2nd International Conference on
Spoken Language Processing (ICSLP 92)
Banff, Alberta, Canada
October 12-16, 1992

FACTORS AFFECTING VOICING DISTINCTION OF STOPS FOR THE HEARING IMPAIRED

Hideaki Seki[1)], Akiko Hayashi[2)], Satoshi Imaizumi[3)]
Takehiko Harada[3)] and Hiroshi Hosoi[4)]

1)Dept. Information Eng., Chiba Inst. Tech., Narashino, Chiba, 275 Japan
2)Faculty of Education, Tokyo Gakugei Univ., Koganei, Tokyo, 184 Japan
3)Faculty of Medicine, Univ. Tokyo, Hongo, Bunkyo, Tokyo, 113 Japan
4)School of Medicine, Kinki Univ., Osaka-sayama, Osaka 589 Japan

ABSTRACT

In order to develop speech enhancement methods for the hearing impaired, acoustical and auditory factors affecting on the voicing distinction for stop consonants in V_1CV_2 contexts were examined. The effects of the level of V_1 and the silent interval between V_1 and C on the voicing distinction for stops in V_1CV_2 contexts were examined in Experiment I. And, the effect of the fundamental frequency on the voicing distinction for stops in V_1CV_2 contexts was examined in Experiment II. The following results were obtained. 1) The identification rate for the voiced stop /b/ decreased when the level of V_1 decreased when the *VOT* was fixed at 30ms. 2) The identification rate for the voiced stop /b/ decreased when the silent interval between V_1 and C increased when the *VOT* was fixed at 30ms. 3) Sensori-neural hearing-impaired subjects needed a longer silent interval compared to normal subjects to identify /p/. 4) Normal subjects needed a longer silent interval to identify /p/ when the presentation level was lower. 5) When the lower the F_0 was, the longer the *dt* needed to detect unvoiced stops.

10.21437/ICSLP.1992-295

I. 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. 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 constraints on speech perception, but impairments in auditory functions may impose significant constraints on speech perception.

In this study, we examined following hypotheses for the voicing distinction for stop consonants in V_1CV_2 sequences. H_1) The lower the level of V_1, the higher the identification rate for an unvoiced stop should be. H_2) The longer the silent interval between V_1 and C, the higher the identification rate for an unvoiced stop should be. H_3) For hearing-impaired subjects who need longer silent interval to recover from temporal masking by the preceding vowel V_1, a longer silent interval between V_1 and C should be required to identify unvoiced stop consonants. H_4) At a lower/softer presentation level, even normal subjects might need a longer silent interval to identify unvoiced stops, because their temporal resolution becomes poorer. H_5) The lower the fundamental frequency of V_1 and V_2 F_0 is, the longer silent interval is needed to identify unvoiced stops, because temporal resolution becomes poorer for lower frequency.

In this study, to test H_1)-H_4), the effects of the level of V1 and the silent interval between V_1 and C on the voicing distinction for stops in V_1CV_2 contexts are examined in Experiment I. And, to test H_5), the effect of the fundamental frequency on the voicing distinction for stops in V_1CV_2 contexts is examined in Experiment II.

II. METHOD

2.1 Experiment I

2.1.1 Subjects

Thirteen normal hearing subject (ages:21-35 years) and 3 hearing-impaired subjects (4 ear's) with postlingual sudden deafness participated in this experiment.

2.1.2 Stimuli

In this experiment, the stimuli were V_1CV_2 (/apa-aba/) sequences with a *VOT* of 30ms. It was confirmed in our preliminary study that a bilabial plosive with a *VOT* of 30ms was perceived as /pa/ in a CV context but as /aba/ in a V_1CV_2 context (with no silent interval between the V_1 and CV_2) by normal and hearing-impaired subjects [1,2].

The stimuli were generated using a Klatt type formant synthesizer[3] with synthetic parameters basically similar to those used by Kuhl[4]. Fig.1 illustrates the stimulus configuration (a) and the contour of each formant transition (b).

