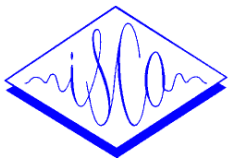

First International Conference on
Spoken Language Processing (ICSLP 90)
Kobe, Japan
November 18-22, 1990

EFFECTS OF TEMPORAL FACTORS ON THE SPEECH PERCEPTION OF THE HEARING IMPAIRED

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

Faculty of Medicine, University of Tokyo, Hongo, Bunkyo, Tokyo, 113 Japan
* Chiba Institute of Technology, Narashino, Chiba, 275 Japan
**School of Medicine, Kinki University, Osaka-sayama, Osaka, 589 Japan

ABSTRACT

The width of ear's temporal window and the values of VOT at the phoneme boundary between voiced versus unvoiced consonants in CV (/ba-pa/) and VCV (/aba-apa/) stimuli were measured for 7 normals and 6 sensori-neural hearing-impaired subjects. The measurements were made at the most comfortable level for each of the hearing-impaired subject, and at three levels, 40, 60 and 80 dB SPL for the normal hearing subjects. The temporal window was generally wider for the hearing-impaired subjects than for the normal subjects. The VOT phoneme boundary was longer for the VCV than for the CV contexts. The VOT phoneme boundary in the VCV context tended to correlate to the width of the temporal window. Based on these results, we concluded that the poor temporal resolution of the ear affects to some extent the VOT perception for the hearing-impaired subjects.

1. INTRODUCTION

We have focused our interest on the effects of the temporal acuity of the ear on the speech perception. Several studies have shown that the temporal resolution of the ear generally declines in hearing-impaired subjects, although individual variability is large. However, it has not been sufficiently explained how the declining temporal resolution affects speech perception.

For a patient with poor temporal resolution of the ear, we assumed the following difficulties. D1) Brief speech sounds may be harder to recognize. D2) Speech sounds in conversation at high speaking rate may be harder to recognize. D3) The temporal acoustic cues used for identification of speech sounds may be harder to detect. Consequently, the speech recognition capability of such a hearing-impaired patient may be reduced. We have already examined difficulties D1 and D2 [1],[2]. The results of our previous experiments indicated that most of the hearing-impaired subjects tested needed a longer vowel duration than normal subjects to identify vowels. And the hearing-impaired subjects needed a longer inter-vowel silent interval than normals to identify two-vowel sequences.

To examine D3, in this paper, we have investigated the relationship between the ear's temporal window, measured using non-speech stimuli, and the VOT (Voice Onset Time) perception, measured using Japanese bilabial plosive consonants. For VOT perception, the phoneme boundaries between voiced versus unvoiced consonants in CV (/pa-ba/) and VCV (/apa-aba/) contexts were measured.

We formulated the following hypotheses, H1-4, for a hearing-impaired subject who had a wider ear temporal window or poorer temporal acuity than normal listeners. H1) The VOT value at the phoneme boundary would be longer than for normal subjects (a longer VOT would be necessary to identify an unvoiced consonant). H2) Because of VOT masking by the preceding vowel, the VOT phoneme boundary would shift to a longer value in VCV context than in CV context. H3)Such VOT boundary shift between CV and VCV contexts would be larger for a patient with a longer temporal window. If H1, H2 and H3 are valid, it should be also true that H4) the temporal window has a close relation to VOT perception.

2. MEASUREMENT OF THE TEMPORAL WINDOW

2.1 Subjects

Seven normal hearing subjects (aged 22-30 years), and 6 sensori-neural hearing-impaired subjects took part in the study. Table 1 shows the absolute thresholds for the tested ears of the hearing-impaired subjects.

2.2 Procedure

The shape of the temporal window was measured on the basis of the method proposed by Moore et al.[3]. Fig. 1 shows the stimulus configuration and the shape of the hypothetical temporal window. Given an outline of this measurement, the threshold was measured for a sinusoidal signal (S) presented in a temporal gap between two bursts of noise (N), as shown in Fig. 1. The duration of the intervals between the signal and the two noise bursts (T1 and T2) were symmetrically and asymmetrically varied. The signal was a 2kHz sinusoid with 10ms duration having 5ms onset/offset ramps (no steady-state portion). The noise masker was bandnoise restricted between 1kHz and

4kHz, having 204ms duration with 2ms onset/offset ramps.

The noise masker was presented at the most comfortable level for individual hearing-impaired subjects. Four normal subjects were tested at the three levels of 80, 60 and 40dBSPL and three normal subjects were tested at two levels of 80 and 40dBSPL. The thresholds were measured using an adaptive, two-alternative, forced-choice procedure controlled by a micro-computer.

