Ann.Bull.RILP(1984)No.18, 197-210

VOICE ONSET TIME PERCEPTION IN NORMAL AND APHASIC SUBJECTS

Motonobu Itoh*, Itaru F. Tatsumi*, Sumiko Sasanuma*
and Yoko Fukusako**

Voice onset time (VOT) has been found to be an important acoustic cue in the voiced-voiceless distinction of stop consonants (Lisker, 1975; Summerfield and Haggard, 1977). The present study was designed to examine VOT perception in normal Japanese adults. In addition, the effects of aging and aphasia on VOT perception were investigated.

1. Method

1.1 Subjects

Three groups of subjects took part in the study. They were 10 normal young adult subjects, 10 normal aged subjects, and 26 aphasic subjects. The aphasic subjects were patients of the Tokyo Metropolitan Geriatric Hospital. Table 1 summarizes the age, sex, cause of brain damage, site of lesion, type of aphasia and post-onset month for each aphasic subject. The normal young adult subjects were staff members of the Tokyo Metropolitan Institute of Gerontology and the Tokyo Metropolitan Geriatric Hospital, whose ages ranged from 24 to 34 years (with a mean age of 28.7 years). The normal aged subjects were volunteers from local senior-citizens groups and their ages ranged from 61 to 79 years (with a mean age of 69.6 years). All subjects, with the exception of the aphasic subjects, had normal speech for their ages. No subjects had a mean hearing level of more than 50 dB (JIS T1201-1982).

1.2 Stimuli

Two types of synthetic stimuli (one speech and one non-speech) were prepared using a computer simulation with a terminal analog synthesizer with a noise source, a buzz source, formant circuits and a radiation circuit. The speech stimuli consisted of 10 monosyllables in which only the interval between the release of the noise burst in a stop consonant and the onset of glottal pulsing (VOT) was varied by 8 msec steps from -16 msec [ga] to +56 msec [ka] (Fig. 1, top). The non-speech stimuli were prepared in the same way as the speech stimuli, with the exception that the glottal source wave was used without any modification through the formant circuits (Fig. 1, bottom). The intensity ratio of noise to glottal pulse was the same both for the speech and non-speech stimuli.

*Tokyo Metropolitan Institute of Gerontology
**Tokyo Metropolitan Geriatric Hospital

Table 1. The age, sex, cause of brain damage, site of lesion, type of aphasia and post-onset month for the aphasic subjects.

Patient	Age	Sex	Cause of brain damage	Site of lesion*	Type of aphasia	POM**
1	44	M	CVD***	Left frontal, temporal and parietal regions	Broca	24
2	19	M	Trauma	Left fronto-parietal region	Broca	21
3	70	F	CVD	Left frontal region	Broca	22
4	46	M	CVD	Left frontal and capusulo-putaminal regions	Broca	7
5	63	M	CVD	Unknown	Broca	5
6	65	F	CVD	Left anterior limb of the internal capusula and the left frontal lobe	Broca	8
7	46	F	CVD	Left superior temporal gyrus	Wernicke?	15
8	44	M	CVD	Left temporo-parietal region	Wernicke	25
9	70	M	CVD	Left temporo-parietal region (near the Sylvian fissure)	Wernicke	36
10	61	M	CVD	Left temporal and parietal regions	Wernicke	39
11	69	F	CVD	Left and right posterior temporal regions (Bilateral lesions)	Wernicke	8
12	76	M	CVD	Left and right temporo-parietal region (Bilateral lesions)	Wernicke	48
13	83	M	CVD	Left frontal region	Amnesic	5
14	55	F	CVD	Left frontal, temporal, and parietal regions	Amnesic	71
15	46	M	CVD	Left frontal and capusulo-putaminal regions	Amnesic	14
16	36	M	Herpes-encephalitis	Unknown	Amnesic	15

(Table 1, Cont.)

