Plasticity of the Human Auditory Cortex Induced by Discrimination Learning of Non-Native, Mora-Timed Contrasts of the Japanese Language

Hans Menning,[1,5] Satoshi Imaizumi,[3] Pienie Zwitserlood,[2] and Christo Pantev[4]

[1]Center for Biomagnetism, Institute of Experimental Audiology, and [2]Department of Psychology, University of Münster, Münster, Germany; [3]Department of Communication Sciences and Disorders, Faculty of Health Sciences, University of Hiroshima, Hiroshima, Japan; [4]Rotman Research Institute for Neuroscience, University of Toronto, Toronto, Canada

In this magnetoencephalographic (MEG) study, we examined with high temporal resolution the traces of learning in the speech-dominant left-hemispheric auditory cortex as a function of newly trained mora-timing. In Japanese, the "mora" is a temporal unit that divides words into almost isochronous segments (e.g., *na-ka-mu-ra* and *to-o-kyo-o* each comprises four mora). Changes in the brain responses of a group of German and Japanese subjects to differences in the mora structure of Japanese words were compared. German subjects performed a discrimination training in 10 sessions of 1.5 h each day. They learned to discriminate Japanese pairs of words (in a consonant, *anni–ani*; and a vowel, *kiyo–kyo*, condition), where the second word was shortened by one mora in eight steps of 15 msec each. A significant increase in learning performance, as reflected by behavioral measures, was observed, accompanied by a significant increase of the amplitude of the Mismatch Negativity Field (MMF). The German subjects' hit rate for detecting durational deviants increased by up to 35%. Reaction times and MMF latencies decreased significantly across training sessions. Japanese subjects showed a more sensitive MMF to smaller differences. Thus, even in young adults, perceptual learning of non-native mora-timing occurs rapidly and deeply. The enhanced behavioral and neurophysiological sensitivity found after training indicates a strong relationship between learning and (plastic) changes in the cortical substrate.

Our brain is a highly adaptive learning device, reorganizing itself in accordance with environmental constraints. Nevertheless, adult people experience difficulties when learning a language that has a completely new and unfamiliar phonemic structure. Language perception is altered by the linguistic experience we make throughout our life. We perceive and produce speech through the filter of our native language. Whereas neonates are able to discriminate most phonetic categories in the first months of life (Dehaene-Lambertz and Pena 2001), they develop prototypical phoneme representations of their native language in the first year of life (Cheour et al. 1998). Exposure to a specific language in the first half year of life directs infants' phonetic perception toward this language and forms prototypes for each category. A prototype functions like a perceptual magnet for all members of that category and attracts all variations of a phoneme into "native" perception (Kuhl 1991). Once established, these phoneme categories form a stable basis for consecutive speech perception and speech production (Kuhl et al. 1992). As a consequence, adults have difficulties in distinguishing non-native phone-

mic contrasts. For instance, Finnish and Estonian adults perceive their native phoneme prototypes more sensitively than non-native phonemes (Näätänen et al. 1997). Japanese listeners have difficulties in distinguishing /l/ and /r/ and perceive both phonetic categories as one (Goto 1971; Yamada and Tohkura 1992). Nevertheless, after extended training, adult Japanese listeners improved their perceptual identification as well as their production performance of English /r/ and /l/ (Bradlow et al. 1997). In another experiment, the behavioral training of just-perceptible differences in speech stimuli resulted in a significant change in the duration and magnitude of the cortical potentials (Kraus et al. 1995).

It is generally assumed that representational maps in the brain are subject to plastic changes subsequent to altering sensory input. Intensive frequency-discrimination training enhances the area of representation of the trained frequency range in the primary auditory cortex of owl monkeys (Recanzone et al. 1993). In humans, changes in the perceptual acuity of frequency discrimination correlate with enhancements in the neuromagnetic responses to these frequencies (Menning et al. 2000). Moreover, intensive phonetic experience of speech sound discrimination is apt to alter the neural activity that underlies coding of these events in infants (Kuhl et al. 1992; Kuhl 1994), as well as in

[5]Corresponding author.
E-MAIL hans.menning@uni-muenster.de; FAX 49-251-8356882.
Article and publication are at http://www.learnmem.org/cgi/doi/10.1101/lm.49402.

adults (Logan et al. 1991). Winkler et al. (1999) showed that learning a new language is accompanied by the development of new cortical representations for unknown phonemes of the new language. Hungarians, who learned to speak Finnish fluently in adulthood, developed cortical memory representations for Finnish phoneme categories that do not exist in their native language. Phonemes are the smallest units of speech that affect meaning. A small difference in the phonemic structure of a word entirely changes its meaning.

Each language has its own "music," that is, languages differ in their rhythm and stress patterns. Language has a certain rhythm or timing because phonological units (i.e., syllables or segments like moras) are organized into rhythmic sequences. Whereas English is generally considered as a stress-timed language and French as a syllable-timed language, Japanese and even Finnish are often cited as mora-timed languages (Aoyama 2001). In Japanese, the "mora" is a temporal unit that divides words into almost isochronous segments (e.g., *na-ka-mu-ra* and *to-o-kyo-o* each comprises 4 mora). For Japanese and Finnish, the relative duration of a vowel or consonant is a decisive feature for differentiating among meaningful words. Nevertheless, in stress-timed languages, for instance in English, the difference between *heed* and *hid*, *bead* and *bid*, or *wooed* and *wood* is sometimes only encoded by the length of the vowel (Clark and Yallop 1995). In the German language, vowel duration can also be decisive for the meaning of a word (as in *biete* and *bitte*, *Miete* and *Mitte*), but is additionally accompanied by a different accentuation or consonant duration that facilitates the perception of the difference. However, in languages such as Finnish, Estonian, Hungarian, and Japanese, every short vowel has a corresponding long counterpart. Similarly, the duration of some consonants has both a short and long variant. These durational contrasts are used to encode different meanings. For instance, in Finnish, all variants of these words, *tule-tuule-tuulle-tuulee-tulee-tullee-tulle-tuullee*, are to be distinguished accurately by their duration quantity because they encode different meanings. Mora-timing might also be relevant in Finnish (Iivonen 1998). Enhanced processing of speech sounds, especially for duration differences as compared with equivalent changes in tones, was found for Finnish speakers (Jaramillo et al. 2001). The Hungarian language also has many long/short differences to be distinguished (e.g., *fülel–füllel*, *sör–sö:r*, *megy–meggy*). Recently, the moraic/nonmoraic distinction was postulated even in natural spoken Danish (Grønnum and Basboll 2001).

Japanese is a mora-timed language par excellence (Warner and Arai 2001); its temporal units consist mostly of consonant–vowel (CV or CVV) syllables, single vowels, or the nasal /n/. Each unit takes about the same length of time (Ladefoged 1993). Long syllables (*ko-o*) may consist of two moras, but short syllables (*ko*) only consist of one. These units are expressed in the two phonetic alphabets hiragana and katakana with maximally 111 possible Japanese CV (*ka*, *ba*, *ra*) or CyV monomoraic (*kyo*, *byo*, *ryo*) syllables, five vowels, and the consonant /n/, and form the entire set of elementary phonetic units of Japanese. Thus, the rhythm of Japanese is controlled by a regular temporal sequence, mostly isochronous.

The Japanese language has a restricted number of moras, which results in a large number of homonyms. These homonyms are often differentiated only by the duration of a mora when the context is not available. For instance, the sound sequence /hoshu/ exists in four variants: with a long /o/, with a long /u/, with both a long /o/ and a long /u/, or with both a short /o/ and a short /u/. The words *hoshu*, *hooshu*, *hoshuu*, and *hooshuu* have completely different meanings: *hoshu* means "a catcher" or "conservative"; *hooshu* means "a gunner" or "artilleryman"; *hoshuu* means "mend, repair"; and *hooshuu* means "remuneration, reward." The same is true for a large number of other Japanese words. Some of these distinctions as in *oji-isan* (grandfather) and *ojisan* (uncle) or *obaasan* (grandmother) and *obasan* (aunt) can lead to embarrassing misunderstandings, if the duration of the vowel is not considered precisely.

Cutler and Otake (1994) studied the role of mora-timing in the recognition of spoken Japanese words. They showed that Japanese listeners are sensitive to the moraic structure of speech and that it is easier for them to respond to moras than to phonemes. Kakehi et al. (1996) even concluded that Japanese listeners parse the speech signal into units larger than phonemes, namely, moras, because they mainly use the prevailing moraic consonant–vowel (CV) unit in current speech. The mora-unspecific VC cues are only used in natural VCV words where the consonant is the onset of a CV mora. This has led to several questions: (1) What are the effects of long-term experience with one's native language, when this language affords accurate temporal discrimination? (2) What kind of changes can be produced by intensive auditory discrimination training of non-native durational differences, and what are the related neurophysiological correlates? (3) What is the difference between native and non-native perception and discrimination learning of durational differences within one mora? To answer these questions, we compared the perception and processing of Japanese words in German and Japanese subjects. The duration quantity in bi- and trimoraic words was manipulated in 8 steps over the length of one mora. German subjects (test group) were trained to discriminate these differences. The control group was composed of Japanese native speakers.

Accurate perception of non-native phonemic categories can be quantified by means of an automatic, preattentive "change detector" component of the Auditory Evoked Response (AER), called the Mismatch Negativity (MMN)

and its magnetic counterpart, the Mismatch Field (MMF; Schröger 1997; Näätänen 1999). Phonemic memory traces, revealed by the MMN/MMF, have shown language specificity by showing that there is an enhancement of MMN/MMF for native but not for non-native phoneme categories (Näätänen et al. 1997). During the processing of speech sounds, the amplitude of the MMN is higher in the left than the right hemisphere (Alho et al. 1998). Recently, MMN was found to be larger for words than for pseudowords (Pulvermüller et al. 2001). There is increasing evidence that speech and language in right-handed subjects are processed and produced preferentially by the left cerebral hemisphere (Altenmüller et al. 1993; Eulitz et al. 1995; Diesch et al. 1996, 1998; Gootjes et al. 1999; Tervaniemi et al. 1999; Shtyrov et al. 2000). Also, a left hemispheric dominance for speech has been shown for a majority of left-handed subjects (Jung et al. 1984). For these reasons, the present study's recordings were carried out over the left temporal cortex.

