ANN.Bull.RILP(1990)No.24, 145-156

RELATIONSHIPS BETWEEN THE EAR'S TEMPORAL WINDOW AND VOT PERCEPTION FOR NORMAL AND HEARING-IMPAIRED LISTENERS

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

1. Introduction

Although most sensori-neural hearing-impaired listeners have difficulties in their speech perception, the details of such difficulties are vary greatly among individuals and are not necessarily accounted for simply by the auditory functions measured in ordinal pure tone audiometry. This may be due to the fact that there are various factors which are complexly related to speech perception.

For instance, pure tone sensitivity, frequency selectivity and the temporal acuity of the ear are possible factors in the auditory process affecting speech perception. Furthermore, factors representing how concretely a patient has phonemic categories and how correctly he/she can use such categories are also very important. Although it seems very important and reasonable to combine such factors in the speech perception, or to try to explain difficulties in speech perception based on such factors, no such theories have been advanced. If we could develop such a theory, it would be very useful for the diagnosis of hearing impairment, hearing-aid design and fitting, and training for the hearing-impaired.

To establish such a theory, however, the effects of each factor mentioned above upon speech perception have to be clarified first. Here, 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. In most of these reports, the temporal resolution was examined for non-speech stimuli based on measuring techniques such as the gap-detection task or the temporal masking paradigm[1,2]. However, it has not been sufficiently explained how the declining temporal resolution measured for non-speech stimuli 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

* Department of Otolaryngology, Faculty of Medicine, University of Tokyo; ** Department of Computer Science, Chiba Institute of Technology; *** Depertment of Otolaryngology, School of Medicine, Kinki University

of such a hearing-impaired patient may be reduced. We have already examined difficulties D1 and D2 [3,4]. 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, and the difference limens of VOT in a CV context 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 healthy 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 contexts than in CV contexts. H3)Such VOT boundary shifts 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

The shape of the temporal window was measured on the basis of the method proposed by Moore et al.(1988)[5]. Fig. 1 shows the stimulus configuration and the shape of the hypothetical temporal window. Given an outline of this measurement, the threshold for a brief sinusoidal signal presented in a temporal gap between two bursts of noise was measured as a function of the length of the gap. 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.

2.1 Procedure

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, and at the three levels of 80, 60 and 40dBSPL for normal hearing subjects.

The thresholds were measured using an adaptive, two-alternative, forced-choice procedure controlled by a micro-computer. One trial consisted of two stimuli intervals separated by a silent interval of 500ms. The onset of each trial was indicated by a small lamp which was set on a response box. The two masker bursts were presented in both intervals, but the signal occurred randomly in either the first or second interval. The subjects had to decide in which interval the signal occurred. The signal level was decreased after two consecutive correct responses, and increased after each incorrect response. The initial step size for changes in the signal level was 10dB, and was reduced by 2dB after each reversal at which a transition from decreasing to increasing level (or vice versa) occurred and was fixed at 2dB after three reversals. Testing continued until 12 reversals had occurred, and the average level at the last 8 reversals was taken as the estimate of the threshold.

2.2 Subjects

Six normal hearing subjects (aged 22-30 years), and three sensori-neural hearing-impaired subjects (PA: a 65-year-old man whose average hearing level of the tested ear was 55dBHTL; PB:a 16-year-old woman with 80dBHTL; PC: a 53-year-old man with 60dBHTL) took part in the study. Fig. 2 shows the audiograms and the ages of the hearing-impaired subjects.

