# Effect of Fundamental Frequency and Silent Intervals on the Voicing Distinction for Stop Consonants in VCV Contexts

Akiko Hayashi*, Satoshi Imaizumi, Takehiko Harada**, Hideaki Seki*** and Hiroshi Hosoi****

## 1. Introduction

Many studies have indicated that the voicing distinction for stop consonants in $V_1CV_2$ contexts is affected by factors such as *VOT*, $F_0$, the length of $V_1$ and the length of the silent interval between $V_1$ and C, and so on. The effects of these factors for healthy-hearing subjects have been well explained based on the motor theory, in which peripheral auditory functions do not play important roles[1]. In this paper, we try to extract possible effects of peripheral auditory functions of the hearing-impaired. Our presumption is that healthy auditory functions are so well developed/organized that they impose only negligible or no constraint on speech perception, but impairment in auditory functions may impose significant constraint on speech perception.

In this study, two hypotheses were examined. First, to verify the relationship between speech perception and the function of the peripheral auditory system, the effects of an auditory filter ringing on the voicing judgment for stops in $V_1CV_2$ contexts was examined.

This examination was based on the following consideration. Fig.1 shows the relationship between the auditory filter band width, the traditional critical band width and the frequency. This figure was made based on the results reported by Moore et al. (1986)[2], and modified partly by us[3,4,5]. The auditory filter band width (ERB) increases when the frequency becomes higher as shown in Fig.1. In addition, Shailer and Moore (1983)[6] have shown that the temporal resolution measured by the gap detection threshold was in inverse proportion to the auditory filter band width, when the frequency was under 3kHz. Thus, the temporal resolution deteriorated in the low frequency range. It has been hypothesized that this poor temporal resolution in the low frequency range may result from the prolonged auditory filter ringing caused by the narrow filter band width in the low frequency range.

If this explanation is correct, it can be expected that the auditory filter ringing, which is longer for lower frequencies, will affect the voicing judgments for stop consonants in $V_1CV_2$ contexts by shortening the perceptual length of the silent interval. Ac-

---

*The Research institute of the Education for Exceptional Children, Tokyo Gakugei University; **Department of Otolaryngology, Faculty of Medicine, University of Tokyo; ***Department of Computer Science, Chiba Institute of Technology; ****Department of Otolaryngology, School of Medicine, Kinki University

cordingly, we formulated the following hypotheses; $H_1$) For voicing judgments of stop consonants in $V_1CV_2$ contexts, when the fundamental frequency of $V_1$ is low, a longer silent interval between $V_1$ and $CV_2$ may be needed for the unvoiced detection of C, because the perceptual length of the buzz bar of C is prolonged by the auditory filter ringing.

There are other phenomena to be taken note of such as comodulation masking release and the cocktail party effect. Recent experiments demonstrate that output from other auditory filters, tuned to frequencies remote from the signal frequency, can both enhance and degrade signal detection and discrimination in noise (Moore, 1992)[11]. It is worth examining the effect these phenomena have on speech perception in relation to the speech perception of hearing-impaired listeners with wider auditory filters. Accordingly, we formulated the following second hypotheses; $H_2$) When the fundamental frequency of $V_1$ differs from that of $V_2$ to an unusual extent, the effect of the silent interval on voicing judgments for stop consonants in $V_1CV_2$ contexts can be decreased, because perceptual segregation may arise and $V_1$ and $CV_2$ may heard separately.

In this study, the two hypotheses $H_1$ and $H_2$ were examined in two experiments.

