Neural signatures of phonetic learning in adulthood: A magnetoencephalography study

Yang Zhang [a,*], Patricia K. Kuhl [b], Toshiaki Imada [b,c], Paul Iverson [d], John Pruitt [e], Erica B. Stevens [b], Masaki Kawakatsu [c], Yoh'ichi Tohkura [f], Iku Nemoto [c]

[a] Department of Speech-Language-Hearing Sciences and Center for Neurobehavioral Development, University of Minnesota, Minneapolis, MN 55455, USA
[b] Institute for Learning and Brain Sciences, University of Washington, Seattle, Washington 98195, USA
[c] Research Center for Advanced Technologies, Tokyo Denki University, Inzai-shi, Chiba 270-1382, Japan
[d] Department of Phonetics and Linguistics, University College London, London NW1 2HE, UK
[e] Microsoft Corporation, Redmond, WA 98052, USA
[f] National Institute of Informatics, Tokyo 101-8430, Japan

ARTICLE INFO

ABSTRACT

The present study used magnetoencephalography (MEG) to examine perceptual learning of American English /r/ and /l/ categories by Japanese adults who had limited English exposure. A training software program was developed based on the principles of infant phonetic learning, featuring systematic acoustic exaggeration, multi-talker variability, visible articulation, and adaptive listening. The program was designed to help Japanese listeners utilize an acoustic dimension relevant for phonemic categorization of /r–l/ in English. Although training did not produce native-like phonetic boundary along the /r–l/ synthetic continuum in the second language learners, success was seen in highly significant identification improvement over twelve training sessions and transfer of learning to novel stimuli. Consistent with behavioral results, pre–post MEG measures showed not only enhanced neural sensitivity to the /r–l/ distinction in the left-hemisphere mismatch field (MMF) response but also bilateral decreases in equivalent current dipole (ECD) cluster and duration measures for stimulus coding in the inferior parietal region. The learning-induced increases in neural sensitivity and efficiency were also found in distributed source analysis using Minimum Current Estimates (MCE). Furthermore, the pre–post changes exhibited significant brain-behavior correlations between speech discrimination scores and MMF amplitudes as well as between the behavioral scores and ECD measures of neural efficiency. Together, the data provide corroborating evidence that substantial neural plasticity for second-language learning in adulthood can be induced with adaptive and enriched linguistic exposure. Like the MMF, the ECD cluster and duration measures are sensitive neural markers of phonetic learning.

Published by Elsevier Inc.

Introduction

A fundamental question in cognitive neuroscience is the degree of neural plasticity as a function of age and experience. Classic studies and arguments on the putative "critical" or "sensitive" period for language acquisition highlight the superiority of learning a second language prior to puberty, and data support both maturation and experience as mechanistic explanations for the effect (Flege et al., 1999; Hernandez and Li, 2007; Johnson and Newport, 1989; Kuhl et al., 2008; Lenneberg, 1967; Mayberry and Lock, 2003). In the phonetic domain, there is clear evidence that early language learning does not involve a permanent loss of perceptual sensitivity to all the nonnative distinctions (Best et al., 2001; Werker and Tees, 2005). Furthermore, adults' perception of nonnative speech can be improved by using a variety of short-term intensive training methods (Akahane-Yamada et al., 1997; Bradlow et al., 1999; Hazan et al., 2006; Iverson et al., 2005;

Jamieson and Morosan, 1986; Logan et al., 1991; McCandliss et al., 2002; Pruitt et al., 2006; Strange and Dittmann, 1984; Tremblay et al., 1997; Wang et al., 2003; Zhang et al., 2000). These training studies, among others, have not only provided important empirical data for reevaluating the "critical period" hypothesis but also revealed key factors that facilitate second language learning independent of age. However, as epitomized by the classic problem of the /r–l/ phonemic contrast for adult Japanese speakers, neither intensive training nor prolonged naturalistic exposure has led to native-like mastery (Callan et al., 2003; Iverson et al., 2005; McCandliss et al., 2002; Takagi, 2002; Takagi and Mann, 1995). The experiential mechanisms that enhance or limit neural plasticity in adulthood are not well understood.