Patient	Age	Sex	Cause of brain damage	Site of lesion	Type of aphasia	POM
17	59	M	CVD	Left basal ganglia	Amnesic	12
18	55	M	CVD	Left putaminal region	Amnesic	8
19	62	M	CVD	Left Sylvian fissure and left posterior temporal region	Amnesic	7
20	45	M	CVD	Left temporo-parietal region	Trans-cortical?	4
21	47	M	CVD	Right putaminal region	Crossed?	14
22	50	M	CVD	Left frontal, temporal and parietal regions	Global?	7
23	62	M	CVD	Left temporo-parietal region and right parietal region (Bilateral lesions)	Unknown	39
24	56	M	CVD	Right frontal, temporal and parietal regions	Crossed?	14
25	65	M	CVD	Left frontal, temporal and parietal regions	Global	15
26	47	M	CVD	Left frontal, temporal and pariental regions, and left thalamus/basal ganglia	Global?	52

* Identified by CT scan

** POM = post-onset month

*** CVD = cerebro-vascular disease

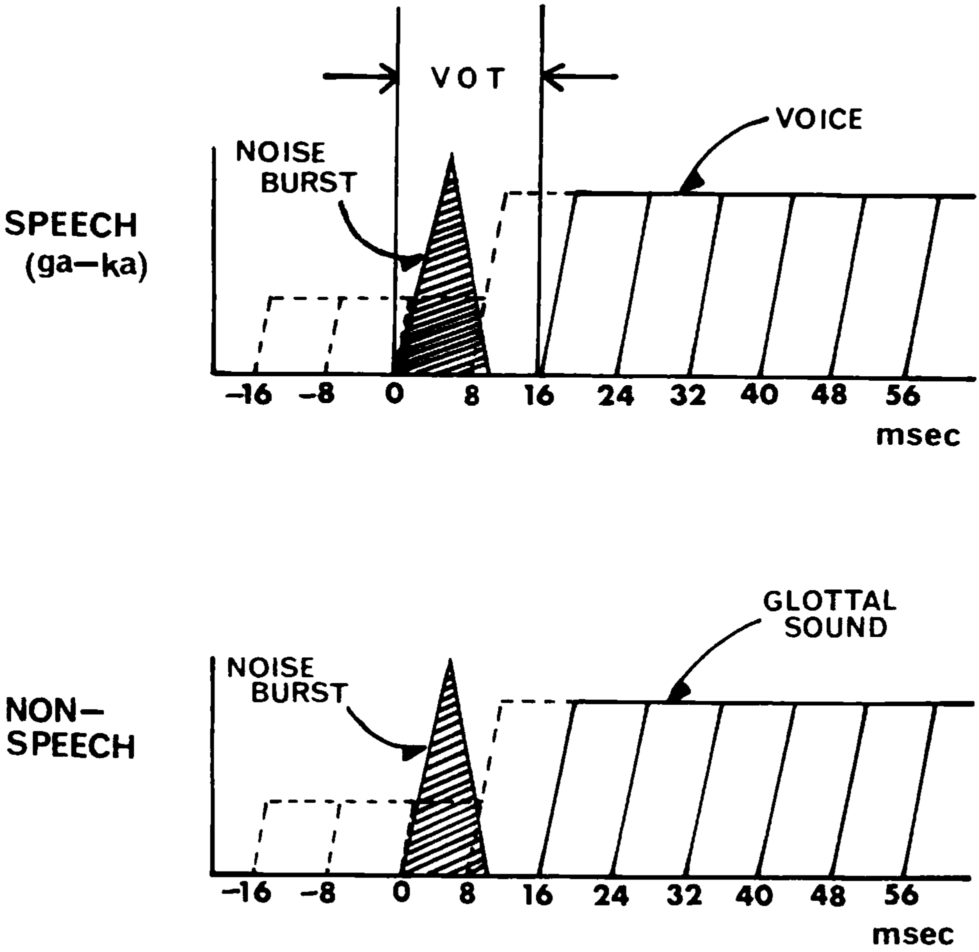

Fig. 1 Schemata of speech stimuli(top) and non-speech stimuli(bottom).

1.3 Procedure

Each subject was tested individually in a soundproof room. The taperecorded stimuli were presented through a loudspeaker at a comfortable listening level of 60 to 80 dB (C scale on a RION Noise Meter NA-09). The interstimulus interval was 4 sec.

