Ann.Bull.RILP(1989)No.23, 39-49

NORMALIZATION OF FREQUENCY-COMPRESSED VOICELESS FRICATIVES

Sotaro Sekimoto

1. Introduction

The efficiency of hearing aids with frequency-lowering has been investigated by the present authors.[1] The result of a hearing test, where the frequency axis of speech was compressed downward by a PARCOR speech analysis-synthesis method and the characteristics of a hearing impairment were simulated by a low-pass filter, showed that considerable improvement was observed for vowels, whereas little improvement was seen for consonants, and especially for voiceless consonants. It is not apparent, however, why such an improvement was or was not obtained. For vowels, it was observed in a subsequent study by the author that the speech of which the frequency axis was compressed or expanded was identified as original over a wide frequency expansion or compression ratio, especially when the fundamental frequency was concurrently raised or lowered in the same ratio.[2] Namely, the difference of the frequency axis was compensated for in the perceptual process, and perceptual frequency normalization occurred. These results coincide considerably with those for vowels described above, and suggest that perceptual frequency normalization plays an important role in identifying frequency compressed speech correctly. For consonants, on the other hand, although it is supposed that the characteristics of perceptual normalization again play a role in the perception of frequency-compressed consonant, few studies have been made on the normalization of consonant speech.

In the present study, perceptual experiments were performed to elucidate the characteristics of frequency normalization on frequency-compressed voiceless consonants.

2. Method

It is well known that voiceless fricative consonants can be synthesized from a single-pole noise followed by a vowel portion. In this case, the noise portion does not have a particular structure, such as a formant structure which is seen in vowel speech. Accordingly, if perceptual normalization occurs in the perception of voiceless fricatives, it can be assumed that the cue for the frequency normalization is not present in the noise portion itself but is in the relation between the noise portion and the following vowel portion.

Synthetic speech samples in which a single-pole frication noise portion was followed by a vowel portion were adopted to the hearing test to determine the perceptual boundary between /s/ and /ʃ/ on the continuum of resonant frequencies of the noise pole. The boundaries were compared for various ratios of the frequency compression.

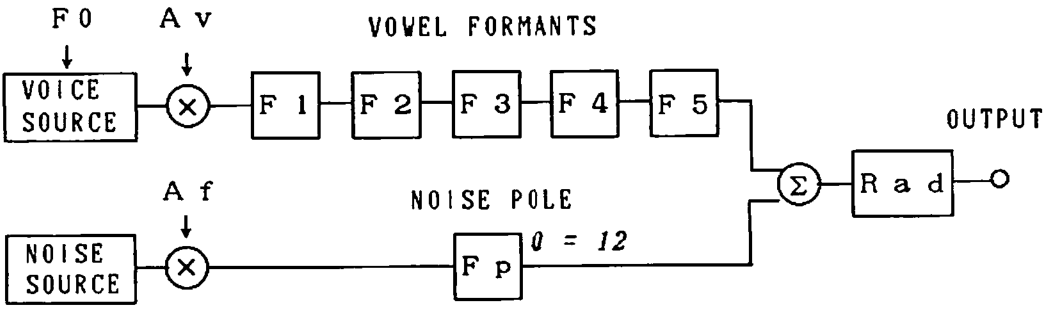

Fig. 1. A block diagram of the software speech synthesizer.

