Ann. Bull. RILP (1983) No. 17, 97-106

FREQUENCY SPECTRUM DEVIATION FOR JAPANESE VOWELS BETWEEN SPEAKERS

*Shuzo Saito and Fumitada Itakura**

1. Introduction

The speech waves uttered by the vocal organs contain both linguistic and personal information and the aural identification of this information is reliable for normal listeners. The recognition by machine of these two kinds of information, on the other hand, is not reliable despite the fact that a vast amount of acoustic features can be analyzed and used for such recognition processing.

The results of machine recognition experiments have suggested that personal and temporal deviations in the acoustic features of speech affect the recognition of linguistic and personal information.[1,2] In the present study, the five Japanese vowels uttered by nine speakers over three years were analyzed in terms of frequency spectrum and their personal and temporal deviations were investigated using a variance analysis.

2. Speech Analysis Procedures

The speech materials analyzed in the present study were four Japanese words: /bakuoN, kogeN, namae, umi/. These were uttered six times every six months over a period of about three years. The speakers of the speech samples were nine male adults.

The speech waves were sampled at 10 kHz and the speech amplitudes were digitized into twelve bits. The coded speech samples were fed into an Eclipse-230 electronic computer, by means of which the PARCOR speech analysis was carried out.[3] The PARCOR coefficients on the order of 12 were analyzed using Hamming-windowed speech samples with 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;

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

while inspecting the stable vowel portions of the analyzed data.

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

Table 1 *Results of the analysis of variance for the four formant frequencies*

vowel / formant	/i/		/e/		/a/		/o/		/u/	
	Speaker	Temporal diff.	Speaker	Temporal diff.	Speaker	Temporal diff.	Speaker	Temporal diff.	Speaker	Temporal diff.
F 1	4.91**	2.49*	30.52**	1.53	17.19**	3.73**	4.86**	1.38	7.18**	1.25
F 2	3.28**	2.1	34.73**	1.28	5.62**	1.14	8.21**	1.41	24.87**	2.60*
F 3	10.33**	<1	9.41**	2.18	35.81**	1.13	14.28**	3.03*	8.11**	1.78
F 4	5.19**	2.88*	32.52**	2.28	9.25**	1.16	13.25**	1.11	19.52**	<1

Using the analyzed results of the three successive intervals, each of which was a windowed speech sample of 25.6 milliseconds, the frequency spectrum envelopes for the five Japanese vowels were derived. Then the averaged four formant frequencies and also the averaged cepstrum coefficients on the order of 20 were calculated. This procedure was executed for the six speech samples uttered every six months by each of the nine speakers. Thus, 54 pieces of data were collected for each formant frequency and also for each cepstrum coefficient for each vowel.

3. Variance Analysis of the Formant Frequencies

Using the formant frequency data of the six utterances made at different times by the nine speakers, an analysis of variance was undertaken for each of the four formant frequencies of the five vowels. The main factors in the analysis of variance were speakers (nine levels) and temporal difference (six levels).

The results of the analysis of variance are shown in Table 1. It can be seen that the speaker factor was highly significant for all of the four formant frequencies of the five vowels, but there are also a few occasions where the temporal factor was significant.

The contribution rate of the speaker variance (CR) was calculated from the following equation.

$$CR = \frac{(\text{ss for sp.}) - 6 \times (\text{ms for res.})}{\text{total ss}} \qquad (1)$$

where ss for sp.: sum of squares for speaker factor
ms for res.: mean square for residual
total ss: total sum of squares for means.

The CR values and the standard deviation of the residual (σ), in which the variance of the temporal factor was included, are shown in Fig. 1. The following can be seen.

(a) The CR of the speaker factor differed for the five vowels. The CR value for the vowel /i/ was smaller than for the other vowels; and that for the vowel /e/ was larger in general.
(b) The CR of the speaker factor differed, too, among the formant frequencies. The CR values of the higher formant frequencies were larger than those of the lower ones.
(c) The σ of the residual ranged about 30 ~ 170 Hz. The σ value of the vowel /i/ was larger than that of the other vowels in general.

Similar analysis of variance were performed for the ratios of the pairs of formant frequencies F2/F1, F3/F1, F4/F1, F3/F2 and F4/F2. The results are shown in Table 2 and are rather similar to those shown in Table 1. The speaker factor was highly significant for almost all of the formant frequency ratios, excepting the ratios of F3/F2 and F4/F2 for the vowel /e/. Also similar to Table 1, there were a few occasions where the temporal factor was significant.

The contribution rate of the speaker factor was calculated and is shown in Fig. 2 together with the standard deviation of the residual (σ). The following can be seen.