There were three objectives for the present magnetoencephalographic (MEG) study. The first objective was to assess if learning and experience induce cortical plasticity in the "phonemotopic" maps of the brain. The second objective was to determine if intensive discrimination training of very small differences in non-native speech affects the discrimination acuity. The third objective was to measure the affect of discrimination training on the brain's automatic change-detection process as reflected by the MMF or other neuromagnetic components. A group of German subjects, unfamiliar with Japanese, were trained to discriminate eight durational variants of four Japanese pairs of words (*anni–ani, itte–ite, kiyo–kyo,* and *kiyou–kyou*). The length of a consonant or a vowel was varied in eight small durational steps. Subjects performed 10 training sessions. To test the behavioral discrimination performance, each training session was accompanied by discrimination tests without feedback. Before the first and after the last training session, MEG recordings using three of the eight duration variants were performed. A group of Japanese subjects received two consecutive training sessions and were tested in the same manner as the German subjects. This was done to test whether the cortical networks that encode speech duration in Japanese native speakers were also subject to short-term plasticity changes, or whether these well-established representations in the cortical maps are not subject to further reorganization processes.

RESULTS

Behavioral Results

Reaction Time

Reaction time (RT) was recorded in all training sessions and referred to the onset of the second stimulus of each stimulus pair. Within German test subjects, training resulted in a clear decrease in RT. After longer training, subjects responded faster to pairs of standard stimuli, as well as to pairs of standard-deviant stimuli (Fig. 1a). The logarithmic function characterizing the decrease in RTs over 10 training sessions showed smaller values for the standard as compared with the deviant stimuli. The corresponding equation for the results of the standards was $y = -0.0366 \ln(x) + 0.7667$, with $R^2 = 0.9274$, and for the results of the deviants was $y = -0.027 \ln(x) + 0.7746$, with $R^2 = 0.8623$, where R^2 represents the variance in RT attributable to the variance in training days. These equations show that there is a smaller decrease for deviants than stan-
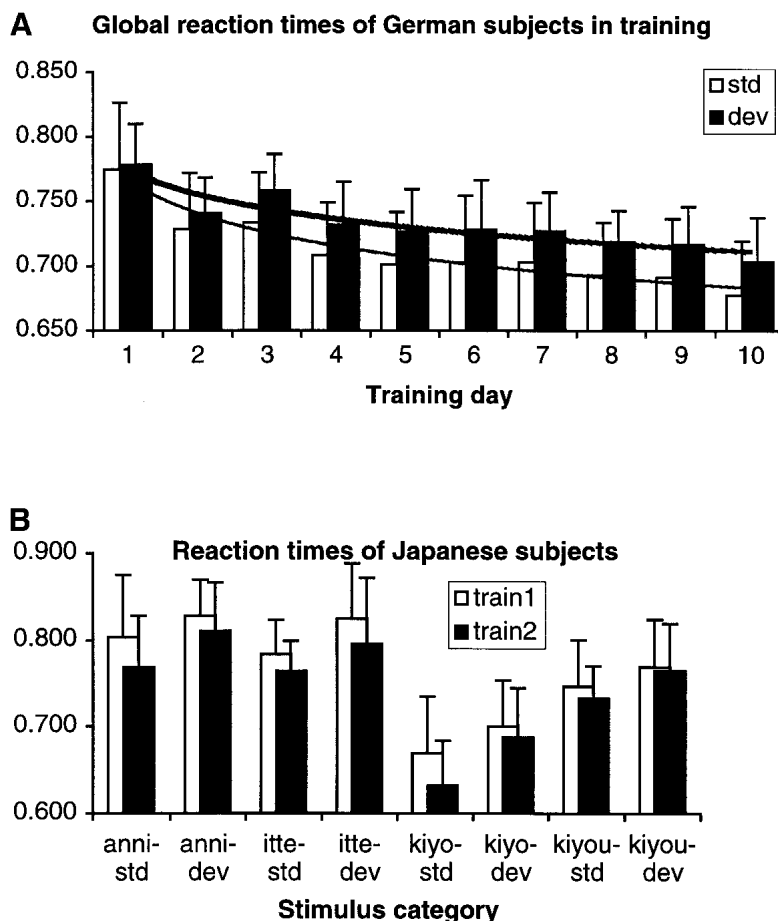

Figure 1 (*a*) Grand averaged RTs and SEM across German subjects for each training session. A constant decrease approximated by a logarithmic trend (thick line for deviants, thin line for standards) occurs from training day 1 to 10. (*b*) Grand averaged RTs and SEM of Japanese subjects for each stimulus condition.

dards. German subjects' RTs to deviant stimuli were, on average, 21.4 msec (SEM = 0.54 msec) longer than their RTs to standards (Fig. 1a). From the first to the last training session, a decrease of 97 msec for standard stimuli and of 76 msec for deviant stimuli was obtained. These decreases for both standards and deviants were significant ($P = 0.0026$ for standards and $P = 0.0038$ for deviants), as shown by a paired two-tailed Student's *t*-test. The RTs of the Japanese subjects also decreased significantly from the first to the second training session for each stimulus condition ($P = 0.0221$ for standards and $P = 0.0595$ for deviants in a paired two-tailed Student's *t*-test) by 20.86 msec (SEM = 9.75), on average (Fig. 1b). The results from Japanese subjects were similar to

Figure 2 Category boundaries for one exceptional subject (S4) with a varying duration concept, a subject (S8) with a nonmonotonic category boundary between the long and short variants of stimuli, and the averaged responses. A clear category boundary is observed between differences 4 and 5.

those from German subjects, showing that RTs for deviant stimuli were longer than for standard stimuli, in the first (by 29.65 msec, SEM = 6.45), and in the second training session (by 40.45 msec, SEM = 8.75). Statistical comparison (unpaired 2-tailed *t*-test) of RTs in the first and second training sessions between Japanese and German subjects revealed no significant differences. Both groups started with similar RTs, but German subjects showed a markedly greater decrease from the first to the second training session (46 msec for standards and 37 msec for deviants) in comparison with Japanese subjects (26 msec for standards and 15 msec for deviants).

Categorization

In natural speech, humans are able to realize the variation in the duration of a speech signal and make categorical judgments. Perception of speech occurs categorically rather than as a matter of continuous, monotonic variation (Eimas 1985). In a separate test, 10 subjects classified eight variants of four stimuli in a two-alternative-forced-choice task as being "long" or "short." The purpose of this test was to determine whether subjects form phonetic categories in a long versus short manner or whether they recognize the differences in duration in a monotonic increasing or decreasing way.

The results of the categorization test are illustrated in Figure 2. A set of 120 stimuli of each condition of stimuli was presented in a semirandomized order with an equal number of each difference variation. Subjects perceived the long versus short version of the stimuli in a categorical,
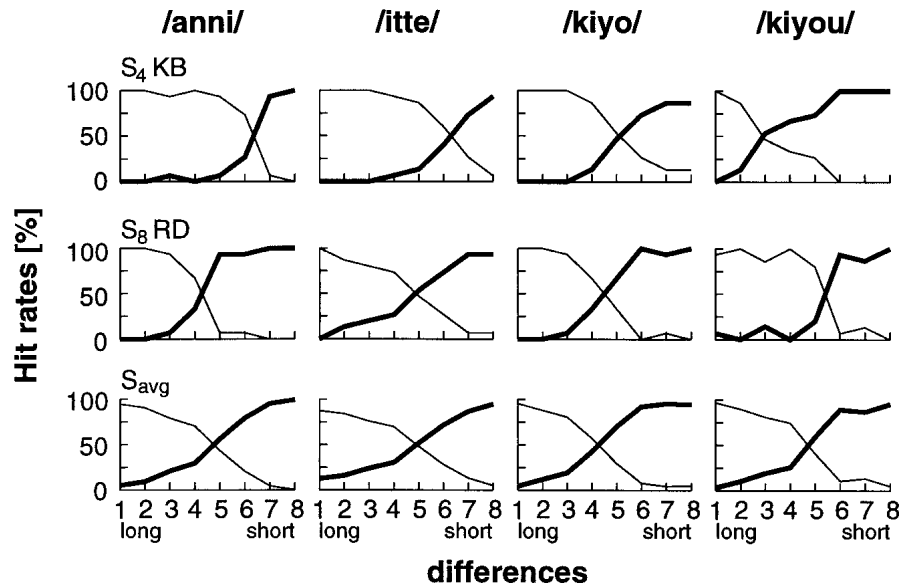
nonmonotonic manner. Most subjects recognized the variants 1–3 as long and 6–8 as short. During the variants 4–5, there was a clear change in the recognition pattern, pointing to a categorical perception at this duration. Nevertheless, some subjects (S4, S6, S9) varied these category boundaries to more extreme positions. They seemed to construct wider categories than the other subjects.

Discrimination Performance

Improvements in the training occurred rapidly (already apparent in the first training session) and evolved continuously until reaching an asymptotic saturation level. This is shown by the hit rate, or true positive rate, in Figure 3. Before training, German subjects discriminated the deviant stimuli differing in consonant duration (*anni*, *itte*) at a hit rate of 43%–45% in tests without feedback. They improved their performance by up to 35%, reaching a level of ~80% at the end of training. For deviant stimuli differing in vowel duration (*kiyo*, *kiyou*), the starting discrimination performance began at 55%–63% and improved by 20%–27%, reaching a level of ~80% at the end of training. Initial discrimination was better for vowel duration than for consonant duration, and reached a higher end level after training. Therefore, it was easier for German subjects to detect differences in vowel duration than in consonant duration. The correct recognition of pairs of standard stimuli was very high from the first training session, and reached a saturation level by the second or third training session. In training sessions with feedback, improvement of hit-rate performance grew continuously by ~15% for each stimu-
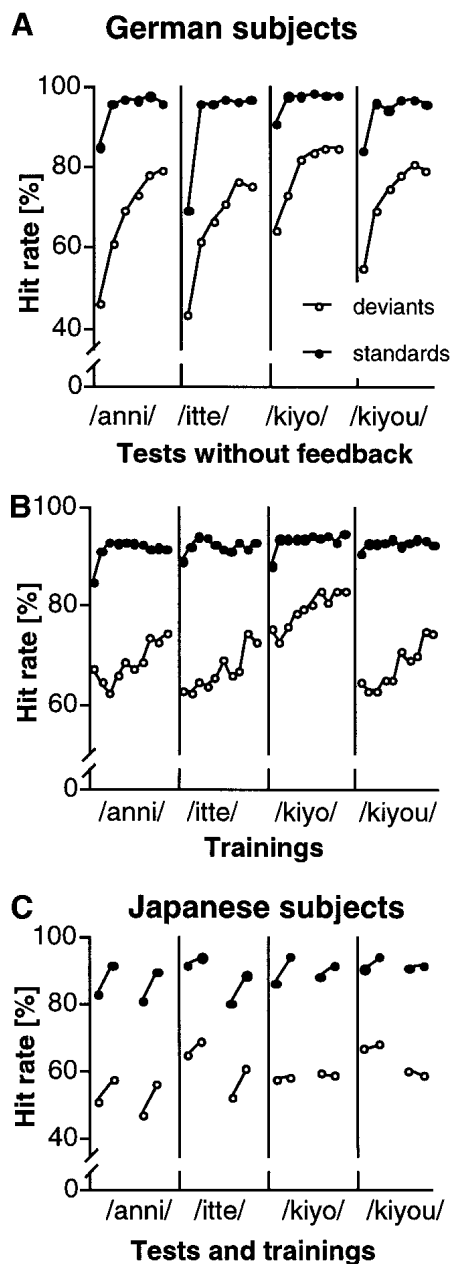
A German subjects

