Differentiation of English Utterances of Japanese and Native Speakers by Several Prosodic Parameters

Hiroyuki Obari, Ryousuke Tomiyama*,Mikio Yamamoto*, Shuichi Itahashi***
*Graduate School of Systems & Information Engineering, University of Tsukuba
**National Institute of Advanced Industrial Science and Technology (AIST)
hobari@jcom.home.ne.jp

Abstract

This paper proposes important parameters of prosody that differentiate the native English speakers from Japanese English speakers.

In this research we explored the features of English pronunciation of both Japanese and native English speakers from a prosodic point of view. We carried out two experiments: 1) a discriminant analysis focusing on speech fundamental frequency(*F*o), speech length, pause, duration and power of consonants and vowels. 2) Principal component analysis with the polygonal line approximation showing the prosodic feature of *F*o. Then English utterances of Japanese English speakers and English native speakers were compared to identify effective prosodic parameters.

The discriminant analysis with the proposed parameter this time could classify Japanese English speakers and English native speakers with correct discrimination rates of 77.5%. We found the power ratio of a vowel and a consonant useful to characterize the prosody of Japanese English speakers.

Then the parameter was extracted from the polygonal line which approximated *F*o, and principal component analysis was conducted. As a result, Japanese English speakers and native English speakers were able to be classified into two categories about 80 percent by using the lower principal components to the 3rd principal components.

1. Introduction

We are now living in the age of ICT and internationalization. English is considered to be an important common international language. So in Japan English education plays an important role to make Japanese English speakers intelligible enough to be properly understood in international communication. In this regard prosodic features such as intonation and the rhythm of the language are crucial to comprehensible speech.

There exist many various factors in evaluating non-native speakers of English. It was pointed out in previous research that rhythmic accent and pauses are more important than segmental features in English utterances to make speech intelligible [1-3].

In the first experiment we carried out a discriminant analysis by focusing on speech length, pause, duration and power of consonants and vowels to explore the features of English pronunciation of both Japanese and native English speakers from a prosodic point of view. Secondly principal component analysis with the polygonal line approximation showing the prosodic features of *F*o was conducted to identify effective prosodic parameters with comparing utterances of Japanese English speakers and English native speakers.

As data of this research, we chose 10 different sentences from English Learners' Speech Database [4] for two experiments. A total of 20 persons' data consists of every five male and female Japanese and English native speakers.

2. Extraction of prosodic features of speech

The fundamental frequency corresponds to pitch of speech which is one of the important prosodic features. There are many methods to extract *F*o among which simple method called "AMDF", one of the correlation processing methods, was used. We extracted the value of *F*o every 10 [msec] using the square window of 30 [msec].

Generally, *F*o for men ranges between 100Hz to 200Hz, and *F*o for women ranges between from 200Hz to 400 Hz. In order to normalize the difference among individuals, we use the logarithmic value of time series *F*o (*t*) of extracted *f*o pattern by the equation (1) where the bar "—" indicates the mean value.

$$F_0(t) = \log_2 fo(t) - \overline{(\log_2 fo(t))} \qquad (1)$$

Actually, we will use the following equation (2) for the display of *Fo*.

$$F_0(t) = 100 \times (\log_2(1 + f_0(t)) - \overline{(\log_2(1 + f_0(t)))} + 2) \qquad (2)$$

3. Discriminant analysis by basic prosodic parameters

It is said that in speaking English, intonation of Japanese English speakers generally is very flat compared with that of English native speakers. There is little change of *F*o value with flat intonation among Japanese English speakers. Moreover, Japanese English speakers tend to make unnecessary pauses and utterance length become long. The following parameters were extracted.

3.1 Sentence utterance time length

We tried to find out a sentence utterance time length of each sentence from 10 different English sentences. Then the average of each Japanese speaker and English native speaker was computed and compared.

3.2 Pause length (Silent Section)

-The rate of pause length over utterance length was computed.
-The number of the pauses in the whole sentence was calculated.

3.3 Power ratio of consonant and vowel

Consonants and vowels were extracted from utterances, and power and length of each segment measured. Segments were aggregated to give:

- The rate between consonant duration and vowel duration in utterance of the whole sentence.
- The rate of the consonant power and vowel power to the average power of the whole sentence.
- The ratio of the consonant power to the vowel power per frame.