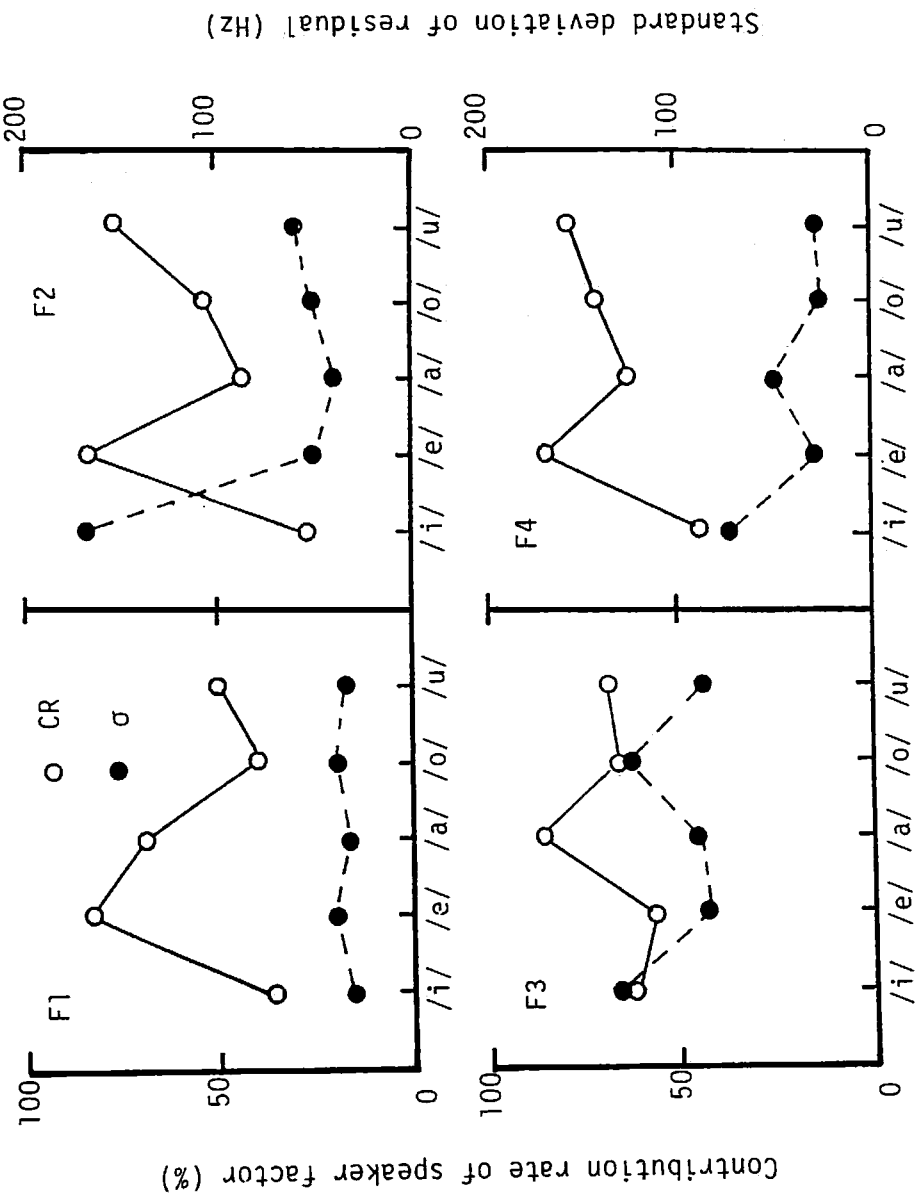

Fig. 1 *Contribution rates (CR) of the speaker factor and standard deviations (σ) of the temporal difference for the four formant frequencies in the five Japanese vowels.*

Table 2 *Results of the analysis of variance for the ratios of the pairs of formant frequencies*

formant \ vowel	/i/		/e/		/a/		/o/		/u/	
	Speaker	Temporal diff.	Speaker	Temporal diff.	Speaker	Temporal diff.	Speaker	Temporal diff.	Speaker	Temporal diff.
F2/F1	3.81**	1.07	36.96**	1.11	13.5**	2.25	6.65**	<1	9.92**	<1
F3/F1	6.12**	2.31	23.68**	1.06	29.32**	3.11*	8.43**	1.68	9.88**	2.13
F4/F1	4.92**	3.68**	32.77**	<1	4.11**	2.69*	9.55**	1.49	10.75**	1.45
F3/F2	2.21*	2.0	1.05	<1	<1	<1	11.23**	1.65	28.39**	4.13**
F4/F2	2.94*	3.06*	1.03	1.04	1.32	<1	9.24**	1.1	24.72**	<1

