Ann. Bull. RILP (1982) No. 16, 73-79

PERSONAL CHARACTERISTICS OF THE FREQUENCY SPECTRUM FOR VOWELS

*Shuzo Saito and Fumitada Itakura**

Introduction

In speech signal are contained both the linguistic and personal information necessary for human communication. The acoustic features of this information have been studied comparatively by several researchers, and it has been reported the acoustic features corresponding to personal information are less stable temporally than linguistic ones and a vast number of speech samples uttered by a speaker over more than one year was needed for speaker-recognition by machine.[1-3]

Such results were derived from experimental data on speaker recognition by machine based on a distance measure of the acoustic spectrum of speech, whereas the temporal variation characteristics of the frequency spectrum were not expressed clearly. In the present study, the temporal variation in the frequency spectrum of the five Japanese vowels was measured for speech samples extending over about three years to investigate the personal information characteristics contained in the frequency spectrum.

Procedures

The speech samples used were four Japanese words /bakuoN, kogeN, namae, umi/, which were uttered six times every six monthes over about three years. Nine male adults were employed as speakers.

The speech wave was sampled at 10 kHz and the amplitude was digitized into 12 bits. The digitized speech was then fed to an Eclipse-230 electronic computer, with which the PARCOR speech analysis was carried out. The PARCOR coefficients on the order of 12 were analyzed from a Hamming windowed speech signal of a 25.6-millisecond time interval. This analysis was repeated over the speech samples successively. Then, three successive intervals for the five Japanese vowels were extracted from the four words as follows.

/u/ in /bakuoN/
/o/, /e/ in /kogeN/
/a/ in /namae/
/i/ in /umi/

Averaging the analyzed results for the three time intervals, each of which was a windowed speech sample of 25.6 milliseconds, the frequency spectrum envelopes for the five Japanese vowels vowels were determined, and the three formant fre-

* Musashino Electrical Communication Laboratory, Nippon Telegram and Telephone Public Corporation

quencies were estimated for each vowel. Such a procedure was executed on six speech samples uttered every six monthes for each of the nine speakers.

Results

There are 54 spectral envelope patterns of six repetitions by nine speakers for each of five vowels. Averaging the six vowel spectra of the repeated utterances for each speaker, the five vowel spectra are shown in Fig. 1 for the nine speakers. It seems that the pattern difference among the vowels was more than that among the speakers for vowels in the same context.

To estimate the temporal variation of the vowel frequency spectrum for each speaker, an analysis of variance was undertaken for each of the three formant frequencies of the five vowels. Main factors for the analysis of variance were speakers (nine levels) and number of repetition of utterance (six levels).

An example of the analysis of variance for the vowel /e/ is shown in Table 1. Summarizing the results of the analysis for the five vowels, the F ratios of the two main factors, the contribution rate of the factor of speaker and the standard deviation of the error are included in Table 2. There were a few occasions where the factor of repetition was significant, but the variance of the error was calculated including that for the number of repetition. The speaker factor was significant at all of formant frequencies of the five vowels.

Table 1 *Results of the analysis of variance for F1, F2 and F3 of the vowel /e/ uttered by nine speakers.*

a) F1

	d.f.	S.S.	M.S.	F
S_A(Speaker)	8	363741.67	45467.78	30.5**
S_B(Time)	5	11393.06	2278.61	1.53
E	40	59486.11	1489.65	

b) F2

	d.f.	S.S.	M.S.	F
S_A(Speaker)	8	746831.48	93353.94	34.7**
S_B(Time)	5	17142.59	3428.52	1.28
E	40	107524.08	2688.10	

c) F3

	d.f.	S.S.	M.S.	F
S_A(Speaker)	8	473445.37	59180.67	9.4**
S_B(Time)	5	68575.93	13715.19	2.2
E	40	251681.7	6292.04	