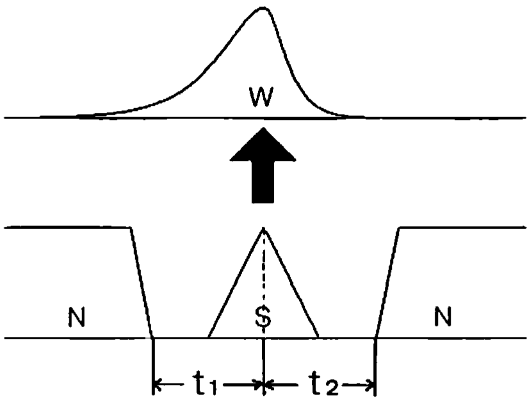

Fig. 1 Stimulus configuration for measurement of the temporal window (W)

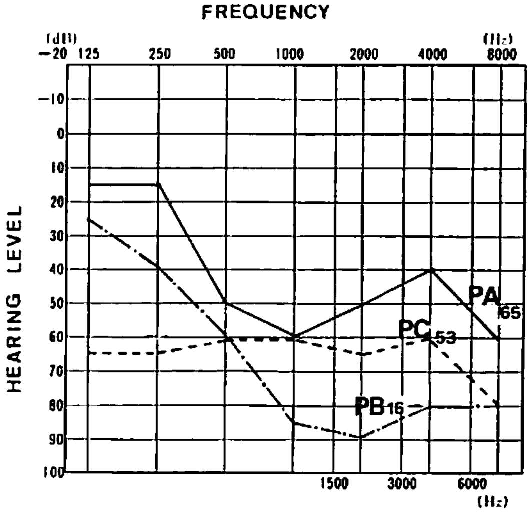

Fig. 2 Audiograms of the tested ears of the hearing-impaired subjects

2.3 Results

Fig. 3 shows the shapes of the temporal windows derived from the data for each hearing-impaired subject, and Fig. 4 shows those derived from the data with a masker noise level of 80dB or 40dBSPL for a normal subject. 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. To clearly indicate the characteristics of the asymmetry in the window shape, the ERD for the left side (t<0) of the window (ERDN) and the ERD for right side (t>0) of the window (ERDP) were also calculated separately. Fig. 5 shows the relationship between the ERD and the masker noise level.

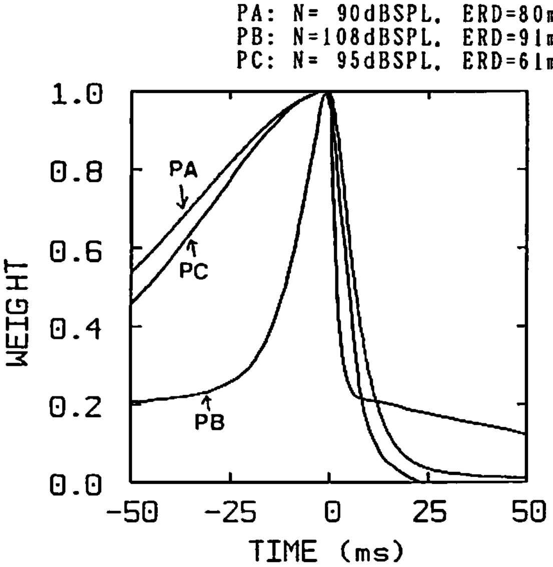

Fig. 3 The temporal windows of the heaing-impaired subjects.

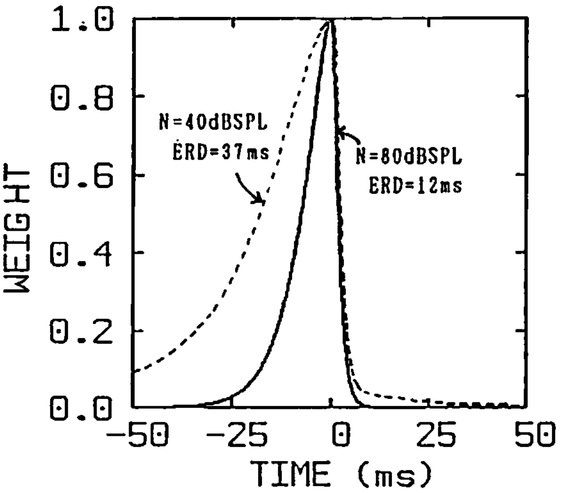

Fig. 4 The temporal windows of a normal subject at 40dB and 80dBSPL of masking noise.