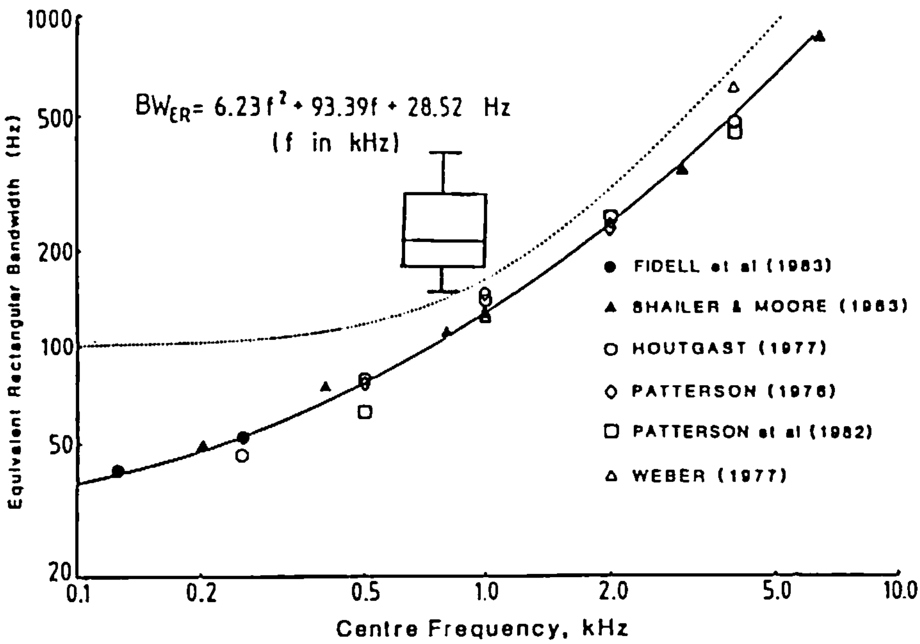


Fig.1 Equivalent rectangular bandwidths (ERBs) of the auditory filter and the traditional critical band as a function of frequency. This figure is derived from Moore et al.(1986). The dotted line indicates the critical bandwidth by Scharf (1970). The rectangle with bar indicates the data at 800Hz for hearing-impaired subjects by Dubno and Dirks (1989).

## 2. Method

### 2.1 Experiment I

#### 2.1.1 Subjects

Six healthy hearing subjects and three sensori-neural hearing-impaired subjects took part in this experiment. Fig. 1 shows the audiograms and ages of the hearing-impaired subjects with postlingual sudden deafness.

#### 2.1.2 Stimuli

The stimuli were generated using a klatt-type formant speech synthesizer (Klatt, 1980)[8]. The synthetic parameters were similar to those used by our preliminary study (Hayashi et al.,1990, 1991)[9,10]. The stimuli were $V_1CV_2$ (/apa-aba/) sequences with a *VOT* of 30ms and three fundamental frequencies ($F_0$) of 70, 150 and 300Hz. It was confirmed in our preliminary study (Hayashi et al., 1990)[11] that a bilabial plosive with a *VOT* of 30ms was perceived as /pa/ in a CV context but as /ba/ in a VCV context (with no silent interval between the $V_1$ and $CV_2$) by healthy-hearing and hearing-impaired subjects. The fundamental frequencies of $V_1$ and $V_2$ were the same. The variable parameter was the silent interval (*dt*) between $V_1$ and $CV_2$.

#### 2.1.3 Procedure

The phoneme boundary between /apa/ and /aba/ was measured separately for various silent intervals between $V_1$ and $V_2$ (*dt*) at three $F_0$ conditions. The *dt* was changed from 0ms to 100ms in 20ms steps. Accordingly, 18 stimuli were made from three $F_0$ and 6 silent interval conditions, which were then presented randomly 10 times each.

The presentation level was set at the most comfortable listening level individually for the hearing-impaired, but was set at 80 and 60 dBSPL for the healthy-hearing subjects.

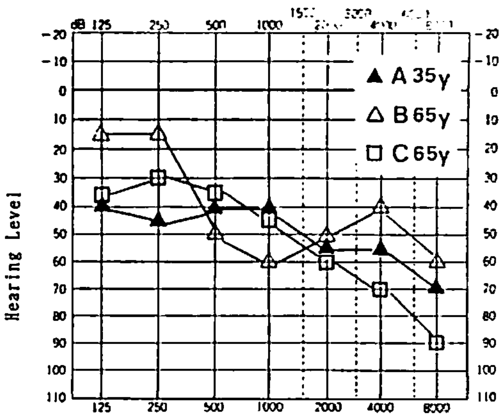


Fig.2 Audiograms and ages of the hearing-impaired subjects.