Following the release of the burst (at 0ms), the *VOT* was conducted by a cutback of the first formant and an excitation of the higher formants with a noise source simulating aspiration instead of the periodic source during the cutback. The amplitude of this noise source fell linearly until the *VOT* as shown in Fig.1(a). The preceding and following vowels were 300ms in duration with rise/fall times of 10ms. The first V_1C transition contour was set symmetrically to the following CV_2 transition as shown in Fig.1(b).

The variable parameters were the stimulus level differences between V_1 and V_2 (*da*) and the silent interval between V_1 and C (*dt*). For the *da* changing condition, the silent interval(*dt*) was fixed at 0ms.

2.1.3 Procedure

The phoneme boundary between /apa/ and /aba/ was measured separately for various levels of V_1(*da* condition), and various silent intervals (*dt* condition). For the *da* condition, the *da* was changed in 5dB SPL steps. For the *dt* condition, the *dt* was changed in 5ms steps for the normal subjects and in 10ms steps for the hearing-impaired subjects. All stimuli were randomly presented ten times each.

The stimuli were presented monaurally using earphones. The subjects identified the consonant in each stimulus as either /b/ or /p/. The 50% response level *da* (in dB) and *dt* (in ms) were used as the phoneme boundary.

For two experimental conditions (*da* and *dt* conditions), the level of V_2 was set at the most comfortable level individually for the hearing-impaired, but was set at 50dBSPL and 80dBSPL for the normal-hearing subjects.

2.2 Experiment II

2.2.1 Subjects

Six normal subjects and eighteen sensori-neural hearing-impaired subjects with postlingual sudden deafness took part in this experiment.

2.2.2 Stimuli

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. 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.

2.2.3 Procedure

The phoneme boundary between /apa/ and /aba/ was measured separately for various silent intervals between V_1 and C (*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 normal 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.

III. RESULT

3.1 Experiment I

For the *da* condition, Fig. 2(a) shows the percent of /b/ responses at the two presentation levels for a normal subject. The level of V_2 was fixed at 50 and 80dBSPL. When the stimulus level difference (V_1 level minus V_2 level) increased, in other words, when the level of V_1 decreased, the percent of /b/ responses decreased for both presentation levels.

Fig. 2(b) shows the percent of /b/ responses at one level for the hearing-impaired subjects. Also as shown in this figure, the percent of /b/ responses decreased when the level of V_1 decreased.

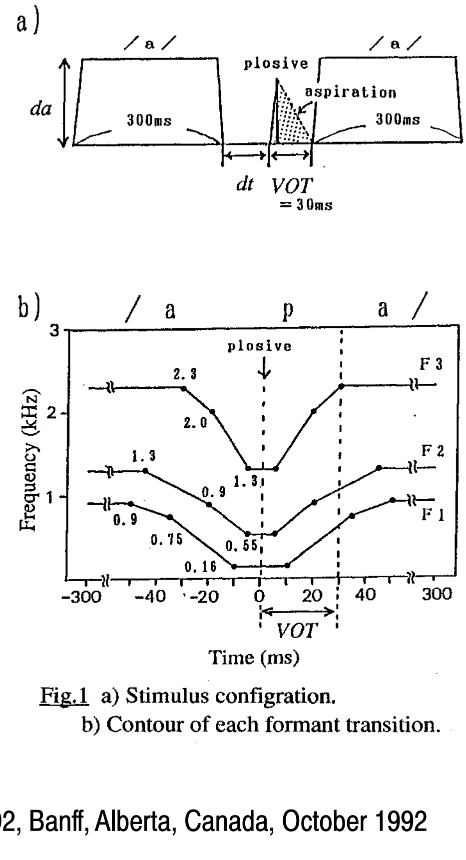

Fig.1 a) Stimulus configration.
b) Contour of each formant transition.

For the *dt* condition, Fig.3 (a) shows the percent of /b/ responses for a normal subject. The percent of /b/ responses decreased *dt* increased. The *dt* phoneme boundary shifted from 20ms to 42ms when the presentation level was changed from 80dBSPL to 50dBSPL.

Fig.3 (b) shows the percent of /b/ responses for the hearing-impaired subjects. The percent of /b/ responses decreased when *dt* increased. The *dt* phoneme boundaries tended to be longer than those of the normal subjects.