Preceding the main experiment, subjects were given a preliminary session to insure that they could understand the procedure. The stimuli selected for the preliminary session were unambiguous ones that occupied one or the other of the two extremes of the VOT range. For the speech stimuli, subjects were asked to point to the appropriate card printed with [ga] or [ka] in Japanese kana upon hearing the stimuli. For the non-speech stimuli, subjects were told that on some trials two kinds of sounds (a click and a buzz) would be present, whereas on the other trials only one kind of sound (a buzz) would be present. Subjects were asked to point to the card on which "2" was printed for two kinds of sounds or the card on which "1" was printed for one kind of sound.

When the examiner judged that the subjects understood the task and could respond appropriately, they were given 30 trials with stimuli which occupied one or the other of the two extremes of the VOT range in random order. If subjects responded correctly to 20 trials out of the 30 trials, they were included in the study. Two aphasic subjects failed to reach this criterion for the speech task, and thus were eliminated from the main experiment.

In the case of the non-speech task, all of the normal young adult subjects reached the criterion level while in two cases the normal aged subjects appeared not to understand the task. Furthermore, there were at least three normal aged subjects who seemed to understand the procedure and reached the criterion level, yet performed at a near chance level in a practice session (described shortly). Based on these observations, the experimenter decided not to collect data for the non-speech stimuli from the normal aged subjects nor from the aphasic subjects. Throughout the preliminary and screening sessions, immediate feedback as to the correct response was provided by the experimenter. After the screening session, subjects were given a practice session of 30 trials using stimuli taken from the main experiment series. During the practice session, the experimenter said "That's correct" immediately after the subject's responses only when they were adequate responses to the unambiguous stimuli with VOT values of -16, -8, +48 and +56 msec. For the rest of the stimuli that were more or less ambiguous, the experimenter said "Yes" regardless of the adequacy of the responses.

The main experiment was conducted in two to four sessions on separate days for each of the speech and non-speech tasks. Each of the 10 stimuli was presented singly, in a predetermined random order. There were 76 replications of each stimulus, requiring a total of 760 judgments from each subject*.

1.4 Data analysis

According to Tatsumi et al.'s model of perception (Tatsumi, Sasanuma and Fujisaki, 1980), values of three parameters(μ,σ and qe) were obtained for each subject. In essence, the values of these parameters were estimated by means of the least squares method. μ and σ represent the category boundary and an index of accuracy of identification, respectively. qe indicates the probability of occurrence of attention and/or response errors. Fig. 2 shows examples of the approximate identification curve and estimated values of μ, σ and qe for the speech stimuli in an normal young adult subject (top), and an aphasic subject (bottom). The abscissa indicates the stimulus value of VOT in msec and the ordinate shows the probability of making [ka] judgments. Circles represent the measured data, and the solid curve represents the closest approximation to such data. The vertical line indicates the category boundary (μ).

2. Results and discussion

2.1 Identification of non-speech stimuli in normal young adult subjects

Table 2 presents mean, minimum and maximum values of μ, σ and qe for the three subject groups. It is apparent from the table that these values differ between non-speech and speech stimuli in normal young adult subjects. That is, μ of non-speech stimuli is smaller than that of speech stimuli, while the σ and qe of the former stimuli are larger than those of the latter stimuli. The discrepancy in the location of the category boundaries between non-speech and speech stimuli with varying VOT has already been reported by Miller, Wier, Pastore, Kelly and Dooling (1976). The difference in the position of the category boundaries of non-speech and speech stimuli may be attributable mainly to the difference in the acoustical parameters of stimuli other than VOT since systematic changes in the position of the voicing boundary can be accomplished by manipulations of certain acoustical parameters such as formant locus (Lisker, 1975; Summerfield and Haggard, 1977). Our non-speech stimuli were indeed different from our speech stimuli in terms of several acoustical parameters other than VOT.

*Some subjects, especially aphasic subjects, exhibited a near chance level performance during the early session(s) in the main experiment despite the fact that they had had the practice session and then improved in performance in the following sessions. In these cases, data from the early session(s) were eliminated from analysis, resulting in less than 760 judgments.

