Different Brain Strategies Underlie the Categorical Perception of Foreign and Native Phonemes

Yasuyo Minagawa-Kawai[1,2], Koichi Mori[2], and Yutaka Sato[2]

Abstract

■ The present study using near-infrared spectroscopy examined the neuronal correlates of Japanese long/short vowel contrast discrimination and its relationship with behavioral performance by comparing native Japanese (L1) subjects and Korean subjects learning Japanese as a second language (L2). Phoneme-specific responses were predominantly observed in the left auditory area only in the L1 subjects, although the behavioral scores of the L2 subjects indicated categorical perception (CP) that was indistinguishable from that of the L1 subjects. These inconsistent relationships were more evident in the correlation coefficients between the brain recording and behavior. However, slower reaction times and nonspecific brain responses in the L2 listeners suggest differences in their cortical processes from those of the L1 subjects. These findings suggest that the CP of L2 phonemes as determined by behavioral scores alone does not always predict a language-specific neural processing as employed by the L1 listeners. ■

INTRODUCTION

Languages differ in the constitution of their phoneme inventories and in the distinctiveness of those phonemes. Such phonemic repertories are acquired through the linguistic experience of the first language (L1) that modifies the innate perceptual ability to optimally adapt to the phonological system in L1 (Jusczyk, 1997; Strange, 1995; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Aslin & Pisoni, 1980). Cross-linguistic behavioral experiments such as identification and discrimination tests have shown that, as a consequence of the linguistic experience, an adult's perception of the L1 phonemic contrast is both categorical and efficient, whereas that of certain contrasts in nonnative languages (L2) turns out to be inaccurate and difficult (e.g., English /r/ and /l/ by Japanese listeners; Yamada & Tohkura, 1992; Miyawaki et al., 1975). Furthermore, brain imaging studies have demonstrated neurophysiological evidence for language-specific phoneme perception (Rivera-Gaxiola, Csibra, Johnson, & Karmiloff-Smith, 2000; Sharma & Dorman, 2000; Simos et al., 1998; Dehaene-Lambertz, 1997; Näätänen et al., 1997). Some of these studies using magneto-encephalography (MEG) and near-infrared spectroscopy (NIRS) showed that the large responses that are derived from phonemic differences and not acoustic physical ones in L1 are observed in the left auditory area (Minagawa-Kawai, Mori, Furuya, Hayashi,

& Sato, 2002; Rivera-Gaxiola et al., 2000; Dehaene-Lambertz, 1997; Näätänen et al., 1997).

Phonemic contrast of English /r/ and /l/ is frequently referred to as an example of difficult phonemes for Japanese adults to distinguish. There have been extensive behavioral studies that examined possible factors influencing the identification of /r/–/l/, such as positions of the phonemes within a word, age of learning L2, duration of learning experience, and methods of training (Takagi, 2002; Bradlow & Pisoni, 1997; Lively, Logan, & Posoni, 1993; Yamada & Tohkura, 1992; Logan, Lively, & Pisoni, 1991). Training with only a restricted word or synthetic /r/–/l/ continuum in limited word positions did not result in learners' perceptual improvement because the acquired ability of identification/discrimination, if any, did not generalize to the perception of phonemes in various word or speaker contexts (Lively et al., 1993; Logan et al., 1991; Strange & Dittomann, 1984). Instead, it was found that the more effective stimuli for inducing generalization were naturally spoken tokens of /r/–/l/ occurring in various word positions pronounced by multiple talkers (Lively et al., 1993; Yamada & Tohkura, 1992). However, most of the training studies failed to demonstrate learners' perceptual improvement to the level of natives even after the intensive training (Takagi, 2002).

Neurophysiological studies on the training processes revealed neural plasticity associated with L2 phoneme learning by way of short but intensive training sessions (Callan et al., 2003; McClelland, Fiez, & McCandliss, 2002; Menning, Imaizumi, Zwitserlood, & Pantev, 2002; Tremblay & Kraus, 2002; Imaizumi, Itoh, Tamekawa,

