

slopes of the identification curve were the steepest. There was no significant difference in either the PB or the slope value between the Japanese and the Korean subjects (PB: $p = .072, t = 1.64$; slope: $p = .152, t = 1.11$). In the discrimination test, the two stimuli whose final vowel differed by 11 msec in length were presented sequentially in each trial. The subjects were asked to judge which of the two words had a longer final vowel. Discrimination functions (Figure 2) in both groups had the highest discrimination ratio in the across-category trial (BC) and poorer within-category discrimination (AB and CD).

The results of the behavioral identification and discrimination tests indicate that the L2 learners had acquired categorical percepts of Japanese long/short vowels that were indistinguishable from those of the L1 subjects. This result is also evident when the present subjects were compared with the low-proficiency learners (data analyzed from Minagawa-Kawai, 2000) who had never been exposed to Japanese before 16 years old (average learning period; 1.7 years, $SD = 1.2$) nor had ever stayed for more than 1 week in a Japanese-speaking environment. Their identification curve does not show as steep a change in the BC segment (Figure 1). There is a significant difference in the slope of the present high-proficiency L2 learners ($p = .008, t = 2.36$).

NIRS Results

Automatic change-detection responses were recorded in the L1 and L2 subjects by presenting them with baseline stimuli and deviant stimuli under three conditions, namely, Sessions AB, BC, and CD. All these conditions presented the baseline and deviant stimuli differing by the same length; however, only the stimulus pair under the BC condition had a phonemic length difference

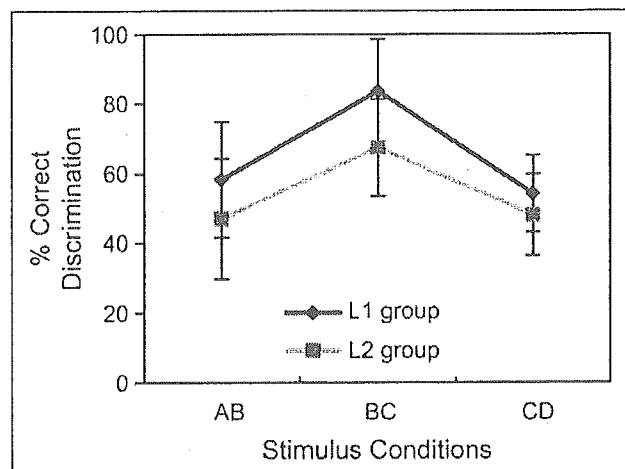


Figure 2. Average discrimination rate of /mama-/mama/ continuum. Both the L1 group (solid line) and the L2 group (gray line) show higher discrimination rates in the stimulus pair near the phoneme boundary.

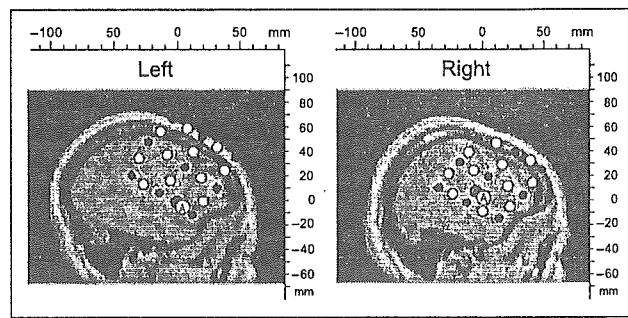


Figure 3. Locations of the NIRS probes and recording channels. NIRS probe positions (small white circles) and the centers of the measurement channels (white circles) are superimposed onto the parasagittal MR brain images. The lateral posterior borders of the Heschl's gyri are labeled "A." The channels with the maximal responses are shown with gray circles. Positive is towards the anterior and towards the vertex in the scales.

of the long/short vowels (across-category condition). Among the 12 recording sites on either side, the largest hemodynamic changes to the target block containing the deviant stimuli were most frequently observed in the auditory area (Figure 3), whereas the temporal and amplitude patterns of the recorded responses in the posterior frontal and inferior parietal areas (Figure 3) were smaller for all the subjects and varied considerably in both groups. The auditory channel on each side that presented maximal peak responses in the total hemoglobin (Hb) was chosen for statistical analysis. The averaged coordinates for the channels chosen were $y = -11.0$ ($SE = 2.9$) and $z = 11.5$ (2.2) on the left side and $y = -9.7$ (1.9) and $z = 8.9$ (5.6) mm on the right side for the Japanese group (only y and z coordinates are measured on parasagittal projection because the precise recording depths or x are not exactly known). Those for the Korean group were $y = -14.0$ ($SE = 1.6$) and $z = 9.9$ (2.7) on the left side and $y = -12.0$ (1.7) and $z = 11.7$ (2.4) mm on the right side with no statistical difference from those of the L1 group ($p > .3$). These areas are more superior and posterior to the lateral border between Heschl's gyrus

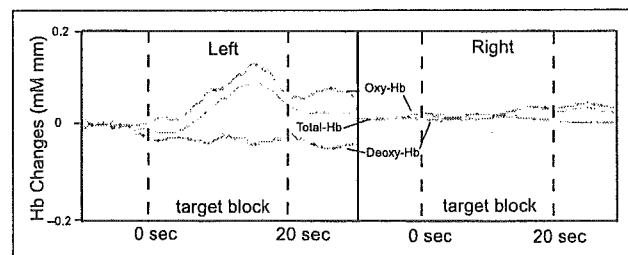


Figure 4. Time course of NIRS response in an L1 subject. The average total Hb changes in response to the phonemic contrast (BC) recorded at the left and right channels. A marked increase in Hb was observed 20 sec after the deviant stimulus was presented in the target block (vertical dotted line) in the left auditory channel.

and planum temporale (PT) in the superior temporal gyrus (STG), corresponding to BA 22 (part of Wernicke's area). The hemodynamic responses in those channels were observed to peak at 5–14 sec after the target stimulus onset (Figure 4). There was no significant difference in the peak latency for the L1 and L2 groups (L1: 9.7 sec, $SD = 3.1$; L2: 9.9 sec, $SD = 2.3$; $p = .45$, $t = 0.139$). Figure 5 shows the averaged peak responses of the total Hb to the target stimuli under the three conditions. For the Japanese subjects (L1), the across-category BC condition induced prominent responses compared with those in the within-category conditions, however, no such difference was observed for the Korean subjects (L2). A three-way analysis of variance, conducted with Conditions (AB, BC, and CD), Groups (L1 and L2), and Sides (left and right) as factors, revealed a significant Condition \times Language interaction [$F(2,84) = 7.28$, $p = .0012$]. A post hoc test (Fisher's PLSD) indicated that this interaction was due to the BC condition in the Japanese group. That is, the hemodynamic change in the BC condition was significantly larger than that in the other conditions for the L1 group (BC vs. AB: $p = .0014$; BC vs. CD: $p = .0044$), whereas no significant differences were found among the three conditions for the L2 group. It should be noted that the lack of significant brain responses for the L2 listeners was not due to a measurement failure or artifacts, because the same L2 listeners showed significant brain responses to the Japanese (L2) phonemic contrast /a–e/, which is also a distinctive contrast in their native language (Minagawa-Kawai, Mori, Sato, & Koizumi, 2004).

The lateralization observed for this BC condition in the L1 group showed a significantly larger response in the left hemisphere than in the right ($t = 2.65$, $p = .016$). The laterality index calculated from the formula $(L - R)/(L + R)$, where L and R represent the peak total Hb changes in the left and right auditory channels, respectively, was 0.31 for the BC condition in the L1 group. The same indices for the AB and CD conditions were -0.33 and 0.04 , and those for the AB, BC, and CD conditions in the L2 group were 0.03 , 0.17 ,

and -0.27 , respectively. The only index that significantly differed from zero was that for the BC condition in the L1 group ($p = .01$, zero test).

The correlation between the behavioral test and the NIRS measurements was analyzed (Figure 6). The segmental slopes of the identification function reflect the sensitivity to the durational judgment. The representative data of one Japanese subject (Figure 6A and C) shows that the slope values of identification correlated to the peak hemodynamic changes under the three conditions of AB, BC, and CD ($r = .88$, $p < .05$). Conversely, the data of a Korean subject show no such correlation ($r = -.73$, $p > .05$) (Figure 6B and D), although this subject showed a typical categorical perception (CP) for the Japanese long/short vowel contrast, judging from the identification function. A similar tendency is also observed in the correlation between the discrimination ratios and the peak hemodynamic changes (Figure 6C and D). Correlation coefficients between the slope values in the identification functions and the hemodynamic changes, calculated from all the pooled data, were $.51$ ($p < .001$) for the L1 group and $-.28$ ($p > .05$) for the L2 group. The values for the discrimination scores were 0.57 ($p < .001$) for the L1 group and 0.07 ($p > .05$) for the L2 group.

DISCUSSION

Comparison of the hemodynamic responses to the long/short Japanese vowel contrast between L1 subjects and high-proficiency L2 listeners revealed differential patterns of activation. The L1 subjects showed larger responses in the left auditory area only for the across-category condition, even though each condition contained two stimuli differing by the same length. This suggests that the linguistic cerebral representation of the Japanese durational contrast is encoded in the left auditory area, which presumably corresponds to Wernicke's area (Figure 3). In contrast, the L2 group does not have such specific neural representations as the L1 group, because this group exhibits neither

Figure 5. Averaged peak responses to durational changes for the L1 (A) and L2 (B) groups are indicated according to the three stimulus conditions (AB, BC, and CD). The responses in the left hemisphere (L, filled bars) and the right hemisphere (R, striped bars) are shown separately. Error bars indicate one standard error of the mean.