The stimuli were presented monaurally using earphones. The subjects identified the consonant in each stimulus as either /b/ or /p/. The 50% response *dt* (in ms) was used as the *dt* phoneme boundary between the voiced and unvoiced stop consonants.

### 2.2 Experiment II

#### 2.2.1 Subjects

Six healthy-hearing subjects took part in this experiment.

#### 2.2.2 Stimuli

The stimuli were similar to those used in Experiment I, but with two *VOT*s of 0ms (perceived as /b/ in the CV context) and 30ms (perceived as /p/ in the CV context), and two $F_0$s of 70 and 150Hz.

#### 2.2.3 Procedure

In this experiment, the $F_0$s of $V_1$ and $V_2$ were varied independently. So, the $F_0$s of $V_1$ and $V_2$, *VOT* and the silent interval between $V_1$ and $V_2$ (*dt*) were varied randomly. The *dt* was changed from -60ms to +60ms in 30ms steps. The negative values of *dt* indicate when $V_1$ temporally overlapped with $CV_2$. Accordingly, 40 stimuli were made from the two $F_0$s of $V_1$ (70, 150Hz), the two $F_0$s of $V_2$ (70, 150Hz), the two *VOT*s of C (0, 30ms) and the five *dt*s. The stimuli were presented 10 times each.

The stimuli were presented monaurally using earphones. The subjects identified the consonant in each stimulus as either /b/ or /p/. The percentage of correct detections in each *dt* condition was obtained. The presentation level was set at 80 dBSPL.

## 3. Results

### 3.1 Experiment I

Fig.3 shows the percentage of /b/ responses as a function of the silent interval (*dt*) for the three $F_0$s. Fig. 3(a) shows the results for a healthy-hearing subject at 80dBSPL and Fig. 3(b) shows those for a hearing-impaired subject. The 50% response *dt* at each $F_0$ was used as the *dt* phoneme boundary.

Fig. 4 shows the *dt* phoneme boundary distribution for each $F_0$. For the healthy-hearing subjects, the medians and quartile ranges at the two presentation levels (80 and 60 dBSPL) are shown. For the hearing-impaired subjects, the *dt* phoneme boundaries are plotted individually.

Additionally, the gap detection threshold curve (dotted line) as a function of the

$F_0$ is shown in Fig. 4 for reference. These gap detection thresholds were estimated using the formula for calculating auditory filter band widths derived by Moore et al. (1983)[12]. There is much evidence to suggest that these estimates fit closely actually measured values in the frequency range above 200Hz. However, below 200Hz, the estimates are no more than approximate values, since no studies have ever been tried in this frequency range.

In $V_1CV_2$ contexts, the voiced /b/ responses for the stimuli which were perceived as unvoiced /p/ in CV contexts decreased when the *dt* increased.

The *dt* phoneme boundaries significantly shifted depending upon the $F_0$. When the $F_0$ was low, the voiced responses tended to increase, while a longer *dt* was needed for the unvoiced judgment.