B

C Japanese subjects

Figure 3 (*a*) Hit rates of German subjects for deviants and standards in tests without feedback and in (*b*) 10 training sessions with feedback. (*c*) Hit rates of Japanese subjects in 2 tests (each condition *left*) and trainings (each condition *right*).

lus, indicating that training performance was independent of stimulus type, but initial performance differed because of stimulus difficulty. Japanese subjects started to discriminate deviants at a higher performance level than German subjects (average hit rate = 61%). After only two training sessions, Japanese subjects' hit-rate performance improved by up to 7% for consonant duration (*anni*, *itte*) but remained almost unchanged for vowel duration (*kiyo*, *kiyou*).

Psychometric functions of the hit-rate performance for German subjects during training are illustrated in Figure 4. The largest increase in the percentage of correctly recognized deviants occurred in the variants 2–4 ($\Delta t = 30$–60 msec). Small increases were found for variants that were very similar to the standards (variant 1 being the smallest difference, $\Delta t = 15$ msec). In variants 5–7 ($\Delta t = 75$–105 msec), which differed greatly from the standards (variant 7 being the largest difference, $\Delta t = 105$ msec), the best performance was reached early and was maintained until the end of training. The middle difference variant 4 ($\Delta t = 60$ msec), where the phoneme boundary is expected to be, showed improvement throughout the training. This improvement was 30%–45% for /anni/, /itte/, and /kiyou/; whereas for /kiyo/, which is less difficult to discriminate, the improvement was <10%. These findings indicate that the categorical boundary, determined by a hit rate of 50%, was at the beginning of the training for consonants (/anni/, /itte/) at the difference variant 4, and for vowels (/kiyo/, /kiyou/) at the difference variant 3. Over the course of the training this boundary shifted for consonants by about 2 ($\Delta t = 30$ msec) and for vowels by about 1.5 ($\Delta t = 22$ msec) difference degrees nearer to the standard. This shift is depicted in Figure 4. The fitted psychometric curves develop from a rather flat, logistic function before training to a higher and steeper function at the end of training. The areas between the two fitting curves map the "learning gain" throughout training, that is, the hit rates of the small difference degrees significantly increased after training. These results are paralleled by the performance obtained in the training sessions. The steep decrease in difference magnitude for the first three training sessions was followed by a slow, but constant improvement until the end of the training.

Evoked Neuromagnetic Responses

Amplitude Differences

To assess differences in perceptual abilities and the effects of training, the results from a group of native Japanese speakers were compared with those from a non-native German group. The assumption was that language-dependent memory traces exist and will be activated in the native but not in the non-native group. Effects of training in the non-native group should reveal the "trainability" of these memory traces. This hypothesis is based on the notion of language-dependent memory traces that are activated when the stimuli are perceived as native speech and not activated when the stimuli are perceived as non-native, "phonetic" material. Because the mismatch response is considered to be an automatic, preattentive response to stimulus change, it should be sensitive to the ability to discriminate fine perceptual differences. For evaluation of the mismatch fields, the averaged responses to the standard stimuli were sub-
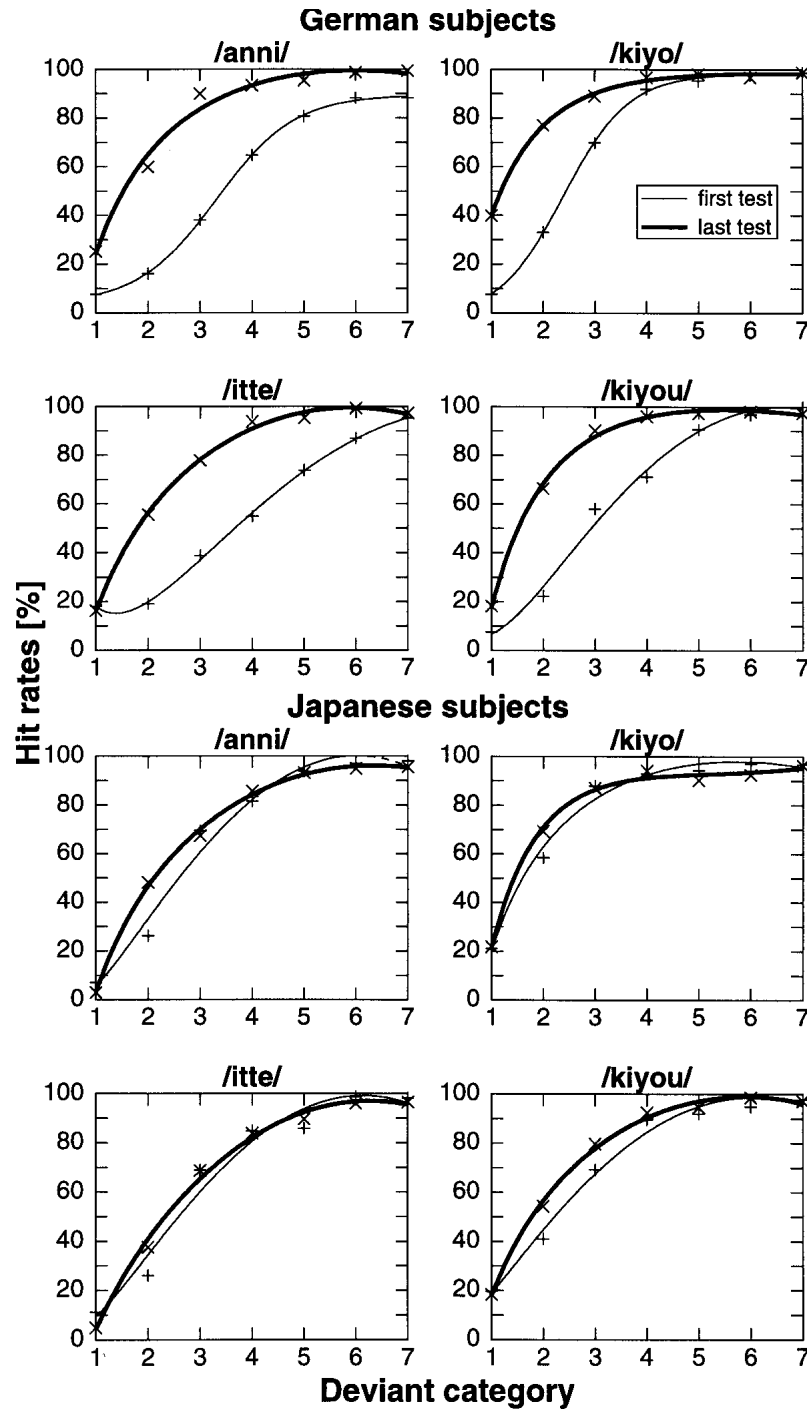
Figure 4 Psychometric functions of German and Japanese subjects for each deviant stimulus in the tests before (thin line) and after (thick line) training. Before training, the deviant variant 4 exceeds the random level (50%). After training in each condition, the deviant variant 2 is recognized as different. German subjects improved maximally in the difference variants 2–4 from the first test before training to the last test after training. Japanese subjects already recognized the deviant variant 2 in almost each condition above random level.

tracted from averaged responses to deviant stimuli. A source analysis based on a moving equivalent current dipole (ECD) in a spherical model fitted to the head surface was performed separately for the derived (deviant minus standard) waveforms as well as for the standard and deviant responses for each subject and condition. The origin of the head-based coordinate system was set at the midpoint of the mediolateral axis (y-axis). The Cartesian coordinates (x,y,z) of the source location, (the posteroanterior x-axis connecting the inion and nasion, positive toward nasion; the mediolateral y-axis connecting the two center points of the entrances to the acoustic meatuses of the left and the right ear, positive toward the left ear; and the superoinferior z-axis projecting perpendicular to the x–y plane, positive toward the vertex), the orientation of the ECD (ϕ_x, ϕ_y, ϕ_z), as well as the global field power or root mean square (RMS) amplitudes and the source strength (Q) were calculated. When the goodness of fit between calculated and measured waveforms was >90%, these parameters were used to calculate individual source waveforms for each stimulus condition under the assumption that the source space projection (SSP) on the supratemporal plane explains most of the field variance. Thus, the MMFs to the deviant responses established with the method described above appear to originate from distinct neural representations in the supratemporal plane (see Table 1).