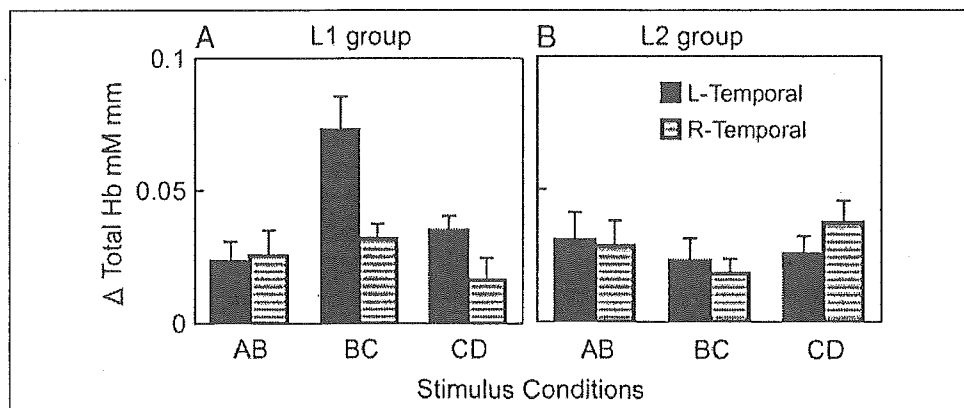
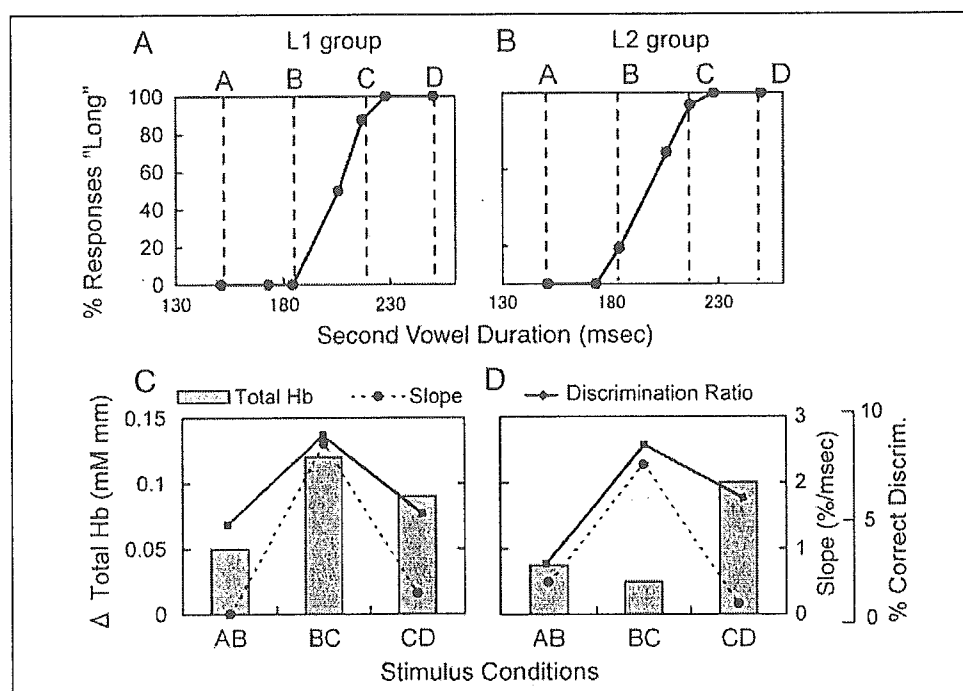


Figure 6. Correlation of behavioral tests and brain responses. Individual data of the L1 (A,C) and L2 (B,D) subjects are shown. Identification rate of /mama-/mama/ continuum and durations of the second vowel in the Stimuli A–D (dotted line) are indicated (A,B). The slopes of identification function between two stimuli for each condition and discrimination ratios (right axis) were plotted onto the total Hb responses in the left auditory area (left axis) (C,D).



category-specific nor left-dominant responses, but only small, nondifferential responses. The behavioral results of the identification and discrimination tests indicated that the L2 learners had acquired categorical percepts of Japanese long/short vowels that are indistinguishable from those of the L1 subjects. However, the correlation coefficients between the brain recording and behavior revealed inconsistency between the L1 and L2 groups. The identification slopes for the L1 group were well correlated with the peak amplitude of hemodynamic changes, whereas those for the L2 group were not. A reason for this discrepancy may lie in the response time for the behavioral test. Because the averaged latency to identify phonemic types was longer for the L2 group (586 msec) than for the L1 Japanese group (482 msec), it is likely that the L1 listeners could identify long/short vowels more promptly ($p = .039$, $t = 2.11$) with the use of linguistic neural representations encoded in the left auditory area. This was indicated by the cortical hemodynamic response exclusive to the across-category stimuli. The L2 listeners' longer latency suggests that their cortical processes for it were different and less efficient than that of the L1 subjects, due to the apparent lack of native-like phoneme representations in the Wernicke's area. These results indicate that the behavioral scores of the L2 listeners, indistinguishable from those of the L1 listeners, did not predict language-specific cerebral processing, which is recruited by the L1 listeners.

Although the hemodynamic responses to the contrastive stimuli were observed similarly as previous studies with fMRI and PET (Schall, Johnston, Todd, Ward, & Michie, 2003; Tervaniemi, Medvedev, et al., 2000), they

could result from any difference in the stimuli, such as duration and intensity in ERP and MEG studies (Todd & Michie, 2000; Tervaniemi, Lehtokoski, et al., 1999). The similar levels of responses in all conditions, except under the BC in L1 subjects in the left, are most likely due to the physical durational contrast of the same amount (33 msec) across conditions, as well as its concomitant changes in the average level. On the other hand, the timing differences of the stimulus onsets and offsets, rather than duration, should not have contributed to the results because hemodynamic measurements cannot resolve timings of less than 1 sec (Watanabe, Yamashita, Maki, Ito, & Koizumi, 1996). The increased left responses only under the BC condition as observed in the L1 subjects are attributable to experience- or language-dependent neural memory traces for phonemes of the native language (Näätänen et al., 1997), because the largest responses were evoked specifically by the contrast which had a phonemic difference and only in the L1 group. The larger and more leftward responses in the L2 subjects to one of their native phonemic contrasts (Minagawa-Kawai, Mori, Sato, et al., 2004) are consistent with this view, and together highlight the lack of phoneme-specific brain responses to the L2 contrast in this group. The peak activations found in our study were slightly superior and posterior to the lateral boundary between Heschl's gyrus and the PT in the STG, from which Jacquemot, Pallier, LeBihan, Dehaene, and Dupoux (2003) also observed strong activations associated with the processing of phonological grammar. The present results would also support the view that the PT relates to the

language- or experience-dependent phonological processing, in addition to the general acoustic processing (Jacquemot et al., 2003; Scott & Johnsrude, 2003).

Listening to stories in L1 is shown to consistently activate the left temporal lobe (Perani, Paulesu, et al., 1998; Dehaene et al., 1997; Perani, Dehaene, et al., 1996; Bottini et al., 1994). A cross-language study using PET (Perani, Paulesu, et al., 1998) showed that high-proficiency bilinguals did not differ in their cortical activation between L1 and L2 listening, regardless of their age at the time of L2 acquisition. However, differences were reported in the cortical activities of the low-proficiency L2 learners while listening to L1 and L2 stories (Perani, Dehaene, et al., 1996; Dehaene et al., 1997). In those subjects, variable activations from complete right-lateralization to weak left-lateralization were demonstrated during L2 story listening (Dehaene et al., 1997). This inconsistent lateralization pattern may be similar to that observed in our study for the L2 learners in their responses to the across-category stimuli. However, because listening to a story involves phonological, semantic, and syntactic processing plus prosodic feature decoding, the observed variable brain activations (Perani, Paulesu, et al., 1998; Dehaene et al., 1997; Perani, Dehaene, et al., 1996; Bottini et al., 1994) could have resulted from the variance in any or all of those linguistic processing levels. Although the activation at the individual levels for L1 processing has already been studied with functional brain mapping (e.g., Friederici, Ruschemeyer, Hahne, & Fiebach, 2003; Bookheimer, 2002), similar analytical work has been sparse for the L2. The present work revealed the brain activation related to a specific level of linguistic perception (i.e., L2 phonemes), which turned out to be different from that for L1. This in turn explained, at least partially, the variable brain activity regarding L2 story listening and further showed a discrepancy between brain activation and behavior. These results suggest that the neuronal substrates for L2 processing could be even more complicated and interesting by considering respective linguistic levels than shown previously.

The identification and discrimination tests, known as the CP paradigm (Strange, 1995), have been employed by cross-language behavioral researches to empirically verify the perceptual discontinuities of CP by evaluating the slope of the identification function, the phonemic boundary, and the discrimination peak. The conceptual definition of CP was that discrimination of certain speech sounds would be limited by classification (identification); thus, two different stimuli could be discriminated solely to the extent that they are classified differently (Liberman, Harris, Hoffman, & Griffith, 1957). This translates to the discrimination rate profile of pairs of sounds differing by a constant physical amount that peaks around the phonemic boundary of CP, as was observed in the present behavioral results. Although it is now widely accepted that CP

is not specific to human speech perception (Dooling, Okanoya, & Brown, 1989; Kuhl & Miller, 1975; Locke & Kellar, 1973), the identification and discrimination tests and some variations of these tests (Beddor & Gottfried, 1998) are still used as the basic methods for examining whether L2 learners acquire L2 phonemes. This is partly because CP is rarely observed in the perception of untrained L2 phonemes. However, the neuronal correlates of behavioral results given from these measures have not been thoroughly examined as yet. The present study revealed that the behavioral identification and discrimination scores that are identical to the natives do not always indicate an involvement of language-specific cerebral processes as employed by the native listeners.

In the current experiment, the CP of the L2 contrast was achieved even though L2 listeners recruited a neural circuit different from that of the L1 listeners. This is partly due to the artificial experimental situation wherein L2 listeners were required to identify phonemes in an isolated word. However, such a circuit, which requires longer latency, may not be efficient enough in processing L2 phonemes that are contained in continuous running speech in actual conversation. This may explain a fact well known to Japanese instructors, that is, distinguishing and producing Japanese long/short phonemic contrasts is challenging even for the high-proficiency learners (Minagawa-Kawai, Maekawa, et al., 2002; Toda, 1994; Han, 1992).