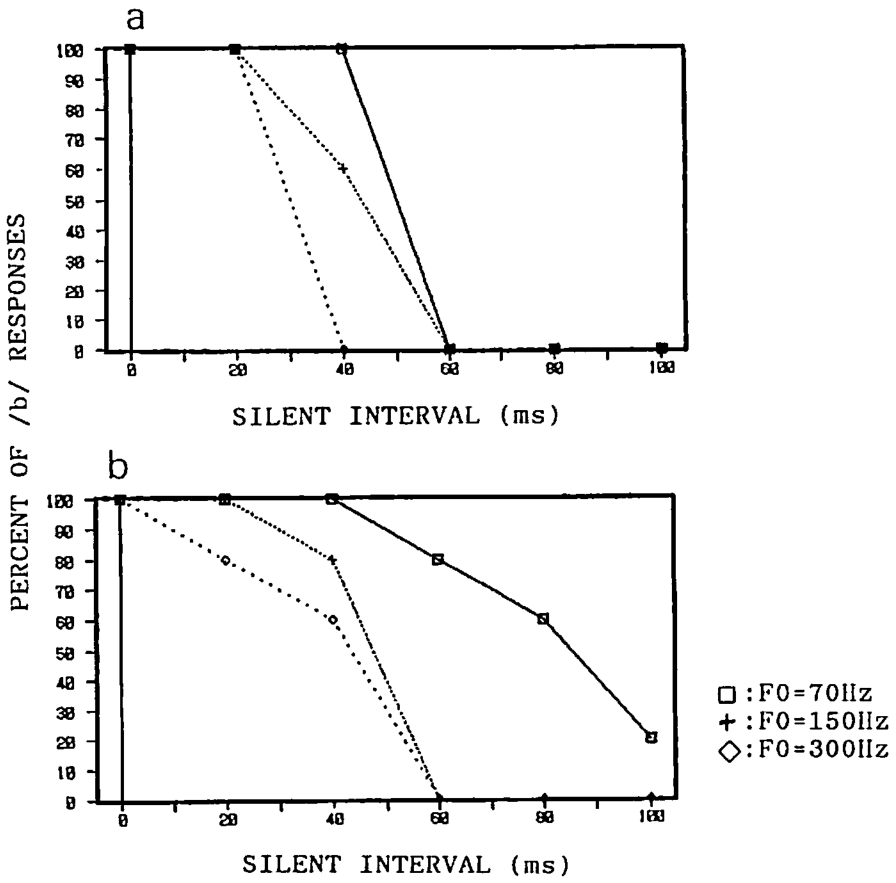


Fig.3 The percent of /b/ responses as a function of the *dt* for the three $F_0$s.
a) A healthy-hearing subject at 80dBSPL.
b) Hearing-impaired subjects at their most comfortable level.

For the healthy-hearing subjects, the *dt* phoneme boundaries were also affected by the presentation level. When the presentation level was low, the voiced responses tended to increase, while a longer *dt* was needed for the unvoiced judgment.

For the hearing-impaired subjects, there were large inter-subject differences. For subject C, the results were analogous to the healthy-hearing subjects' at 80 and 60 dBSPL. For subjects A and B, however, the *dt* phoneme boundaries were significantly longer than those for the healthy-hearing subjects, especially at the lower $F_0$.