3.4 Discriminant analysis results

In each parameter, the significance test was performed about the difference of both a Japanese speaker and an English native speaker. Homogeneity of variance was checked for *F* statistics for each parameter, and since differences were found to be insignificant, the *t* test was used to check for level differences.

In the rate of a pause the null hypothesis was rejected at 5% of significance levels, and the null hypothesis was rejected at 1% of significance levels in other parameters. Therefore, the significance of the difference of this data was shown.

Discriminant analysis was conducted using the parameters determined in the foregoing paragraph, and the distinction rate was computed. Discriminant analysis can judge to which group each data belongs based on some variables, and can measure the usefulness of the variable. Correct discrimination rates became 77.5% as a result of discriminant analysis. Among those, Japanese English speakers: 73% right and misjudgment 27%, whereas, the native English speakers: 82% right and misjudgment 12% of rates. Especially the following four leading parameters are found to be useful:

- Speech length
- The ratio of consonant time length to utterance length
- The ratio of the consonant power to vowel power
- The ratio of the consonant power to the average power of the whole sentence

Among the 27% of Japanese utterances misjudged to be by native speakers, an English teacher regarded 50% of its data, Japanese English speakers as closer to English native speaker's pronunciation; the validity of this discrimination result can be considered to be higher. About 72% of 18% misjudgment data was a sentence containing a comma, and the data which consists of two sentences. It can be considered that there is an individual difference in how to make a pause between sentences with or without a comma.

4. Polygonal line approximation of *F*o and principal component analysis

4.1 Approximation model of *F*o

If a fundamental frequency pattern is treated as it is, an amount of data becomes so large that it can't be easily processed. Therefore, in this research, we decided to use the simplest polygonal line approximation to show overall tendency in the *Fo* changes.

4.2 Approximation by polygonal line

When the *F*o pattern is approximated, approximation by a polygonal line has most often been used. A polygonal line can be represented with four parameters of *x* and *y* coordinates of the approximation starting point and ending point. This method requires fewer data than using the original *F*o data.

The approximation starting point was set as the voiced section starting point, the ending point was set as voiced section ending point, and *k*

approximation straight lines were given to each voiced section. The value of K makes ten the maximum and adopts $k \leq K$ from which the mean square error with an actual measurement became below a threshold value.

We used the dynamic programming method for the determination of the boundary of polygonal lines [5].

$$g(x)^{(k)} = a^{(k)}(x - x_k) + b^{(k)} \qquad (3)$$
$$0 \leq k \leq 10$$

Each parameter is obtained by searching for the slope a and intercept b when making a mean square error E into the minimum.

4.3 Parameter extraction

There is little change of *F*o value with flat intonation among Japanese English speakers. Speech length time becomes longer with making unnecessary pauses among Japanese English speakers.

The following 18 parameters are extracted in order to make a clear distinction of and explain the features in detail.

1. Amount of average change of *F*o value
2. Amount of average change of slope.
3. Amount of average change of *F*o value (positive)
4. Amount of average change of *F*o value (negative)
5. Ratio of length of positive sloe and negative slope (positive)
6. Ratio of length of positive slope and negative slope (negative)
7. Number of polygonal lines per unit time (positive)
8. Number of polygonal lines per unit time (negative)
9. Average of slope (positive)
10. Average of slope (negative)
11. Standard Deviation of slope (positive)
12. Standard Deviation of slope (negative)
13. Average of duration-weighted slope (positive)
14. Average of duration-weighted slope (negative)
15. The degree of skewness of slope (positive)
16. The degree of skewness of slope (negative)
17. The degree of kurtosis (positive)
18. The degree of kurtosis (negative)

5. Speech Data

As data of this research, we chose 10 different sentences from English Learners' Speech Database [4] as shown below, and we extracted 18 parameters from the model of approximate polygonal lines in the *F*o trajectory to do a principal component analysis. A total of 20 persons' data consists of every five male and female Japanese and English native speakers.

As a result, we could use up to the third principal component score with the cumulative contribution rate of 80%, and the important feature parameters were checked.

(1) That's from my brother who lives in London.
(2) The superintendent says that the teacher is a fool.
(3) The superintendent, says the teacher, is a fool.
(4) I haven't seen you before, have I?
(5) Did you find my camera? Did you leave it in Edinburgh?
(6) Have you locked the front door?
(7) Is this elevator going up or down?
(8) He drank, he stole, he was soon despised.
(9) When I came, he greeted me warmly.
(10) What difference does it make?