Deguchi, & Mori, 1998). Because what most of those studies revealed, however, were neurophysiology based on the training of a phonemic contrast in a certain limited word and speaker context, it could have been that the neural changes found in these training studies might be different from those acquired in natural linguistic context. One way to reveal neurophysiologic correlates to the naturally acquired novel L2 phonemes is to examine bilinguals (or high-proficiency L2 learners). Several imaging studies explored neuronal basis of bilingualism (Chee, Soon, Lee, & Pallier, 2004; Dehaene et al., 1997; Kim, Relkin, Lee, & Hirsch, 1997; Perani, Dehaene, et al., 1996). Brain activity during listening stories in L1 and L2 was different between low-proficiency and high-proficiency bilinguals but not between early- and late-acquired bilinguals (Perani, Paulesu, et al., 1998; Perani, Dehaene, et al., 1996), suggesting that the age of learning alone does not influence the neuronal activity related to listening to L2. Although their studies were not specific about the abilities of L2 phoneme identification, their results do not relate well to the difficulties of distinguishing L2 phonemes for adults without pre-adolescent L2 exposure (Takagi, 2002) in contrast to the ease for children (Kuhl, Tsao, & Liu, 2003; Cheour, Shestakova, Alku, Ceponiene, & Näätänen, 2002). In fact, neuroimaging studies on phonemic processing in L2 learners (or late bilinguals) have been limited (Chee et al., 2004), and more analytical work is lacking on the correlation of the neurophysiological evidence with the behavioral results of the L1 and L2 phonemic contrast. Until now, such correlations have been assumed rather than fully elucidated (cf. Winkler et al., 1999). Does the similar behavioral performance on L2 phonemes by L2 listeners to those by L1 listeners in fact denote the use of a neuronal circuit similar to that for L1 listeners? The present study directly addresses this question by examining the neural substrates for processing the Japanese phonemic contrast as L1 and L2 and its relationship with the behavioral performance.

Japanese is referred to as a ''mora-timed language'' in which the duration of each mora (a subsyllabic unit) is kept roughly constant; thus, the duration of geminate vowels and consonants is psychologically perceived for about twice as long as that of single ones. Because these types of long/short phonemes (e.g., kado [corner] and kadoo [flower arrangement]) that occur in any syllable position can be distinguished solely by their durational difference, many learners of Japanese whose native language does not possess a similar contrast find it very difficult to distinguish this contrast (Minagawa-Kawai, Maekawa, & Kiritani, 2002; Toda, 1994). Cerebral representations of this type of opposition for L1 listeners were reported in Japanese (Minagawa-Kawai, Mori, Furuya, et al., 2002) and in Thai (Gandour et al., 2002). The present study measured hemodynamic responses to this Japanese long/short vowel contrast in fluent L2 speakers (Koreans) without specific phonetic training and L1

speakers with NIRS using a contrastive paradigm (Furuya & Mori, 2003). To assess whether the Korean (L2) and also Japanese (L1) subjects exhibit similar neural responses specific to the phonemic difference in Japanese language, three contrastive conditions were prepared, comprising AB (a short vowel pair of Stimulus A and Stimulus B), BC (a short and long vowel pair of Stimulus B and Stimulus C), and CD (a long vowel pair of Stimulus C and Stimulus D). Although each stimulus pair differs equally in length, only the BC condition has a phonologically distinctive difference in Japanese linguistic system, and thus, the L1 group is expected to show large neural response to the B and C contrast. In contrast, the comparisons of AB and CD are control conditions where no language-specific responses would be observed. Finally, the relationship between the hemodynamic and behavioral results was then analyzed in each group.

RESULTS

Behavioral Results

The stimulus word used in both the behavioral and NIRS experiments was the pseudoword ''mama'' whose final vowel ranges from /mama/ (short vowel) to /mama:/ (long vowel). The subjects in the identification test of long/short vowels judged if the final vowel of the stimulus is a long or short phoneme. The results of the test showed that the averaged values of the phoneme boundary (PB) was 192 msec ($SD = 7.5$, $n = 8$) for the Korean group and 198 msec ($SD = 6.9$, $n = 8$) for the Japanese group (Figure 1). The PB for both groups was located between the Stimuli B and C, where the
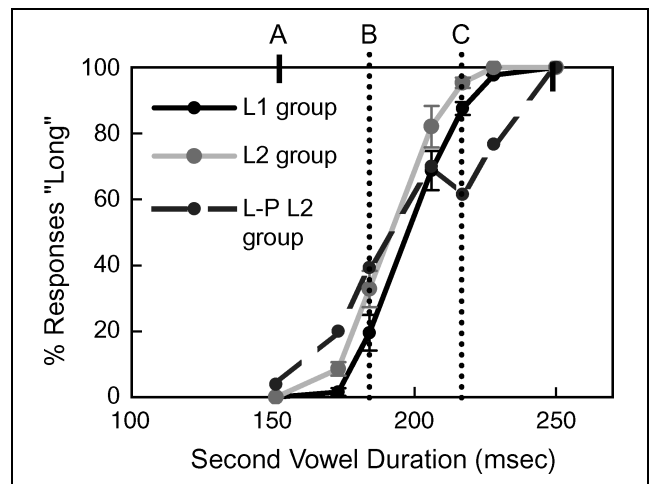