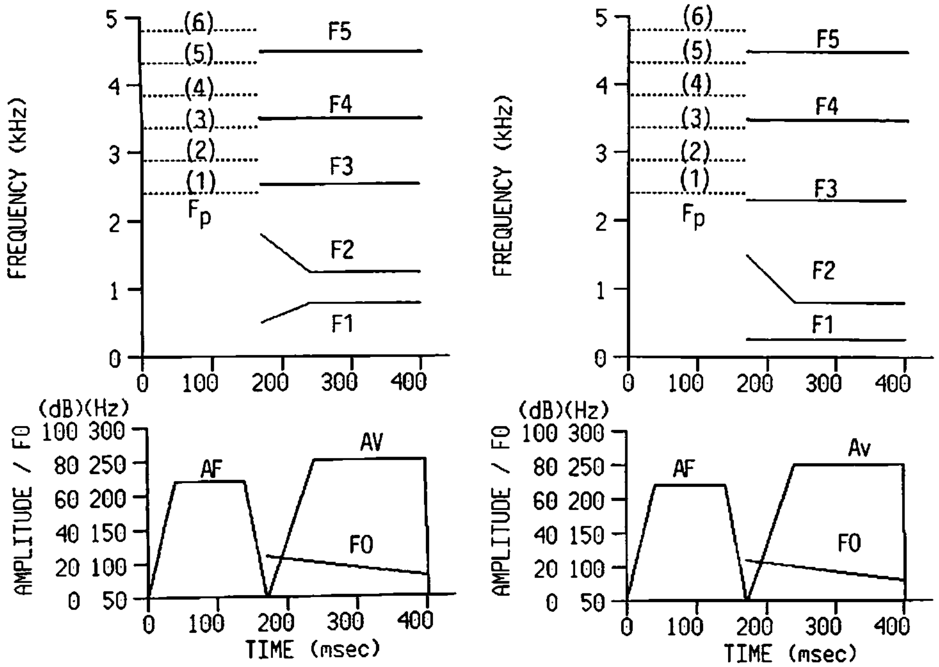

Fig. 2. The time patterns of the control parameters for synthesizing /sa/-/ʃa/ and /su/-/ʃu/ are shown in Figs (a) and (b), respectively. The same fundamental frequency pattern was used in both vocalic contexts.

Stimuli

Stimuli were synthesized with a software terminal-analog speech synthesizer. A block diagram of the synthesizer is shown in Fig. 1. The synthesizer had five poles for vowel production and a pole for frication noise production. Rosenberg's C-waveform was used as a voice source.[3] Q, which represented the sharpness of the resonance of the noise pole was determined from data reported by Fujisaki and Kunisaki.[4] Q=12 was used as an ordinary condition. The time patterns of the control parameters of the synthesizer when the vowels /a/ and /u/ followed the noise are shown in Figs. 2 (a) and (b), respectively. The formant pattern was close to that used in the experiment by Mann and Repp[5], and the same fundamental frequency patterns were used for both /a/ and /u/. The output level was set so that the level of the stationary portion of the noise was 12dB lower than that of the vowel for each stimulus. The compression of the frequency axis was accomplished by lowering the sampling frequency of the synthesizer filters. As a result, the spectrum envelope was compressed toward zero analogously. The sampling frequency without frequency compression was 10 kHz. The speech waveform synthesized on the Apollo DN-4000 workstation was D/A-converted at 12-bit precision and low-pass filtered with a cutoff of -135 dB/oct and was recorded on a DAT(Digital Audio Tape). The cutoff frequency of the low-pass filter was dynamically changed in proportion to the sampling frequency by a factor of 0.45.

Subjects

Six paid students participated.

Procedure

The speech material was presented through binaural headphones (STAX SR-Lambda Signature) in a soundproof room. The presentation level was about 75 dB SPL. The hearing test was carried out with the constant method. The speech samples were presented in a random order. Subjects were requested to identify the synthetic stimuli as one of the following Japanese syllables: /sa/, /su/, /so/, /ʃa/, /ʃu/, /ʃo/, /ha/, /fu/, /ho/. Each token was presented 20 times for each subject.

Experiments

The following four conditions were examined.

1) The effect of the different following vowels. Vowels which were heard as /a/ and /u/ in the uncompressed condition were examined.

Table 1. Pole frequencies of fricative noises. (in Hz)

Stimulus Number					
1	2	3	4	5	6
2400	2880	3360	3840	4320	4800

2) The effect of the lowering of the fundamental frequency of the following vowel. The boundary between /s/ and /ʃ/ was compared between the condition where the fundamental frequency was not lowered (100%) and where the fundamental frequency was lowered by a factor of 0.7 (70%). These fundamental frequencies were kept independent of the frequency compression ratio. This condition was examined because the lowering of the fundamental frequency had shown a remarkable effect on the frequency-normalization of the vowels in the previous study[2].

3) The effect of the formant transition. The boundary between /s/ and /ʃ/ were compared between the conditions where the formant transition was present and absent. When the formant transition was absent, the onset values of the formant frequencies were held at the same value as in the stationary portion.