6. Analysis Results

Figures 1 and 2 show examples of normalized *F*o trajectory and the approximated polygonal lines of the first sentence spoken by a female Japanese English speaker and a female native English speaker. .

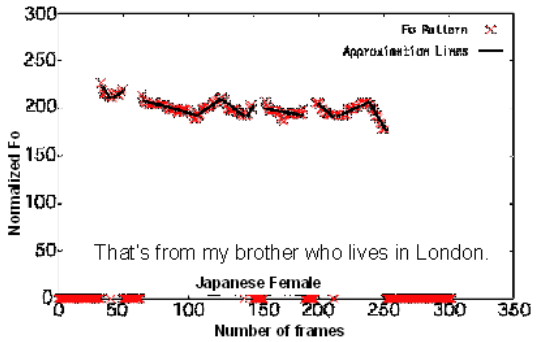

Figure 1. Example of *F*o pattern and approximated polygonal lines (Female Japanese English speaker)

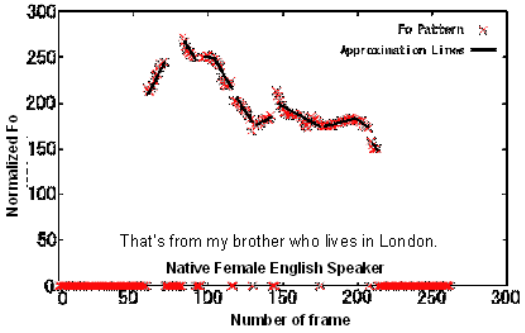

Figure 2. Example of Fo pattern and approximated polygonal lines (Female native English speaker)

We checked to what extent principal component analysis score was able to classify both Japanese and native English speakers with discriminant analysis shown in Table 1.

Table 1: Results of discrimination

	Japanese	Native	Average
1-2 Principal components	87%	64%	75.5%
1-3 Principal components	95%	70%	82.5%
1-6 Principal components	95%	70%	82.5%
1-18 Principal components	94%	75%	84.5%

As an example of a result of principal component analysis, first and third principal component score distribution is shown in Fig. 3. The contribution rate and the cumulative contribution rate are shown in Fig. 4. Principal component score coefficient is shown in Figures 5 to 7. In these figures, the larger the absolute value of a score coefficient is, the more important element of principal components becomes.