Because of the anatomical differences between subjects and restrictions of the method, these source coordinates have to be judged carefully. The location and orientation of the generators of responses to deviant stimuli (MMF) were calculated and compared with those of the standard stimuli. The standard stimuli were located mainly in secondary auditory areas (i.e., planum temporale). In contrast, the main sources of the MMF were found to be located more medial and more anterior to the sources of the standard stimuli. With respect to the x, y, and z coordinates, Student's t-tests with a corrected α level of 0.008 revealed no significant differences for pre- versus posttraining localizations or for Japanese versus German subjects. How-

Table 1. *Source Coordinates of Responses to Standard and Deviant Stimuli*

	Responses to Standards					
	anni8 pre			anni8 post		
Native	x (cm)	y (cm)	z (cm)	x (cm)	y (cm)	z (cm)
German	0.59 ± 0.59	4.07 ± 0.73	6.05 ± 0.75	0.69 ± 0.47	4.13 ± 0.81	5.62 ± 1.29
Japanese	1.23 ± 0.91	4.72 ± 0.87	5.46 ± 0.58	0.85 ± 0.84	4.90 ± 0.94	5.70 ± 0.83
	kiyo8 pre			kiyo8 post		
German	1.02 ± 1.3	4.14 ± 0.79	5.76 ± 0.66	0.99 ± 1.04	4.12 ± 0.69	6.11 ± 0.93
Japanese	1.22 ± 0.92	4.47 ± 0.85	5.43 ± 0.41	1.35 ± 1.00	5.10 ± 0.77	5.61 ± 0.35
Average	1.01 ± 0.96	4.35 ± 0.82	5.68 ± 0.64	0.97 ± 0.86	4.56 ± 0.89	5.76 ± 0.91
	anni6 pre			anni6 post		
German	0.52 ± 0.50	4.07 ± 0.73	5.91 ± 1.04	0.64 ± 0.54	4.51 ± 0.45	6.10 ± 1.06
Japanese	1.17 ± 0.93	4.96 ± 0.51	5.60 ± 0.75	1.01 ± 0.63	4.86 ± 0.69	5.67 ± 0.93
	kiyo6 pre			kiyo6 post		
German	1.51 ± 0.96	4.53 ± 0.88	5.78 ± 1.02	0.88 ± 0.85	4.57 ± 0.74	5.52 ± 1.15
Japanese	0.90 ± 0.93	5.36 ± 0.62	5.57 ± 0.48	1.09 ± 1.11	5.11 ± 0.69	5.73 ± 0.48
Average	1.03 ± 0.89	4.73 ± 0.87	5.71 ± 0.83	0.91 ± 0.79	4.77 ± 0.67	5.75 ± 0.93
	anni4 pre			anni4 post		
German	0.47 ± 0.66	4.20 ± 1.00	6.13 ± 0.84	0.51 ± 0.61	4.35 ± 0.49	6.01 ± 0.86
Japanese	1.20 ± 0.92	4.89 ± 0.70	5.40 ± 0.76	0.88 ± 0.69	4.86 ± 0.98	5.81 ± 1.13
	kiyo4 pre			kiyo4 post		
German	1.01 ± 1.21	4.26 ± 0.79	5.88 ± 0.83	0.64 ± 1.10	4.45 ± 0.62	5.56 ± 1.47
Japanese	0.93 ± 0.96	4.61 ± 1.08	5.59 ± 0.50	0.97 ± 1.07	5.07 ± 0.99	5.57 ± 0.45
Average	0.90 ± 0.95	4.49 ± 0.91	5.75 ± 0.77	0.75 ± 0.88	4.68 ± 0.83	5.74 ± 1.02
	Responses to Deviants					
	anni8 pre			anni8 post		
German	1.08 ± 1.53	3.92 ± 0.51	5.44 ± 1.12	0.98 ± 0.70	4.18 ± 0.84	5.69 ± 1.45
Japanese	1.28 ± 0.95	4.50 ± 0.68	5.26 ± 0.56	0.98 ± 1.15	4.77 ± 0.63	5.71 ± 1.08
	kiyo8 pre			kiyo8 post		
German	1.25 ± 1.53	4.14 ± 0.76	5.59 ± 0.68	0.52 ± 0.94	3.84 ± 0.97	5.85 ± 1.43
Japanese	0.77 ± 0.95	4.11 ± 0.92	5.23 ± 0.71	0.54 ± 0.88	4.15 ± 0.63	5.70 ± 0.60
Average	1.10 ± 1.24	4.17 ± 0.74	5.38 ± 0.78	0.75 ± 0.92	4.24 ± 0.83	5.74 ± 1.16
	anni6 pre			anni6 post		
German	0.14 ± 1.54	3.62 ± 0.63	5.26 ± 0.98	0.66 ± 0.89	4.07 ± 0.98	5.61 ± 1.47
Japanese	1.16 ± 1.03	3.84 ± 1.13	5.05 ± 0.59	1.01 ± 0.95	4.59 ± 0.80	5.80 ± 1.32
	kiyo6 pre			kiyo6 post		
German	0.94 ± 0.76	4.05 ± 1.01	5.50 ± 0.75	0.80 ± 1.19	3.59 ± 1.07	5.48 ± 1.78
Japanese	0.98 ± 1.60	4.35 ± 0.95	5.53 ± 1.28	1.19 ± 1.08	4.15 ± 1.01	5.06 ± 0.93
Average	0.81 ± 1.29	3.96 ± 0.95	5.34 ± 0.92	0.91 ± 1.01	4.10 ± 0.99	5.49 ± 1.39
	anni4 pre			anni4 post		
German	0.09 ± 1.08	3.34 ± 1.39	5.62 ± 0.99	1.28 ± 0.90	4.05 ± 1.16	5.44 ± 0.56
Japanese	1.06 ± 0.87	3.57 ± 0.87	5.32 ± 0.52	1.36 ± 1.09	4.62 ± 0.51	5.54 ± 0.76
	kiyo4 pre			kiyo4 post		
German	0.95 ± 1.51	3.09 ± 0.49	5.65 ± 0.89	0.94 ± 1.34	3.90 ± 1.40	5.23 ± 0.80
Japanese	1.72 ± 1.06	4.53 ± 1.01	4.77 ± 0.93	1.79 ± 0.92	4.79 ± 0.56	4.76 ± 3.34
Average	0.95 ± 1.25	3.63 ± 1.13	5.34 ± 0.89	1.34 ± 1.07	4.34 ± 1.01	5.24 ± 1.17

Average Cartesian coordinates (mean and standard deviation) of the dipoles localized for the standard stimuli and for the deviant stimuli in German and Japanese subjects before and after training.

ever, Japanese subjects had their sources consistently more anterior, more lateral, and more inferior than did German subjects. On average, the MMF sources were found consistently more anterior, more medial, and more inferior than the corresponding sources of the standards. The MMF for /anni/ is detected in the time window from 370 to 450 msec and for /kiyo/ from 170 to 250 msec after the stimulus onset. The MMF is also detected for both /anni/ and /kiyo/ in the time window from 150 to 230 msec after the "point of uniqueness" of the stimuli, which is the time point within

the deviant stimulus from where the response is unequivocally identified as different from that of the standard stimulus (see Fig. 5).

The global field power (or RMS) of the rare deviants as compared with that of the frequent standard stimuli is shown in Figure 6 for all conditions and groups. Before and after training, Japanese subjects showed a "narrower" time window of the derived MMF amplitudes in the first and second MEG sessions as compared with the time window for the German subjects. The responses to duration deviants in /kiyo/ increased in German subjects after training, but showed a flatter, more scattered MMF in Japanese subjects as compared with German subjects. After only two training sessions, the MMFs of Japanese subjects were clearer and concentrated in a narrower time band.

The MMF was originally defined as a component resulting from the difference between responses to (frequent) standard stimuli and (rare) deviant stimuli. For statistical analysis, the peak RMS values of the subtraction waveforms were calculated within a latency window of 20 msec. A repeated-measure analysis of variance (ANOVA) was separately calculated for each group of subjects (German vs. Japanese) with the factors "training" (before/after), "phonetic category" (*anni/kiyo*) and "difference degree" (8, 6, 4) as the repeated within-subjects factors.

For German subjects, who performed the extended training, a significant overall effect of training ($F_{(1,9)} = 15.772$, $P = 0.0032$) was observed. The interaction of training and phonetic category (*anni/kiyo*) also reached significance ($F_{(1,9)} = 8.982$, $P = 0.015$). The main effect for the phonetic category (*anni/kiyo*) was highly significant ($F_{(1,9)} = 34.35$, $P = 0.0002$), and, typically for MMN studies, the difference degrees (8, 6, 4) evoked highly differential responses ($F_{(2,9)} = 71.11$, $P < 0.0001$). Within-group comparisons of the training effect (before/after) were significant for *anni8* ($F_{(1,9)} = 8.42$, $P = 0.017$), for *kiyo8* ($F_{(1,9)} = 10.25$, $P = 0.01$), *kiyo6* ($F_{(1,9)} = 13.43$, $P = 0.005$), and *kiyo4* ($F_{(1,9)} = 14.83$, $P = 0.003$). In conditions *anni6* and *anni4* the differences, though not significant, also showed an overall increase.

For Japanese subjects, only the difference degrees (8, 6, 4) reached significance ($F_{(2,9)} = 11.00$, $P = 0.0008$). No significant training effects were observed.

Comparisons between groups (German vs. Japanese) showed significant differences for *anni8* ($F_{(1,18)} = 5.99$, $P = 0.02$), for *kiyo8* ($F_{(1,18)} = 6.82$, $P = 0.017$), for *kiyo6* ($F_{(1,18)} = 7.76$, $P = 0.012$), and for *kiyo4* ($F_{(1,18)} = 5.64$, $P = 0.028$). Thus, in summary, German subjects showed higher MMF amplitudes after training than before training, whereas Japanese subjects (in four of six cases) started at a significantly higher MMF level and maintained this level throughout the training. The grand-averaged global field power of the MMF peak amplitudes in the left part of Figure 7 illustrates these effects.