Fig.4 shows the *da* phoneme boundary distribution and Fig.5 the *dt* phoneme boundary distribution for the normal and the hearing-impaired subjects.

The results shown in Figs.4 and 5 indicate that the voicing distinction for stop consonants in V_1CV_2 contexts was affected by the level of V_1 and the silent interval between V_1 and C. The identification rate for the voiced stop /b/ decreased when the level of V_1 decreased and the silent interval became longer.

3.2 Experiment II

Fig.6 shows the percentage of /b/ responses as a function of the silent interval (*dt*) for the three F_0s. Fig. 6(a) shows the results for a normal subject at 80dBSPL and Fig. 6(b) shows those for a hearing-impaired subject. The 50% response *dt* at each F_0 was used as the *dt* phoneme boundary.

Fig. 7 shows the *dt* phoneme boundary distribution for each F_0. For the normal subjects, the medians and quartile ranges at the two presentation levels (80 and 60 dBSPL) are shown. For the hearing-impaired subjects, the results of three typical subjects are also plotted individually. Additionally, the gap detection

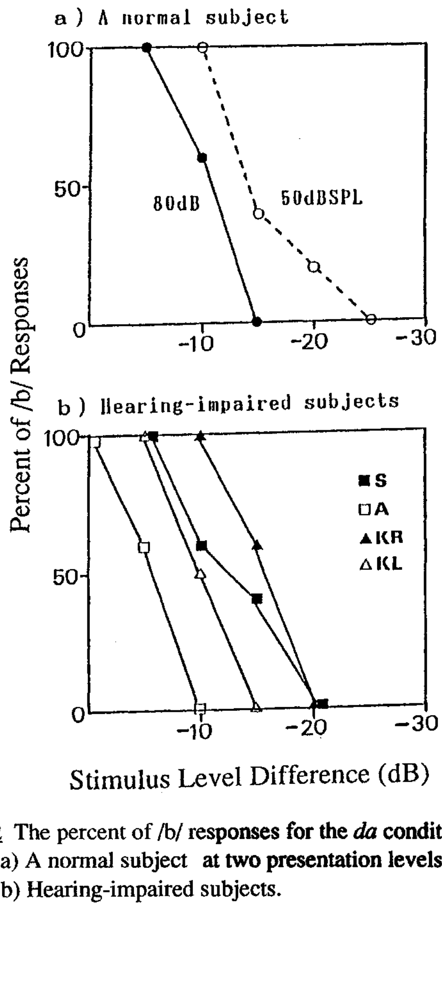

Fig.2 The percent of /b/ responses for the *da* condition.
a) A normal subject at two presentation levels.
b) Hearing-impaired subjects.

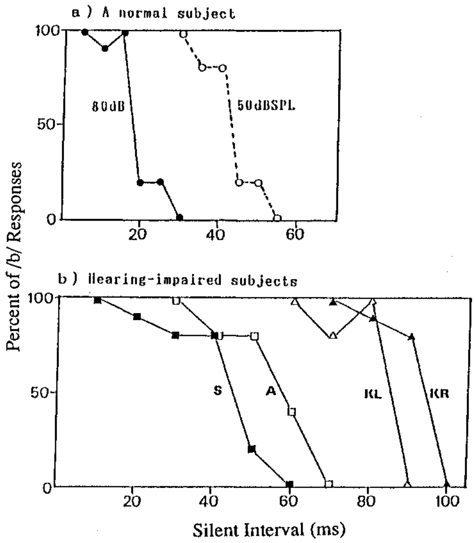

Fig.3 The percent of /b/ responses for the *dt* condition.
a) A normal subject at two presentation levels.
b) Hearing-impaired subjects.

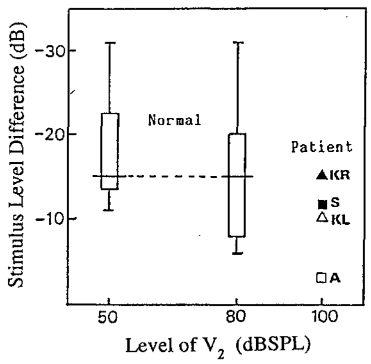

Fig.4 The *da* phoneme boundary distribution.