The data were used for an estimation of the intensity weighting function (W) describing the amount of interference on the perception of the sinusoidal signal by noise bursts placed both before and after the signal. Although this method of measuring temporal resolution has some restrictions, it has advantages in making a model to explain the relationship between speech reception and the temporal acuity of the ear.

2.3 Results

Fig. 2 shows the samples of the shapes of the temporal windows derived from the data for a hearing-impaired subject(No.4), and a normal subject with a masker noise level of 80dB or 40dBSPL. In order to represent the width of the temporal window, the equivalent rectangular duration (ERD) of each window was calculated. The equivalent rectangular duration was defined as the area divided by the height of the window, which was 1.0. Figures. 3a and 3b show the relationships between the ERDs and the masker noise levels at sound pressure level(dBSPL) and sensation level(dBSL).

Table 1. Audiometric data for the hearing-impaired subjects

Sub. No.	Sex	Age	Threshold, dB HL 125	250	500	1K	2K	4K	8K(Hz)
1	F	34	30	35	35	30	45	50	60
2	F	34	40	45	40	40	55	55	70
3	M	65	35	30	35	45	60	70	90
4	M	63	15	15	50	60	50	40	60
5	M	53	65	65	60	60	65	60	80
6	F	17	25	40	60	85	90	80	80

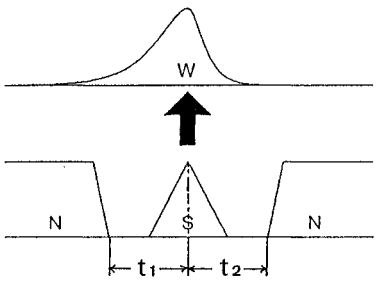

Fig.1 Stimulus configuration for measurement of the temporal window(W)

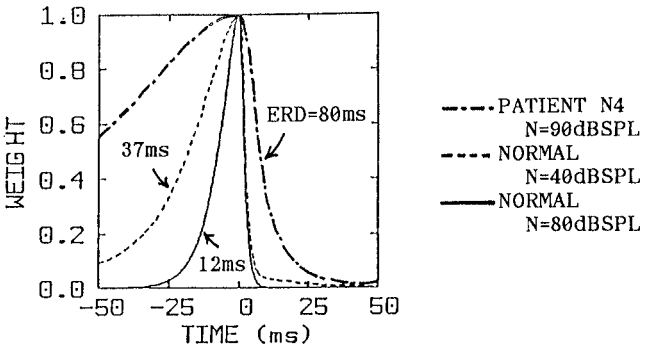

Fig.2 The temporal windows for a normal subject at two levels of masking noise and a patient No.4 at one level.

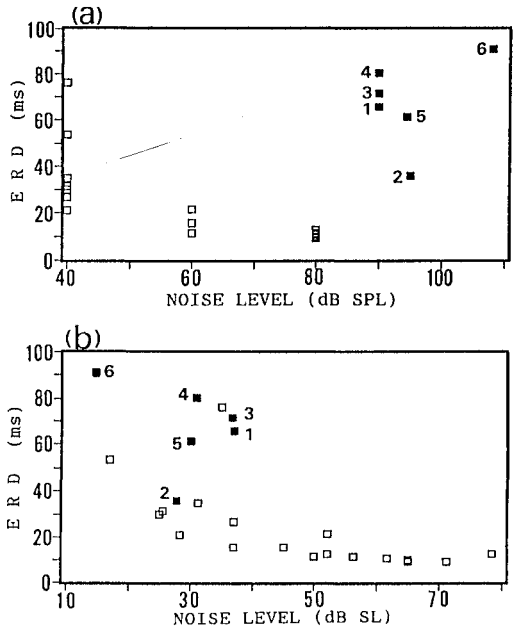

Fig.3 The relationships between the ERDs and the masking noise levels at SPL (a) and at SL (b).
□:Normal, ■:Patient

3. MEASUREMENT OF VOT PERCEPTION

3.1 Stimuli

The stimuli were generated using a Klatt-type formant speech synthesizer (Klatt,1980)[4]. The synthetic parameters were basically identical to those used by Kuhl (1978)[5] with some modifications. Following the release of the burst (at 0ms), the change in the VOT entailed both a cutback in 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 VOT. The VOT value of the

stimuli changed in 10ms step, from 0 to 120 ms. Accordingly, 13 stimuli were synthesized. The VCV stimuli were made by adding the vowel /a/ before the CV syllables synthesized above. The first VC transition contour was set symmetrically to the following CV transition shown in Fig. 6. The stimuli were presented at the most comfortable level for individual hearing-impaired subjects and at the three levels of 80, 60 and 40dBSPL for the normal subjects.