The N1m component (at ~100 msec) disappeared almost completely in most subjects. This disappearance resulted from a smaller recovery time for the N1m because of the very small onset-to-onset ISI (900 ± 100 msec) and the remark-
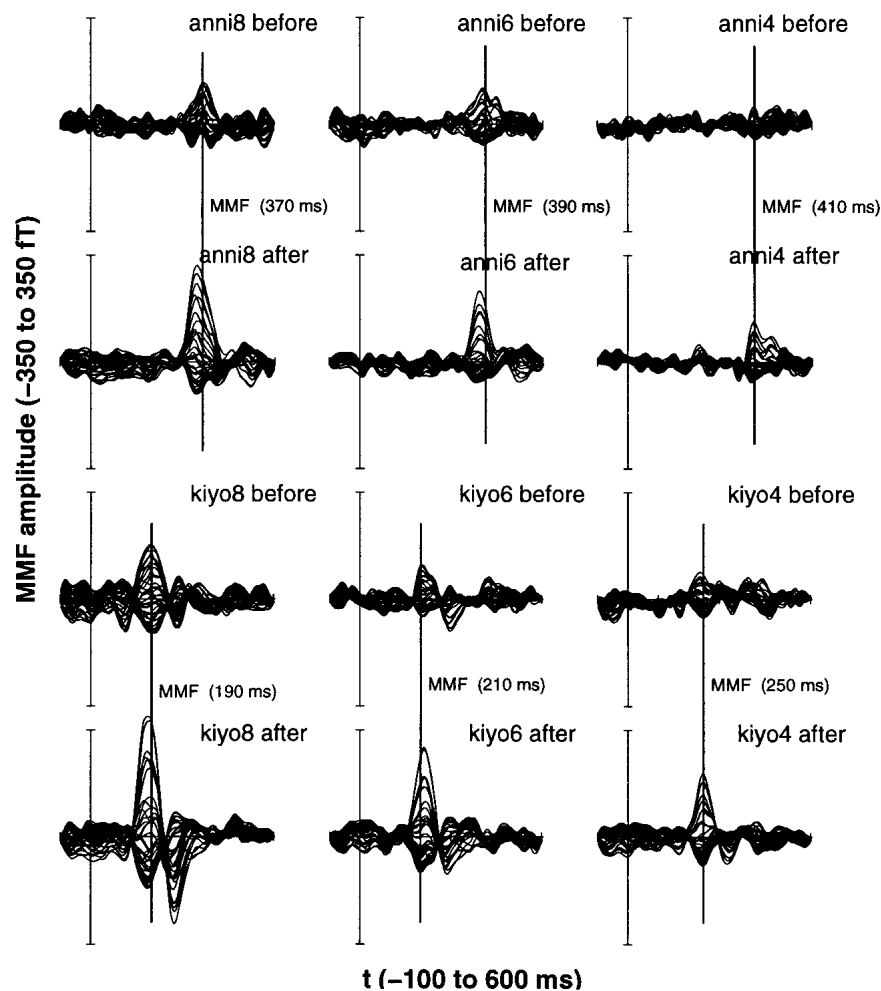

Figure 5 MMF amplitudes of one exemplary German subject. A remarkable increase of the MMF amplitudes is seen after training (second and fourth rows) compared with the amplitudes before training (first and third rows).
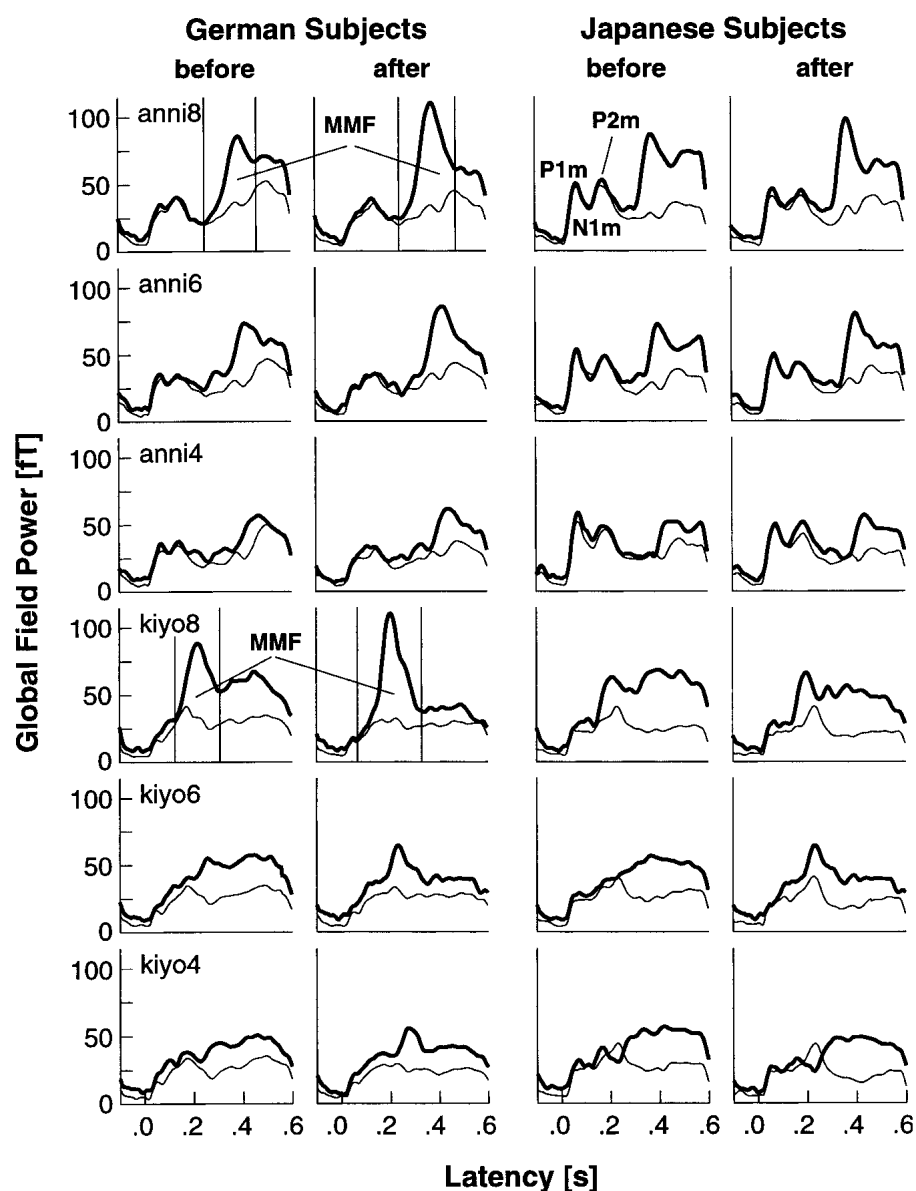
higher peak amplitudes than did the N1m. However, in the responses to larger differences (*kiyo8* and *kiyo6*), these peaks merged with the MMF response. For the responses to the small-difference *kiyo4*, the P1m–N1m–P2m complex was again, as in the *anni* conditions, more pronounced. Because the duration of the deviant stimulus was longer, the latency of MMF was later (at ~250 msec) and separated from the P1m–N1m–P2m complex.

Latency Differences

The same procedure for analyzing amplitudes was also used for analyzing latencies of the MMF (see Fig. 7). For German subjects, within-group comparisons of the training effect (before/after) were significant for *anni8* ($F_{(1,9)} = 20.85$, $P = 0.001$) and for *kiyo8* ($F_{(1,9)} = 18.18$, $P = 0.002$). In all other conditions (except *kiyo4*) there were small decrements in the latency of the MMF after training. For Japanese subjects, there were small decreases in latency for *anni8*, *kiyo8*, and *kiyo6*. Japanese subjects had shorter latency mismatch responses in each condition (except *kiyo4*) as compared with German subjects. These differences were very pronounced for *anni* but reached significance only in condition *anni8* ($F_{(1,18)} = 4.82$, $P = 0.04$).

DISCUSSION

In several studies (Dehaene-Lambertz 1997; Näätänen et al. 1997; Sharma and Dorman 2000), enhanced neurophysiological responses were found native compared to non-native speakers. These studies indicated that at a neurophysiological level, the mismatch negativity is differentially altered by lifelong linguistic experience with a particular language. On this basis, it was suggested that a language-specific phonemic code has a separate neural representation in sensory memory. In this study we showed the same result for mora contrasts. Japanese people perceive [l]
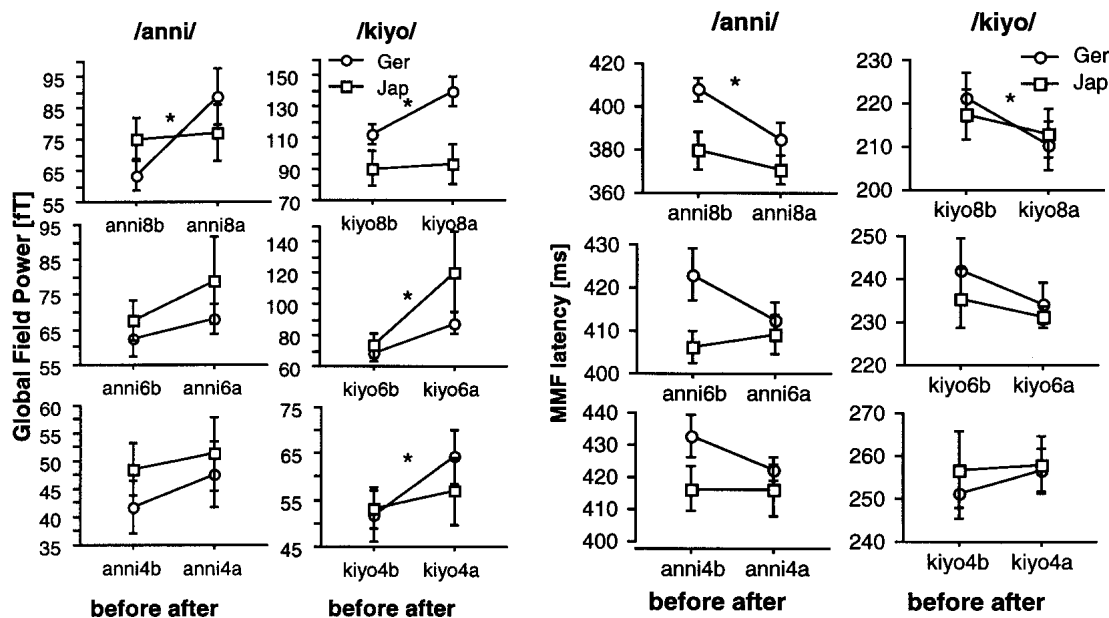

Figure 6 (*Left*) Grand averaged responses of German subjects before and after 10 d of training in three difference degrees of consonant and vowel duration. (*Right*) Japanese subjects before and after two training sessions. German subjects elicited higher responses before training than Japanese subjects and showed a remarkable improvement throughout training. In Japanese subjects the responses did not change much in the course of the training, but showed a narrowed MMF response compared with the German subjects. Thick lines denote the deviant, thin lines the standard response. The time domain marked by the two thin vertical lines denotes the area of the MMF.