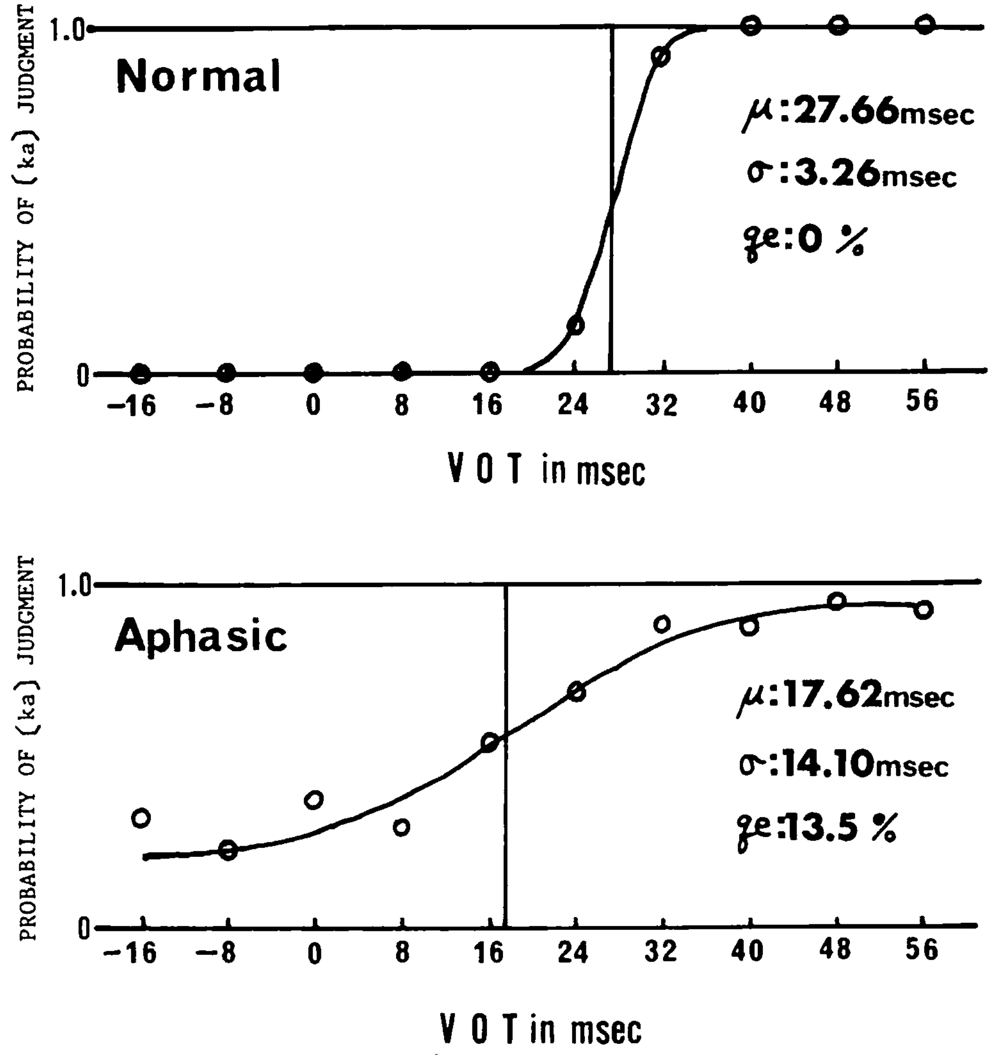

Fig. 2 Examples of the approximate identification curve and estimated values ofμ,σ and qe for the speech stimuli in a normal young adult subject(top) and an aphasic subject(bottom). Circles represent the measured data and the vertical line indicates the category boundary(μ).

Table 2. Mean, Minimum and maximum values of μ, σ and qe for the three subject groups.