In our language acquisition model, adults' difficulty with nonnative languages stems from an early strong neural commitment to the statistical and spectral patterns in the language input during infancy (Kuhl et al., 2008). The effects of *native language neural commitment* (NLNC) are self-reinforcing and bidirectional — it enhances the detection of higher-order linguistic patterns, such as words, that utilize learned phonetic patterns, while at the same time hindering

the detection of non-conforming patterns contained in foreign languages, as shown behaviorally (Iverson et al., 2003) and neurally (Zhang et al., 2005). We further theorize that second language acquisition in adulthood can be improved by manipulating the language input to incorporate the basic principles underlying infants' acquisition of the sound patterns of their native language (Kuhl et al., 2001; Zhang et al., 2005).

To address the underlying mechanisms of brain plasticity for phonetic learning in adulthood, we designed a training software program in a preliminary single-subject MEG study to test its success (Zhang et al., 2000). The program incorporated features that were motivated by studies of infant-directed speech (IDS) or "motherese" (Burnham et al., 2002; Fernald and Kuhl, 1987; Kuhl et al., 1997; Liu et al., 2003), including adaptive signal enhancement, visible articulation cues, a large stimulus set with high variability, and self-initiated selection. The preliminary results suggested that rapid improvement could be achieved on the difficult nonnative phonemic contrast. Approximately 12 h of training for the Japanese adult subject showed an overall 22% improvement in identification accuracy with remarkable transfer of learning — there was a 27% improvement in recognizing the /r–l/ tokens by untrained voices. The training effect was also shown in enhanced neural sensitivity for the /r–l/ distinction, particularly in the left auditory cortex. Compared with other /r–l/ training studies with equivalent amounts of behavioral improvement, transfer of learning and sustained effect tested six months after training (e.g., Bradlow et al., 1999; Callan et al., 2003), our program reduced the total training hours by over 70%. As the preliminary results were based on a single subject, more subjects needed to be tested in order to evaluate the training methodology and investigate the neural mechanisms that reflect phonetic learning at both the individual and group levels.

There are two main objectives in the present training study: (a) to test the efficacy of our IDS-motivated training program in adults' learning of second language phonetic categories, and (b) to examine two hypothetical neural markers of learning in terms of brain-behavior correlates: *neural sensitivity*, as measured by the mismatch field response for phonetic discrimination (Näätänen et al., 1997), and *neural efficiency*, as measured by the focal degree and duration of brain activation during phonetic perception in terms of equivalent current dipole (ECD) clusters (Zhang et al., 2005). Previous neurophysiological studies have shown strong evidence of learning-induced enhancement in neural sensitivity to support phonetic categorization in adults as well as in children (Cheour et al., 1998; Imaizumi et al., 1999; Kraus et al., 1995; Menning et al., 2002; Näätänen et al., 1997; Nenonen et al., 2005; Rivera-Gaxiola et al., 2000; Tremblay et al., 1997; Winkler, 1999; Zhang et al., 2000). There is also evidence for learning-induced shift toward left hemisphere dominance in terms of enhanced neural sensitivity for linguistic processing (see Näätänen et al., 2007 for a review.). In line with the neural efficiency idea, fMRI studies have reported more focal activation for learned auditory stimuli particularly in native speakers or more advanced learners (Callan et al., 2004; Guenther et al., 2004; Wang et al., 2003). Cross-language MEG data have additionally indicated a shorter duration of bilateral activation for native speech processing in specific brain regions — the superior temporal and inferior parietal cortices (Zhang et al., 2005).

The central question of our study is whether substantial behavioral improvement in second language phonetic learning can be achieved in adulthood and simultaneously reflected by the spatiotemporal markers of neural sensitivity and neural efficiency, resulting in native-like perception and native-like brain activation patterns for learning the difficult speech contrasts in a second language. To cross-validate the brain activation patterns shown by the ECD cluster analysis approach and investigate the relationship between neural sensitivity and efficiency, we also employ distributed source analysis using

minimum current estimates with fundamentally different assumptions about the source activity (Uutela et al., 1999; Zhang et al., 2005). We predict that our IDS-motivated training program would help circumvent interference from neural networks that have been shaped by native language experience, yielding significant brain-behavior correlations in both domains of sensitivity and efficiency.