able length of stimuli (400–505 msec for /anni/ and 250–355 msec for /kiyo/). An N1m-off at ~100 msec after the end of the stimulus also limited the refractoriness of the following N1m and, instead, a very clear P1m (at ~60 msec) and P2m (at ~160 msec) were elicited. The P1m and P2m for Japanese subjects in their responses to /anni/ were higher than the corresponding components for German subjects. In the responses to /kiyo/, the P1m and P2m also reached

Figure 7 (*Left*) Grand average of the Global Field Power of the MMF (mean, SEM) of German (circle) and Japanese (square) subjects before and after training for the consonant condition in /anni/ (*left*) and the vowel condition in /kiyo/ (*right*). Significant increases occur after training in conditions *anni8, kiyo8, kiyo6,* and *kiyo4.* In all (but *kiyo8*) conditions, the MMF of Japanese subjects started at larger values than that of German subjects. (*Right*) Mean (±SEM) latency differences of mismatch responses for German subjects (circle) and Japanese subjects (square) grand averaged over all subjects. Latency decreases in all conditions for German subjects, except for the *kiyo4* condition. Before training, Japanese subjects have mismatch responses at earlier latencies than German subjects in all conditions (except *kiyo4*) and maintain this without significant changes.

and [r] as allophones of the same sound, whereas Western people hardly place them in the same phonemic category (Goto 1971). In contrast, Japanese people have to differentiate, in their native language, durations between similar words and to extract where this difference actually occurs. For instance, they have to categorize instantaneously whether a perceived word consists of one mora or two (e.g., /o/ or /o:/) because the number of moras determines the word's meaning. Long-lasting experience with phonemes of slightly different durations is thought to facilitate perception of mora-timing. Therefore, we suggest, that the neuronal networks responsible for this ability are functionally well organized into representational maps, even possibly in phonemotopic maps.

The first question of this experiment was whether intensive daily training of non-native speech features, which leads to a remarkable shift of the phonetic boundary toward smaller differences, has the power to trigger reorganization within these cortical maps. The second question was whether native speakers show a different behavioral and neuromagnetic response to these stimuli than do non-native speakers.

Behavioral Results

The behavioral results showed clear learning effects across all stimuli. In the first test, the mean RT was at a similar level for both Japanese and German subjects. German subjects showed a markedly greater decrease in RT from the first to the second training than Japanese subjects. Both groups showed a smaller decrement of RTs for deviants than for standards. It is generally assumed that decreases in RT reflect the facilitation and automation of task performance, implying that additional processing is required for detecting a deviant.

The discrimination performance, as reflected by the hit rates obtained in tests without feedback, indicated that after a relatively short training time, subjects learn to discriminate phonemes differing only in small durational differences. The German subjects reached an improvement in performance up to 52% for the smallest duration difference after the last training as compared with their performance before the first training. The performance of Japanese subjects started in the same range as that of German subjects, but showed smaller improvements and a higher rate of errors than that of German subjects in the course of two training sessions. Apparently, the meaning of the stimulus interfered with the phonetic processing of the stimuli in Japanese subjects. Japanese subjects had to categorize the stimuli based on their phonemic and semantic content; whereas German subjects categorized phonetically different, but meaningless non-words. Their decision was based on a phonetic category boundary effect. The categorization

test (Fig. 2) showed that German subjects formed long/short categories out of the 8 variants of each stimulus. The training sessions alleviated this process and resulted in a more automated and accelerated processing that shifted the phonetic boundary toward the smaller variants (Fig. 4). Japanese subjects made their decisions based on phonetic and further (semantic) processing. This requires time-consuming and extensive reprocessing (Winkler et al. 1999).

Neurophysiological Changes

With regard to neurophysiological changes, training had an enhancing effect on the MMF, which reflects the brain's sensitivity to small differences in auditory input. Even small differences are perceived easier and faster and result in a higher activation of the underlying neuronal networks. MMF responses for Japanese subjects were more "narrow" than for German subjects, which may be an indicator that fewer subcomponents were required for the mismatch response. Moreover, most of the Japanese subjects had MMF responses of slightly higher amplitudes and earlier latencies than did German subjects.

In most subjects, the amplitude of the N1m component was diminished. However, the very clear appearance of the surrounding P1m and P2m components raised the question, whether the neuronal networks underlying these responses were also subject to plastic changes caused by the training. For /anni/, the P1m and P2m of Japanese subjects were higher than German subjects. Whereas there is no remarkable change between the two measurements in Japanese subjects, German subjects developed an explicit N1m component throughout training. This N1m component merged in the waveform of the global field power with the P1m and P2m components. In conclusion, speech-related plasticity of cortical structures seems to occur as a consequence of discrimination learning. This implies that the cortical representations of phoneme features can be remodeled dynamically through intensive training, even in adulthood. In this study, the observed effects of amplitude increment and latency decrement are similar to the results of frequency-discrimination learning reported previously (Menning et al. 2000). Intensive training improved the discrimination performance of the subjects and changed the neural representation of speech sounds as reflected by the automatic change-detection response (i.e., MMF). Assuming that a German listener can learn to perceive the duration quantity of phonemes as a matter of different phonemic categories, the shift of the category boundary to smaller differences—achieved in the training—is a clear behavioral indicator of the changes found in the neuromagnetic measures.

The main components (P1m, N1m, P2m) of the standard stimuli showed locations in secondary areas of the auditory cortex (parts of Heschl's gyrus and planum temporale), which are consistent with previous reports (cf. Csepe et al. 1992). The main sources of the MMF were found in the supratemporal plane (see Table 1). This result is in accordance with previous assumptions that the cerebral generators of the MMN/MMF have one source in the auditory cortex and one in the frontal cortex. The auditory component of the MMN/MMF is assumed to get its main contribution from the auditory memory trace. It is also assumed that this auditory component reflects the preperceptual deviance from the auditory memory trace. The other component located in the frontal lobe might be related to involuntary switching of attention to a stimulus change (Alho 1995). Results showed small, insignificant differences between conditions. However, the major generator in the temporal lobe could represent a more distributed network of neurons. Owing to the restricted area of our measuring unit, additional sources (e.g., in the frontal lobe) cannot be excluded. Furthermore, contributions of subcortical sources from the hippocampus and thalamus have to be reconsidered in further studies.

Based on the results of this study, we propose a more continuous view between perception and memory that allows for interactions between memory and sensory systems in the processing of small perceptual differences. If this processing does not differ between native and non-native speech perception, we can exclude a contribution of experience or memory to perceptual organization. However, there were significant differences between these two groups and we should, therefore, conclude that perceptual organization is affected by experience or memory. Because perceptual skills appear to be learned gradually and specifically, human subjects cannot generalize a perceptual discrimination skill to solve similar problems with different attributes. Typically, subjects start off with poor discrimination and improve substantially with training. But when tested with another stimulus with a new attribute, their performance starts again from the baseline. In other words, learning does not transfer from the first to the second attribute.

Further studies have to clarify whether a well-tuned neuronal network in a native speaker is able to reach better performance than an untrained network. The total number of cells responding to this specific feature could decrease as cells specialize during the training more and more for the trained feature. In the present study, semantic information was absent in the non-native group and present in the native Japanese group. The native group showed high automaticity of lexical processing, pointing out that fewer brain resources were required. The non-native group uses an explicit, more laborious processing system, whereas the native group processes the mora differences in a "flatter," more implicit procedural fashion.

Deficits in learning foreign languages as well as temporal processing deficits can be assessed with methods similar to the procedure used in these two experiments. A method is needed wherein subjects' discrimination abilities are

tested with small difference stages without feedback and trained with a staircase method of diminishing differences. Temporal-perception impairments are often associated with an inability to discriminate small temporal lags in speech. The training procedure presented here could help to improve their temporal resolution.

MATERIALS AND METHODS

Subjects

Ten German subjects (aged between 23 and 33 yr, mean age: 25.6 yr) and 10 Japanese subjects (aged between 21 and 40 yr, mean age: 26.5 yr) participated in this study. Nine of the subjects in each group were right-handed, and one in each group was left-handed. Each of the right-handed subjects showed a laterality score higher than 85% on the Edinburgh Handedness Questionnaire (Oldfield 1971). All German subjects and most Japanese subjects were students at the University of Münster with no history of neurological, otological, or speech disorders. None of the German subjects had any knowledge of a mora-timed or duration-sensitive second language. Japanese subjects grew up in Japan and had Japanese as their native language. In agreement with the Declaration of Helsinki, informed consent (approved by the Ethics committee of the University of Münster) was obtained from each subject after the nature of the experimental procedure had been fully explained to them.

Stimuli

The Japanese word stimuli were based on naturally spoken Japanese words and were generated with 8 degrees of duration for each critical vowel or consonant within one mora (cf. Fig. 8). These stimuli were produced at the Department of Speech and Cognitive Sciences at Tokyo University using the STRAIGHT algorithm (Kawahara et al. 1999)

For the consonants (/anni/–/ani/), the continuum was resolved into 8 different durations. The stimulus duration decreased from 505 msec to 400 msec, with the consonant /n/ decreasing in 15-msec steps. In the /itte/–/ite/ continuum, the silent period between the onset and offset of /t/ also decreased in 15-msec steps, resulting in 8 duration variants of /itte/ from 505 msec to 400 msec. Additionally, the duration of the vowel /i/ was manipulated in 15-msec steps. The continuum for /kiyo/–/kyo/ ranged from 355 msec to 200 msec and for /kiyou/–/kyou/ from 530 msec to 425 msec. All stimuli were low-pass-filtered at 4.5 kHz to eliminate high-frequency noise portions. The first and last 20 msec of the stimulus were smoothed with a cosine-shaped envelope to avoid sharp beginnings and endings.