4) The effect of the envelope of the frication noise spectrum. For this purpose, the sharpness of the resonance (Q) was changed. The overall spectrum was different, and the amplitude difference between noise and vowel portion was kept constant. The boundary between /s/ and /ʃ/ was compared between the conditions when Q=12 (ordinary condition) and Q=7 (damped).

Throughout these four experiments, the following three frequency compression ratios were adopted: 100% (uncompressed), 80% and 60%. The resonant frequency of the noise pole was varied from 2400 Hz to 4800Hz in 480Hz steps as shown in Table 1. The synthetic speech stimuli whose following vowel was /a/, the formant transition and the original fundamental frequency, with a Q=12 fricative pole, were used as a reference condition for each experiment.

3. Results and discussion

Results are shown in Figs. 3 - 6. In these figures, for convenience of comparison, the figures from the reference condition are shown in Figs. (a) and (c), and the results which are specific to each experiment are shown in Figs. (b) and (d).

The results of Experiment 1 are shown in Fig. 3. The answers of six subjects are averaged. Figs. 3 (a) and (c) show the identification rates when the noise portion was followed by /a/ and /u/, respectively. The ordinate shows the absolute noise pole frequency after frequency compression was made. The abscissa shows the identification rates. The results for the three fre-

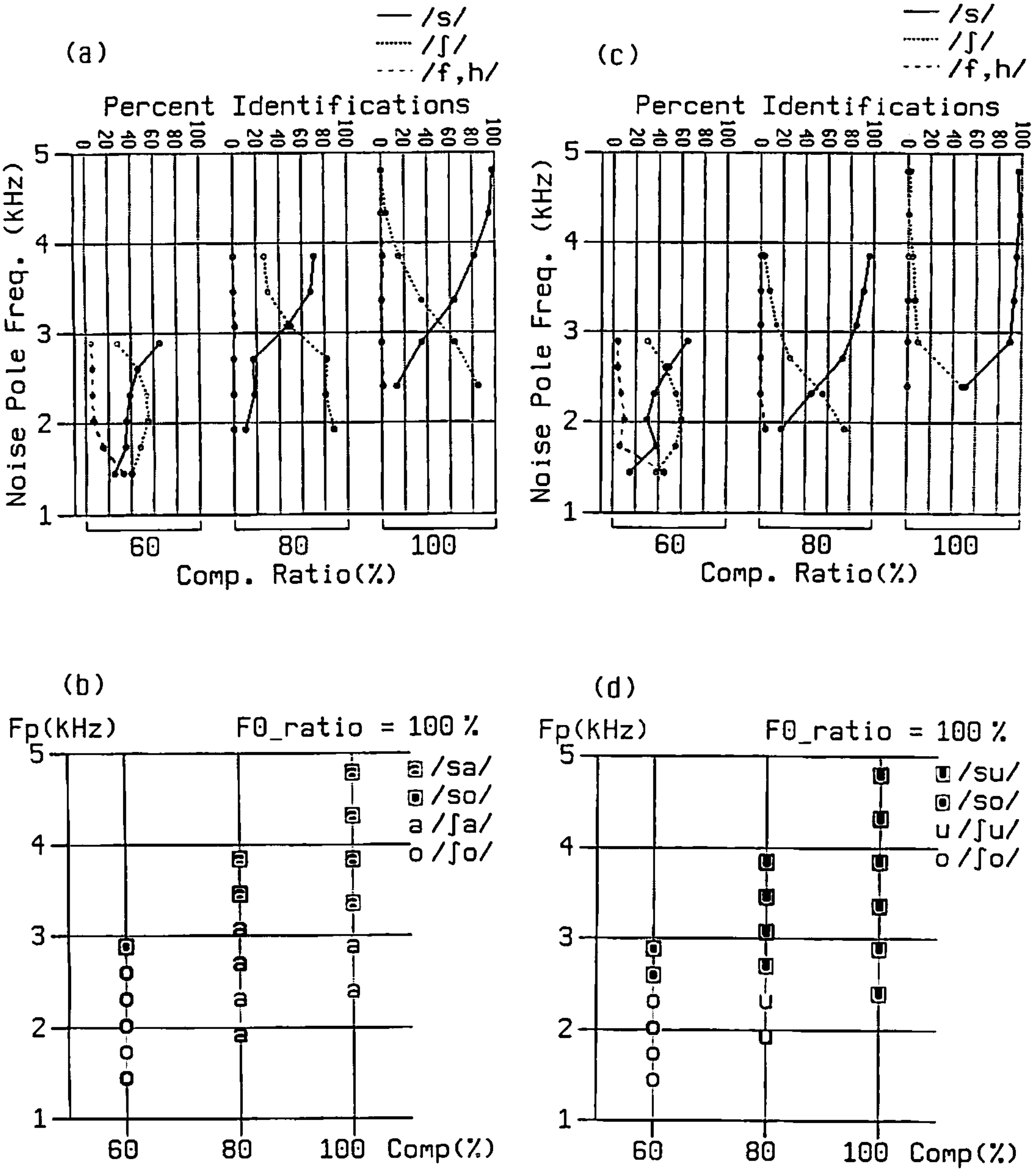

Fig. 3. Results of Experiment 1.
The identification rates when the noise portion was followed by /a/ and /u/, are shown in Figs. (a) and (c), respectively.
The ordinate shows the absolute noise pole frequency after the frequency-compression was made. The abscissa shows the identification rates. The results for the three frequency compression ratios 60%, 80% and 100% (uncompressed) are shown in a same figure. The identification rates for /s/, /ʃ/ and /f/ are shown by solid lines, dots and dashes, respectively. In Figs. (b) and (d), the identified vowels are shown. The ordinate shows the absolute noise pole frequency after frequency-compression was made. The abscissa shows the frequency-compression ratio. The consonant which showed the highest identification rate is shown by inverted and normal character for /s/ and /ʃ/, respectively.

quency compression ratios 60%, 80% and 100% (uncompressed condition) are shown in the same figure for the sake of comparison. The identification rates for /s/, /ʃ/ and /f/ are shown by solid lines, dotted lines and dashed lines, respectively. Figs. (b) and (d) show the identification of the following vowels.