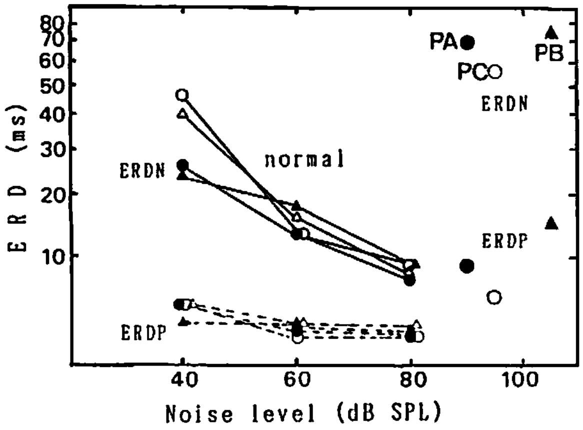

Fig. 5 The relationship between the equivalent rectangular duration (ERD) of the temporal window and the noise level for the normal subjects and the impaired (PA, PB and PC) subjects.
(------)ERDP:ERD of right-side
(———)ERDN:ERD of left-side

The results show that the left-side 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. For the normal subjects, the lower the masker level was, the broader the temporal window was. For the hearing-impaired listeners, the temporal windows were generally broader than for the normals, although the inter-subject differences were large.

3. Measurement of VOT perception

3.1 Stimuli

Temporal changes in some acoustic features may be used as critical cues for consonant perception. For example, voiced and unvoiced consonants are distinguished based on the value of the VOT, the voice onset time. In this experiment, we examined the VOT values at the phoneme boundaries between the voiced and unvoiced bilabial plosive consonants in CV (/pa-ba/) and VCV (/apa-aba/) sequences .

The stimuli were generated using a Klatt-type formant speech synthesizer (Klatt, 1980)[6]. The synthetic parameters were basically identical to those used by Kuhl (1978)[7] with some modifications. Fig. 6 illustrates the contour of each formant transition (F1-F3). 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 1ms step, from 0 to 120 ms. Accordingly, 121 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. Furthermore, there were two versions of the VCV stimuli, one with a 20-ms silent interval between the first vowel and the release of the burst for the stop consonant, and another without such an interval. Consequently, three contexts, CV, V_1CV (with a 20-ms silent interval between V_1 and C) and V_2CV (without any silent interval between V_2 and C), were examined. 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.

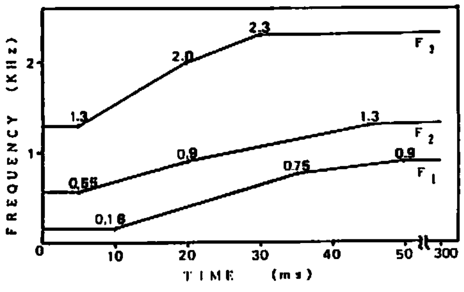

Fig. 6 The formant contour of the stimulus in the CV context. The burst release at 0ms.

3.2 Measurement of the phoneme boundary

3.2.1 Procedure

The three sets of stimuli (CV, V_1CV with a 20-ms silent interval between V_1 and C, and V_2CV without any silent interval) were tested. From each set, the stimuli in steps of 10ms along the VOT continuum were selected and 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.

3.2.2 Results

The results are shown in Fig.7(a) for the normal subjects and Fig.7(b) for the hearing-impaired subjects .

The following tendencies were observed from the results of the normal subjects . 1) The VOT values at the phoneme boundary in the VCV contexts were significantly longer than those in the CV context. 2) The lower the level was, the longer the VOT value at the phoneme boundary. 3) Although the inter-subject differences were not small, for some normal subjects, the VOT values at the boundary in the V_1CV context with a 20-ms silent interval between V_1 and C were shorter than in the V_2CV sequence without any silent interval.