The Japanese stimuli were chosen such that the first mora was intonated high (H) and the following moras low (HLL → HL). There was one mora difference between the long and the short variant of the stimulus. In geminates (/an-ni/, /it-te/, and /kiyo-

o/), the doubled phonemes were counted as two moras (Warner and Arai 2001). Because Japanese, in contrast to Western languages, has intonation shifting between a high and low level within the word and no stress on single syllables, the fundamental frequency for the high-intonated beginning vowel portion of all stimuli was chosen at ~300 Hz, decreasing to the low-intonated
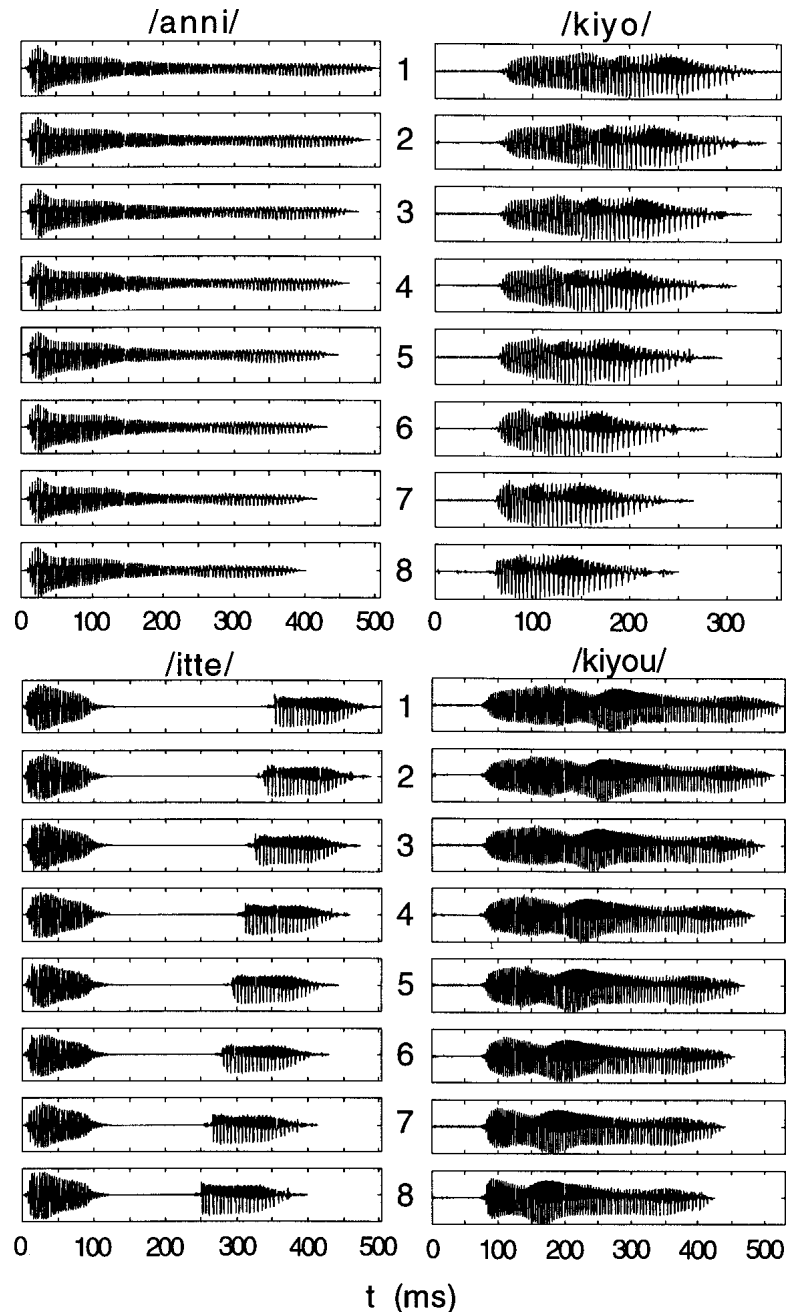

Figure 8 (*Left*) The duration of one mora in /anni/ and /itte/ is varied in 8 steps of 15 msec. (*Right*) The duration of one mora in /kiyo/ and /kiyou/ is varied in 8 steps of 15 msec. In MEG measurements, the differences 8, 6, and 4 of /anni/ and /kiyo/ were tested. During the behavioral discrimination training, all 8 differences were trained and tested.

Table 2. *Phonetic Structure and the Most Salient Meaning of the Stimuli*

Long variant	Intonation	No. of moras	Meaning	Short variant	Intonation	No. of moras	Meaning
/anni/	HLL	3	"Indirectly"	/ani/	HL	2	"Elder brother"
/itte/	HLL	3	"Roast it!"	/itc/	HL	2	"Shoot"
/kiyo/	HL	2	"Contribution"	/kyo/	H	1	"Nothingness"
/kiyou/	HLL	3	"Skill"	/kyou/	HL	2	"Today"

Left side: long variant of the stimulus continuum; right side: short variant of the stimulus continuum, H, high; L, low intonation.

ending vowel at ~150 Hz. The starting (high) fundamental frequency, the formants, and the amplitude envelope were kept constant for all 8 stimuli in each condition. The stimuli were sampled at 20 kHz with 16-bit resolution and were delivered during training at 70 dB$_{nHL}$ through AKG (K240-DF) studio headphones with circumaural shells.

For Japanese listeners the first (long) and the last (short) variant of each word formed separate perceptual categories, which corresponded to two meaningful words. Obviously, this was not the case for German listeners, who could not associate any meaning to the long or short variant. Table 2 gives an overview of the stimuli, their intonation, and their meaning.

All difference variants were used in the behavioral training and test sessions. In the MEG measurements, three difference variants of the word pairs /anni/–/ani/ and /kiyo/–/kyo/ were selected, where the duration of either a consonant or a vowel was varied. The long version of the stimuli served as the standard stimulus and differed from the large deviant (*anni8/kiyo8*) by 105 msec, from the middle deviant (*anni6/kiyo6*) by 75 msec, and from the small deviant (*anni4/kiyo4*) by 45 msec. The duration of /nn/ (long) in /anni/ ranged from 150 msec to 310 msec and of /n/ (short) from 150 msec to 205 msec. The duration of /i/ in kiyo ranged from 75 msec to 180 msec (long) and from 70 msec to 75 msec (short). Depending on these time relations, the MMF was expected to occur at ~100–150 msec after the "point-of-uniqueness" of the deviant

stimulus (i.e., the point in the stimulus from where the deviant unequivocally differs from the standard).

Discrimination Training

Ten German subjects were trained on 10 consecutive workdays for ~1.5 h each day. A self-adjusting staircase method with two-alternative forced-choice judgments and visual feedback was used for training. Stimulus pairs were distributed in a pseudorandom order with an interstimulus interval (ISI) of 1 sec and an interblock interval (IBI) of 2.5 sec. A stimulus pair consisted of either two standard stimuli or of a standard and a deviant stimulus. For training purposes, the probability of occurrence for a deviant was set at $P = 0.28$. The feedback interval was 0.2 sec (cf. Fig. 9). When the subject pressed the right mouse button, correctly classifying a pair of standard and deviant stimuli as "different," a green square was generated on the computer monitor, signaling a "hit" or "true positive." When the subject pressed the left mouse button falsely, classifying a standard and deviant pair as the "same," a red square was generated on the screen, signaling a "miss" or "false negative." Similarly, when a pair of two standard stimuli was correctly classified as the "same," a "correct rejection" or "true negative" was signaled by a green square. In the case that it was incorrectly classified as "different," a red square indicating a "false alarm" or "false positive" was shown as feedback (Fig. 9).

Standard stimuli were the long versions of the stimuli /anni/,
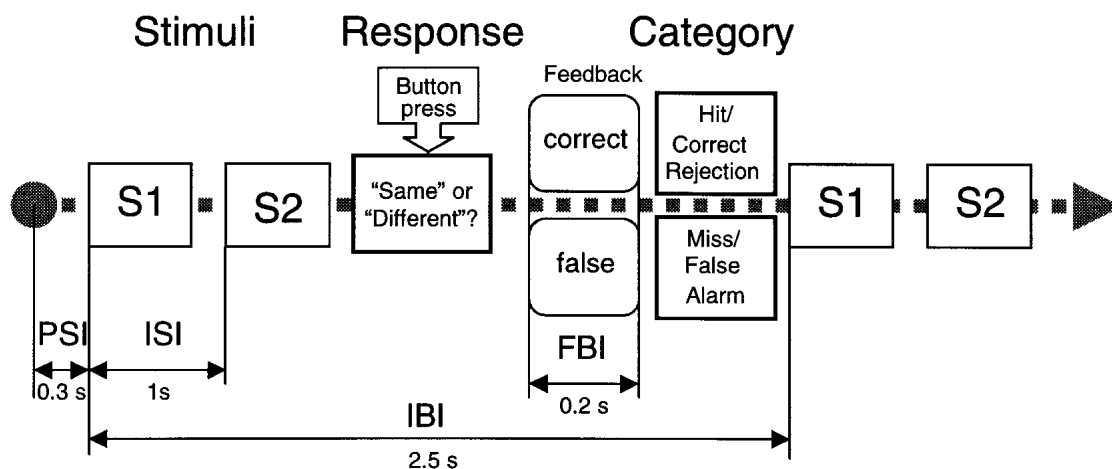

Figure 9 Stimulation diagram. Pairs of stimuli with an interstimulus interval (ISI) of 1 sec and an interblock interval (IBI) between pairs of 2.5 sec were distributed in a pseudorandom order with a probability of occurrence of a deviant, $P = 0.28$. The feedback interval (FBI = 0.2 sec) is followed by a short prestimulus interval (PSI = 0.3 sec) before the next stimulus pair. Subjects press a button if the pair of stimuli is perceived as "same" or a different button if the pair of stimuli is perceived as "different." Hits and correct rejections are fed back as a green color flash, whereas misses and false alarms are signaled as a red color flash.

/itte/, /kiyo/, and /kiyou/, whereas the deviant stimuli were one of the seven shorter versions (Fig. 1). The training started at the highest durational difference (105 msec) and continued with progressively smaller differences, corresponding to discrimination performance, until the smallest durational difference relative to the standard stimulus (15 msec) was reached. After a correct response to a deviant, the magnitude of the deviance was reduced in the next pair of stimuli by one step. After a false response to a deviant, the deviance magnitude was increased by one step. Each training performance was cross-validated by a discrimination test without feedback. This test consisted of 50% pairs of standard–standard stimuli of /anni/, /itte/, /kiyo/, or /kiyou/ and 50% pairs of shorter, standard–deviant stimuli with an equal amount of the seven difference variants. Thus, within a total of 210 pairs of stimuli, there were 105 pairs of standard stimuli and 105 pairs of deviant stimuli, with 15 stimuli from each difference variant. The test stimuli were delivered in a pseudorandomized order, and no visual feedback was provided. In these tests, subjects had to recognize the difference strictly by comparison with the standard. The timing of the stimulus presentation was the same as in the training session. Each test session began with a standard stimulus for orientation. After each training session, subjects were informed about their best performance in the session.