The current study revealed the differential relationships between the behavioral scores and cerebral representations of a Japanese vowel contrast for L1 listeners and proficient late L2 learners. The behavioral scores and the corresponding NIRS responses showed a clear correlation only in the L1 group. Along with the longer latency in identifying phonemes for the L2 group, a cerebral processing strategy that was probably different from that of the L1 group was proposed. The findings suggest that a language-specific neural processing as employed by the native listeners is only predictable when L2 listeners are assessed by neurophysiological measures in addition to the observations of the CP evaluated by the conventional behavioral tests.

METHODS

Subjects

Eight Korean subjects who had been learning Japanese (3 men and 5 women, 27–40 years, mean age: 32 years, $SD = 4.3$) and eight native speakers of standard Japanese (3 men and 5 women, 28–39 years, mean age: 31 years, $SD = 4.3$), who grew up in the Tokyo area, participated in the NIRS recording after giving their written informed consent. Some of the subjects were the same as those reported in the previous study (Minagawa-Kawai, Mori, Sato, et al., 2004; Minagawa-

Kawai, Mori, Furuya, et al., 2002). All the subjects had normal hearing and were right-handed according to the Edinburgh inventory (Oldfield, 1971). Although four more subjects participated in a pilot experiment with different stimuli, they were not included in the present subjects because cerebral activations could not be steadily measured from some channels.

The present Korean subjects were judged as high-proficiency, late-acquisition, second-language learners (bilinguals), according to the criteria employed by Perani, Paulesu, et al. (1998), Kim et al. (1997), and Perani, Dehaene, et al. (1996) in terms of language exposure, education, and residence in an L2 environment. They started learning Japanese after 24.6 years ($SD = 3.2$) on average and were highly proficient in Japanese with an average learning period of 7.6 years ($SD = 4.1$). They had been living in Japan for more than 3 years (average: 5.4 years, $SD = 2.5$) during which time they used Japanese in their daily communication. Five of the subjects were graduate students of the University of Tokyo and the others were employees of Japanese companies at the time of the experiment.

Procedures

The changes in Hb concentration and its oxygenation level in the bilateral temporal areas and infero-posterior frontal and lower parietal lobes were measured with a 24-channel NIRS system (ETG-100, Hitachi Medical, Japan). The NIRS system emits continuous near-infrared lasers whose wavelengths are approximately 780 and 830 nm, modulated at different frequencies and detected with the sharp frequency filters of lock-in amplifiers (Watanabe et al., 1996). It can assess localized cortical responses with 24 channels closely arranged two-dimensionally and recorded simultaneously. The temporal resolution of the NIRS system is 1–2 sec (Watanabe et al., 1996) after a hemodynamic latency of approximately 2 sec (Malonek & Grinvald, 1996). The recording channels exist in the brain's optical path, between the nearest pairs of emission and detection probes, which were separated by 3 cm on the scalp surface. This separation enables us to measure hemodynamic changes in the brain 2.5–3 cm deep from the head surface, which corresponds to the gray matter on the outer surface of the brain (Fukui, Ajichi, & Okađa, 2003). Five emission and four detection probes arranged in a 3 × 3 square lattice were fitted on each lateral side of the head. This resulted in a total of 12 recording channels on either side. After the optical measurement, the positions of the optical probes were recorded with a three-dimensional digitizer (Polhemus, Colchester, Vermont). The 3-D coordinates were superimposed onto T1-weighted MR brain images for each subject to identify the centers of recording sites (Figure 3). T1-weighted anatomical images were acquired in 80 contiguous axial slices with thickness of 2.0 mm using a

1.5-T scanner (Excelart, Toshiba Medical, Japan) (repetition time/echo time [TR/TE] 15/3.4 msec, flip angle [FA] 20°, matrix 256 × 192, field-of-view [FOV] 22 × 22 cm). The channels close to the lateral end of the border between the transverse temporal gyrus and the PT, as projected onto a parasagittal MRI, were presumed to be the “center” of the auditory area. This procedure selected the recording channels whose centers were within a 1.5-cm radius of the abovementioned border; thus, the channels should include the signals in the auditory cortex due to the spread of the laser in the brain tissue (Yamashita, Maki, & Koizumi, 1996). The averaged coordinates of the center of the auditory area as defined above were $x = -63.2$ ($SE = 1.4$), $y = -7.2$ (2.5), and $z = 5.0$ (1.5) mm in the left side and $x = 62.0$ (1.3), $y = -5.8$ (3.8), and $z = 3.9$ (1.6) mm in the right side (Figure 3), where positive y is to the anterior, and z is superior (coordinate system of Talairach and Tournoux). There were no statistical significant differences between the two groups.

Stimuli

Pseudoword stimuli consisting of two “ma” syllables were synthesized to obtain a low–high pitch pattern (a Japanese unaccented pattern) with a stable pitch contour and steady formant structure in the final vowel by a PARCOR (partial autocorrelation) analysis–resynthesis procedure (Markel & Gray, 1976) with Onsei-Rokubunken (Imagawa & Kiritani, 1989; Datel, Japan: www.datel.co.jp/onsei/). The original speech source for the analysis–synthesis was recorded by a female Japanese speaker. The long/short target vowel was set at the final position, where nonnative speakers find it most difficult to identify phonemic length (Minagawa-Kawai, Maekawa, et al., 2002). The first syllable was 110 msec and intervocalic /m/ was 90 msec. The acoustical parameters including pitch and duration were determined from the results of a word production experiment performed beforehand with seven native standard Japanese speakers. The averaged segment durations of /mama/ and /mama:/ words were as follows: the first syllable /ma/ = 113 msec, the second consonant /m/ = 92 msec, the final short vowel /a/ = 129 msec, and the final long vowel /a:/ = 313 msec. The duration of the final syllable is set to more than twice of that of the first syllable, because the word-final syllable are 1.5–2.5 times longer in duration than the nonfinal ones universally in any languages (Hoequist, 1983; Klatt, 1976), including Japanese (Minagawa-Kawai, Kagomiya, & Maekawa, 2003), particularly in an isolated word context. Four stimuli (A–D) differing in the final vowel duration in 33-msec steps were selected from the /mama/ to /mama:/ continuum. The durations of the final vowels were 151 msec (A), 184 msec (B), 217 msec (C), and 250 msec (D). Because a pilot behavioral experiment showed that the phonemic boundary of long and short

vowels was located between Stimuli B and C, the across-category condition included the B and C stimuli, and the two within-category conditions included A and B, and C and D stimulus pairs. In Session AB (a within-category condition), Stimulus A was repeated for 20 sec as a baseline block. Then, Stimuli A and B were presented in a pseudorandom order with equal probabilities of another 20-sec period as a target block with an SOA of 1.25 sec. The baseline and target blocks were alternated and repeated at least five times. Similar procedures were carried out for Sessions BC (an across-category condition) and CD (a within-category condition). These three sessions and the two extra sessions for the other experiments were assigned to each subject in a random order. The subject seated in a chair was instructed to listen to the stimuli delivered through insert earphones (EarTone 3A) at a comfortable level (≈ 70 dB SPL).

Behavioral Test

The subjects participated in identification and discrimination tests of Japanese long/short vowel categories after the NIRS recording. The stimuli used in the identification test were the same pseudowords A–D as that in the NIRS recording and the three extra stimuli whose second vowels were 173, 206, and 228 msec in duration. The subjects were instructed to listen to these stimuli repeated 16 times each (Jamieson & Morosan, 1986; Fujisaki, Nakamura, & Imoto, 1985) in random order played back by a PC through a pair of headphones. They were asked to decide whether the second vowel was a long vowel or not. In the discrimination test, the two stimuli, whose final vowel duration differed by 11 msec, were presented sequentially in each trial (interstimulus interval = 500 msec), and the subjects were asked to judge which of the two words had a longer final vowel. The durations of the final vowels for the pseudoword pair used were 151 msec (A) versus 162 msec, 184 msec (B) versus 195 msec, 206 msec versus 217 msec (C), and 239 msec versus 250 msec (D). These were repeated 14 times in random order (Fujisaki et al., 1985; Jamieson, & Morosan, 1986), and the stimulus order within each pair was balanced to avoid a positional effect. The response time was measured in both tests. The PB between the long and short vowels was defined as the vowel length, where the fitted identification curve crossed 50% of the responses “long.” The curve was fitted by the least square method.

Data Processing

The concentrations of oxygenated, deoxygenated, and total Hb were calculated from the absorbance changes of 780 and 830 nm laser beams sampled at 10 Hz. After discarding the blocks with artifacts, the Hb concentrations of the remaining blocks were averaged five or six times synchronously to the target blocks and smoothed

with a 5-sec moving average. The response peaks of the averaged target blocks were measured against the 10-sec pre-stimulus period.

Acknowledgments

We thank I. Furuya, R. Hayashi, and T. Koizumi for comments and assistance, and T. Kagomiya for his technical assistance. This work was supported by a grant from the Ministry of Health Welfare and Labor of Japan (H10-Kankaku-011 and H13-Kankakuki-017, K. M.) and CREST of Japan Science and Technology Agency (Y. M.).

Reprint requests should be sent to Yasuyo Minagawa-Kawai, Keio University, Centre for Integrated Research on the Mind, 3-1-7 Mita, Minato-ku, Tokyo 108-0073, Japan, or via e-mail: myasuyo@bea.hi-ho.ne.jp.