Figure 1. Identification rates of /mama/–/mama:/ continuum. Averaged responses of the L1 group (solid line), the L2 group (gray line), and the low-proficiency (L-P) L2 learners from the previous data (dotted line) are indicated. Durations of the second vowel of the Stimuli A through D are indicated in the graph. The phoneme boundary lies between Stimuli B and C. Error bars indicate one standard error of the mean.

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
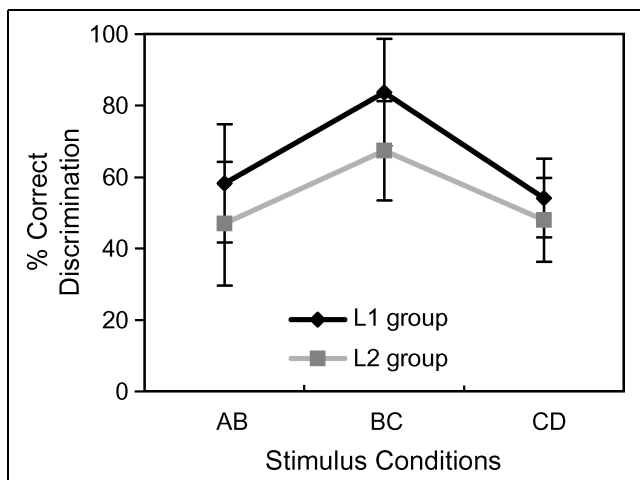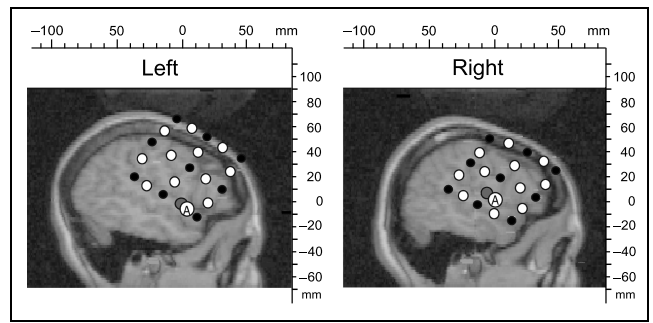

Figure 3. Locations of the NIRS probes and recording channels. NIRS probe positions (small white circles) and the centers of the measurement channels (white circles) are superimposed onto the parasagittal MR brain images. The lateral posterior borders of the Heschl's gyri are labeled "A." The channels 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.