Materials and methods

Subjects

A pretest–intervention–posttest design was implemented to assess initial capability and the training effects. Nine right-handed Japanese college students (6 males and 3 females) participated in the study (21–23 in age). Subjects were volunteers under informed consent. They were recruited after screening for hearing, handedness, and language background. The subjects had no history of speech/hearing disorders, and all showed clear N1m responses to a 1000 Hz tone. All had received nine years of English as a second language education with limited exposure to spoken English. Seven of the nine subjects received training, and the other two subjects (KI and KO, one male and one female) who did not participate in the training were the top and bottom scorers in the pre-training behavioral assessment of /r–l/ identification. In effect, the seven trainees served as their own controls in pre–post tests. The two additional high-performing and low-performing controls who did not go through training were included in the test–retest to address two basic concerns: (1) whether spontaneous learning would take place from being exposed to hundreds of trials of /r–l/ stimuli, inducing significant learning effects, and (2) whether the MEG responses in the test–retest procedure would show large changes in the absence of training.

Training stimuli, program and protocol

The pre–post test sessions used auditory /r–l/ stimuli recorded from eight native speakers of American English (4 males, 4 females) producing five vowels (/a/, /i/, /u/, /e/, /o/) in the Consonant–Vowel (CV) and Vowel–Consonant–Vowel (VCV) contexts. Other English vowels were excluded to minimize compounding effects from unfamiliar vowels to Japanese listeners (Akamatsu, 1997). The training sessions used audiovisual /r–l/ stimuli from five talkers (3 males, 2 females) and three of the five vowels (/a/, /e/, /u/) in the two syllabic contexts (CV and VCV). The untrained auditory stimuli were included in the pre–post tests to assess transfer of learning. Adaptive training was implemented by using acoustic modification on the training stimuli with four levels of exaggeration on three parameters of the critical F3 transition for the /r–l/ distinction (Table 1) (Zhang et al., 2000). Specifically, the recorded tokens were submitted to an LPC (linear predicative coding) analysis–resynthesis procedure to exaggerate the formant frequency differences between pairs of /r–l/ tokens, and to reduce the bandwidth of F3. The LPC technique analyzed the speech signal by estimating the formants, removing their effects from the speech signal by inverse filtering, and estimating the intensity and frequency of the residual signal. Temporal exaggeration of the /r–l/ stimuli was made using a time warping technique — pitch synchronous overlap and add (Moulines and Charpentier, 1990).

Table 1
Levels of acoustic exaggeration on third formant for /r–l/ training tokens

Parameter exaggeration	Level 3	Level 2	Level 1	Level 0
F3 transition duration	4	2.5	1.67	1
F3 frequency	124%	116%	108%	1
F3 bandwidth	25%	50%	75%	1

Note. Minimum F3 difference = 80 Hz, Minimum F3 bandwidth = 250 Hz, Level 0 = naturally recorded tokens.

The training software program (see Supplemental Fig. 1 for a screenshot) incorporated the following key features:

1. *Self-directed listening.* Trainees selected the sounds by clicking on iconic buttons that indicated talkers and vowel/syllabic contexts.
2. *Visible articulation cues.* Photographic facial images and visual animation effects of each talker articulating /r/ or /l/ were provided for each sound presentation.
3. *Large stimulus sets with high variability.* A total of 120 different tokens were used for training. A message would prompt the participant to select a different icon if one icon had been clicked on for 20 times.
4. *Adaptive scaffolding.* There were 12 training sessions, starting from the most exaggeration sounds with one speaker only (see Table 2). Each session consisted of 10 listening blocks with 50 tokens in each block. Each listening block was followed by an identification quiz of 10 sounds to monitor progress. Difficulty level was increased when the quiz score was 90% or above. The scaffolding system worked by first adding talkers (up to five talkers) and then reducing exaggeration level (down to Level Zero).
5. *Feedback outside of listening blocks for training.* Each correct answer in the quiz accumulated monetary reward of two cents in US currency for the participant. Incorrect answers were prompted with a one-time playback of the sound. No feedback was given during the listening blocks in training or pre–post tests.