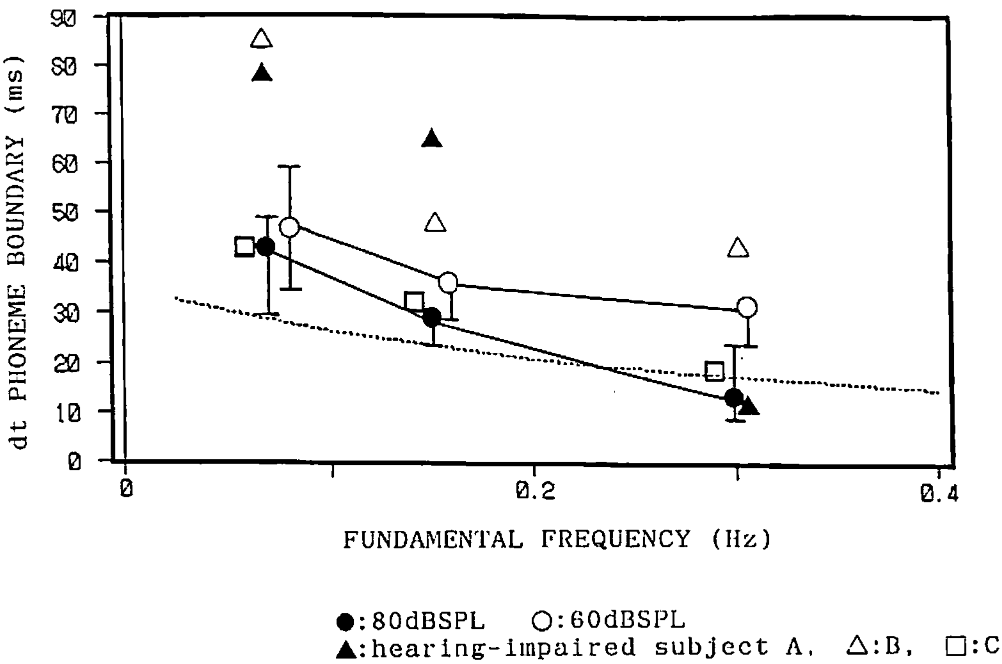


Fig.4 The *dt* phoneme boundary distribution for each $F_0$. the medians and quartile ranges at 80 and 60dBSPL are shown for the healthy-hearing subjects.
For the hearing-impaired subjects, the dt phoneme boundaries are plotted individually.

### 3.2 Experiment II

Fig. 5 shows the median of the percentage of correct detections as a function of the *dt* and $F_0$s of $V_1$ and $CV_2$. As this figure indicates, the voicing judgment, for stops in the $V_1CV_2$ contexts were affected by the *dt*, whether the $F_0$s of $V_1$ and $CV_2$ were the same or not. Generally, regardless of the $F_0$, the correct detections for the /p/ stimuli (*VOT*=30ms) with a shorter *dt* tended to decrease. On the other hand, the correct detections for the /b/

stimuli (*VOT*=0ms) tended to decrease when the *dt* became longer.

As shown Fig. 5(a), however, when the $F_0$ of $V_1$ was 70Hz and that of $V_2$ was 150Hz, the percentage for the correct detection of the unvoiced stop /p/ did not decrease even when the *dt* was shorter than 30ms. And this percentage was 80% even though a part of $V_1$ overlapped with $CV_2$ (with a negative *dt*). But, this tendency was not noted either when the $F_0$s of $V_1$ and $V_2$ were the same or when the $F_0$ of $V_1$ was 150Hz and that of $V_2$ was 70Hz.