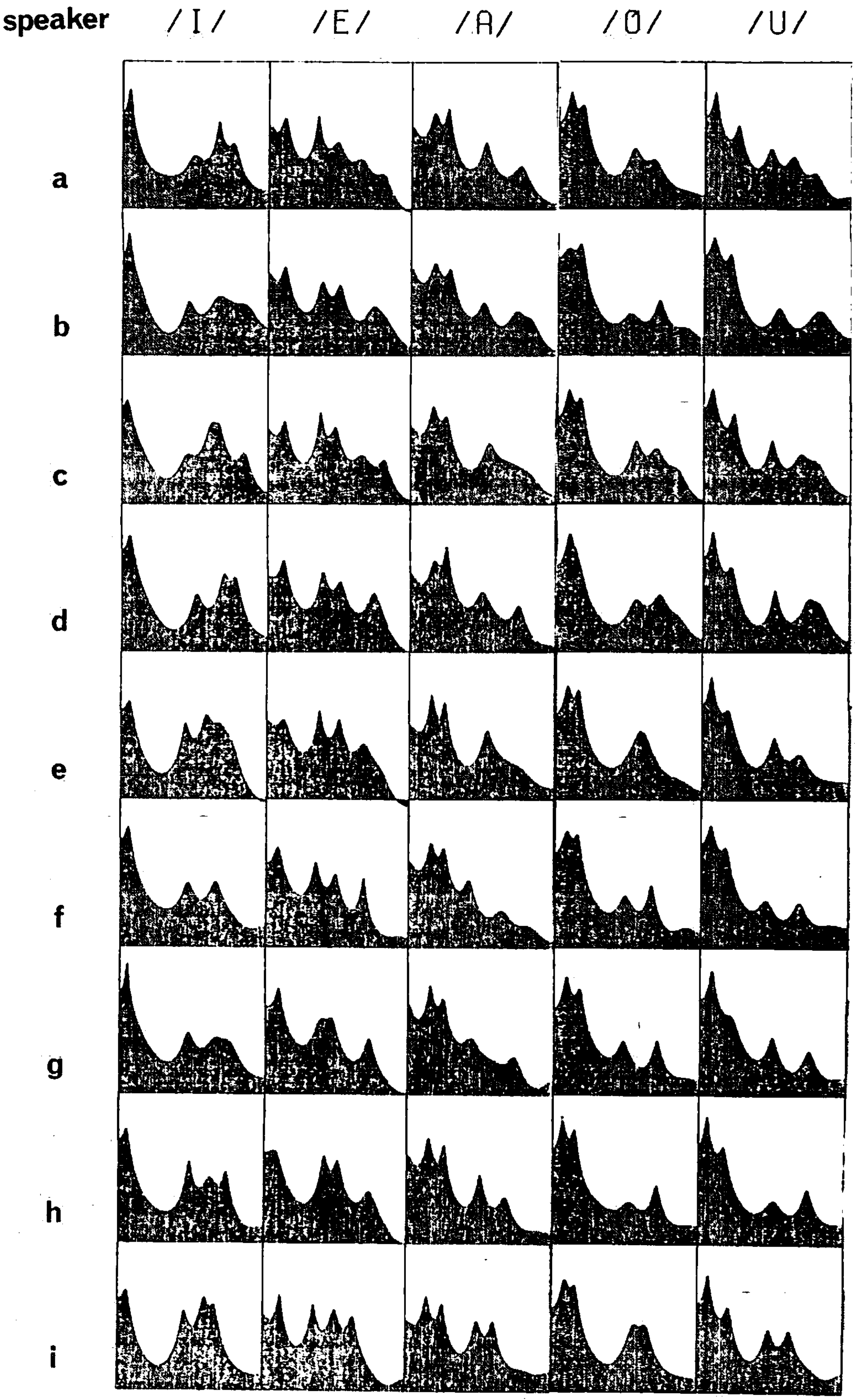

Fig. 1 *Frequency spectrum envelope patterns for the five Japanese vowels uttered by nine speakers*

Table 2 *Summarized results of the analysis of variance for the three formant frequencies of the five vowels.*

