Ann.Bull.RILP(1984)No.18, 107-112

# FURTHER STUDY ON THE FREQUENCY SPECTRUM DEVIATIONS BETWEEN SPEAKERS

Shuzo saito and Fumitada Itakura*

## 1. Introduction

Personal and temporal deviations of the acoustic characteristics of speech between speakers are essential problems to be solved for speech research, particulary for further advance of speech recognition by machine.

Several studies on frequency spectrum deviations between speakers for Japanese vowels have been made and significant deviations of vowel frequency spectra between speakers have been reported, although magnitudes of spectrum deviation between speakers differed among the five vowels. A similar study on frequency spectrum deviation between speakers in three Japanese nasal consonant sounds was performed and is repoted in this paper.

## 2. Speech Analysis Procedures

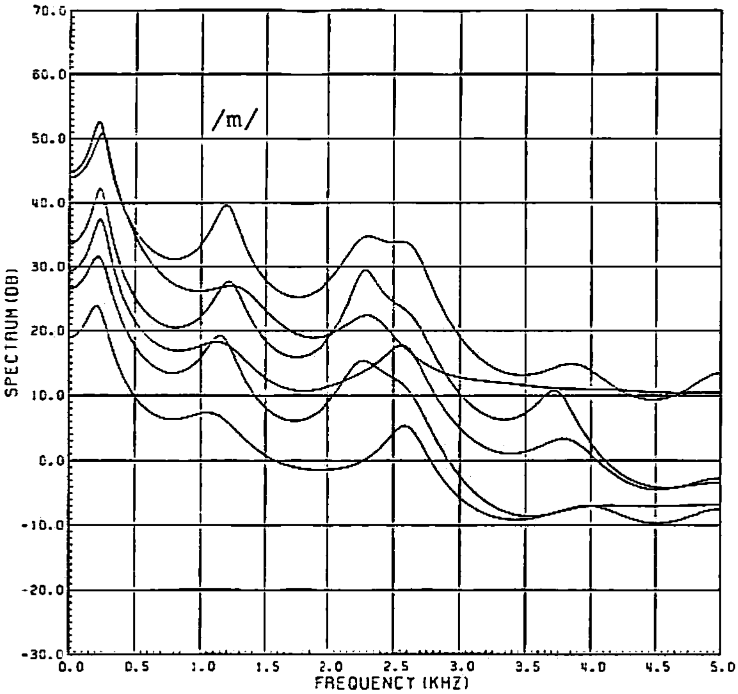
The speech materials analyzed in the present study were two Japanese words, /kogeN, namae/, the same as those used in the preceding study on frequency spectrum deviation for Japanese vowels between speakers.(3) These words were uttered six times every six months over a period of about three years. The speakers were nine male adults.

The speech waves were sampled at 10 kHz and the speech amplitudes were digitized with twelve-bit resolution. The coded speech samples were fed into an Eclipse-230 electronic computer, by means of which a PARCOR speech analysis was carried out. The PARCOR coefficients of order 12 were analyzed using Hamming-windowed speech samples with a 25.6 millisecond time interval. This analysis was repeated over successive frames. Then, three successive intervals for the three Japanese nasal consonants were extracted from the speech materials of two words as follows: /m/ in /namae/, /n/ in /namae/, /N/ in /kogeN/. Samples were chosen from the stable consonant portions of the analyzed data.

Using the analyzed results from all sets of three intervals, frequency spectrum envelopes for the three Japanese nasal consonants were derived. Examples of the frequency spectrum envelopes in the six repetitions for the three nasals are illustrated in Fig. 1. Figs. 1 (a),(b) and (c) are results for consonants /m/, /n/ and /N/, respectively, uttered by a

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

(a)




(b)