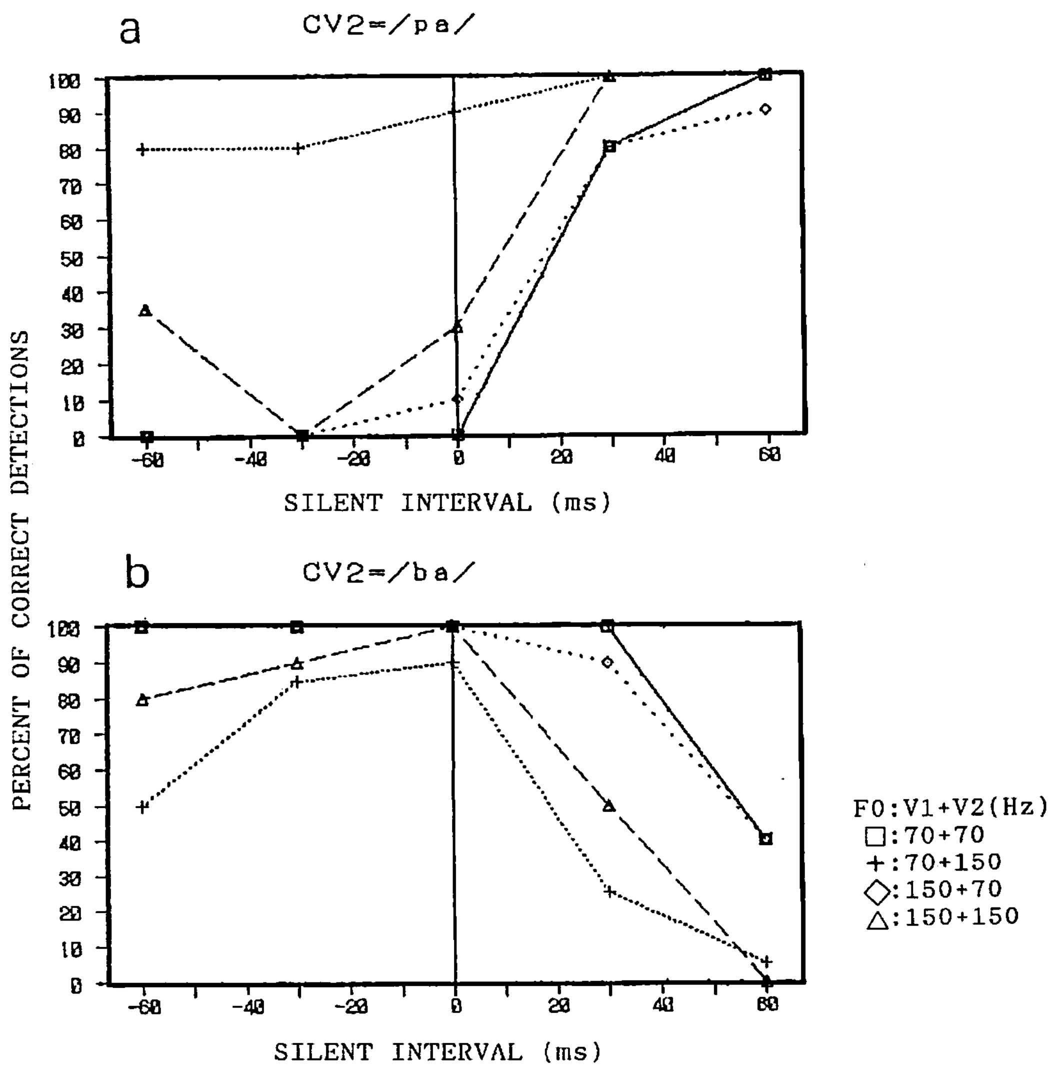


Fig.5 The medians of the percent correct detections as a function of the *dt* and $F_0$s of $V_1$ and $CV_2$.

## 4. Discussion

### 4.1 Experiment I

Several studies have shown that the band-width of the auditory filter narrows either when the frequency is low (see Fig. 1) or when the presentation level is low. The results from Experiment I shows that the effect of the silent interval, *dt*, on the voicing judgments for stops in $V_1CV_2$ contexts varied with the $F_0$. In other words, when the $F_0$ was low, a longer *dt* was needed to identify unvoiced stops for both the healthy-hearing and the hearing-impaired subjects. In addition, for the healthy-hearing subjects, when the presentation level was low, a longer *dt* was needed to identify unvoiced stops. From these results, it seems reasonable to suppose that the auditory filter ringing, which was longer for the lower frequencies, may affect voicing judgments. Accordingly, $H_1$ was confirmed.

For the hearing-impaired subjects, while their auditory filter was thought to be broader than that of the healthy-hearing subjects, the *dt* phoneme boundaries tended to be longer than those of the healthy-hearing subjects. If $H_1$ is true, for the hearing-impaired subjects with a wider auditory filter, the ringing effect of $V_1$ will disappear faster, and thus this *dt* phoneme boundary must be expected to be shorter than in the healthy-hearing subjects. However, the results for the hearing-impaired subjects did not agree with $H_1$. Therefore, their results can not be explained by the auditory filter ringing effect. It is not to be denied that factors other than the auditory filter ringing may be reflected in these results, or even that a different factor may be reflected in each subject group. In this respect, further investigation is needed.

### 4.2 Experiment II

In this experiment, the range for *dt* was set from -60ms to +60ms. When a part of $V_1$ overlapped with $CV_2$, the following results were expected. a) Under such a condition, the percentage of correct detections in the two alternative forced-choice tasks may be about 50%, because the amplitude gap during the silent interval and *VOT*, which are the important acoustic cues for the voicing judgment of stops, can not be perceived clearly. b) The percentage of correct detections of /p/ may be about 0%, because the silent interval and the *VOT* of C are filled by $V_1$. c) When the $F_0$s of $V_1$ and $CV_2$ are different, the percentage of correct detections of /p/ may be more than 50%, because of the effect of perceptual segregation. Under our hypothesis $H_2$, we expect result c).