The training program was implemented using Macromedia Authorware on an Apple PowerPC. A STAX SR-404 Signature II electrostatic headphone system (model SRM-006t, STAX Limited, Japan) was used to ensure high-fidelity audio quality. The sounds were played back at approximately 70 dB SPL to both ears. The seven subjects completed 12 training sessions at their own pace in an acoustically treated booth over a two-week period. Each session lasted 45–60 min.

Pre–post behavioral tests

The trainees took the same behavioral and MEG tests under identical experimental settings one week before and one week after training. The two subjects who did not receive training took the pre- and post-tests in the same time frame spaced four weeks apart. Behavioral tests used natural as well as synthetic /r–l/ stimuli. Unlike the training sessions, no visual articulation cues were given in the pre–post tests. The natural stimuli recorded from all eight speakers in all vowel and syllable contexts were tested in four identification blocks. The synthetic speech syllables were tested using both identification and AX discrimination tasks to assess whether training produced native-like phonetic boundary in the /r–l/ continuum in the Japanese listeners (Zhang et al., 2005). The stimuli were a grid of /ra–la/ sounds created by using the Klatt SenSyn speech synthesizer (Sensimetrics, Inc, Massachusetts, MA) (Figs. 1a, b). The stimulus grid consisted of three levels (C1, C2, C3), each with different starting F2 values covering the range of 744–1031 Hz. Each continuum level contained six stimuli (1, 3, 5, 7, 9, and 11) that varied only in F3 transition with starting frequencies in the range of 1325–3649 Hz. The syllables were 400 ms in duration, containing an initial 155 ms of steady formants, a 100 ms F3 transition that ended in a steady third formant of 2973 Hz, appropriate for /a/. Additional details for the /ra–la/ synthesis can be found in Iverson et al. (2003). The identification task presented
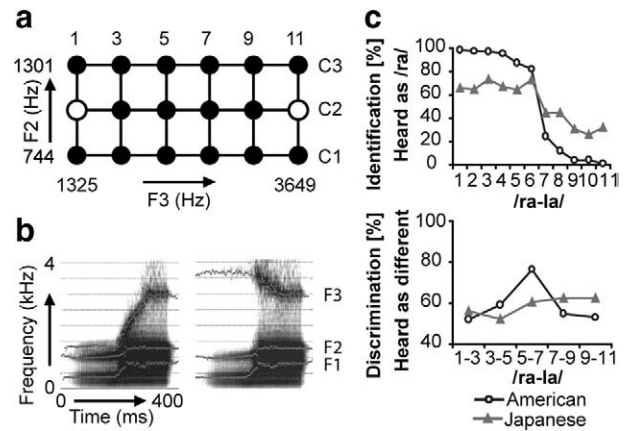

Fig. 1. (a) Schematic illustration of carefully controlled synthetic /ra–la/ stimulus grid. The physical distances for the interval steps on the second and third formants (F2 and F3) were respectively equated on the mel scale. The rows in the grid were labelled C1, C2 and C3 from bottom to top, indicating the three /ra–la/ continua that differ in F2. (b) Spectrograms of the endpoint stimuli in C2 continuum. F1 and F2 were kept the same and the critical /r–l/ difference was in F3. (c) Behavioral results from a previous cross-language study (Zhang et al., 2005): American listeners showed a clear phonetic boundary in the middle of the C2 continuum whereas Japanese listeners did not.