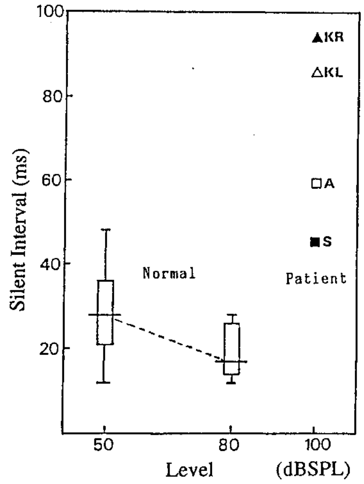

Fig.5 The *dt* phoneme boundary distribution.

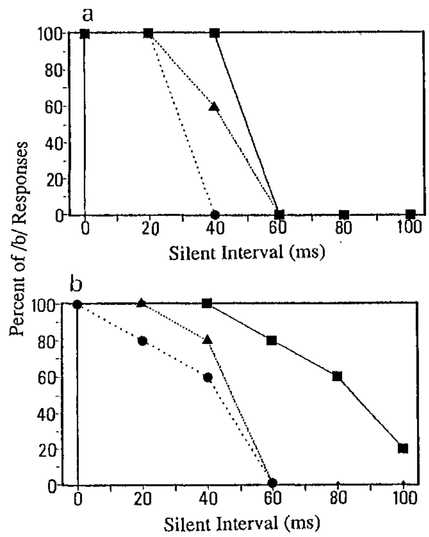

Fig.6 The percent of /b/ responses as a function of the *dt* for the three F_0s (■:F_0=70Hz, ▲:F_0=150Hz, ●:F_0=300Hz)
a) A normal subjects at 80dBSPL.
b) A hearing-impaired subjects.

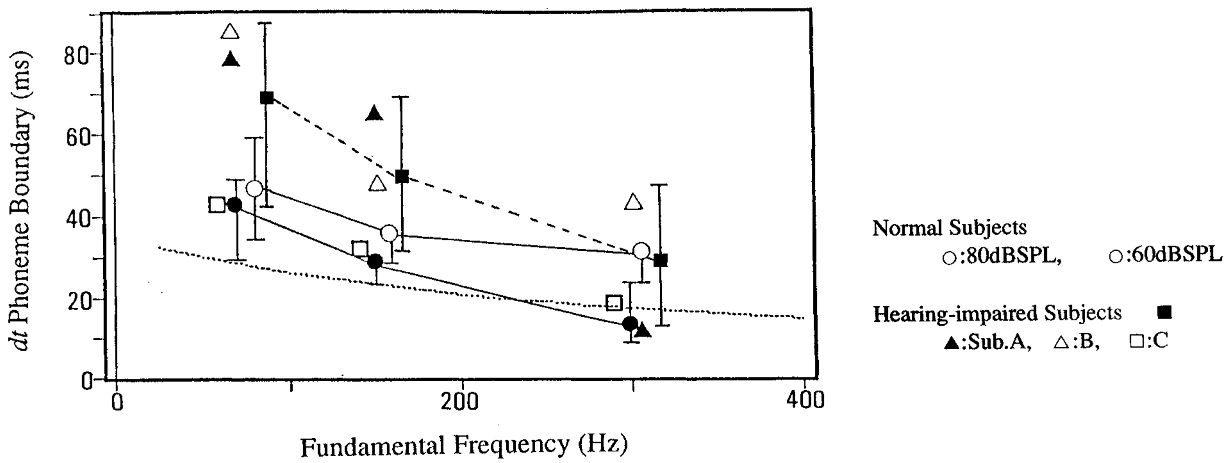

Fig.7 The *dt* phoneme boundary distribution for each F_0.

threshold curve (dotted line) as a function of the F_0 is shown Fig.7 for reference. These gap detection thresholds were estimated using the formula for calculating auditory filter band widths derived by Moore et al. (1983) [6]. There are some data to suggest that these estimates fit closely actually measured values in the frequency range above 150Hz. However, below 150Hz, 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.