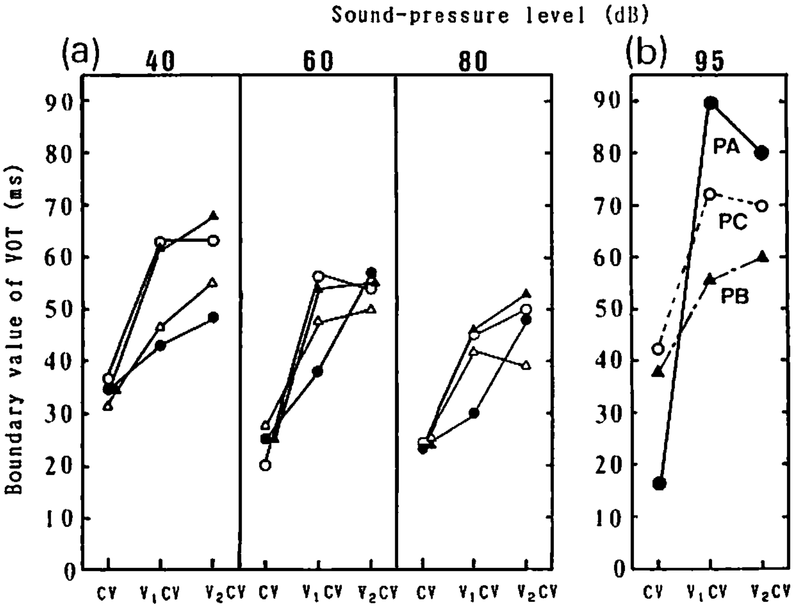

Fig. 7 a) The VOT values at the phoneme boundary between /ba/ and /pa/ with or without a preceding /a/ at three sound-pressure levels for the normal subjects. b) Those for the impaired subjects at one level. V_1=V_2, however V_1C has a 20ms silent interval between V_1 and C, and V_2C has no silent interval.

The results obtained from the hearing-impaired subjects were generally similar to the results observed for the normal subjects at 40dBSPL, although the inter-subject differences were larger than for the normals. The differences between the phoneme boundaries in the CV context and in the VCV contexts were larger than those for the normal subjects.

3.3 Measurement of the DLs of the VOT

3.3.1 Procedure

The difference limens for the VOTs were measured as the smallest detectable increments in the VOT at several reference values using the adaptive procedure of ABX discrimination task. The reference values of the VOTs were set in the region around the phoneme boundaries for the individual subjects. The adaptive procedure of the ABX task was a similar method used to measure the temporal window. First, a reference VOT was specified, for instance, at VOTr. Two stimuli whose VOT were VOTr and VOTr+dVOT (dVOT > 0) were presented as A or B. Then one of these two stimuli was presented as X. The subject had to select A or B, which was felt to be the same as X. The value of dVOT at which the subject could not select correctly more than at the chance level was specified as the DL of the VOT.

The stimuli used were only the CV syllables and were presented at the most comfortable level for the individual subjects.

3.3.2 Results

Fig. 8 shows the relationship between the DLs of the VOT and the reference values. This figure shows the individual results for each hearing-impaired subject, but only the average values for the normals. The vertical lines indicate the phoneme boundaries.

For the normal subjects, the following tendencies were observed. 1) The DL curves had a "V" shape, and the minimum values of the DLs were about 10ms. 2) The VOT values at which the DLs had their minimums tended to correspond with the VOT values of the phoneme boundaries.

For the hearing-impaired subjects, there were certain individual differences. For subject PA and PB, their results were analogous to the normal subjects, however their DLs tended to become larger than those of the normals when the reference VOT values were longer than the phoneme boundary values. For subject PC, the DLs were remarkably larger than for the other subjects, and the reference VOT values at which the DLs had their minimums did not correspond with the phoneme boundaries.

4. Discussion

The results presented above indicate that the width of the temporal window for the hearing-impaired subjects was broader than for the normals, and the inter-subject differences were larger than for the normals. This means that the ear's temporal resolution of the hearing-impaired subjects tended to be wider than for the normals. For the normal subjects, the width of the temporal window became broader for the lower levels of the masking noise. This means that the ear's temporal resolution of the normal subjects tended to be lower for the lower levels of the masking noise.