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.
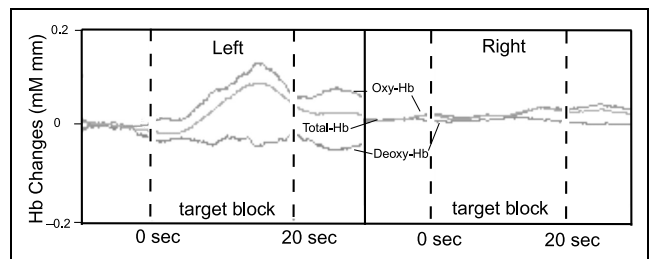
and planum temporale (PT) in the superior temporal gyrus (STG), corresponding to BA 22 (part of Wernicke's area). The hemodynamic responses in those channels were observed to peak at 5–14 sec after the target stimulus onset (Figure 4). There was no significant difference in the peak latency for the L1 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 × 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 contrasts. The total Hb 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.
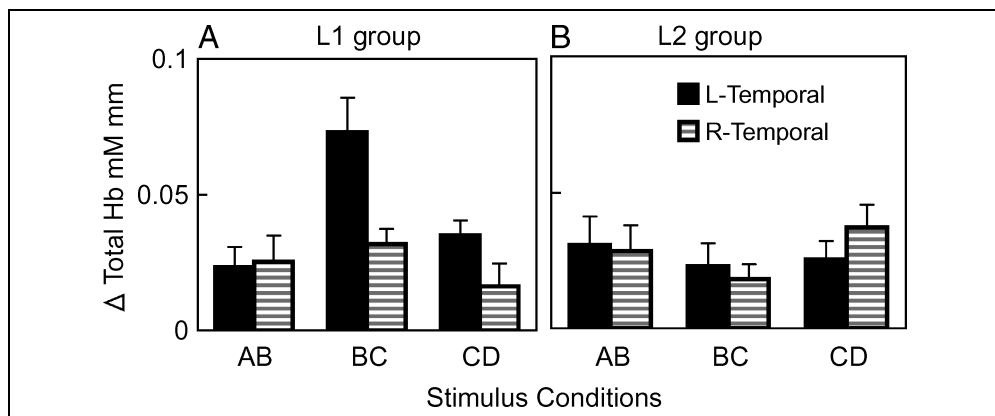
Figure 6. Correlation of behavioral tests and brain responses. Individual data of the L1 (A,C) and L2 (B,D) subjects are shown. Identification rate of /mama/–/mama:/ continuum and durations of the second vowel in the Stimuli A–D (dotted line) are indicated (A,B). The slopes of identification function 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).
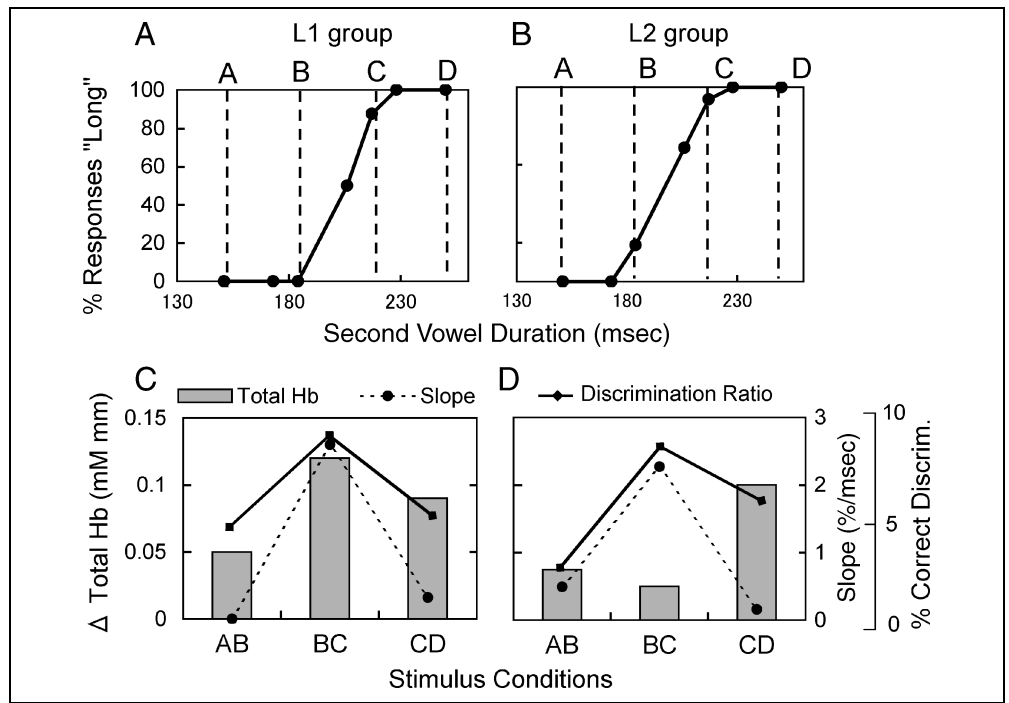
category-specific nor left-dominant responses, but only small, nondifferential responses. The behavioral results of the identification and discrimination tests indicated that the L2 learners had acquired categorical percepts of Japanese long/short vowels that are indistinguishable from those of the L1 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, & Okada, 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 ($\approx$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.

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., Satthamnuwong, 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 Phoniatrics, 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 & Psychophyisics, 40,* 205–215.

Juscyzk, 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.

Liberman, 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 handed-ness: 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., Sivonen, 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.