		F ratio		Contribution rate of the speaker factor (%)	Standard deviation of the Error (Hz)
		Speaker	Number of repetition		
/i/	F1	4.9**	2.5*	36.3	32.5
	F2	3.3**	2.1	26.4	170.2
	F3	10.3**	1	60.9	134.2
/e/	F1	30.5**	1.5	81.6	39.7
	F2	34.7**	1.3	83.8	52.6
	F3	9.4**	2.2	54.9	84.4
/a/	F1	17.2**	3.7**	67.0	127.5
	F2	5.6**	1.1	43.0	40.2
	F3	35.8**	1.1	84.5	90.3
/o/	F1	4.9**	1.4	38.3	39.4
	F2	8.2**	1.4	52.9	53.5
	F3	14.3**	3.0*	63.9	127.8
/u/	F1	7.2**	1.3	49.6	35.0
	F2	24.9**	2.6*	76.6	60.0
	F3	8.1**	1.8	66.6	86.7

It can be seen in Table 2 that

(1) The contribution rate of the variation based on the speaker difference differed for the five vowels. That for the vowel /i/ was the lowest and that for the vowel /e/ was the highest.

(2) The contribution rate of the variation differed also in the formant frequencies; that of the higher formant frequencies were higher than those of the lower ones.

(3) The standard deviation of the error ranged from 30 to 170 Hz: that of the higher formant frequencies was larger than that of the lower ones in general.

It is well known that the correlation of the two formant frequencies, F1 and F2, is the principal phonetic feature of vowels. To investigate the correlation of the formant frequencies in our data, two diagrams of the formant frequencies, F1-F2 and F1-F3 were derived and are illustrated in Figs. 2 and 3, respectively.