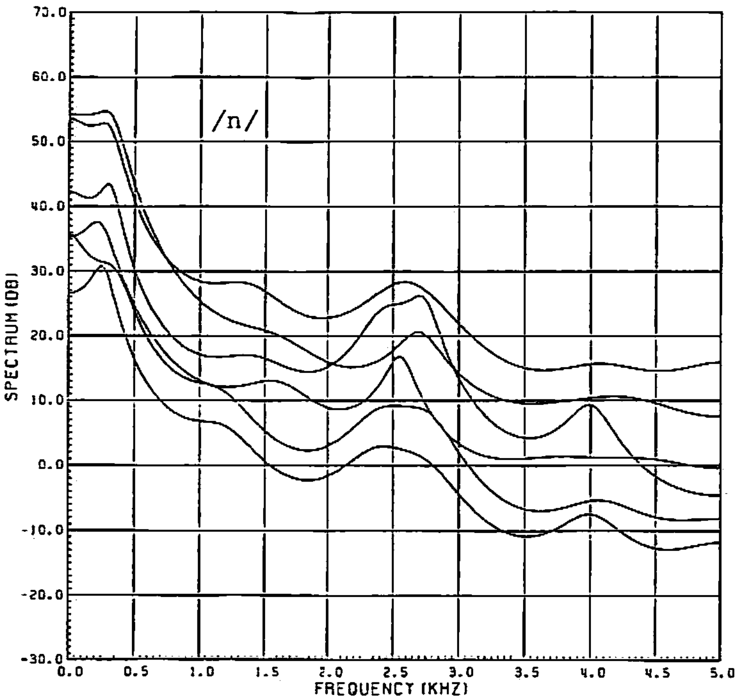
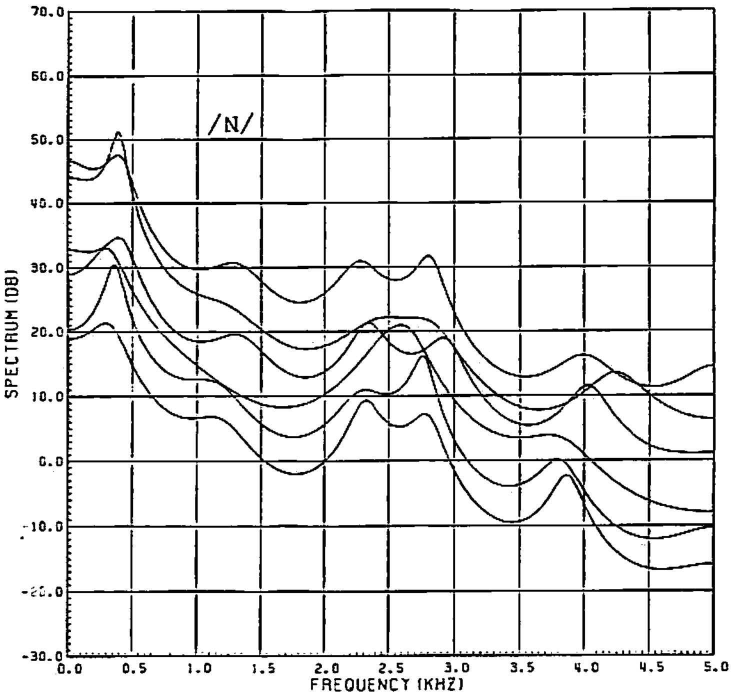



Fig. 1 Frequency spectrum envelopes of the three nasal consonants /m/, /n/ and /N/, uttered in six repetitions.

(c)




single speaker. It is seen that the formant frequencies for the three nasal consonants are relatively stable as are those for vowels. Then, the three formant frequencies in the successive intervals were averaged for each of the six utterance by each of the nine speakers. Thus, 54 data samples are collected for each of the four formant frequencies of each nasal consonant.

## 3. Variance Analysis of the Formant Frequencies

Using 54 samples of formant frequency data of the six utterances by the nine speakers, an analysis of variance was performed for each of the four formant frequencies of the three nasal consonants. The main factors in the analysis of variance were speakers (nine levels) and temporal differences (six levels).

The results of the analysis of variance are shown in Table 1. It is seen that for /m/ and /n/, the speaker factor was relatively highly significant for higher formant frequencies of the three nasal consonants, for /N/ it was highly significant for all of the four formant frequencies. None of the temporal factors or speaker factors of the lower formant frequencies of the nasal consonants /m/ and /n/ were significant.

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

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

where ss for speaker: 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 ($\sigma$), in which the variance of the temporal factor was included, are shown in Fig. 2. It is seen that (1) CR of the speaker factor exceed 40% for the cases of F3, F4 of /m/, F4 of /n/ and F1, F2, F3, F4 of /N/, (2) the $\sigma$ of the residual ranged from about 35 to 300Hz, the higher values being obtained for the higher formant frequencies.

Table 1 Results of the analysis of variance for the four formant frequencies of the three nasal consonants.

| formant \ vowel | /m/ | | /n/ | | /N/ | |
|---|---|---|---|---|---|---|
| | Speaker | Temporal diff. | Speaker | Temporal diff. | Speaker | Temporal diff. |
| F1 | 1.51 | 0.66 | 1.40 | 1.12 | 5.63** | 1.30 |
| F2 | 1.70 | 0.74 | 2.16 | 0.82 | 23.18** | 2.28 |
| F3 | 8.72** | 1.72 | 2.95* | 0.18 | 7.43** | 1.85 |
| F4 | 15.75** | 0.83 | 8.74** | 2.08 | 22.80** | 2.15 |

## 4. Comparison with the Personal Deviations in the Vowels

Results of the analysis of variance for the frequency spectrum deviation for Japanese nasals between speakers were compared with the results for Japanaese vowels that were described in reference (3).