When the compression ratios were 100% and 80%, the response for /f/ was almost zero. The boundary pole frequencies, where the responses of /s/ and /ʃ/ cross, were almost identical for the frequency compression ratios 80% and 100% for both /a/ and /u/; however, their absolute boundary pole frequencies were different. The boundary frequency when the noise portion was followed by /u/ shifted downward compared with /a/. This result suggests the existence of a context effect from the following vowel on the identification of the prevocalic voiceless fricative consonant. This result agrees with that of Kunisaki and Fujisaki.[4]

In the case where the frequency compression ratio was 60%, the response for /f/ increased, and the maximum identification rates for /s/ and /ʃ/ became lower. When the noise portion was followed by /a/, the boundary pole frequency between /s/ and /ʃ/ for the compression ratio 60% was lower than that for the compression ratios 80% and 100%. On the other hand, when the noise portion was followed by /u/, the boundary noise pole frequency was similar among the three frequency-compression ratios. Note that the following vowel was identified as /o/ for both vowel context conditions when the frequency compression ratio was 60%. In Kunisaki and Fujisaki,[4] the perceptual boundary of the noise pole frequency between /s/ and /ʃ/ was affected by the following vowel and the extent of the frequency shift was similar between /u/ and /o/ in comparison with /a/ and /e/. Thus, it can be cpncluded that the downward boundary shift when the noise portion was followed by /a/, which was exclusively observed for the frequency-compression ratio of 60%, is explained by the context effect of the following perceived vowel /o/.

These results suggest that the perceptual boundary is not changed by the frequency-compression as long as the identification of the vowel does not change, namely, the identification for voiceless fricative consonants is performed based on the absolute noise pole frequency.

The results of Experiment 2 are shown in Fig. 4. The boundary frequencies between /s/ and /ʃ/ for each frequency-compression ratio did not shift when the fundamental frequency was lowered. This result suggests that the identification of voiceless fricative consonants is irrelevant to the lowering of the fundamental frequency of the following vowel.

