - F5	French
G1 - G5	German

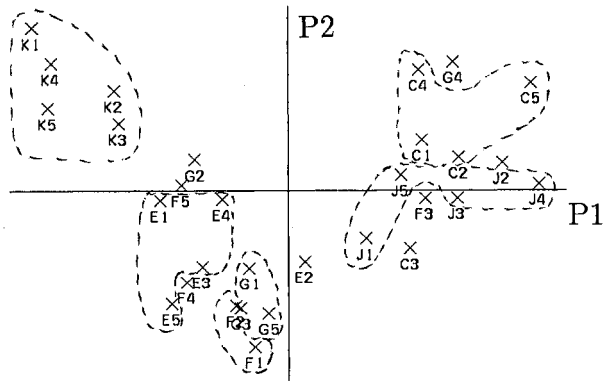

Fig. 4 Six languages on P1 - P2 plane. (Slope occurrence frequency ratio parameters)

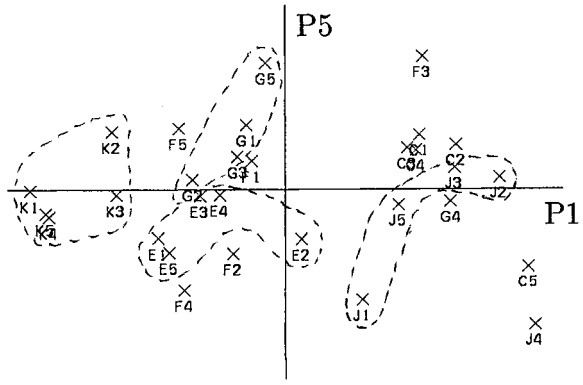

Fig. 5 Six languages on P1 - P5 plane. (Slope occurrence frequency ratio parameters)