REFERENCES

- Aslin, R. N., & Pisoni, D. B. (1980). Some developmental processes in speech perception. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson, (Eds.), *Child phonology* (Vol. 2, pp. 67–96). New York: Academic Press.
- Beddor, P. S., & Gottfried, T. L. (1995). Methodological issues in cross-language speech perception research with adults. In W. Strange (Eds.), *Speech perception and linguistic experience* (pp. 207–232). Baltimore: York Press.
- Bookheimer, S. (2002). Functional MRI of language: New approaches to understanding the cortical organization of semantic processing. *Annual Review of Neuroscience*, 25, 151–188.
- Bottini, G., Corcoran, R., Sterzi, R., Paulesu, E., Schenone, P., Scarpa, P., Frackowiak, R. S., & Frith, C. D. (1994). The role of the right hemisphere in the interpretation of figurative aspects of language. A positron emission tomography activation study. *Brain*, 117, 1241–1253.
- Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/. IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America*, 101, 2299–2310.
- Callan, D. E., Tajima, K., Callan, A. M., Kubo, R., Masaki, S., & Akahane-Yamada, R. (2003). Learning-induced neural plasticity associated with improved identification performance after training of a difficult second-language phonetic contrast. *Neuroimage*, 19, 113–124.
- Chee, M. W. L., Soon, C. S., Lee, H. L., & Pallier, C. (2004). Left insula activation: A marker for language attainment in bilinguals. *Proceedings of the National Academy of Sciences, U.S.A.*, 101, 15265–15270.
- Cheour, M., Shestakova, A., Alku, P., Ceponiene, R., & Näätänen, R. (2002). Mismatch negativity shows that 3–6-year-old children can learn to discriminate non-native speech sounds within two months. *Neuroscience Letters*, 325, 187–190.
- Dehaene, S., Dupoux, E., Mehler, J., Cohen, L., Paulesu, E., Perani, D., van de Moortele, P. F., Lehericy, S., & Le Bihan, D. (1997). Anatomical variability in the cortical representation of first and second language. *NeuroReport*, 8, 3809–3815.
- Dehaene-Lambertz, G. (1997). Electrophysiological correlates of categorical phoneme perception in adults. *NeuroReport*, 8, 919–924.
- Dooling, R. J., Okanoya, K., & Brown, S. D. (1989). Speech perception by budgerigars (*Melopsittacus undulatus*): The voiced–voiceless distinction. *Perception & Psychophysics*, 46, 65–71.

- Friederici, A., Ruschemeyer, S., Hahne, A., & Fiebach, C. (2003). The role of left inferior frontal and superior temporal cortex in sentence comprehension: Localizing syntactic and semantic processes. *Cerebral Cortex*, *13*, 170–177.
- Fujisaki, H., Nakamura, K., & Imoto, T. (1985). Auditory perception of speech and non-speech stimuli. In G. Fant & M. A. A. Tatham (Eds.), *Auditory analysis and perception of speech* (pp. 197–219). London: Academic Press.
- Fukui, Y., Ajichi, Y., & Okada, E. (2003). Monte Carlo prediction of near-infrared light propagation in realistic adult and neonatal head models. *Applied Optics*, *42*, 2881–2887.
- Furuya, I., & Mori, K. (2003). Cerebral lateralization in spoken language processing measured by multi-channel near-infrared spectroscopy (NIRS). *Brain and Nerve*, *55*, 226–231.
- Gandour, J., Wong, D., Lowe, M., Dzemidzic, M., Saththamnuwong, N., Long, Y., & Lurito, J. (2002). Neural circuitry underlying perception of duration depends on language experience. *Brain and Language*, *83*, 268–290.
- Han, M. (1992). The timing control of geminate and single stop consonants in Japanese: A challenge for non-native speakers. *Phonetica*, *49*, 102–127.
- Hoequist, C. (1983). Syllable duration in stress-, syllable- and mora-timed languages. *Phonetica*, *40*, 203–237.
- Imaizumi, S., Itoh, H., Tamekawa, Y., Deguchi, T., & Mori, K. (1998). Plasticity of non-native phonetic perception and production: A training study. *Proceedings of the 5th International Conference on Spoken Language Processing (ICSLP98), Sydney Australia*, *6*, 1887–1890.
- Imagawa, H., & Kiritani, S. (1989). High speed speech analysis system using a personal computer with DSP and its application to pronunciation training. *Annual Bulletin of Research Institute of Logopedics and Phoniatics, University of Tokyo*, *23*, 175–164.
- Jacquemot, C., Pallier, C., LeBihan, D., Dehaene, S., & Dupoux, E. (2003). Phonological grammar shapes the auditory cortex: A functional magnetic resonance imaging study. *Journal of Neuroscience*, *23*, 9541–9546.
- Jamieson, D. G., & Morosan, D. E. (1986). Training non-native speech contrasts in adults: Acquisition of the English /delta-/theta/ contrast by francophones. *Perception & Psychophysics*, *40*, 205–215.
- Jusczyk, P. W. (1997). *The discovery of spoken language*. Cambridge: MIT Press.
- Kim, K. H., Relkin, N. R., Lee, K. M., & Hirsch, J. (1997). Distinct cortical areas associated with native and second languages. *Nature*, *388*, 171–174.
- Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of Acoustical Society of America*, *59*, 1574–1585.
- Kuhl, P. K., & Miller, J. D. (1975). Speech perception by the chinchilla: Voiced–voiceless distinction in alveolar plosive consonants. *Science*, *190*, 69–72.
- Kuhl, P. K., Tsao, F. M., & Liu, H. M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences, U.S.A.*, *100*, 9096–9101.
- Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, *255*, 606–608.
- Lieberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*, 358–368.
- Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/: II. The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America*, *94*, 1242–1255.
- Locke, S., & Kellar, L. (1973). Categorical perception in a non-linguistic mode. *Cortex*, *9*, 355–369.
- Logan, J. S., Lively, S. E., Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America*, *89*, 874–886.
- Malonek, D., & Grinvald, A. (1996). Interactions between electrical activity and cortical microcirculation revealed by imaging spectroscopy: Implications for functional brain mapping. *Science*, *272*, 551–554.
- Markel, J. D., & Gray, A. H., Jr. (1976). *Linear prediction of speech*. Berlin: Springer Verlag.
- McClelland, J. L., Fiez, J. A., & McCandliss, B. D. (2002). Teaching the /r/-/l/ discrimination to Japanese adults: Behavioral and neural aspects. *Physiology and Behavior*, *77*, 657–662.
- Menning, H., Imaizumi, S., Zwitserlood, P., & Pantev, C. (2002). Plasticity of the human auditory cortex induced by discrimination learning of non-native, mora-timed contrasts of the Japanese language. *Learning and Memory*, *9*, 253–267.
- Minagawa-Kawai, Y. (2000). Perception of vowel duration in native and non-native languages. Doctoral dissertation submitted to University of Tokyo.
- Minagawa-Kawai, Y., Kagomiya, T., & Maekawa, K. (2003). Durational variations of Japanese long/short vowels in different speaking rates: Analysis of a spontaneous speech corpus. *Proceedings of International Conference of Phonetic Science 03, Barcelona*, 2127–2730.
- Minagawa-Kawai, Y., Maekawa, K., & Kiritani, S. (2002). Effects of pitch accent and syllable position in identifying Japanese long and short vowels: Comparison of English and Korean speakers. *Journal of the Phonetic Society of Japan*, *6*, 88–97.
- Minagawa-Kawai, Y., Mori, K., Furuya, I., Hayashi, R., & Sato, Y. (2002). Assessing cerebral representations of short and long vowel categories by NIRS. *NeuroReport*, *13*, 581–584.
- Minagawa-Kawai, Y., Mori, K., Sato, Y., & Koizumi, T. (2004). Differential cortical responses in second language learners to different vowel contrasts. *NeuroReport*, *15*, 899–903.
- Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. L., Jenkins, J. J., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics*, *18*, 331–340.
- Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R. J., Luu, A., Allik, J., Sinkkonen, J., & Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, *385*, 432–434.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*, 97–113.
- Perani, D., Dehaene, S., Grassi, F., Cohen, L., Cappa, S. F., Dupoux, E., Fazio, F., & Mehler, J. (1996). Brain processing of native and foreign languages. *NeuroReport*, *7*, 2439–2444.
- Perani, D., Paulesu, E., Galles, N. S., Dupoux, E., Dehaene, S., Bettinardi, V., Cappa, S. F., Fazio, F., & Mehler, J. (1998). The bilingual brain: Proficiency and age of acquisition of the second language. *Brain*, *121*, 1841–1852.

- Rivera-Gaxiola, M., Csibra, G., Johnson, M. H., & Karmiloff-Smith, A. (2000). Electrophysiological correlates of cross-linguistic speech perception in native English speakers. *Behavioural Brain Research, 111*, 13–23.
- Schall, U., Johnston, P., Todd, J., Ward, P. B., & Michie, P. T. (2003). Functional neuroanatomy of auditory mismatch processing: An event-related fMRI study of duration-deviant oddballs. *Neuroimage, 20*, 729–736.
- Scott, S. K., & Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception. *Trends in Neurosciences, 26*, 100–107.
- Sharma, A., & Dorman, M. F. (2000). Neurophysiological correlates of cross-language phonetic perception. *Journal of the Acoustical Society of America, 107*, 2697–2703.
- Simos, P. G., Diehl, R. L., Breier, J. I., Molis, M. R., Zouridakis, G., & Papanicolaou, A. C. (1998). MEG correlates of categorical perception of a voice onset time continuum in humans. *Cognitive Brain Research, 7*, 215–219.
- Strange, W. (1995). Cross-language studies of speech perception: A historical review. In W. Strange (Eds.), *Speech perception and linguistic experience* (pp. 3–45). Baltimore: York Press.
- Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception & Psychophysics, 36*, 131–145.
- Takagi, N. (2002). The limits of training Japanese listeners to identify English /r/ and /l/: Eight case studies. *Journal of the Acoustical Society of America, 111*, 2887–2896.
- Tervaniemi, M., Lehtokoski, A., Sinkkonen, J., Virtanen, J., Ilmoniemi, R. J., & Näätänen, R. (1999). Test-retest reliability of mismatch negativity for duration, frequency and intensity changes. *Clinical Neurophysiology, 110*, 1388–1393.
- Tervaniemi, M., Medvedev, S. V., Alho, K., Pakhomov, S. V., Roudas, M. S., Van Zuijen, T. L., & Näätänen, R. (2000). Lateralized automatic auditory processing of phonetic versus musical information: A PET study. *Human Brain Mapping, 10*, 74–79.
- Toda, T. (1994). Interlanguage phonology: Acquisition of timing control in Japanese. *Australian Review of Applied Linguistics, 17*, 51–76.
- Todd, J., & Michie, P. T. (2000). Do perceived loudness cues contribute to duration mismatch negativity (MMN)? *NeuroReport, 11*, 3771–3774.
- Tremblay, K. L., & Kraus, N. (2002). Auditory training induces asymmetrical changes in cortical neural activity. *Journal of Speech Language and Hearing Research, 45*, 564–572.
- Watanabe, E., Yamashita, Y., Maki, A., Ito, Y., & Koizumi, H. (1996). Non-invasive functional mapping with multi-channel near infra-red spectroscopic topography in humans. *Neuroscience Letters, 205*, 41–44.
- Winkler, I., Kujala, T., Tiitinen, H., Sivenon, P., Alku, P., Lehtokoski, A., Czigler, I., Csepe, V., Ilmoniemi, R. J., & Näätänen, R. (1999). Brain responses reveal the learning of foreign language phonemes. *Psychophysiology, 36*, 638–642.
- Yamada, R. A., & Tohkura, Y. (1992). The effects of experimental variables on the perception of American English /r/ and /l/ by Japanese listeners. *Perception & Psychophysics, 52*, 376–392.
- Yamashita, Y., Maki, A., & Koizumi, H. (1996). Near-infrared topographic measurement system: Imaging of absorbers localized in a scattering medium. *Review of Scientific Instruments, 67*, 730–732.