	Stimulus	parameters	Mean	Minimum	Maximum
Young Adult N = 10	Non-speech	μ	18.53	12.91	27.40
		σ	4.96	2.58	6.59
		qe	1.5	0	6.0
	Speech	μ	24.38	16.47	29.98
		σ	3.85	3.01	5.95
		qe	0.3	0	1.0
Aged N = 10	Speech	μ	25.49	18.66	32.61
		σ	3.98	3.01	5.06
		qe	0.8	0	5.0
Aphasic N = 22	Speech	μ	29.54	13.97	40.14
		σ	6.35	2.95	14.10
		qe	5.0	0	14.0

μ and σ (in msec), qe (in %)

The mean value of the category boundary for the non-speech stimuli in Miller et al.'s study was +15.13 msec which was slightly smaller than +18.53 msec in the present study. This discrepancy may be due to the difference in subjects. Their subjects were experienced listeners, while ours were inexperienced. The results of the peresent study concerning the category boundary of non-speech stimuli are consistent with the results of Hirsh (1959), Hirsh and Sherrick (1961), Stevens and Klatt (1974) and Pisoni (1977) who found that "20 msec is about the minimal difference in onset time needed to identify the temporal order of two distinct events" (Pisoni, 1977, PP. 1359-1360).

The difference in σ and qe between the two kinds of stimuli simply indicates that the non-speech task was more difficult than the speech task. It should be pointed out, however, that the overall performances for the non-speech stimuli are quite good, i.e., they are never at a near chance level. This result serves as another demonstration of the existence of categorical perception for non-speech stimuli (Miller et al. 1976; Pisoni, 1977). Furthermore, as Pisoni (1977) pointed out, it appears that the processing of temporal order information may underlie the perception of voicing distinctions in stop consonants.

2.2 Identification of speech stimuli in normal young adult subjects

As Table 2 shows, the mean VOT value of the categorical boundary for the speech stimuli [ka/ga] in the normal young adult subjects was +24.38 msec which is similar to the VOT value (+26 msec) reported by Shimizu (1977) who investigated the perceptual boundary of [ka/ga] in 10 Japanese college students. These values of Japanese speakers are lower than those (+30∼+40 msec) of English speakers but are similar to those (+20∼+30 msec) of Spanish speakers (Lisker and Abramson, 1967).

Our previous study (Itoh, Sasanuma, Tatsumi, Murakami, Fukusako and Suzuki, 1982) which examined the VOT production of normal young adult Japanese speakers indicated that the maximum VOT value of /ge/ was +20 msec, whereas the minimum VOT value of /ke/ was +30 msec. Thus, the perceptual boundary found in the present study matches the productive crossover zone very well.

2.3 Effect of aging on VOT perception

It can be seen in Table 2 that the normal aged subjects showed values for μ, σ and qe almost identical with those of the normal young adult subjects. Actually, there is no statistical difference in the values between the two subject groups. This result suggests that aging does not affect VOT perception.

2.4 VOT perceptin of aphasic subjects

As described in the procedure section, two aphasic subjects (Patient Nos. 12 and 25 in Table 1) failed to perform the speech task and were eliminated from the main experiment. In addition, for two other aphasic subjects (Patients Nos. 9 and 10 in Table 1) no approximation to the measured data could be obtained since the probability that their responses to a given stimulus belong to one of the two categories remained at a chance level regardless of the VOT values. These aphasic subjects are judged to be incapable of identification. Table 2 does not contain the data for these four aphasic subjects.

The distribution of the μ values for the speech stimuli in the three subject groups along with the distribution of the μ values for the non-speech stimuli in the normal young adult group are plotted in Fig. 3. Fig. 4 shows a scattergram of qe and σ for all subjects. A circle, a triangle and a filled square represent the normal young adult, the normal aged and the aphasic subjects, respectively.