The speaker factor was not significant at the first and second formant frequencies of each of the nasals /m/ and /n/, whereas the speaker factor was highly significant in all of the four formant frequencies of the five Japanese vowels. The speaker factor of the nasal consonant /N/ was highly significant in all of the four formant frequencies and CR value was comparable to that of the vowel /e/, the sound that was previously estimated as the most appropriate for extracting personal differences from the acoustic characteristics.

The standard deviation of the residual was more than twice that for the vowels in general. It seems that the frequency spectra of the nasal consonants are unstable and susceptible to change relative to those of the vowels. It is concluded from these results that the frequency spectrum of the nasal consonants /m/ and /n/ may not be available to estimate the personal differences, whereas that of the nasal /N/ is useful to extract these characteristics.

## 5. Conclusion

From the present study on the frequency spectrum deviation for Japanese nasal consonants between speakers, the following can be concluded.

(1) The speaker factor of deviations differed in each of the three Japanese nasals /m/, /n/ and /N/. The speaker factor of the nasal sound /N/ is highly significant in all of the four formant frequencies. Whereas speaker factors of the nasals /m/ and /n/ are significant at the higher formant frequencies, F3 and F4.

(2) The standard deviations of the residuals that include the temporal difference are large in each of the three nasal sounds relative to those of the vowel sounds.

(3) Frequency spectrum deviation of the nasal sound /N/ between speakers is available to extract personal differences, whereas that of nasals /m/ and /n/ may not be useful to estimate these differences.

## References

1. Saito, S. and F. Itakura (1982); Personal characteristics of the frequency spectrum for vowels, Ann. Bull. RILP, 16, p.73.
2. Saito, S. and F. Itakura (1983); Frequency spectrum deviations between speakers, Speech Communication, 2, p.149.
3. Saito, S. and F. Itakura (1983); Frequency spectrum deviations for Japanese vowels between speakers, Ann. Bull. RILP, 17, p.97.

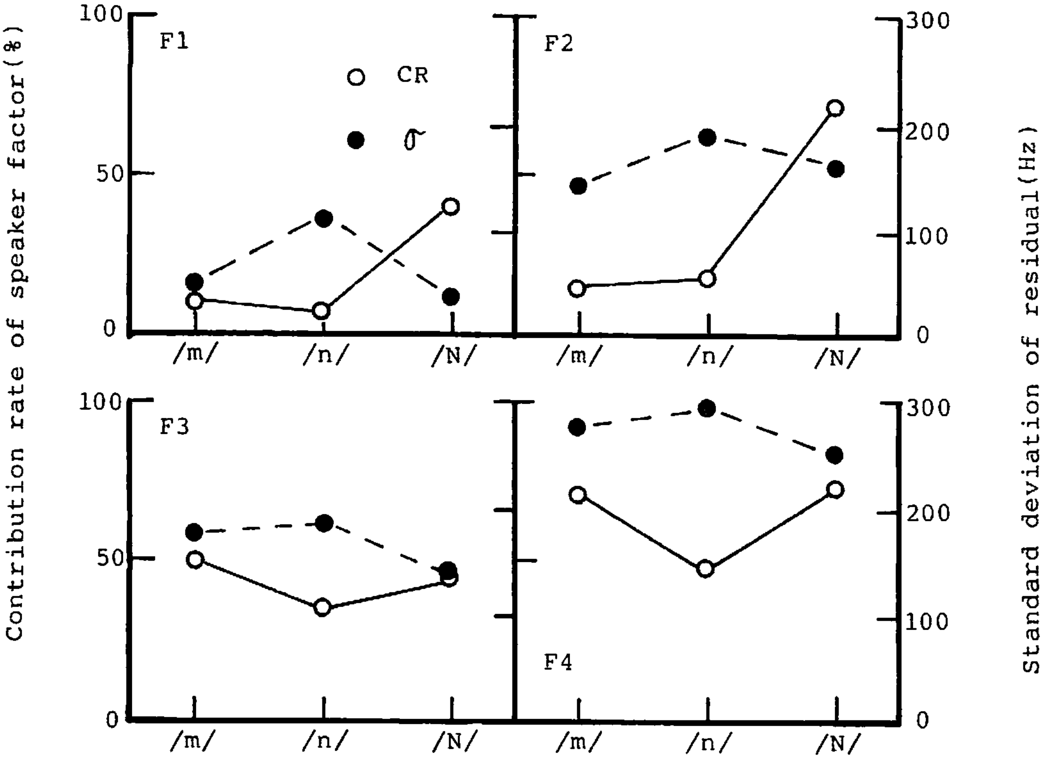



Fig. 2 Contribution rates (CR) of the speaker factor and the standard deviations (σ) of the residuals for the four formant frequencies in the three Japanese nasal consonants.