The results of Experiment 3 are shown in Fig. 5. The effect of the formant transition seems to be small. This result suggests that the formant transition does not affect the boundary noise pole frequency which was obtained in Experiment 1. This result is inconsistent with the results by Whalen[5] and Mann and Repp[6]. These authors reported a considerable effect of the formant tran-

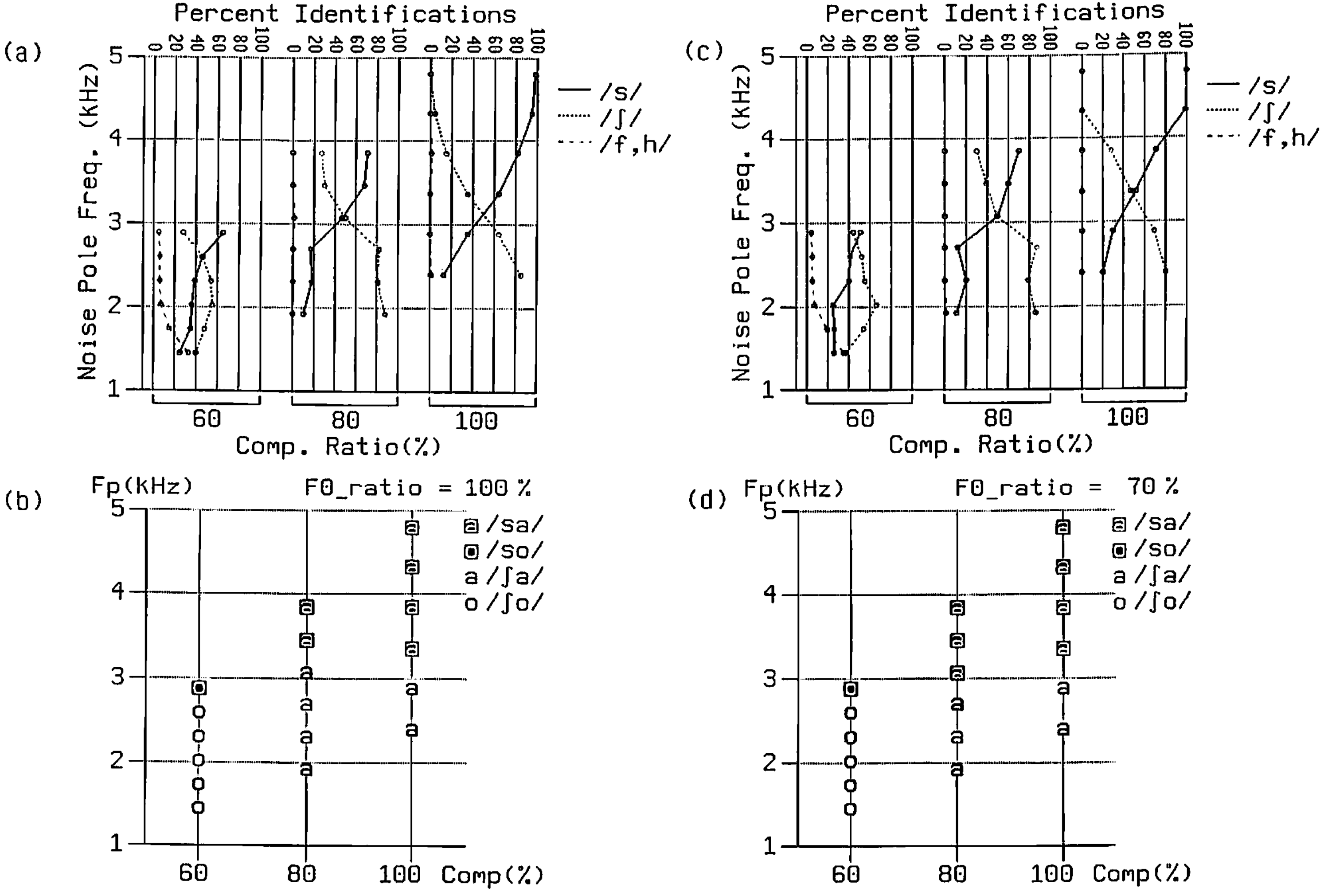

Fig. 4. The effect of the lowering of the fundamental frequency.