For both groups of subjects, the shape of the temporal window was asymmetric. It had a broader width at negative times (t<0) than at positive times (t>0). A temporal window at the negative side of the time axis indicates the duration in which the forward masking lasts, and one at the positive side indicates the duration in which the backward masking lasts.

The results presented above indicate that the effects of the forward masking lasted longer that those of backward masking, and that the effects of forward masking tended to last longer for the impaired ears than for the healthy ears.

In the experiment on VOT perception, the following results were obtained. 1) For the normal subjects, the lower the level was, the longer the VOT value at the phoneme boundary, in both the CV and VCV contexts. And at the lowest measurement level, that is 40dBSPL, the VOT values became nearly identical to those of the hearing-impaired. 2) For both the normal and the hearing impaired, the VOT values of the phoneme boundary were significantly longer in the VCV contexts than in the CV contexts. 3)The differences in the phoneme boundary between the CV and the VCV contexts tended to be larger for the hearing-impaired subjects than for the normal subjects.

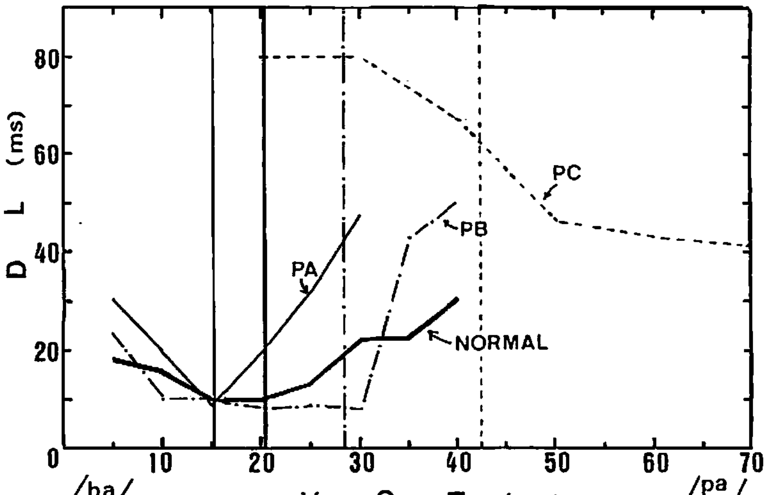

Fig. 8 The difference limens of the VOT. For the normal subjects, the average is shown. The vertocal lines show the phoneme boundary.

It seems that these results can be explained in relation to the temporal resolution of the ear. For the result 1), it is plausible that, because the temporal window became broader or the temporal resolution got worse at a lower masking level, the influence of the preceding or following vowel on the VOT perception was prolonged at the lower stimulus level. Therefore, to identify a stimulus as a unvoiced consonant, a longer VOT value might have been needed.

To explain result 2), we may also relate this result with the characteristics of the temporal resolution. Fig. 9 shows an illustration of the outline of our interpretation.

To identify the unvoiced stop consonants, the VOT had to be longer than a certain value which was the phone boundary Pb. The VOT is usually defined physically or acoustically as the interval between the plosive burst and the voice onset, which is shown in Fig. 9 as VOTa. However, we define it here as a psychoacoustial VOT, VOTp, in the following way.

When the plosive burst input is received at the ear, it takes a certain time for the loudness of the input sound to increase to a level which is loud enough to be perceived, and it also takes a certain time for the loudness of the input sound to decay to have no influence on the perception of following sounds. These temporal characteristics of loudness, increasing and decaying characteristics, may be represented by a moving average filtering of the input signal using the temporal window of the ear as shown in Fig. 9.