It is apparent in Table 2 that the mean values of μ, σ and qe for the aphasic subject group are quite different from those of other subject groups. The larger mean value of μ for the aphasic subject group indicates that the category boundary shifted to a higher region along the VOT continuum. From Fig. 3, it is also clear that two aphasic subjects exhibited μ values much smaller than the mean μ value of the aphasic group, resulting in a greater distribution of μ values of the group in comparison with the distributions of the normal subject groups. The larger mean values of σ and qe may indicate that the aphasic subjects as a group are less accurate in their perception of VOT and that they made more errors in the attention and/or response process than the normal subjects. As Figs. 3 and 4 indicate, however, the values of approximately half of the aphasic subjects fell within the normal range of distribution. In order to determine what factors are contributing to the deterioration of VOT identification in half of the aphasic subjects, eight kinds of correlation coefficients were obtained (Table 3). It can be seen in the table that in general, the magnitudes of the correlations are weak with the exception of a moderate negative relationship between the total score on our diagnostic aphasia test (the Roken Test for the Differential Diagnosis of Aphasia - RTDDA) and σ. This latter finding may suggest that the more severe the aphasic impairment is, the less accurate the VOT perception is. Analysis of the results further indicates that there seems to be little relation between the clinical type of aphasia and the capability for VOT identification, except that the patients with amnesic aphasia showed comparatively better performances than the patients with other types of aphasia. The latter finding, however, can again be accounted for in terms of the overall severity of aphasic impairment because the impairment of the amnesic patients is in general less severe than that of the other aphasic patients.

At any rate, it is clear that some aphasic subjects do exhibit deterioration in identification of speech stimuli with varying VOT. These aphasic subjects may have difficulty in processing the temporal order information which underlies the

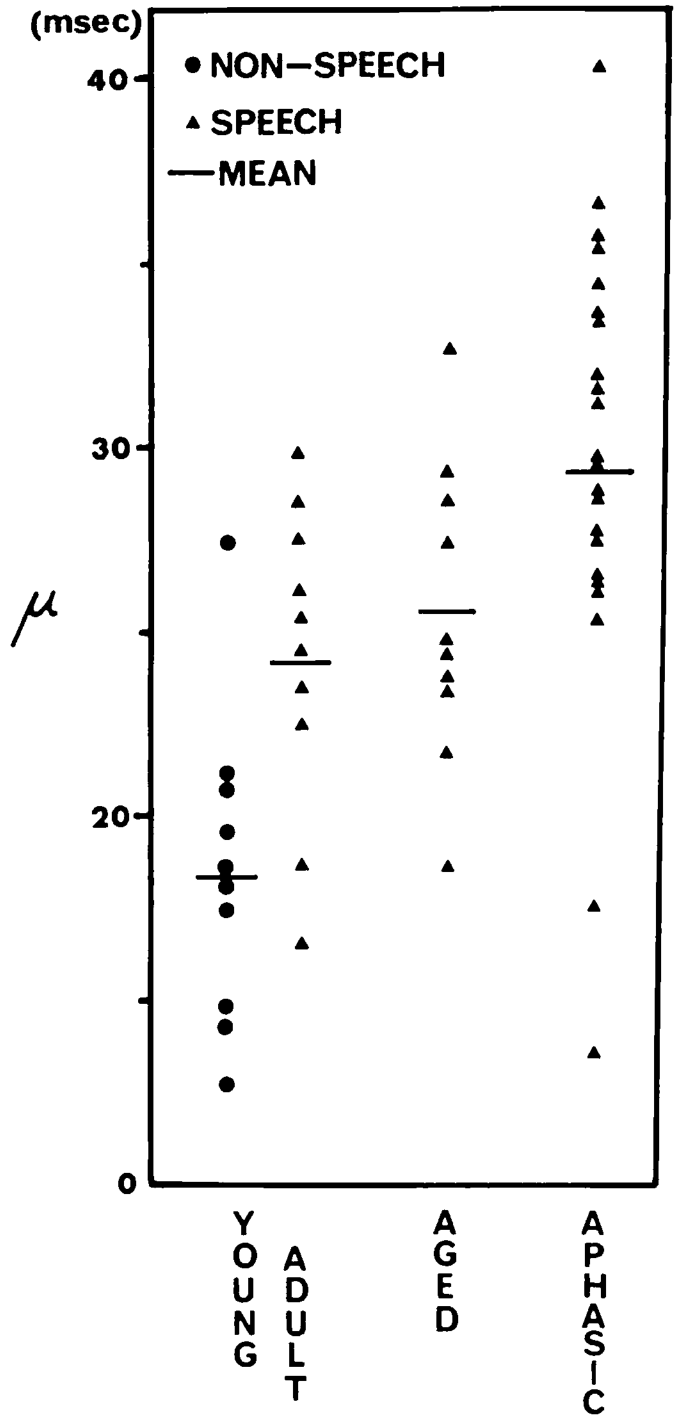

Fig. 3 The distribution of μ values for the speech stimuli in the three subject groups along with the distribution of μ values for the non-speech stimuli in the normal young adult group.