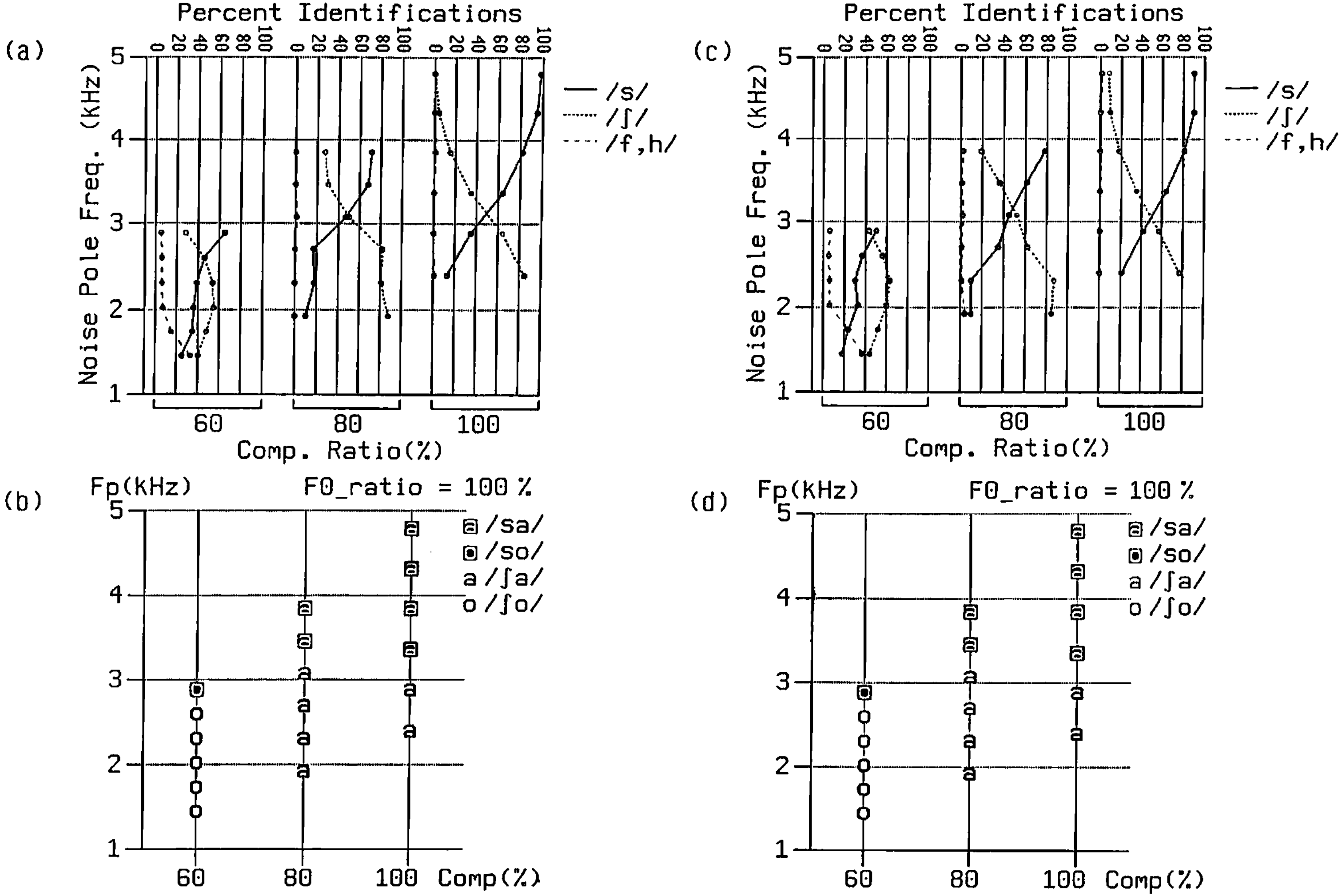

Fig. 5. The effect of the formant transition.