In the results of experiment II, when the $F_0$ of $V_1$ was 70Hz and that of $CV_2$ was 150Hz, the percentage of correct detections for /p/ was significantly higher than at the other $F_0$ conditions, even when the *dt* was short. However, this tendency was not revealed when the $F_0$ of $V_1$ was 150Hz and that of $CV_2$ was 70Hz. Accordingly, $H_2$ was confirmed when the $F_0$ of $V_1$ was lower than that of $CV_2$ alone.

The question to be answered is why the effect of perceptual segregation varies according to the $F_0$ combination of $V_1$ and $CV_2$. This calls for further consideration, although there are some possible explanations, such as in relation to the characteristics of masking release, phonemic restoration and temporal resolution.

## 5. Conclusion

In this study, the perceptual effects of the fundamental frequency ($F_0$) and silent interval length between $V_1$ and C (*dt*) on voicing judgments for C in $V_1CV_2$ contexts were examined. The following results were obtained.

1) When the $F_0$s of $V_1$ and $CV_2$ were the same, the lower the $F_0$ was, the longer the *dt* needed to detect unvoiced stops.

2) When the $F_0$ of $V_1$ was lower than that of $CV_2$, the percentage of correct detections of unvoiced stops did not decrease even when the *dt* was shorter than 30ms. And this percentage was 80% even though a part of $V_1$ overlapped with $CV_2$.

These results indicate that auditory filter ringing, which is longer for lower frequency, may affect voicing judgments by shortening the perceptual length of the silent interval, but the effect can be decreased by perceptual segregation between $V_1$ and $CV_2$.

## Acknowledgement

This work was supported in part by a Grant-in Aid for Scientific Research, the Ministry of Education, Science and Culture, Japan.

## References

1) Liberman,A.M. and Mattingly,I.G.: The motor theory of speech perception revised, Cognition, 21, 1-36, 1985.
2) Moore, B.C.J.: Frequency selectivity in hearing, Academic Press, London, 1986.
3) Imaizumi, S.: Recent topics on hearing characteristics of the hearing-impaired, J. Acoust. Soc. Japan., 754-759, 1991, in Japanese.
4) Scharf, B.: Critical bands, In foundation of modern auditory theory, 1, J. V. Tobias, Ed., Academic Press, New York, 1970.
5) Dubno, J.R. and Dirks, D.D.: Auditory filter characteristics and consonant recognition for hearing-impaired listeners, J. Acoust. Soc. Am., 85, 1666-1675, 1989.
6) Schailer, M.J. and Moore, B.C.J.: Gap detection as a function of frequency, bandwidth, and level, J. Acoust. Soc. Am., 74, 467-473, 1983.
7) Moore, B.C.J.: Across-channel processes in auditory masking, J. Acoust. Soc. Japan

(E), 25-37, 1992.
8) Klatt, D.M.: Software for a cascade-parallel formant synthesizer, J. Acoust. Soc. Am., 67, 971-995, 1980.
9) Hayashi, A., Imaizumi, S., Harada, T., Seki, H. and Hosoi, H.: Relationships between the ear's temporal window and VOT perception for normal and hearing-impaired listeners, Ann. Bull. RILP., 24, 145-156, 1990.
10) Hayashi, A., Imaizumi, S., Harada, T., Seki, H. and Hosoi, H.: Effects of preceding vowels on the voicing distinction for stop consonants among hearing impaired subjects, Ann. Bull. RILP., 25, 91-98, 1991.
11) Hayashi, A., Imaizumi, S., Harada, T., Seki, H. and Hosoi, H.: Effects of temporal factors on the speech perception of the hearing impaired, Proceedings of ICSLP90, Kobe, Japan, 1, 593-596, 1990.
12) Moore, B.C.J. and Glasberg, B.R.: Suggested formulae for calculating auditory-filter band-widths and excitation patterns, J. Acoust. Soc. Am., 74, 750-753, 1983.