It can be seen in these figures that the variation in the formant correlation

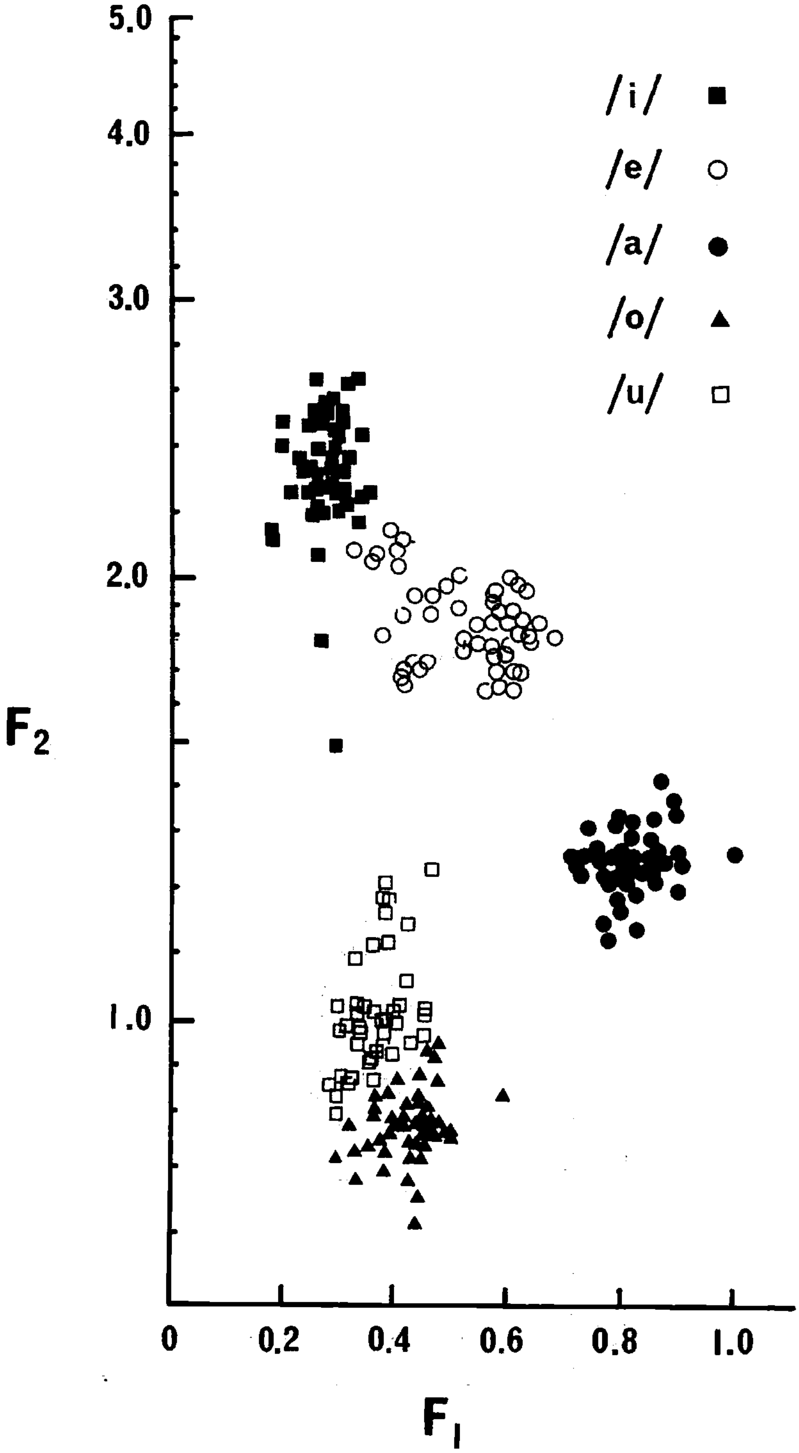

Fig. 2 *F1-F2 diagram for the five Japanese vowels uttered by nine speakers*

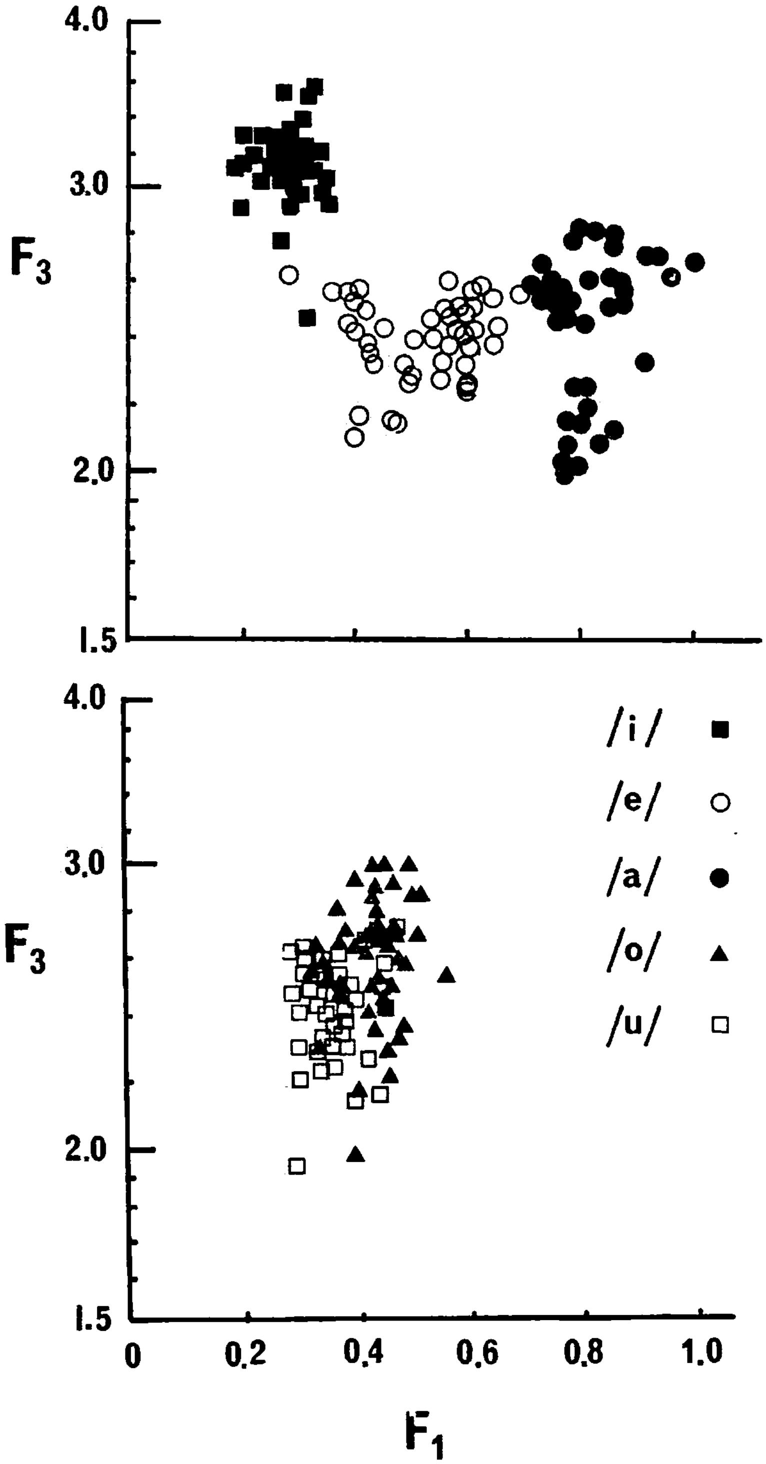

Fig. 3 *F1-F3 diagram for the five Japanese vowels uttered by nine speakers*

depends mainly on the variation in the higher formant frequency in the case of the vowel /i/. Similar results were also observed in the case of the vowels /u/ and /o/. In the case of the vowel /e/, both formant frequencies F1 and F2 varied in the circular direction of the F1-F2 diagram, that is, the ratio of F2/F1 was not held constant but allowed to vary among speakers. A similar relation was also seen in the case of the vowel /a/.

Summarizing the results of Figs. 2 and 3, it can be concluded that

(1) The correlate of the formant frequencies differs vowel by vowel: variation in the formant correlation is higher for F1-F2 of vowel /e/ and F1-F3 of vowel /a/.

(2) In the correlates of the formant frequencies for /i/, /o/ and /u/, variations in the formant correlations are mainly caused by the F2 or F3 changes.

Discussion

The results of our present study show that variations in each formant frequency, and also in the correlates of the formant frequencies among speakers differ vowel by vowel. It is suggested that there are a few preferable vowels for representing speaker differences. To represent it more explicitely, further study is needed using frequency spectrum patterns instead of formant frequencies.