MEG Recordings

Magnetoencephalographic recordings were performed before and after training sessions with a 37-channel neuromagnetometer (Magnes, Biomagnetic Technologies) in an acoustically and magnetically shielded room. In each measurement session, three difference variants of each stimulus pair (/anni/–/ani/, /kiyo/–/kyo/), differing in length by 105 msec (*anni8*, *kiyo8*), 75 msec (*anni6*, *kiyo6*), and 45 msec (*anni4*, *kiyo4*) from standard, were presented. Each variant was presented three times in blocks of 300 sec to ensure test–retest reliability. The three identical blocks of one variant were then averaged. The resulting number of stimulus-related epochs was ~1000 for each condition. After exclusion of artifacts, one measurement session comprised, on average, 146 stimulus-related epochs to deviants and 698 epochs to standards. The MEG session for /anni/ lasted, with pauses, ~1 h and was followed by a short break, where the subjects walked around and relaxed. After this session, the MEG session for /kiyo/ was performed with the same timing. The probability of occurrence for deviants was $P = 0.15\%$, with a randomized stimulus-onset-asynchrony (SOA) of 900 ± 100 msec, and a minimum of 3 and a maximum of 9 standards before a deviant. The intensity of each stimulus was 70 dB_{SL} above the sensation threshold (SL), individually measured for each subject. Continuous MEG recordings were low-pass filtered at 0.1–100 Hz and sampled at a rate of 297.6 Hz. Stimulation was delivered monaurally to the contralateral side of the recorded hemisphere. The sensor array, with gradiometer-type coils (circular concave array with diameter of 144 mm and spherical radius of 122 mm), was centered above the T3 position of the international 10–20 system for EEG electrode placement for the left hemisphere and centered above T4 for the right hemisphere. A sensor-positioning system was used to determine the sensor location relative to the head and to indicate whether head movements occurred during the measurement. The subjects rested in a supine position supported by a vacuum cast to ensure a stable body and head position. They were instructed to stay relaxed and awake during MEG measurements by watching an animated video of their choice. This was expected to keep their state of arousal constant and attention continuously oriented toward the visual distraction. Compliance was verified by video monitoring.

Any stimulus-related epochs contaminated by eye blink or head movement artifacts (usually no more than 20% of the data) were rejected. Stimulus-related epochs of 700 msec with prestimulus intervals of 100 msec were selectively averaged, baseline-corrected, and low-pass filtered from 0 to 20 Hz with a second-order Bessel filter (no phase shift).

REFERENCES

Alho, K. 1995. Cerebral generators of mismatch negativity (MMN) and its magnetic counterpart (MMNm) elicited by sound changes. *Ear Hear* **16:** 38-51.

Alho, K., Connolly, J.F., Cheour, M., Lehtokoski, A., Huotilainen, M., Virtanen, J., Aulanko, R., and Ilmoniemi, R.J. 1998. Hemispheric lateralization in preattentive processing of speech sounds. *Neurosci. Lett.* **258:** 9-12.

Altenmüller, E., Kriechbaum, W., Helber, U., Moini, S., Dichgans, J., and Petersen, D. 1993. Cortical DC-potentials in identification of the language-dominant hemisphere: Linguistical and clinical aspects. *Acta Neurochir. Suppl.* **56:** 20-33.

Aoyama, K. 2001. A psycholinguistic perspective on Finnish and Japanese prosody: Perception, production and child acquisition of consonantal quality distinctions, pp. 7-16, Ph.D. thesis. Kluwer Academic Publishers, Norwell, MA.

Bradlow, A.R., Pisoni, D.B., Akahane-Yamada, R., and Tohkura, Y. 1997. Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *J. Acoust. Soc. Am.* **101:** 2299-2310.

Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., and Näätänen, R. 1998. Development of language-specific phoneme representations in the infant brain. *Nat. Neurosci.* **1:** 351-353.

Clark, J. and Yallop, C. 1995. *An introduction to phonetics and phonology,* pp. 91-99. Blackwell, Oxford, UK.

Csepe, V., Pantev, C., Hoke, M., Hampson, S., and Ross, B. 1992. Evoked magnetic responses of the human auditory cortex to minor pitch changes: Localization of the mismatch field. *Electroencephalogr. Clin. Neurophysiol.* **84:** 538-548.

Cutler, A. and Otake, T. 1994. Mora or phoneme? Further evidence for language-specific listening. *J. Memory & Lang.* **33:** 824-844.

Dehaene-Lambertz, G. 1997. Electrophysiological correlates of categorical phoneme perception in adults. *Neuroreport* **8:** 919-924.

Dehaene-Lambertz, G. and Pena, M. 2001. Electrophysiological evidence for automatic phonetic processing in neonates. *Neuroreport* **12:** 3155-3158.

Diesch, E., Eulitz, C., Hampson, S., and Ross, B. 1996. The neurotopography of vowels as mirrored by evoked magnetic field measurements. *Brain Lang.* **53:** 143-168.

Diesch, E., Biermann, S., and Luce, T. 1998. The magnetic mismatch field elicited by words and phonological non-words. *Neuroreport* **9:** 455-460.

Eimas, P.D. 1985. The perception of speech in early infancy. *Sci. Am.* **252:** 46-52.

Eulitz, C., Diesch, E., Pantev, C., Hampson, S., and Elbert, T. 1995. Magnetic and electric brain activity evoked by the processing of tone and vowel stimuli. *J. Neurosci.* **15:** 2748-2755.

Gootjes, L., Raij, T., Salmelin, R., and Hari, R. 1999. Left-hemisphere dominance for processing of vowels: A whole-scalp neuromagnetic study. *Neuroreport* **10:** 2987-2991.

Goto, H. 1971. Auditory perception by normal Japanese adults of the sounds "L" and "R". *Neuropsychologia* **9:** 317-323.

Grønnum, N. and Basboll, H. 2001. Consonant length, stod and morae in standard Danish. *Phonetica* **58:** 230-253.

Iivonen, A. 1998. Intonation in Finnish. In *Intonation systems: A survey of twenty languages* (eds. Daniel Hirst and Albert Di Cristo), pp 311-327. Cambridge University Press, Cambridge, UK.

Jaramillo, M., Ilvonen, T., Kujala, T., Alku, P., Tervaniemi, M., and Alho, K. 2001. Are different kinds of acoustic features processed differently for speech and non-speech sounds? *Brain Res. Cogn. Brain Res.* **12:** 459-466.

Jung, R., Altenmüller, E., and Natsch, B. 1984. Hemispheric dominance for speech and calculation: Electrophysiologic correlates of left dominance in left handedness. *Neuropsychologia* **22:** 755-775.

Kakehi, K., Kazumi, K., and Makio, K. 1996. Phoneme/syllable perception and the temporal structure of speech. In *Phonological structure and language processing* (eds. T. Otake and A. Cutler), pp. 126-143. Mouton de Gruyter, Berlin.

Kawahara, H., Masuda-Katsuse, I., and de Cheveigne, A. 1999. Reconstructing speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction. *Speech Commun.* **27:** 187-207.

Kraus, N., McGee, T., Carrell, T.D., King, C., Tremblay, K., and Nicol, T. 1995. Central auditory system plasticity associated with speech discrimination training. *J. Cog. Neurosci.* **7:** 25-32.

Kuhl, P.K. 1991. Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Percept. Psychophys.* **50:** 93-107.

———. 1994. Learning and representation in speech and language. *Curr. Opin. Neurobiol.* **4:** 812-822.

Kuhl, P.K., Williams, K.A., Lacerda, F., Stevens, K.N., and Lindblom, B. 1992. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* **255:** 606-608.

Ladefoged, P. 1993. *A course in phonetics.* Harcourt Brace, Fort Worth, TX.

Logan, J.S., Lively, S.E., and Pisoni, D.B. 1991. Training Japanese listeners to identify English /r/ and /l/: A first report. *J. Acoust. Soc. Am.* **89:** 874-886.

Menning, H., Roberts, L.E., and Pantev, C. 2000. Plastic changes in the auditory cortex induced by intensive frequency discrimination training. *Neuroreport* **11:** 817-822.

Näätänen, R. 1999. Phoneme representations of the human brain as reflected by event-related potentials. *Electroencephalogr. Clin. Neurophysiol. Suppl.* **49:** 170-173.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R.J., Luuk, A., et al. 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.

Pulvermüller, F., Kujala, T., Shtyrov, Y., Simola, J., Tiitinen, H., Alku, P., Alho, K., Martinkauppi, S., Ilmoniemi, R.J., and Näätänen, R. 2001. Memory traces for words as revealed by the mismatch negativity. *Neuroimage* **14:** 607-616.

Recanzone, G.H., Schreiner, C.E., and Merzenich, M.M. 1993. Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. *J. Neurosci.* **13:** 87-103.

Schröger, E. 1997. On the detection of auditory deviations: A pre-attentive activation model. *Psychophysiology* **34:** 245-257.

Sharma, A. and Dorman, M.F. 2000. Neurophysiologic correlates of cross-language phonetic perception. *J. Acoust. Soc. Am.* **107:** 2697-2703.

Shtyrov, Y., Kujala, T., Lyytinen, H., Kujala, J., Ilmoniemi, R.J., and Näätänen, R. 2000. Lateralization of speech processing in the brain as indicated by mismatch negativity and dichotic listening. *Brain Cogn.* **43:** 392-398.

Tervaniemi, M., Kujala, A., Alho, K., Virtanen, J., Ilmoniemi, R.J., and Näätänen, R. 1999. Functional specialization of the human auditory cortex in processing phonetic and musical sounds: A magnetoencephalographic (MEG) study. *Neuroimage* **9:** 330-336.

Warner, N. and Arai, T. 2001. Japanese mora-timing: A review. *Phonetica* **58:** 1-25.

Winkler, I., Lehtokoski, A., Alku, P., Vainio, M., Czigler, I., Csepe, V., Aaltonen, O., Raimo, I., Alho, K., Lang, H., et al. 1999. Pre-attentive detection of vowel contrasts utilizes both phonetic and auditory memory representations. *Brain Res. Cogn. Brain Res.* **7:** 357-369.

Yamada, R.A. and Tohkura, Y. 1992. The effects of experimental variables on the perception of American English /r/ and /l/ by Japanese listeners. *Percept. Psychophys.* **52:** 376-392.