耳掛け型メトロノームを用いた吃音訓練
——成人吃音者を対象に——

酒井奈緒美・森 浩一・小澤 恵美・餅田亜希子

音声言語医学 Vol. 47, No. 1 別刷

(2006年1月20日発行)

原 著

耳掛け型メトロノームを用いた吃音訓練

——成人吃音者を対象に——

酒井奈緒美¹⁾ 森 浩一¹⁾ 小澤 恵美²⁾ 餅田亜希子²⁾

要約: 吃音者がメトロノームに合わせて発話すると、流暢に話せることが知られている。その現象を利用し、多くの訓練のなかでメトロノームが利用されてきた。しかしその効果は日常生活へと般化しづらいものであった。そこでわれわれは国内で初めてプログラム式耳掛け型メトロノームを開発し、日常生活場面において成人吃音者へ適用した。耳掛け型メトロノームは、毎分6~200の間でテンポを設定でき、ユーザーによる微調整も可能である。また音量は20~90 dB SPLの間の任意の2点を設定でき、ユーザーが装用中に切り替え可能である。1症例に対し約3ヵ月半、発話が困難な電話場面において適用したところ、電話場面と訓練室場面において吃症状の減少が認められた。本症例は発話が困難な電話場면을避ける傾向にあったが、耳掛け型メトロノームの導入により、積極的に電話ができるようになった。また自己評価の結果から、症例自身は吃頻度以外の面での改善を高く評価していることも認められた。

索引用語: 成人, 吃音訓練, 耳掛け型メトロノーム, 日常生活への般化

Stuttering Treatment with a Behind-the-Ear Type Metronome for an Adult Stutterer

Naomi Sakai¹⁾, Koichi Mori¹⁾, Emi Ozawa²⁾ and Akiko Mochida²⁾

Abstract: Most stutterers can speak fluently when they time their speech to the rhythmic beat of a metronome. Although metronomes have been widely used for the treatment of stuttering, their effect is reportedly difficult to transfer to daily life. We devised a miniaturized portable electronic metronome that can be worn behind the ear, and applied it to an adult who stutters. The metronome can be programmed for a beat rate from 6 to 200 tone pips per minute, around which the user can adjust the rate up to ± 10 or $\pm 20\%$. The device has a switch enabling the user to set two levels of loudness programmed from 20 to 90 dB SPL. Our adult stutterer wore the device when she used the telephone during a period of three-and-a-half months. A reduction in the frequency of disfluency in her speech was observed not only during phone calls but also in the therapy room. The subject had earlier exhibited a tendency to avoid the telephone, but became able to call actively with use of the behind-the-ear metronome. She evaluated the device positively because she felt her stuttering improved in terms of naturalness of speech and reduced avoidance of

国立身体障害者リハビリテーションセンター研究所¹⁾, 同病院²⁾: 〒359-8555 埼玉県所沢市並木4-1

¹⁾Research Institute, ²⁾Hospital of National Rehabilitation Center for Persons with Disabilities: 4-1 Namiki, Tokorozawa, Saitama 359-8555

2005年7月20日受稿 2005年9月6日受理

verbal interaction.

Key words: adult, stuttering treatment, behind-the-ear type metronome, daily life

はじめに

成人吃音者の多くは、メトロノームのような律動的なビートに合わせて発話することで、流暢に話せることが知られている¹⁻³⁾。このメトロノーム効果の原理についてはいくつかの解釈がある。1つは吃音をリズムの崩壊と捉え、吃音者がメトロノームの律動的な刺激に合わせて発話した場合、崩壊していたリズムが回復するという解釈である³⁾。またメトロノームに合わせて発話を音節に分節化して話すことは、発話の運動プログラミングが簡単になることを意味し、その結果発話が流暢になるという解釈もある⁴⁾。他には、メトロノームの外的な刺激が、発話開始の合図を送るキューとしての役割を担っており、そのキューに合わせて発話することで流暢になることもいわれている⁵⁾。このような効果を利用し、メトロノームのクリック音に1モーラや2モーラの音、あるいは単語を合わせて発話する方法が、成人吃音の訓練に使われてきた^{4,6)}。

しかしながら、このような訓練の効果は、日常生活場面のようなメトロノームの刺激音のない場面には般化しづらいという問題があり、そのためメトロノーム原理の訓練への導入は限られたものとなっていた⁴⁾。そこで、電子工学の技術の発展により、訓練室外のさまざまな場面でのメトロノーム使用を可能とする、小型で携帯可能な耳掛け型メトロノームが開発されてきた^{4,6)}。また同時に行動学の発展により、さまざまな場面へメトロノームの使用を拡大することで、訓練室において身につけた流暢性の系統的な般化をねらう行動プログラムも作成されてきた⁴⁾。これにより訓練室と変わらない訓練条件が日常生活において提示されることとなった。

この耳掛け型メトロノームを取り入れた訓練プログラムについては、海外において詳細に示されている^{4,7)}。このような訓練が行われるなかで、その効果について報告がなされてきたものの、それは主観的な報告^{8,9)}や日常生活場面での発話データを示していないものが多く⁴⁾、効果を定量的に示した研究はほとんどないのが現状である¹⁰⁾。また日本では、メトロノームを用いた訓練についての報告はあるものの⁵⁾、携帯できる耳掛け型のメトロノームが存在しなかった。

そこで本研究では、わが国で耳掛け型メトロノーム

を開発し、それを1症例の日常生活場面での訓練に導入したところ、吃症状の改善が認められたので報告する。

方 法

1. 装置

プログラム式の耳掛け型メトロノームを使用した。これは、われわれが厚生労働省の補助金により開発したものである(図1)。

1) 設計方針・仕様

速度や音量等のパラメーターは、先行研究^{4,9,11)}において訓練室あるいは日常生活場面で使用されている値に、若干の余裕をもたせた範囲を設定した。また音量については、夜間の静かな室内からある程度騒音の多い場所でも対応できるものとし、かつ音響外傷を起こさせない範囲を設定した。クリック音については、同一周波数音が続く単調な印象を避けるため、2つの周波数による2拍子を設定できるようにした。さらにコスト面に配慮し、通常の補聴器の外形と補聴器用DSPチップを採用し、補聴器メーカーに生産委託した。

2) プログラム設定

設定は補聴器と同様、HI-PROのインターフェースと接続ケーブルを介してパソコンより行う。パソコンの設定画面を図2に示す。毎分6~200クリックの範囲でテンポ設定が可能である。そして使用者が発話状況により変化させうるテンポの変化幅も±10ないし±

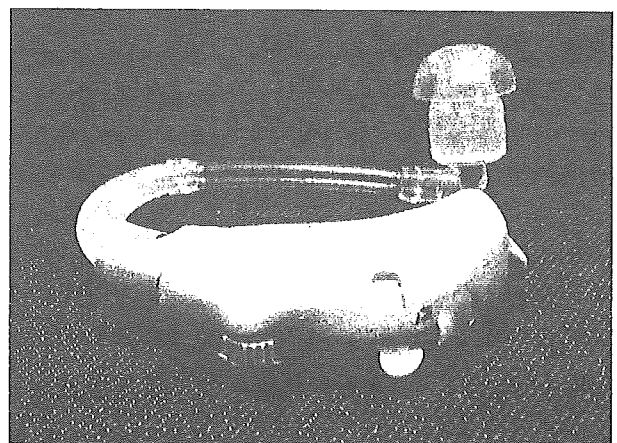


図1 耳掛け型メトロノーム

左側のボリュームによって、ユーザーが速度を変えられる。右側のスイッチにより、音量を2点から選択できる。

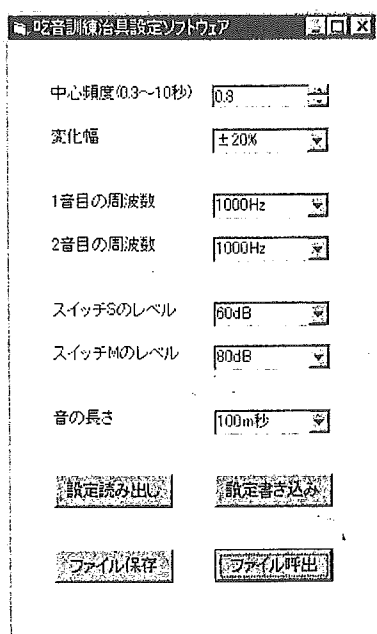


図2 コンピューターのプログラム設定画面

20%に設定できる。音量は20~90 dB SPLの範囲(10 dBステップ)の任意の2点を設定できるようになっている。またクリック音の周波数は、500~4000 Hzの範囲から500 Hzステップで2つの音を選択・設定できるようになっている。その2つの音の周波数を異なるものとした場合、2拍子の設定となる。