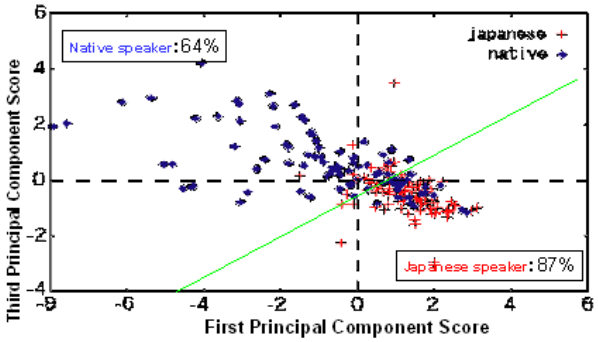

Figure 3. First and third principal component score distribution

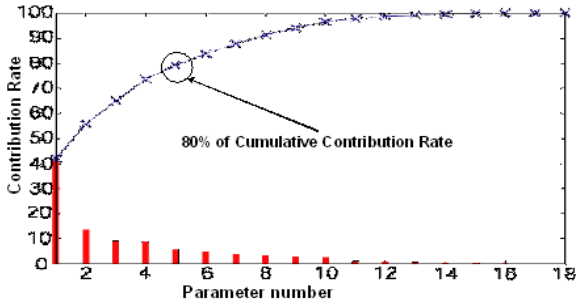

Figure 4. Cumulative Contribution Rate

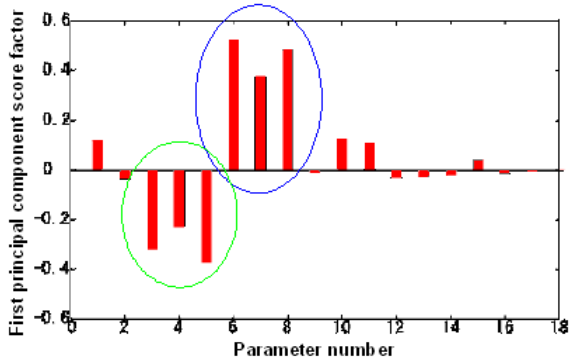

Fig 5. First principal component coefficients

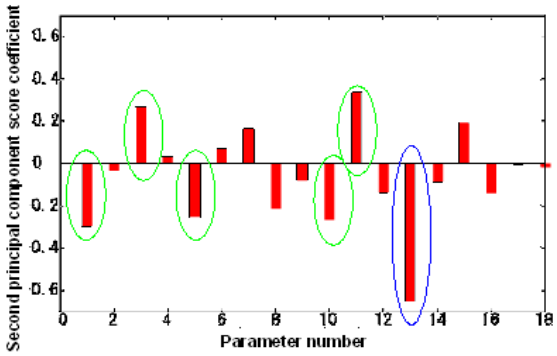

Fig 6. Second principal component coefficients

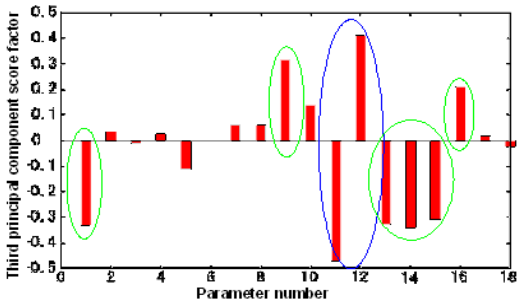

Fig 7. Third principal component coefficients

Although Japanese and English native speakers were able to be classified about 80 percent using from 1st to 3rd principal components, Japanese discrimination rate was higher as shown in Table 1. A result with an English native speaker's low distinction rate was brought into attention. We may have to pay attention to the fact that there is a variety of pronunciation even among native speakers of English. English speakers should be judged whether they speak Standard English or not although it sounds very difficult to decide what Standard English is [6].

A Japanese English teacher did subjective evaluation (1: bad 2: average 3: good) of the data with which the Japanese speaker was distinguished from the English native speaker by discriminant analysis, we found out good correlation between the subjective evaluation and discriminant analysis.

7. Conclusion

The discriminant analysis with the proposed parameter could classify Japanese speakers and English native speakers in the first experiment. We found the power ratio of a vowel and a consonant useful to characterize the prosody of Japanese English speakers. Discriminant analysis can measure the usefulness of the variable. Correct discrimination rates became 77.5% as a result of the discriminant analysis.

In the second experiment, the parameters were extracted from the polygonal line which approximated *Fo* to conduct principal component analysis. As a result, Japanese English speakers and native English speakers were able to be classified into two categories about 80 % by using the lower three principal components.

In comparing two experiments, we can say that the principal component analysis seems to be slightly better, but needs more calculation than the discriminant analysis.

In the future research we will have to test whether important parameters from principal component analysis could classify both native English speakers and Japanese English speakers using more new data.

References

[1] Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. 1996. *Teaching Pronunciation: A Reference for Teachers of English to Speakers of Other Languages.* Cambridge: Cambridge University Press

[2] H. Obari, R. Tomiyama, M. Yamamoto, and S. Itahashi. 2003. “Comparison of the prosodic feature of Japanese English pronunciation and an English native speaker sound”(in Japanese), Autumn Meeting of Acoust. Soc. Japan 2-8-6, pp.255-256.

[3] N.Minematsu, et al. 2003. “CART-based factor analysis of intelligibility reduction in Japanese English,” Proc. EUROSPEECH, pp.2069-2072.

[4] English Learners’ Speech Database (β version) (Grant-in-Aid for Scientific Research on Priority Areas, Funded by Ministry of Education, Culture, Sports, Science and Technology)”Advanced Utilization of Multimedia to Promote Higher Education Reform”. Vol.1-Vol.8, January 2002.

[5] S. Itahashi. 1978. “Description of Speech Data Patterns by Several Functions with Applications to Formant and Fundamental Frequency Trajectories", STL-QPSR 2-3/1978, pp1-22.

[6] R.Tomiyama, A.Shimazaki, H. Obari, M. Yamamoto, and S. Itahashi. 2004. “The feature analysis of the Japanese English student by polygonal line approximation of Fo”, (in Japanese) Autumn Meeting of Acoust. Soc. Japan 2-4-19, pp. 309-310.