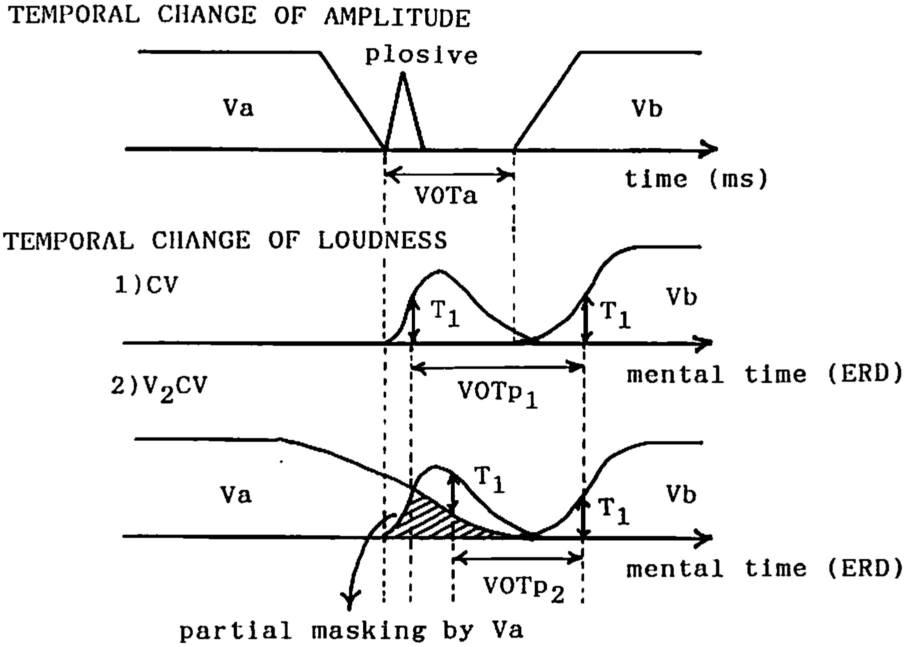

Fig. 9 A model of VOT perception.

Therefore, the psychoacoustical VOT, VOTp, can be defined as the interval between the epoch when the loudness of the plosive burst exceeds a threshold T1 and the epoch when the loudness of the following vowel exceeds the threshold.

Thus, our model of the discrimination between the unvoiced and voiced stops can be simply described as the following equation using VOTp (the psychoacoustical voice onset time), ERD (the width of the temporal window represented by the equivalent rectangular duration) and Pp:the psychoacoustical threshold dividing voiced and unvoiced stops.

$$\begin{array}{lll} \mathrm{VOTp/ERD} < & \mathrm{Pp} : & \text{Voiced stop} \\ \mathrm{VOTp/ERD} >= & \mathrm{Pp} : & \text{Unvoiced stop} \end{array} \qquad (1)$$

In Equation (1), ERD is introduced to take into account the temporal resolution of the ear. Our model predicts that a longer VOTp is required for a subject having poorer temporal resolution (larger ERD) in identifying unvoiced stops. This explains why the VOT value at the phoneme boundary gets longer at lower perceiving levels even for normal subjects.

We would like to discuss next the differences in the VOT perception which depend on context. In the case of CV contexts, the discussion mentioned above is directly valid. In the case of the VCV contexts, because of the partial masking caused by the preceding vowel V_1 or V_2, 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 V_1CV or V_2CV contexts might be shorter than that in a CV context if the acoustically defined VOT or VOTp remains the same. Furthermore, because there is a silent interval between V_1 and C and no such interval between V_2 and C, the VOTp in V_2CV contexts must be shorter than in V_1CV contexts. Therefore, according to Equation (1), our model predicts that the VOT value at the phoneme boundary may be the shortest for the CV contexts,the longest for the V_2CV, and somewhere between these two extremes for the V_1CV contexts. This prediction is correct as shown in Fig. 7 and is listed as result 2).

For the hearing-impaired subjects, although the tendency of the effects of the contexts on the VOT phoneme boundaries were the same as for the normal subjects, there were large intersubject differences in the amount of the boundary shifts depending on the context. For hearing-impaired subject PA, the boundary value of the VOT became significantly longer than for the normal subjects in the VCV contexts. This might have been due to a wider temporal window or a larger ERD than for the normal subjects. To obtain a concrete conclusion concerning this issue, however, we should measure a larger number of patients.