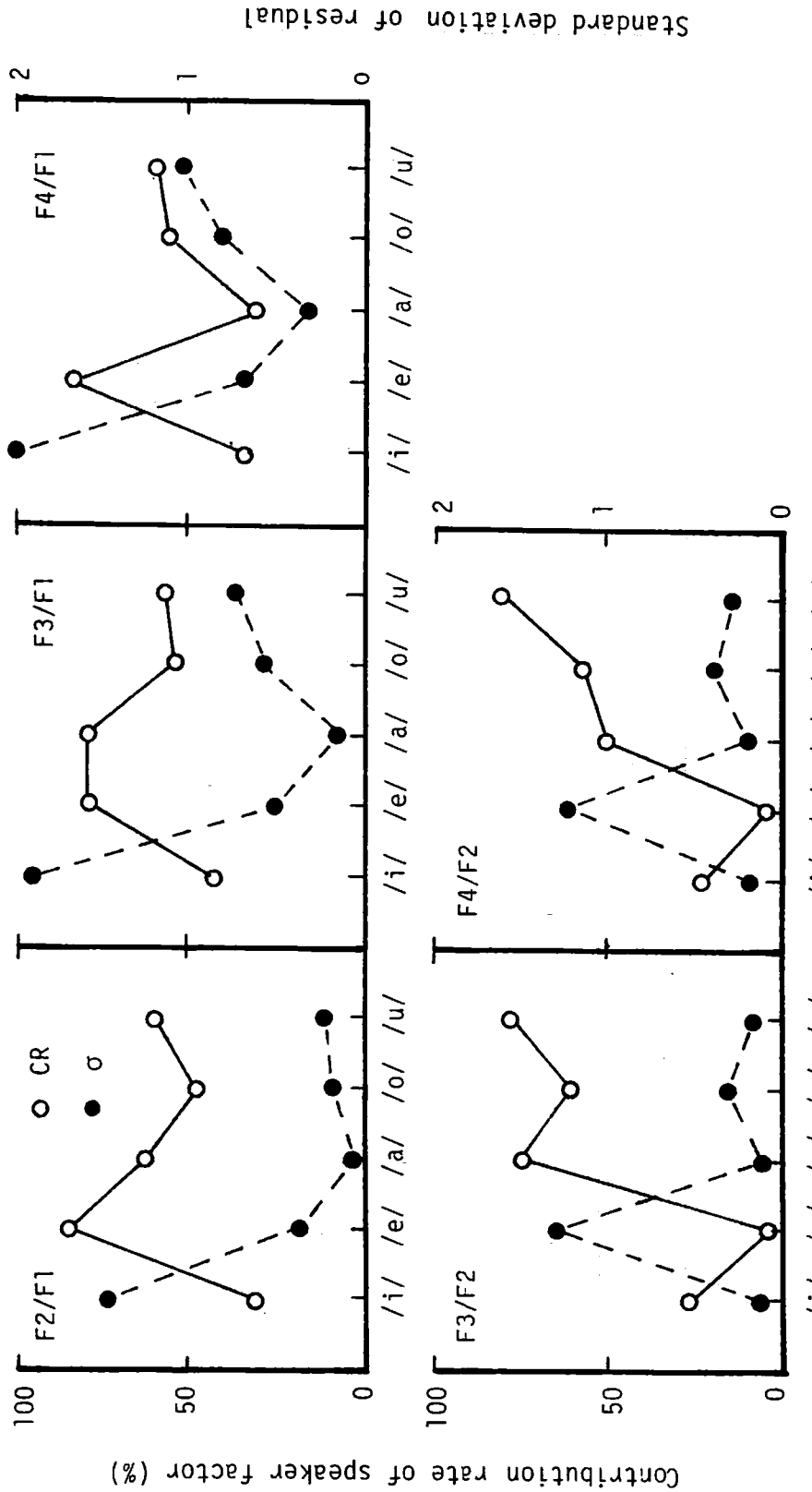

Fig. 2 *Contribution rates (CR) of the speaker factor and standard deviations (σ) of the temporal difference for the two-formant ratios among the four formant frequencies of the five Japanese vowels.*

(d) The CR of the speaker factor for the formant frequency ratios differed among the five vowels. The CR values for the vowel /e/ were larger than for the other vowels in the cases of F2/F1, F3/F1 and F4/F1. Those for the vowel /i/, on the other hand, were smaller than for the other vowels in almost all of the formant ratios. The CR values for the vowels /a/, /o/, /u/ were large in general, especially in the higher formant ratios, such as F3/F2 and F4/F2.