40 trials for each synthetic stimulus in random order. The AX discrimination task required subjects to judge whether pairs of stimuli were the same or different, using the middle row of the continuum (C2) so that the results could be compared to our previous cross-language study using the same stimulus set (Fig. 1c) (Zhang et al., 2005). The inter-stimulus interval for the discrimination pairs was 250 ms. Stimulus presentation was randomized, and both directions of presentation order were tested, each for 20 trials. An equal number of control trials were used, in which the same sound was presented twice, to assess false positives.

Pre–post MEG tests

The MEG measurements were conducted using the classic passive listening oddball paradigm (Näätänen et al., 2007). The stimulus pairs for the MEG tests were taken from the middle continuum (C2) of the /ra–la/ synthetic grid. Stimulus pair 1–11, which was recognized as prototypical /ra/ and /la/ syllables by American listeners (Zhang et al., 2005), was used to assess pre–post changes in neural sensitivity and efficiency. Two additional pairs, a cross-category pair 3–7 and within-category pair, 7–11, which had equated acoustic intervals on the mel scale, assessed whether the mismatch field (MMF) response could reflect native-like phonetic boundary effect.

The MEG measurement settings were identical to our previous study (Zhang et al., 2005). Prior to the MEG experiments, subjects were taken to an MRI facility (Stratis II, a 1.5T Superconductive Magnetic Resonance Imaging System, Hitachi Co, Japan) for structural brain imaging. Individual subjects' MRIs were used to construct head models for MEG-MRI co-registration and source analysis. The data were recorded using a whole-scalp planar 122-channel system (Neuromag Ltd, Finland) in a four-layered magnetically shielded room (Amuneal Manufacturing Corp., USA). Subjects read self-chosen books under the instruction to ignore the stimuli during the recording session. The stimuli were delivered at 80 dB SPL to the right ear via a non-magnetic foam earplug through a non-echoic plastic tube system. Stimulus presentation consisted of two consecutive blocks with the standard and the deviant reversed in the second block; the block sequences were counter-balanced among subjects. The deviant and standard presentation ratio was 3:17. The inter-stimulus interval was randomized between 800 ms and 1200 ms. The MEG signals were bandpass-filtered from 0.03 to 100 Hz and digitized at 497 Hz. Epochs with amplitude bigger than 3000 fT/cm or EOG bigger than 150 μV

Table 2
Initial setup for each adaptive session in the training program

Training sessions	1	2	3	4	5	6	7	8	9	10	11	12
Number of talkers	1	1	2	2	3	3	3	3	4	4	5	5
Exaggeration level	3	3	3	3	3	3	1	1	1	1	1	1

were rejected to exclude data with blinking/movement artifacts or other noise contamination. At least 100 epochs were averaged for each deviant stimulus.

Behavioral data analysis

Behavioral identification and discrimination data were first calculated in percent correct conversion. The overall percent correct measure for the AX discrimination task took into account all possible response categories (hits, correct rejections, misses, and false alarms), allowing the calculation of a bias-free estimate (d') of perceptual sensitivity using the signal detection theory (Macmillan and Creelman, 1991).

MMF analysis for assessing changes in neural sensitivity

MEG analysis procedure essentially followed our previous study (Zhang et al., 2005). The data were digitally low-pass filtered at 40 Hz, and the DC component during the prestimulus baseline of 100 ms was removed. Only the standards immediately before the deviant were used in the analyses to match the epoch numbers of the deviants and standards. To eliminate effects due to the inherent acoustic differences between standard and deviant, the MMF calculation used the identical stimulus when it occurred as the standard versus when it occurred as the deviant in a different oddball presentation block. For instance, the calculation of MMF for /ra/ subtracted the average of the /ra/ standard before /la/ deviant from the average of /ra/ deviants. Waveform amplitude was defined as the vector sum of amplitudes at two orthogonal channels in the same gradiometer sensor location. Individual subjects' head position changes in the MEG device were monitored, and standardized head position, which eliminated possible differences due to head position changes in pre–post tests, was used for the grand average.