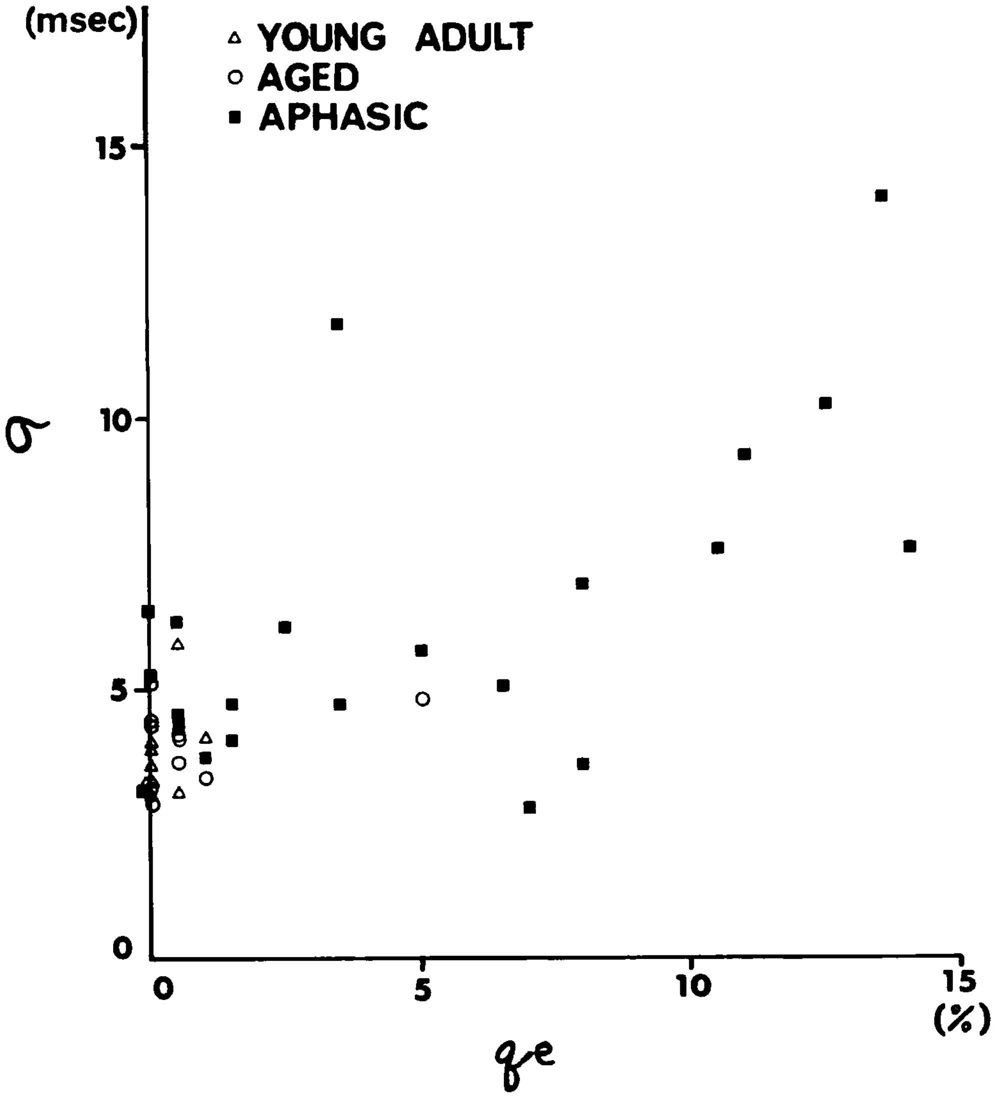

Fig. 4 A scattergram of qe and σ for all subjects.