For the relationship between the DL of the VOT and the phoneme boundaries, the minimum values of the DLs for the normal subjects was almost the same as the phoneme boundaries. It seems plausible that the ability to identify stop consonants may have stabilized because of this correspondence between the most sensitive area for a change in the VOT and the VOT phoneme boundary.

For the hearing-impaired subjects with wide temporal windows, on the other hand, the inter-subject differences were large. For subject PC, whose pure tone audition was flat across a wide frequency range, the DL values were larger than for the normal subjects, and the location of the minimum DL was separate from the phoneme boundary. Such a large discrepancy might be an important factor restricting his ability to identify voiced or unvoiced stops. For the other two hearing-impaired subjects, the results were almost the same as for the normal subjects. It may be that these subjects had a relatively good sensitivity in the low frequency area which is important for the identification of the stop consonants.

As a summary, we may conclude the following for the hearing-impaired subjects who had a wider temporal window of the ear or a poorer temporal acuity than healthy listeners. 1)The VOT value at the phoneme boundary tended to be longer than for the normal subjects (a longer VOT was necessary to identify an unvoiced consonant). 2) Because of masking on the VOT caused by preceding vowels, the VOT phoneme boundary shifted to a longer value in VCV contexts than in CV contexts. 3)Such VOT boundary shifts between the CV and VCV contexts tended to be larger for the patient with a poorer temporal resolution. It seems plausible to conclude that the temporal resolution of the ear has a close relation to the VOT perception.

Of course, we should evaluate a large number of patients before we generalize the tendencies observed in this study for a few subjects.

5. Conclusions

In the experiments reported here, the following results were obtained.

1) For the normal subjects, the lower the measurement level, not only the broader the width of temporal window, but also the longer the VOT value at the phoneme boundary.

2) The width of the temporal window was longer for the hearing-impaired subjects than for the normal subjects. Also, the VOT values at the phoneme boundaries tended to be longer for the hearing-impaired subjects than for the normal subjects.

3) The phoneme boundaries between the unvoiced and voiced stops significantly shifted depending upon the contexts and also upon the perception levels. Such VOT boundary shifts between the CV and VCV contexts tended to be larger for the patient with longer temporal window.

4) 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) Glasberg, R., Moore, B.C.J., and Bacon, S.P.: Gap detection and masking in hearing-impaired and normal-hearing subjects, J. Acoust. Soc. Am., 81(5), 1546-1556, 1988.
2) Tyler, R.S., Summerfield, Q., Wood, E.J. and Fernandes, M.A.: Psychoacoustic and phonetic temporal processing in normal and hearing-impaired listeners, J. Acoust. Soc. Am., 72(3), 740-752,1982.
3) Yamada, A., Imaizumi, S., Harada, T., Mikami, Y., Deguchi, T. and Hosoi, H.:Effects of temporal factors on the Speech perception of the hearing-impaired --A preliminary report--, Ann. Bull. RILP, 21, 131-140, 1987.
4) Hayashi, A., Imaizumi, S., Harada, T. and Hosoi, H.:Effects of stimulus duration and inter-stimulus interaction on vowel intelligibility for normal and hearing-impaired subjects, Ann. Bull. RILP., 23, 163-172, 1989.
5) Moore, B.C.J., Glasberg, B.R., Plack, C.J. and Baswas, A.K.:The shape of the ear's temporal window, J. Acoust. Soc. Am., 83(3), 1102-1106, 1988.
6) Kuhl, P.K. and Miller, J.D.:Speech perception by the chinchilla:Identification function for synthetic VOT stimuli., J. Acoust. Soc. Am., 63, 905-917, 1978.
7) Klatt, D.M.:Software for a cascade/parallel formant synthesizer, J. Acoust. Soc. Am., 67(3), 971-995, 1980.