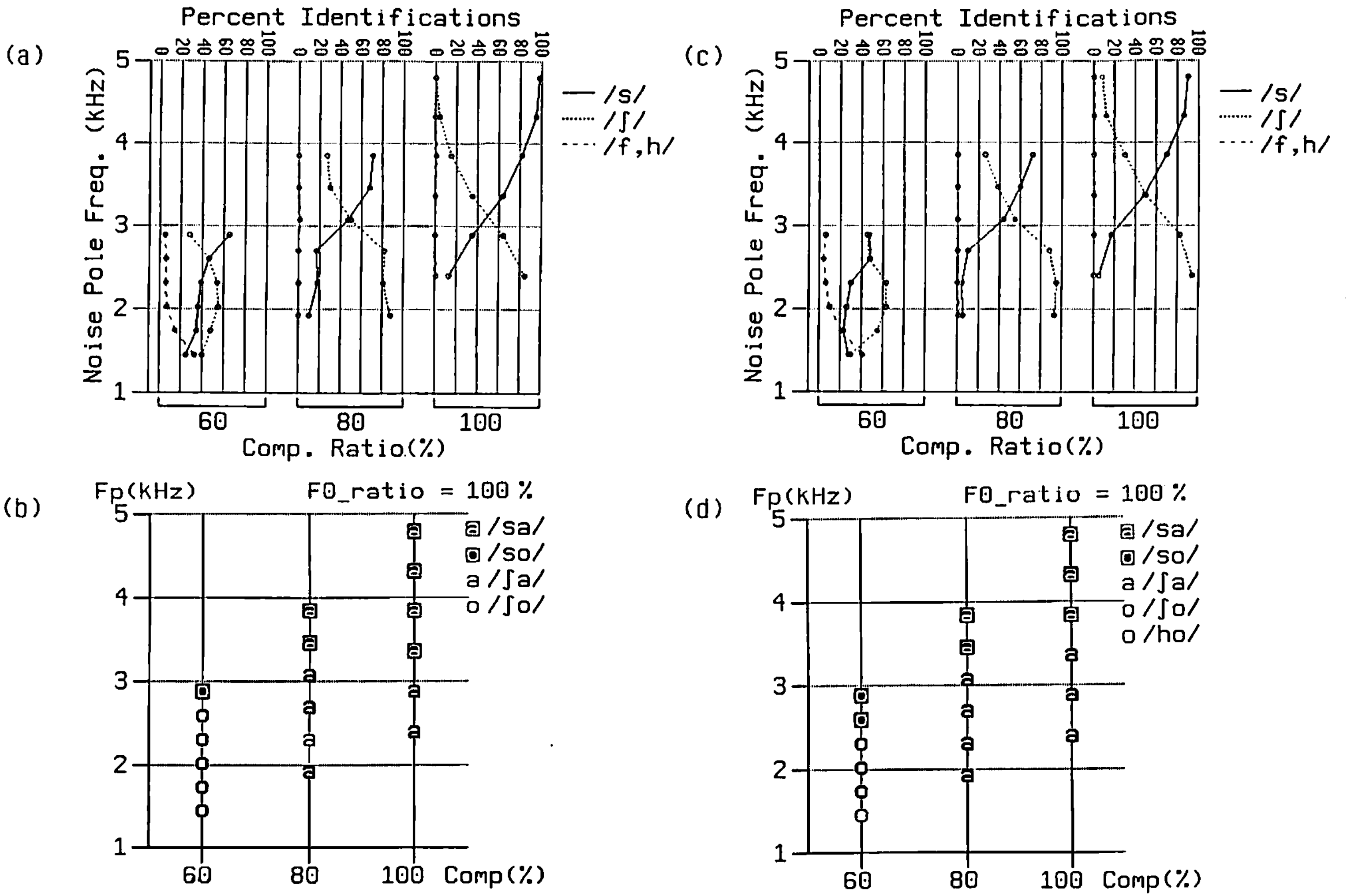

Fig. 6. The effect of the difference in the Q of the noise pole.

sition on the identification of /s/ and /ʃ/. In the present experiment, however, an effect for the formant transition was not found. Possible reasons for this were that an artificial formant transition pattern was used, and a silent gap was present between the noise portion and the following vowel portion in the synthetic stimuli used in this experiment. It was observed in a preliminary experiment that /j/ was always perceived clearly and /ʃ/ was always identified if the noise portion and the vowel portion were quite close. It may be concluded when noise is followed by /j/ and the frequency of the noise is within the /ʃ/ and /s/ region, the existence of /j/ plays a primary role in distinguishing /s/ and /ʃ/. That is, it triggers the perception of /ʃ/ without testing the noise frequency, because there is no other candidate for a voiceless fricative in Japanese which can be followed by /j/.