3.2 Procedure

The stimuli were presented randomly ten times each. The subjects were asked to identify the consonant in each stimulus as either /b/ or /p/. The VOT values at 50% of the responses were estimated as the phoneme boundary.

3.3 Results

The results are shown in Fig.4(a) for the normal subjects and Fig.4(b) for the hearing-impaired subjects. The relationship between the ERD and the VOT phoneme boundary for the CV context was shown in Fig.5(a), and for the VCV context was shown in Fig.5(b).

4. DISCUSSION

The width of the window was broader than the right-side for both the normal subjects and the hearing-impaired subjects. This means that it took longer to overcome the forward masking than the backward masking caused by the noise bursts. The temporal window was generally wider for the hearing-impaired subjects than for the normal subjects when compared at similar sound pressure level (dB SPL) of noise masker. For the normal subjects, the temporal window became wider for the lower levels of noise masker. Some hearing-impaired subjects had almost the same windows as the normal hearing ones when compared at similar sensation level (dB SL).

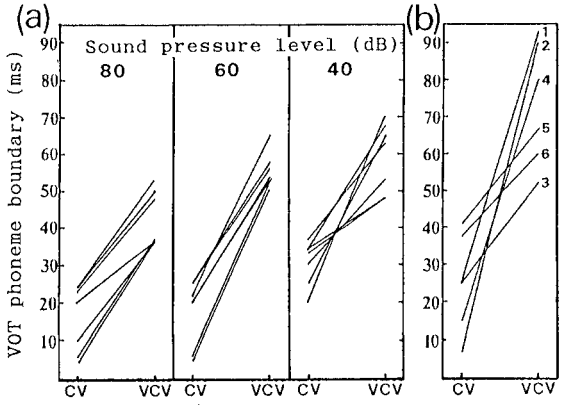

Fig.4 a) The VOT phoneme boundary for the CV and the VCV contexts at three SPL levels for the normal subjects. b) Those for the hearing-impaired subjects.

In the experiment on VOT perception, the following results were obtained as shown in Figs.4(a-b). 1) For the normal subjects, the VOT phoneme boundary tended to become longer with decrease measurement level in both the CV and VCV contexts. 2) For both the normal and hearing-impaired subjects, the VOT phoneme boundary was significantly longer for the VCV context than the CV context. 3) For the hearing-impaired subjects, the results were generally similar to the results

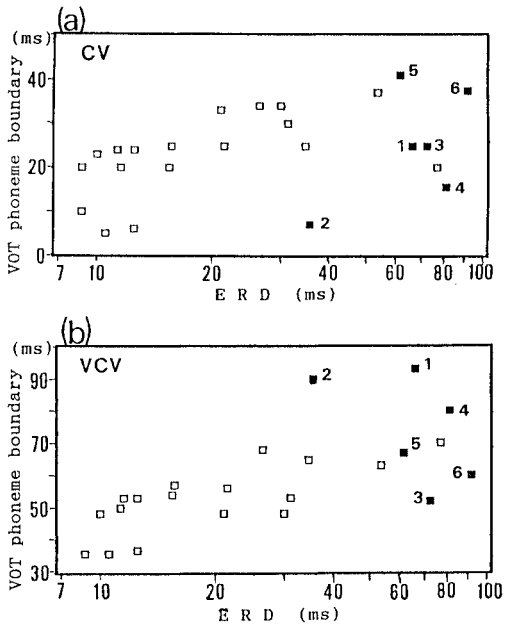

Fig.5 The relationships between the VOT phoneme boundaries and the ERDs for the CV context(a) and the VCV context(b). □:Normal, ■:Patient

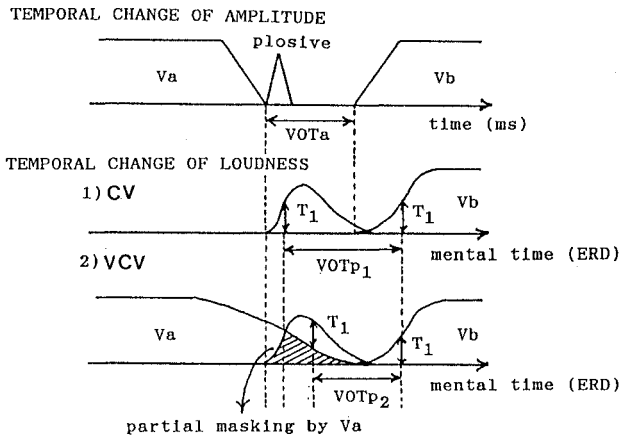

Fig.6 A model of VOT perception

observed for the normal subjects at the lowest measurement level, that is 40 dBSPL. For 3 hearing-impaired subjects, however, the differences between the VOT boundaries in the CV and the VCV contexts were significantly larger than those for the normal subjects.