(e) The σ of the residual differed too among the five vowels. The σ values for the vowel /i/ were larger than for the other vowels in F2/F1, F3/F1 and F4/F1, but were smaller in F3/F2 and F4/F2. Those for the vowel /e/ were, on the other hand, relatively small in F2/F1, F3/F1 and F4/F1, but were larger than for the other vowels in F3/F2 and F4/F2. The values of CR and σ for the vowels /i/ and /e/ were inversely related in the two formant ratio groups F2/F1, F3/F1, F4/F1 and F3/F2, F4/F2.

These results are supported by the orthogonal representation of the two formant frequencies F1-F2, which has been reported in our preceeding paper.[4] As can be seen in those diagrams, the variation of the formant frequency ratios is derived mainly from the variation of the higher formant frequencies in the case of the vowel /i/, which can also be seen in Fig. 1. In the case of the vowel /e/, on the other hand, both formant frequencies F1 and F2 deviate between speakers. In this case, the formant frequency ratio F2/F1 deviates in the circular direction of the F1-F2 diagram. In other words, ratios such as F2/F1, F3/F1 and F4/F1 do not hold a constant value but deviate among speakers for the vowel /e/.

4. Variance Analysis of the Cepstrum Coefficient

A similar analysis of variance was undertaken for the cepstrum coefficient on the order of 20 for each of the five vowels and extracted from the same speech samples used for the formant frequency analysis. The results of the analysis of variance are shown in Table 3.

In this table, the mean squared variance values of the cepstrum coefficients are not shown, but the order numbers of the cepstrum coefficients estimated as significant factors are tabulated. It can be seen that the speaker factor was highly significant for the greater part of the cepstrum coefficients, but there were also a few occasions where the temporal factor was also significant.

The contribution rate of the cepstrum variance of the speaker factor (CR) was calculated for each of the five vowels using equation (1). The results are shown in Fig. 3 (a), (b) for the five vowels. These results show the following.

(a) The order of the cepstrum coefficient making a prominent contribution to the speaker factor differed for each of the five vowels. Higher contributions were observed at 2, 4. 8 for the vowel /i/; at 5, 8, 12 for the vowel /e/; at 3, 5. 10 for the vowel /a/; at 5, 7, 8 for the vowel /o/; and at 2, 9 for the vowel /u/.

(b) The standard deviations of the residual (σ) of the 20 cepstrum coefficients decreased in inverse proportion to the order of the cepstrum coefficients. This result and also the actual values of the σ's for the 20 cepstrum coefficients were very similar among the five vowels.

Table 3 *Results of the analysis of variance for the cepstrum coefficients*

Vowel / Cepstrum	/i/		/e/		/a/		/o/		/u/	
	Speaker	Temporal diff.	Speaker	Temporal diff.	Speaker	Temporal diff.	Speaker	Temporal diff.	Speaker	Temporal diff.
Orders of the cepstrum coefficient in significance	1 2 3 6 10 16	1 2 3 6 10 16	1 19	1 4 5	1 18	1 4 16	1 16 18	2 3 4 7	1 12 14 16 19	3