3) ユーザー操作

テンポは、ボリュームを操作することで設定した変化幅の範囲内で変えられる。音量は、電源スイッチと共用の切り替えスイッチで、設定した2点いずれかを選択できる。どちらも使用者が随時、耳に掛けたまま変更可能である。

2. 手続き

1) 適用者の選択

訓練室内での吃症状は改善傾向を示しているものの、日常生活のある特定の場面での発話が困難な者、つまり日常生活場面での訓練が進まない者、あるいは訓練効果が日常生活場面に般化しづらい者を対象とする。

2) 使用場面・時間

使用場面は、発話が困難で、かつ使用者が実際に使用可能な場面を選択する。使用時間は基本的に最低1日合計15分以上とするが、使用者が負担を感じる際は、必ずしもそれに限らないこととする。

3) メトロノームの設定

使用者が使いやすいと感じる数値に設定する。

4) データ収集と分析

訓練場面と日常生活場面における発話を録音記録し、発話データを収集する。そのなかから吃症状(吃音者に特有の症状)とその他の非流暢性を抽出し、非流暢性総数/発話総文節数×100の式から非流暢性頻度を算出する¹²⁾。また、毎日の吃症状に関する自己評価として、メトロノーム装用時と非装用時それぞれの吃音重症度について、7段階(1. 全く吃らない~4. 文ごとに吃る~7. 非常にひどく吃る)の評価を依頼する。そして訓練時ごとに自身の吃音状態、また耳掛け型メトロノームの効果について、それぞれの質問紙(表1, 2)を用いた7段階評価を依頼する。

症 例

女性。訓練開始時年齢は22歳。主訴は吃音である。初回面接時から1年5ヵ月の間に、全19回の訓練を行っている。訓練開始前に実施した吃音検査<試案1>¹²⁾では、文章音読時に吃頻度が最も少なく20%、単語呼称や情報聴取では最も多く60%の吃頻度が観察された。全訓練経過を通しての目標は、柔らかい起声とともに発話速度を落とし、力の入らない楽な吃り方を目指すというものである。7回目以降卓上型メトロノームを導入し、1クリック音に1モーラあるいは2モーラの音を合わせて発話する訓練を行った。17回目以降からは耳掛け型メトロノームに切り替え、7回目以降と同様の訓練に加えて、ことばが詰まったときにメトロノームのクリック音を発話の合図として使用するという訓練を行った(図3)。

耳掛け型メトロノーム導入直前に再度吃音検査を実施したところ、音読課題や文による絵の説明課題における吃症状の出現頻度が最も少なく0~5%程度であった一方、モノログや情報聴取という日常会話に近い状況では20%程度(非流暢性頻度は30%)となり、比較的高頻度で吃音症状が観察された。訓練開始前の結果と比べると、症例は卓上型メトロノームでの訓練がある程度進んでおり、短い発話や音読では比較的症状が減少してきていることが認められるが、日常生活における会話に近い状況ではまだ発話が困難な場面が存在していることが示された。症例本人は電話場面での発話を困難と感じており、電話を避ける傾向にあったため、電話場面において携帯型メトロノームの適用があると判定された。本症例が実際に携帯型メトロノームを使用した場面は、電話場面(週に1~3回)に加え、家族との会話場面や職場における接客場面であった。

表1 吃音に関する自己評価質問紙

氏名	吃音の自己評価表							ID			
								記入日	年	月	日
あなたの過去1ヶ月*をふり返り、以下の質問に答えて下さい。											
各質問は、あなたの状態について、質問の下に書かれた線の上に○をつけて答えるようになっています。例えば(1)の質問については、文ごとに食べるくらいを中程度と考え、その程度の吃音頻度の場合は数字の4辺りに、それよりは多少良い状態であったと思う場合は、その程度によって4より左側の線の上(数字の2や3の周辺)に○をつけて下さい。数字は目安であり、数字に○をつける必要はありません。それぞれの質問について、あなたの状態を線上に表した場合、どの程度に相当するかを答えて下さい。											
(1)	この1ヶ月の生活の中で、どのくらい頻繁に吃りましたか										
	1	2	3	4	5	6	7				
	全く吃らない						非常に頻繁に吃った				
(2)	この1ヶ月の生活の中で、話すことや、ある場面/状況を避けようとするがありましたか										
	1	2	3	4	5	6	7				
	全く避けなかった/ 避けたいと思わなかった						可能な時は常に避けた				
(3)	この1ヶ月の自分の話し方について、どの程度自然だと感じますか										
	1	2	3	4	5	6	7				
	非常に不自然						非常に自然				
(4)	この1ヶ月の中で電話を使用した時、あなたはどのくらい頻繁に吃りましたか										
	1	2	3	4	5	6	7				
	全く吃らない						非常に頻繁に吃る				
(5)	この1ヶ月の生活の中で、どのくらい積極的に電話を使用しましたか										
	1	2	3	4	5	6	7				
	全く使用しなかった						非常に積極的に使用した				
(6)	この1ヶ月で人と会話した時、あなたはどのくらい頻繁に吃りましたか										
	1	2	3	4	5	6	7				
	全く吃らない						非常に頻繁に吃る				

*ここでの「過去1ヶ月」という表現は、前回来院時からの経過期間である。実際に実施する質問紙は、その都度、前回来院時からの経過期間に書き直したものとする。

本症例の卓上型メトロノームのテンポ設定は、訓練時ごとに症例の希望に合わせて行い、毎分80~100クリックとした。その後の耳掛け型メトロノームの設定は、テンポを毎分75クリック(変化幅±20%)、音量を60dBと80dBに設定した。2ヵ月半後、本人の要望によりテンポを毎分66クリックに変更した。

結 果

1. 訓練室内における耳掛け型メトロノームの効果
毎訓練時にメトロノームなし→あり→なしの順で、セラピストへ日常生活のコミュニケーションについての報告を求め、その発話100文節ごとに非流暢性頻度を算出した。その結果を図4に示す。話の内容が深刻だった第67日目を除いて、それぞれの日はメトロノーム装用時に吃頻度の減少が認められた。そしてメトロ

ノーム装用/非装用時ともに、経過に沿って非流暢性発話頻度の減少が認められ、特にテンポ設定変更後に大きな減少が認められた($p < 0.01$, Mann-Whitney U-test)。さらに非流暢性発話の内訳においては、その他の非流暢性の割合が減少していく傾向が認められた($p < 0.05$, Kruskal Wallis H-test)。

2. 電話場面における耳掛け型メトロノームの効果
耳掛け型メトロノーム装用時の電話場面における発話を録音記録・分析し、非流暢性頻度を算出した。それぞれの日の電話場面における発話総文節数は、13~100の範囲の値をとり、平均は65文節、標準偏差は30.0となったため、極端に発話文節数の少ない71日(文節数13)と92日(文節数24)は除外し38文節/日以上を分析することとした。非流暢性頻度の推移を図5に示す。装用開始直後の2回をメトロノームなし

表2 耳掛け型メトロノームに関する自己評価質問紙

治療機器についての自己評価表

ID _____

氏名 _____ 記入日 _____ 年 _____ 月 _____ 日

あなたの過去1ヶ月*をふり返り、以下の質問に答えて下さい。

各質問は、あなたの状態について、質問の下に書かれた線の上に○をつけて答えるようになっています。例えば(1)の質問に対し、ほとんど変化がなかったということであれば数字の4辺りに、それよりは多少良い状態に変化したと思う場合は、その程度によって4より右側の線の上(数字の5や6の周辺)に○をつけて下さい。数字は目安であり、数字に○をつける必要はありません。それぞれの質問について、あなたの状態を線上に表した場合、どの程度に相当するかを答えて下さい。

(1) 機器を使用してみて、話す時の苦しさに変化がありましたか

1 _____ 2 _____ 3 _____ 4 _____ 5 _____ 6 _____ 7 _____

非常に苦しくなった 非常に楽になった

(2) 機器を使用してみて、これまでのあなたの吃音の症状は変化しましたか

1 _____ 2 _____ 3 _____ 4 _____ 5 _____ 6 _____ 7 _____

非常に重度になった 非常に軽度になった

(3) 機器を使用することで、話すときのスピードをコントロールできる感じがありましたか

1 _____ 2 _____ 3 _____ 4 _____ 5 _____ 6 _____ 7 _____

全くコントロールできる 非常にコントロールできる
感じがしない 感じがする

(4) この機器に満足していますか

1 _____ 2 _____ 3 _____ 4 _____ 5 _____ 6 _____ 7 _____

非常に不満がある 非常に満足している

(5) (4)で答えた不満・満足について、具体的な内容を以下に記入して下さい

*ここでの「過去1ヶ月」という表現は、前回来院時からの経過期間である。実際に実施する質問紙は、その都度、前回来院時からの経過期間に書き直したものとす。

の条件で録音し、その発話データをベースラインとした。ベースラインの非流暢性発話頻度は25%弱であった。その後、耳掛け型メトロノームを連日装用している間、一時的に非流暢性発話頻度は上昇し、装用を開始してから2ヵ月間は20~35%程度の頻度を示していた。しかし装用開始2ヵ月半以降からは、非流暢性発話頻度は多くてもベースラインと同等の20%程度となり、少ないときには5%程度と減少が認められた。なお吃症状の減少傾向が認められた時期は、テンポ設定変更とほぼ同時期であるが厳密には一致していなかった。