Fig. 6 shows the results of Experiment 4. The boundary pole frequencies between /s/ and /ʃ/ were small. In the present study, a single pole had been used to represent the noise characteristic in order to avoid complexity. If multiple noise poles are used, there is no reasonable criterion for changing the pole frequencies systematically and for assessing the relation between the noise and the vowel portion, ie, knowing which noise component is related to which formant component. Kunisaki and Fujisaki[4] have reported that two poles and one zero showed best-match in simulating the noise spectrum envelope of natural voiceless fricative consonants. If it is true that the identification of the voiceless fricative consonants is determined from the absolute noise frequency, the cue to the normalization must be present in the noise spectrum itself. Although the possibility that the relation among multiple noise poles in the frequency domain determines the identification of voiceless fricative consonants can not be neglected, there is another possibility that some spectral template is used for the identification. Experiment should be performed as variations of this experiment, where the overall spectral characteristics are modified without changing the peak frequency in such a way that low-pass or high-pass characteristics is used.

Fig. 7 shows the supplementary result of an informal experiment where the frequency axis was compressed and expanded from 50% to 140% with fundamental frequency lowering and raising at the same ratios. It seems that the boundary pole frequencies between /s/ and /ʃ/ were consistent for a wide range of frequency-compression ratios. These data support the assumption that voiceless fricative consonants are identified from the absolute noise pole frequency.

4. Conclusion

The results of the above experiments suggest that the identification of voiceless fricative consonants in frequency compressed speech are mainly based on the absolute noise pole frequency, that is, that the normalization does not occur within voiceless fricative consonants. At the same time, the identification is affected by the vocalic context.

References

1) Sekimoto, S., S. Kiritani and S. Saito (1980); Intelligibility of frequency compressed speech in low-pass filtered condition, Ann. Bull. RILP, 14, 181-193.
2) Sekimoto, S. (1982); Perceptual normalization of frequency scale, Ann. Bull. RILP, 16, 95-101.
3) Rosenberg, A. E. (1971); Effect of glottal pulse shape on the quality of natural vowels, JASA, 49, 2(Pt. 2), 583-590.
4) Kunisaki, O. and H. Fujisaki (1977); On the influence of context upon perception of voiceless fricative consonants, Ann. Bull. RILP, 11, 85-91.
5) Whalen, D. H. (1979); Effects of vocalic formant transitions and vowel quality on the English [s]-[š] boundary, Haskins Laboratories SR-59/60, 35-48.
6) Mann, V. A. and B. H. Repp (1980); Influence of vocalic context on perception of the [s]-[š] distinction, Perception & Psychophysics, 28 (3), 213-228.
7) May, J. (1976); Vocal Tract Normalization for /s/ and /š/, Haskins Laboratories SR-48, 67-73.

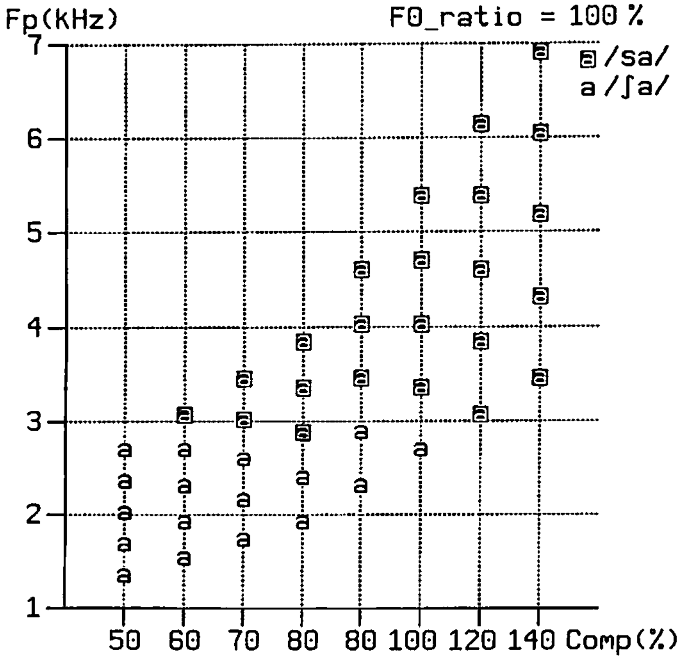

Fig. 7. Identification of /s/ and /ʃ/ at various frequency compression ratios.