Table 3. Correlations between four kinds of scores on RTDDA*, and σ and qe

	Scores on RTDDA			
	The total score	The score on the subtest of following spoken directions	The score on the subtest of repeating monosyllable words	The score on the subtest of pointing to letters(kana) named
σ	-.62	-.42	-.32	-.42
qe	-.46	-.24	-.33	-.22

*The Roken (Tokyo Metropolitan Institute of Gerontology) Test for the Differential Diagnosis of Aphasia

perception of the voicing distinction in stop consonants. In other words, the aphasics' impairment does reflect a deficit at the prelinguistic level as well as the linguistic level. These observations and interpretations are consistent with those of previous experiments with aphasic subjects on auditory identification of signal duration and fundamental frequency patterns (Tatsumi, Sasanuma, and Fujisaki, 1978; Tatsumi, et al., 1980).

3. Summary and conclusion

This study was designed to examine the status of voice onset time (VOT) identification in Japanese normal young adult, normal aged and aphasic subjects. Synthetic speech and non-speech stimuli were constructed using a computer simulation with a terminal-analog synthesizer.

The present data from an identification experiment on non-speech stimuli in the normal young adult subjects showed strong evidence for the existence of categorical perception for such stimuli. The comparison of the performances of the normal aged subjects with those of the normal young adult subjects for the speech stimuli revealed that there was no effect of aging on the VOT perception of speech stimuli. Finally, it was found that half of the aphasic subjects showed deterioration in performance in VOT identification of speech stimuli, suggesting that these subjects may have difficulty in processing the temporal order information which underlies the perception of voicing distinctions in stop consonants.

Acknowledgement

The authors wish to express their thanks to Ms. Hisako Monoi, Ms. Kyoko Endo, and Mr. Tsutomu Suzuki of the Tokyo Metropolitan Geriatric Hospital for their cooperation.

A part of this study was presented at the 21st Meeting of Japanese Association of Rehabilitation Medicine, Matsuyama-city, Japan (July 6, 1984).

References

Hirsh, I.J. (1959); Auditory perception of temporal order. J. Acoust. Soc. Am., 31, 759-769.

Hirsh, I.J., and C.E. Sherrick (1961); Perceived order in different sense modalities. J. Exp. Psychol., 62, 423-432.

Itoh, M., S. Sasanuma, I.F. Tatsumi, S. Murakami, Y. Fukusako and T. Suzuki (1982); Voice onset time characteristics in apraxia of speech. Brain and Language, 17, 193-210.

Lisker, L. and A.S. Abramson (1967); The voicing dimension: Some experiments in comparative phonetics. Haskins Labs. Status Report on Speech Research, SR-11, 9-15.

Lisker, L. (1975); Is it VOT or a first-formant transition detector? J. Acoust. Soc. Am., 57, 1547-1551.

Miller, J.D., C.C. Wier, R.E. Pastore, W.J. Kelly, and R.J. Dooling (1976); Discrimination and labeling of noise-buzz sequences with varying noiselead times: An example of categorical perception. J. Acoust. Soc. Am., 60, 410-417.

Pisoni, D.B. (1977); Identification and discrimination of relative onset time of two component tones: Implications for voicing perception in stops. J. Acoust. Soc. Am., 61, 1352-1361.

Shimizu, K. (1977); Voicing features in the perception and production of stop consonants by Japanese speakers. Studio Phonologica, 11, 25-34.

Stevens, K.N., and D.H. Klatt (1974); The role of formant transitions in the voiced-voiceless distinction for stops. J. Acoust. Soc. Am., 55, 653-659.

Summerfield, Q. and M. Haggard (1977); On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. J. Acoust. Soc. Am., 62, 435-448.

Tatsumi, I.F., S. Sasanuma, and H. Fujisaki (1978); Auditory and visual perception of verbal and nonverbal stimuli in aphasic patients. Ann. Bull. RILP, 12, 157-166.

Tatsumi, I.F., S. Sasanuma, and H. Fujisaki (1980); Perceptual abilities of aphasic patients to identify fundamental frequency patterns and stroke directions in verbal and nonverbal stimuli. Ann. Bull. RILP, 14, 285-298.