It should also be noted from our present study that the normalization of frequency spectral differences by the use of ratios between the formant frequencies,[4] or corrections based on the vocal tract length,[5] will not be sufficient, because variations in the formant correlates are not always in a radial direction, but in a circular one on a two-formant diagram.

Acknowledgments

This study was supported inpart by a Grant-in-Aid for Scientific Research (No. 0054003) of the Japanese Ministry of Education, Science and Culture and by a Grant-in-Aid for Research of the Asahi-Shimbun.

References

1. Furui, S., F. Itakura and S. Saito (1975); Personal information in the long-time averaged speech spectra, Review of the E.C.L., N.T. & T., 23, [9/10], p. 1133.

2. Furui, S. and F. Itakura (1976): Analysis of talker differences in statistical properties of speech spectra, Review of the E.C.L., N.T. & T., 24, [5/6], p. 418.

3. Furui, S. (1977); Analysis of temporal variation of talker dependent features, Review of the E.C.L., N.T. & T., 25, [3/4], p. 231.

4. Kohda, M. and S. Saito (1972); Speech recognition by incomplete learning samples, 1972 Conference on Speech Comm. and Processing, April 24-26, 1972, Mass., p. 311.

5. Wakita, H. (1977); Normalization of vowels by vocal-tract length and its application to vowel identification, IEEE Trans. ASSP, ASSP-25, [2], p. 183.