さらに、吃症状の内訳の推移について図6に示す。引き伸ばしの頻度については、装用日数が経過してもほとんど変化が認められなかった。その一方、繰り返しの頻度については、初頭音・音節2回以上の繰り返しが77日以降から認められなくなり、さらに後半では初頭音・音節1回の繰り返しも認められないことが多かつ

た。またブロック症状(発声の阻止)は77日目以降において、その他の吃症状(強勢、歪み、とぎれ)は40日目以降において認められないようになり、症状の軽快化がうかがわれた。

3. メトロノーム装用/非装用時の吃症状についての自己評価

毎日の吃症状についての自己評価の結果を図7に示す。装用開始直後は、装用時・非装用時ともに症状は同程度に重いと判定していたが、徐々に装用時の症状が軽いと判定し、その傾向は後半明らかとなった。このことから、自己評価においても耳掛け型メトロノームの有効性が認められた。

4. 吃音・装置に関する自己評価

自身の吃音状態については、耳掛け型メトロノーム装用開始前と装用開始3ヵ月半後の結果を表3に、装置の効果については装用開始3ヵ月半後の結果を表4に示す。自身の吃音についての自己評価の結果からは、

日付	X+1年																					
	5月	6月	8月	9月	9月①	9月②	10月①	10月②	11月①	11月②	12月	1月	2月	3月	4月	5月	6月	7月①	7月②	9月	11月	
指導回数																						
メトロノーム使用経過																						
卓上型メトロノーム使用																						
耳掛け型メトロノーム使用																						
吃音検査	*																					
リラクゼーション体操																						
日常コミュニケーション報告																						
単語		*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
文		*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
文章																						
自由会話																						
自己観察																						
ロールプレイ																						
軟起声																						

図3 訓練経過とメトロノーム装用経過
は実施を示す。なお(M)は卓上型メトロノームを使用して実施、*(EM)は耳掛け型メトロノームを使用して実施。

装用前と比べて、電話の使用頻度が急激に増加していることが示された。また回避や自然さという項目についても、良い方向へ変化したことが認められた。一方、さまざまな場面での吃頻度に関しては、訓練期間中あまり変化がないと評価していた。

また装置の効果については、苦しさの変化、吃症状

の変化、発話速度のコントロール、装置への満足度という項目すべてにおいて装置の効果を高く評価していた。

考 察

耳掛け型メトロノームを開発し、実際の日常生活場面に導入して訓練を行った。発話が困難な電話場面を避ける傾向にあった1症例に適用したところ、自己評価の結果から積極的に電話ができるようになったことが示された。そして電話場面における実際の吃症状も本研究の終盤では減少・改善傾向を示した。このことから、耳掛け型メトロノームは日常生活における発話困難場面での訓練の動機付けとなる可能性、また実際の発話困難場面での吃症状の減少・改善に有効である可能性が示唆された。

しかし、客観的評価において吃頻度の減少が認められているものの、症例の自己評価では、さまざまな場面における吃頻度には変化が認められないという結果となった。このような客観的評価と主観的評価の乖離が認められた要因としては、客観的データで示された吃頻度の減少が安定しておらず、本研究の後半にいたっても吃頻度がベースラインと同等の数値を示す日があったことが挙げられる。また装用期間が長期にわたったため、自己評価の判断基準が変化した可能性も考えられる。あるいは、主観的評価がメトロノーム装用時・非装用時の両方を対象とした結果であるのに対し、客観的データは装用時のみを対象としたものであったため、結果の乖離はメトロノーム非装用時の吃頻度に改善が認められていないことを反映している可能性も考えられた。本人が吃頻度の減少を感じていないこと、また客観的にも吃頻度の減少が安定していないことから、本症例については、今後も装用を続け、吃頻度の減少の安定を目指し、また非装用時にも改善が般化するのかが指導を続けながら経過を追っていきたい。

その一方で、本症例はメトロノームの効果を吃音の頻度以外の側面で高く評価していた。回避傾向の減少、吃症状の軽快化、発話の苦しみの軽減、発話速度のコントロール能力の増大を感じており、それを耳掛け型メトロノームの効果として高く評価している。この回避傾向の減少や発話の苦しみの軽減は、耳掛け型メトロノームの装用が吃音の二次的症狀（もがきや回避行動）に効果があったことを示しており、これは先行研究と一致した結果となっている⁴¹⁾。また吃症状の軽快化については、客観的データにおいて吃症状の内訳に変化が認められたこと（図6）からも支持されている。

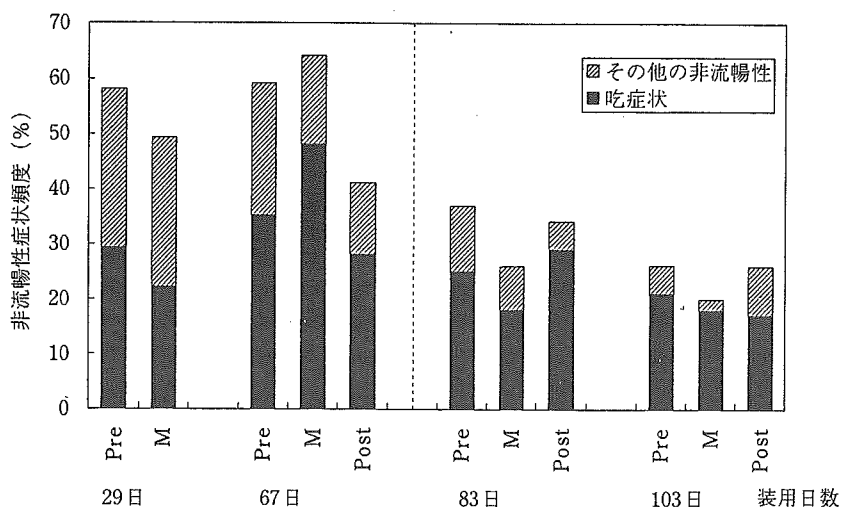


図4 訓練室内における耳掛け型メトロノームの効果
 Pre：プレテスト(メトロノームなし)，M：メトロノーム装用，Post：ポストテスト(メトロノームなし)。点線の時点でメトロノームのテンポ設定を変更している。

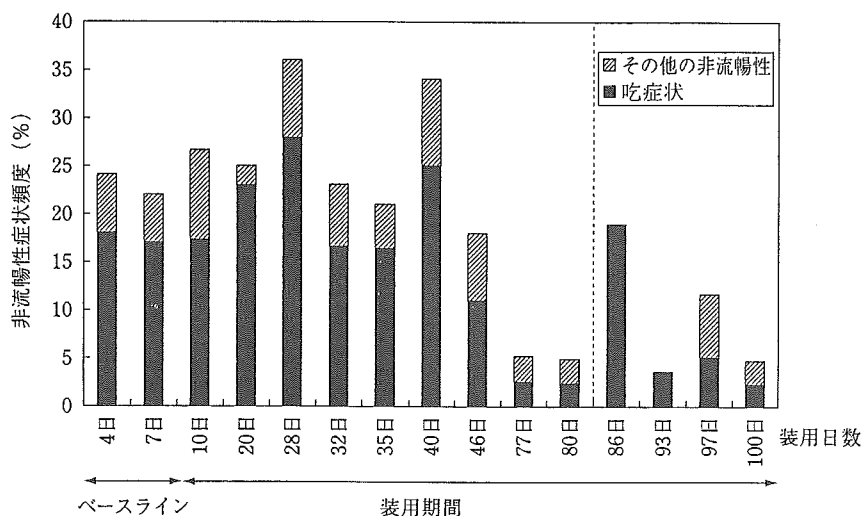


図5 電話場面における耳掛け型メトロノームの効果
 ベースライン：メトロノームなし。点線の時点でメトロノームのテンポ設定を変更している。

これらのことから、少なくともこの症例では、耳掛け型メトロノームが訓練初期には吃頻度以外の面で吃音の問題の軽快化をもたらす可能性が示唆された。

また本研究では、メトロノーム装用後すぐに吃頻度の減少が認められるのではなく、一時的な吃症状の増加の後、その減少が認められている(図5)。そして吃症状に関する自己評価においても、時間経過とともにメトロノーム非装用時と比べて装用時の吃症状のほうが軽いと判定するようになってきている(図7)。これらのことから、メトロノームの装用効果は直接的なものではなくて、装置を有効に利用できるようになるには訓練や慣れなどの要素が介在する可能性が示唆された。

一方、吃頻度が落ち着いてきた時期とメトロノームの速度設定を変更した時期がほぼ同時期であったことから、吃症状の減少にメトロノームの速度を下げたことが関連している可能性がある。本研究では、症例自身から装用経過中にメトロノームの速度を落としたいと要望があり、速度を落とした後はそれに対して肯定的な感想(「速度を落としたほうがゆっくりと落ち着いて話せる」など)が聞かれている。先行研究においては、速い速度のメトロノームを用いても吃症状が減少することから、メトロノームの効果は発話速度の低下のみに帰するものではないことが示されてきたが¹⁾、その一方でメトロノームの速度を落とすと発話速度が