For the normal 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 normal subjects' at 80 and 60 dBSPL. For subjects A and B, however, the *dt* phoneme boundaries were significantly longer than those for the normal subjects, especially at the lower F_0.

IV. DISCUSSION

4.1 Experiment I

The results shown in Figs.2, and 3 indicate that the voicing distinction for stop consonants in V_1CV_2 sequences was affected by vowel V_1 and the silent interval between V_1 and C. The identification rate for the unvoiced stop /p/ increased when the level of V_1 decreased and/or the silent interval was longer. Accordingly, H_1 and H_2 were confirmed.

As shown in Fig.5, the *dt* phoneme boundaries were significantly longer for the hearing-impaired than for the normal subjects. Also, for the normal subjects, the *dt* phoneme boundary was longer at the low presentation level than at the high level. From these results, H_3 and H_4 were confirmed.

It seems that these results can be explained from the effect of the forward masking of V_1 on the detection of silent interval. When the level of V_1 becomes higher, the falling time of the loudness for V_1 is lengthened, and thus the length of perceptual silent interval is shortened and the identification rate of /b/ is increased. In addition, according to our preliminary study, temporal resolution deteriorate for some of the sensori-neural hearing-impaired and also for normal subjects when the presentation level is low. Consequently, subjects having poorer temporal resolution require a longer silent interval (*dt*) to identify an unvoiced stop consonant.

4.2 Experiment II

Several studies have shown that the band-width of the auditory filter narrows either when the frequency is low or when the presentation level is low. The results from Experiment II show 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 normal and the hearing-impaired subjects. In addition, for the normal 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_5 was confirmed.

For the hearing-impaired subjects, while their auditory filter was thought to be broader than that of the normal subjects, the *dt* phoneme boundaries tended to be longer than those of the normal subjects. If H_5 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 normal subjects. However, the results for the hearing-impaired subjects did not agree with H_5. 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.

V. CONCLUSION

The effects of the level of V_1, the silent interval between V_1 and C and fundamental frequency on the voicing distinction for the bilabial stop consonants in V_1CV_2 contexts were examined. The following results were obtained. 1) The identification rate for the voiced stop /b/ decreased when the level of V_1 decreased when the *VOT* was fixed at 30ms. 2) The identification rate for the voiced stop /b/ decreased when the silent interval between V_1 and C increased when the *VOT* was fixed at 30ms. 3) Sensori-neural hearing-impaired subjects needed a longer silent interval compared to normal subjects to identify /p/. 4) Normal subjects needed a longer silent interval to identify /p/ when the presentation level was lower. 5) When the lower the F_0 was, the longer the *dt* needed to detect unvoiced stops. These results agree with the fact that slower speech is more helpful for the hearing impaired to understand speech.

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] A. Hayashi, S. Imaizumi, T. Harada, H. Seki, and H. Hosoi. "Relationships between the ear's temporal window and VOT perception for normal and hearing-impaired listeners," Ann. Bull. RILP., vol. 24, pp.145-156, 1990.

[2] A. Hayashi, S. Imaizumi, T. Harada, H. Seki, and H. Hosoi. "Effects of temporal factors on the speech perception of the hearing impaired," Proceedings of ICSLP90, Kobe, Japan, vol. 1, pp.593-596, 1990.

[3] B. C. J. Moore, B. R. Glasberg, C. J. Plack, and A. K. Baswas. "The shape of the ear's temporal window," J. Acoust. Soc. Am., vol. 83(3), pp. 1102-1106, 1988.

[4] D. M. Klatt. "Software for a cascade/parallel formant synthesizer," J. Acoust. Soc. Am., vol. 67(3), pp. 971-995, 1980.

[5] P. K. Kuhl and J. D. Miller. "Speech perception by the chinchilla: Identification function for synthetic VOT stimuli," J. Acoust. Soc. Am., vol. 63, pp. 905-917, 1978.

[6] B. C. J. Moore and B. R. Glasberg. "Suggested formulae for calculating auditory-filter band-widths and excitation patterns," J. Acoust. Soc. Am., vol. 74, pp.750-753, 1983.