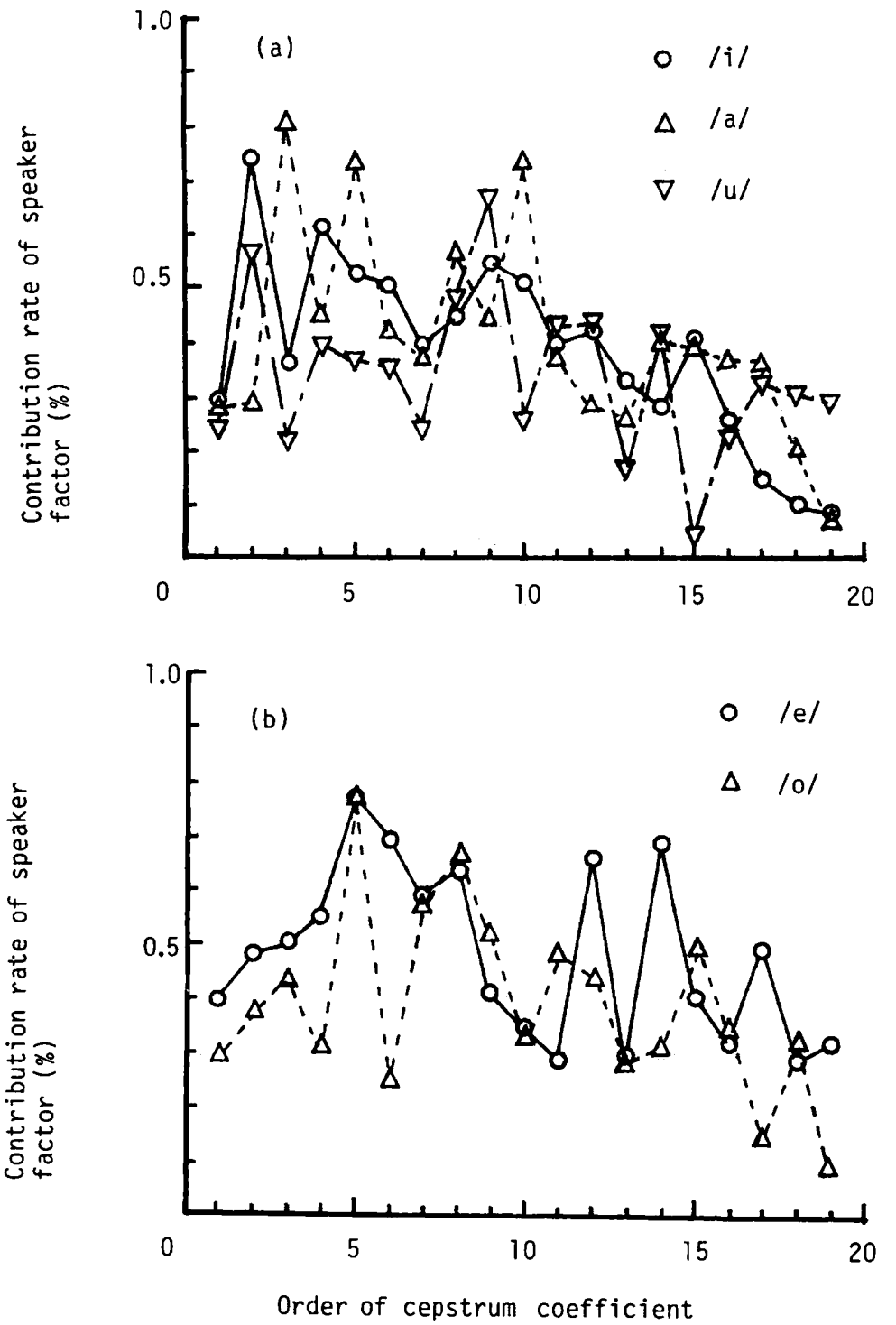

Fig. 3 *Contribution rates (CR) of the speaker factor on the 20 cepstrum coefficients for the five Japanese vowels*

5. Conclusion

From the present study on the frequency spectrum deviation among speakers, the following can be concluded.

1) The deviations of the formant frequency, of the formant frequency ratios and also of the cepstrum coefficients differed from vowel to vowel.
2) The standard deviations of the variance analysis for the formant frequencies and also for the formant frequency ratios differed also from vowel to vowel, whereas the standard deviations of the variance analysis for the cepstrum coefficients were very similar among the five vowels.
3) From the variance analysis of the formant frequency deviations, it seems that the vowel /e/ is suitable for representing speaker differences, but that the vowel /i/ is rather inadequate for this purpose.
4) From the variance analysis of the cepstrum coefficient deviations, it was found that the residuals for the five vowels were very similar to each other, and that all of the whole information about speaker difference is contained in the cepstrum coefficients, although the best vowel for representing speaker differences can not be determined from the present study.
5) It was also noted that the normalization of the frequency spectrum deviation using the formant frequency ratio (that is, a normalization based on the vocal tract length[5]) is insufficient, since the deviation of the formant frequency ratios is not always in a radial direction, but is sometimes in a circular direction as in the two-formant diagram.

Reference

1. 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.
2. Furui, S. and F. Itakura (1976): Analysis of speaker differences in statistical properties of speech spectra, Review of the E.C.L., N.T. & T., 24, [5/6], p. 418.
3. Itakura, F. and S. Saito (1969); Speech analysis-synthesis system based on the partial autocorrelation coefficient, Rep. of the 1969 Autumn Meeting of the Acoust. Soc. Japan, p. 199.
4. Saito, S. and F. Itakura (1982): Personal characteristics of the frequency spectrum for vowels, Ann. Bull. RILP, 16, p. 73.
5. Furui, S. (1977); Analysis of temporal variation of speaker dependent features, Review of the E.C.L., N.T. & T., 25, [3/4], p. 231.