It seems that these results can be explained in relation to the characteristics of the temporal resolution. Fig.6 shows an illustration of the outline of our interpretation. Our preliminary model of the identification between the unvoiced and voiced stops can be simply described as the following equation.

```
VOTp <  VOTb     : Voiced stop
VOTp >= VOTb     : Unvoiced stop      (1)
VOTb = Pp * ERD
```

Here, VOTp is the psychoacoustical voice onset time, ERD the width of the temporal window represented by the equivalent rectangular duration, VOTb the phoneme boundary, and Pp a constant. Equation (1) means that if the psycoacoustical voice onset time VOTp exceeds Pp times of the temporal resolution (Pp*ERD), then the phone is identified unvoiced. Equation (1) predicts that a longer VOTp is required for a subject having poorer temporal resolution (larger ERD) in identifying unvoiced stops.

However, if the VOT phoneme boundaries are so variable among subjects with different ERD, speech communication might not work well. Each subject has to adapt to phonetic restrictions which are somewhat specific to their language. Therefore, for successful speech communication, VOTb cannot be so variable. This limitation may be expressed as Equation(2).

```
VOTmin < VOTb < VOTmax                (2)
```

For the CV context, as shown in Fig.5(a), the VOT boundary values for the normal subjects are limited between VOTmin=5ms and VOTmax=37ms, and tend to be larger when the window width (ERD) is larger.

For the VCV context, as shown in Fig.5(b), the VOT boundary values for the normal subjects are limited between VOTmin=36ms and VOTmax=70ms, and tend to be larger when the window width (ERD) is larger.

The limitation on VOT boundary, VOTmax-VOTmin, is 30ms for both contexts, but it is 30ms longer for the VCV context than that for the CV context. This can be explained based on our model as follows.

In the case of the VCV contexts, because of the partial masking caused by the preceding vowel, the epoch when the loudness of the plosive burst exceeds the threshold T_1 may shift depending on how long the partial masking from the preceding vowel lasts. Therefore, VOTp in VCV contexts might be shorter than that in a CV context, if the acoustically defined VOTa was same. It means that a longer VOTa should be required for the VCV than for the CV context in identifying unvoiced stops.

In Figures 5(a) and 5(b), the results for the hearing impaired subjects were somewhat different from those for the normal subjects. The VOT boundary for subjects 1, 2, and 4 tend to be shorter than VOTmax=37ms for the CV context (Fig. 5a), but longer than VOTmax=70ms for the VCV context (Fig. 5b).

This might be explained as follows. These hearing impaired subjects might be able to utilize other acoustic cues such as formant frequencies or its transition rates at the point when voicing starts. And the use of such cues might be easier for the CV context than for the VCV context, because of the lack of partial masking by the preceding vowel.

5.Conclusions

1) For the normal hearing subjects, the temporal resolution of the ear tended to deteriorate at low sensation level.
2) For the sensori-neural hearing-impaired subjects, the temporal resolution was generally poorer than for the normal subjects when compared at both same SPL and same SL.
3) Accompanying with the deterioration of temporal resolution, a longer VOT was necessary to identify an unvoiced consonant in both the CV and the VCV contexts.
4) The VOT boundary for the normal subjects were between 5ms and 37ms for the VC context, and between 36ms and 70ms for the VCV context. It shifted about 30ms longer for the VCV context that for the CV context.
5) These results can be accounted for in a model which connects the temporal resolution of the ear and the VOT perception.

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.Yamada et al., "Effects of temporal factors on the Speech perception of the hearing impaired -A preliminary report-," Ann. Bull.RILP, Vo.21, PP. 131-140, 1987.
[2] A.Hayashi et al.,"Effects of stimulus duration and inter-stimulus interaction on vowel intelligibility for normal and hearing-impaired subjects," Ann. Bull. RILP., Vo.23, PP.163-172, 1989.
[3] B.C.J.Moore et al., "The shape of the ear's temporal window," J.Acoust.Soc.Am., Vo.83, PP.1102-1106, 1988.
[4] P.K.Kuhl & J.D.Miller, "Speech perception by the chinchilla : Identification function for synthetic VOT stimuli," J.Acoust.Soc.Am., Vo.63, pp.905-917, 1978.
[5] D.M.Klatt," Software for a cascade/parallel formant synthesizer," J. Acoust. Soc. Am., Vo.67, pp.971-995, 1980.