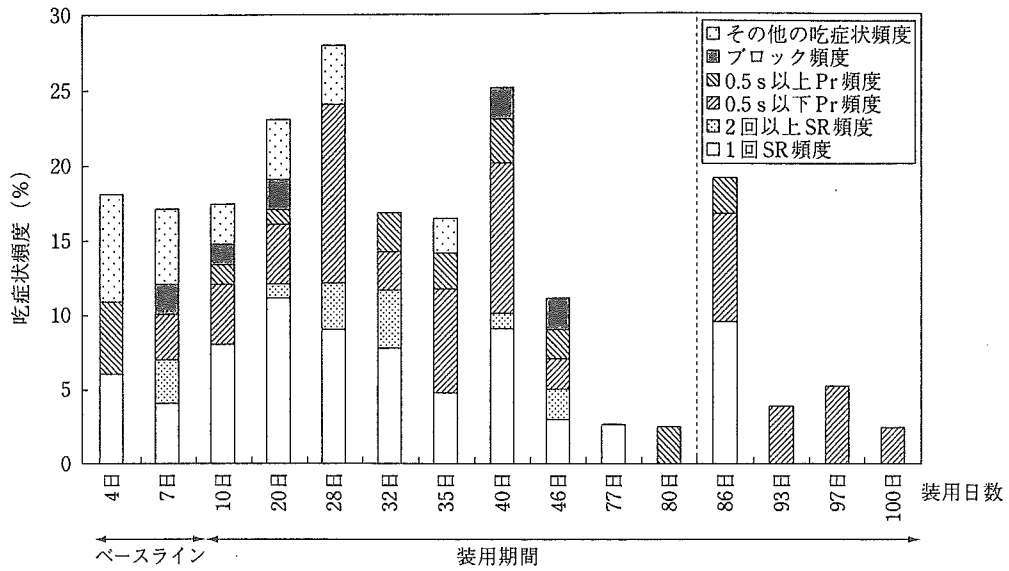


図6 吃症状の内訳の推移

SR：音・音節の繰り返し，Pr：音・音節の引き伸ばし．点線の時点でメトロノームのテンポ設定を変更している。

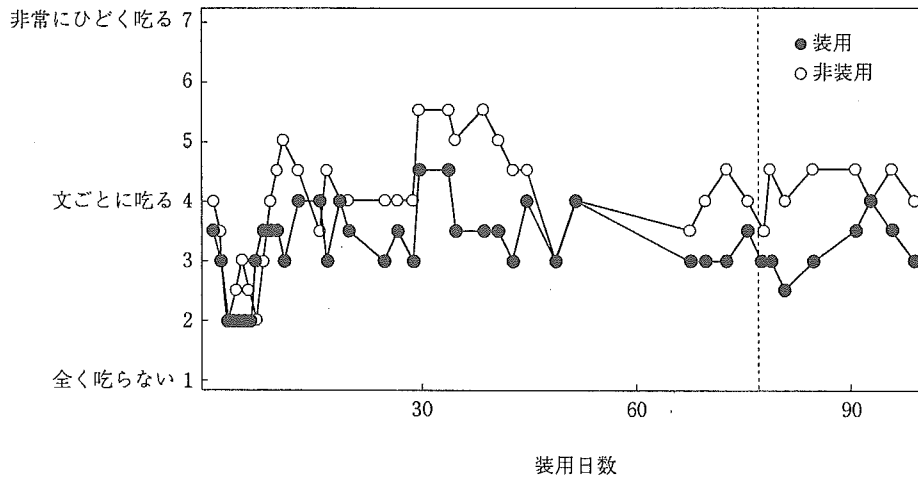


図7 毎日のメトロノーム装用/非装用時の吃症状（自己評価）
点線の時点でメトロノームのテンポ設定を変更している。

表3 吃音に関する自己評価（提示症例）

項目	装用前	装用後（3ヵ月半）
電話使用頻度（5）	2.5	5
回避（2）	5	2.5
自然さ（3）	3	5.5
全体の吃頻度（1）	4	3
電話場面での吃頻度（4）	3	3.5
対面場面での吃頻度（6）	4.5	4

（ ）内は表1の質問項目番号

表4 耳掛け型メトロノームに関する自己評価（提示症例）

項目	装用後（3ヵ月半）
苦しさの改善（1）	6
吃音症状の軽快化（2）	6
発話速度のコントロール（3）	6.5
装置への満足度（4）	6

（ ）内は表2の質問項目番号

低下し、そのことでさらに流暢性が増す傾向があることも示されている¹³⁾。これらのことから本症例においても、メトロノームのテンポ設定を遅くしたことが吃症状の減少を促した一因として考えられた。吃音の治療訓練の一つの方法として、速度のコントロールに基づくものが古くからあることから¹⁴⁾、メトロノームを訓練に用いる意義の大きなものとして速度の調整を挙げておきたい。

最後に、日常生活場面での耳掛け型メトロノームの装用日数が増加するにつれて、訓練室内におけるメトロノーム装用時/非装用時の非流暢性頻度が減少する傾向が認められた。この結果から、訓練時だけでなく日常生活場面において継続的に耳掛け型メトロノームを装用することは、練習効果が大きく、訓練室も含めた生活のすべての場面における非流暢性頻度の減少をもたらすことが示唆された。

結 論

吃音者の日常生活場面での訓練を促進する目的で、小型で携帯可能な耳掛け型メトロノームを開発し、電話場面での発話が困難な1症例に適用した。その結果、電話場面における非流暢性頻度が減少し、症例自身の自己評価においても吃症状に改善が認められた。加えて症例自身が装置に満足していることも示され、耳掛け型メトロノームの有効性が示唆された。

謝辞 本研究は、厚生労働省科学研究費補助金(15130801)の補助を受けて行われた。本論文の要旨は、第49回日本音声言語医学会総会・学術講演会にて発表した。

文 献

- 1) Brady JP: Studies on the metronome effect on stuttering. Behav Res Ther, 7: 197-204, 1969.

- 2) Bloodstein O: A Handbook on Stuttering, Singular Publishing Group Inc, San Diego, pp 354-357, 1995.
- 3) Azrin N, Jones RJ and Flye B: A synchronization effect and its application to stuttering by portable apparatus. J App Behav Anal, 1: 283-295, 1968.
- 4) Brady JP: Metronome-conditioned speech retraining for stuttering. Behav Ther, 2: 129-150, 1971.
- 5) 苺安 誠: 吃音のブロック症状に対するリズム発話と運動制御アプローチの効果. 音声言語医学, 31: 271-279, 1990.
- 6) Meyer V and Mair JMM: A new technique to control stammering: A preliminary report. Behav Res Ther, 1: 251-254, 1963.
- 7) Brady JP: A behavioral approach to the treatment of stuttering. Amer J Psychiat, 125: 843-848, 1968.
- 8) Trotter WD and Silverman FH: Does the effect of pacing speech with a miniature metronome on stuttering wear off? Percept Mot Skills, 39: 429-430, 1974.
- 9) Silverman FH: Long-term impact of miniature metronome on stuttering: An interim report. Percept Mot Skills, 42: 1322, 1976.
- 10) Ingham RJ: Stuttering and Behavior Therapy, College-Hill Press, San Diego, pp 108-113, 1984.
- 11) Mallard AR: The effect of syllable-timed speech on stuttering behavior: An audiovisual analysis. Behav Ther, 8: 947-952, 1977.
- 12) 日本音声言語医学会吃音検査法小委員会, 森山晴之, 小澤恵美, 他: 吃音検査法<試案1>について. 音声言語医学, 22: 194-208, 1981.
- 13) Hanna R and Morris S: Stuttering, speech rate, and the metronome effect. Percept Mot Skills, 44: 452-454, 1977.
- 14) Van Riper C: The Treatment of Stuttering, Prentice-Hall, Englewood Cliffs, pp 62-72, 1973.

別刷請求先: 〒359-8555 埼玉県所沢市並木4-1

国立身体障害者リハビリテーションセンター
研究所

酒井奈緒美

国立身体障害者リハビリテーションセンター

研 究 紀 要

第 25 号

国 リ ハ 研 紀

Nat.Rehab.Res.Bull.Jpn.

平成 16 年

吃音者の発声におけるピッチ制御の聴覚フィードバック特性

佐藤 裕* 森 浩一* 福島 康弘***

Control Characteristics of Voice Pitch by Auditory Feedback in Persons Who Stutter

Yutaka SATO*, Koichi MORI* and Yasuhiro FUKUSHIMA***

Functional abnormality in auditory feedback has been hypothesized to be one of the causes of stuttering, a speech dysfluency. We investigated this possibility by a transformed auditory feedback method (TAF; Kawahara, 1993), which is a method for measuring the effect of auditory feedback onto voice pitch (F0) using parametric perturbations inserted into the feedback loop. The results showed that the stutterers' cross-correlation functions between the F0s of auditory feedback and uttered voice differed from those of non-stutterers. Furthermore, the control characteristics of voice pitch were analyzed in the first and middle parts of phonation separately. The cross-correlation functions between the F0s of auditory feedback and uttered voice differed significantly between the first and middle parts of phonation in nonstutterers. However, the control characteristics for those who stutter were diverse, suggesting that different etiologies or pathophysiology are involved among those who stutter.

キーワード：変換聴覚フィードバック, ピッチ制御, 吃音, F0(基本周波数)

1. はじめに

発話と聴覚が相互に関係している証拠として、Lombard効果[1]や遅延聴覚フィードバック (Delayed Auditory Feedback: DAF)[2]が挙げられる。Lombard効果は、騒音下での発話により自分の声がよく聞こえない環境では、静かな環境に比較して、声が大きくなり、有声音の部分が長くなる等、音声に変化する効果のことである。DAFとは、自分の発話を時間的に(50~200 ms程度)遅らせたものを聴覚にフィードバックさせることであり、吃音に似た症状が誘発されることが知られており、これによる乱れた発話を人工吃音と呼んでいる。また、発話の障害である吃音が聴覚フィードバックに関連していることが知られている。DAFを

吃音者に適用することで吃音がかえって改善する[3]ことや、雑音(白色雑音等)を用いて聴覚遮断した時に吃音が増加するマスキング効果[4]等から吃音者は聴覚フィードバックに障害があるのではないかという説が提唱された(Servo Theory)[5]。その他、音声のピッチをオクターブシフトさせて聴覚にフィードバックさせる手法(Frequency Altered Feedback: FAF)[6]も吃音を減少させる。これらの手法は、吃音の改善にはある程度有効であるが、さらに分析的に検討しようとする、定量性が不十分である。

河原は発声における音声知覚から生成への作用を定量的に明らかにすることを目的とし、聴覚フィードバックされる音声の周波数に摂動を加えた人工的なフィー

* 国立身体障害者リハビリテーションセンター研究所
感覚機能系障害研究部
** 玉川大学工学部

* Department of Rehabilitation for Sensory Functions,
Research Institute of National Rehabilitation Center
for Persons with Disabilities
** Tamagawa University Faculty of Engineering