In particular, group averages for standardized brain positions were obtained using the following procedure. First, the center of the subject's sphere approximating his or her own brain was moved to the origin of the device coordinate system. Second, x, y, and z axes of the subject's head coordinate system were made in parallel with the axes of the device coordinate system respectively. Third, the subject's head was rotated by 15° upward around x axis connecting from the left preauricular point to the right preauricular point. By this procedure, each subject's brain position was moved to the same standardized position. Fourth, the MEG waveforms were recalculated using MNE estimates employing the singular values up to the magnitude of one twentieth of the maximum singular value (Lin et al., 2006; Zhang et al., 2005). Fifth, the virtual waveforms from MNE estimates were averaged channel-wise across all subjects.

ECD cluster and duration analysis for assessing changes in neural efficiency

Unlike the MMF measure, the neural efficiency measures, as defined in Zhang et al. (2005), attempted to characterize cortical engagement for stimulus coding over a long time window without involving subtraction. Neural efficiency was quantified temporally in activation duration and spatially in the extent of focal activation. The shorter the activation duration is (presumably involving higher neural conduction speed and connectivity), the more efficient the neural system. Similarly, the smaller the extent of activation is (presumably involving less computational demand), the more efficient the system.

Specifically, we applied an extended ECD (equivalent current dipole) source localization analysis using a clustering approach to assess the number of ECD clusters and the cumulative ECD duration to assess neural efficiency. The feasibility of our approach using extended ECD analysis and ECD clustering was previously demonstrated in comparison with consistent results obtained from applying two

distributed source models on the same MEG dataset, the Minimum Norm Estimates, and the Minimum Current Estimates (Zhang et al., 2005). Unlike studies that modelled activities for peak components of interest such as MMF derived from subtraction, the ECDs in our extended analysis were sequentially fitted every 2 ms, under a small region of selected channels, for the averaged event-related field responses to the individual syllables (not the subtracted responses). In our extended ECD model, the time-dependent ECD activities were scattered in more than one brain region that was covered by the selected channel configuration. Instead of applying one configuration of channels for the left hemisphere analysis and one for the right for ECD modelling of peak activities, a total of 70 configurations of local channel selection were used for estimating the distributed ECDs over a time window of 20–700 ms. The number of channels in each configuration ranges as follows: 14 channels (11 configurations), 16 channels (33 configurations), 18 channels (25 configurations), and 20 channels (1 configuration) (see Supplemental Fig. 2 for details). These channel configurations were based on ECD simulations for our MEG system, which housed 122 planar gradiometer meters with a short baseline (16.5 mm) (Imada et al., 2001).

The procedure started by localizing a single ECD for a given time point using one of the 70 channel configurations. The ECDs were sequentially obtained every 2 ms for 680 ms from 20 ms after the stimulus onset for all the channel configurations. The ECD selection criteria were as follows: signal to noise ratio (or d') ≥ 2, continuous activation ≥ 10 ms, goodness-of-fit $G_{fit} \geq 80\%$, 95%-confidence-volume $V_{95} \leq 4188.8$ mm^3 (equivalent to a sphere volume with a radius of 10 mm). The locations of the selected ECDs were restricted to those directly beneath the selected channel area and further than 30 mm from the sphere center approximating the brain. Since our MEG system was the planar-type sensor system with a very low ambient noise level, the selected ECDs were found directly beneath the channel with the maximum magnetic amplitude in most cases. The selected ECDs were then grouped into spatial clusters, and the distance between any pair of ECDs at the same temporal sampling point or adjacent sampling points within each cluster was less than or equal to 20 mm. Within each spatial cluster, if there were at least 5 successive sampling points, approximately 10 ms given our sampling frequency at 497 Hz, the sum of continuous ECD activities by sampling points, was counted towards the activity duration of a single ECD cluster. The active sampling points duplicated by more than one ECD within a cluster were eliminated in ECD duration calculation. The duration of ECD activities within an ROI was quantified by a simple sum of the duration of each ECD cluster in the region. The focal extent of ECD activities within a region of interest (ROI) was quantified by the number of ECD clusters in the region. This quantification procedure for the ECD cluster and duration measures eliminated spatial and temporal redundancies of the ECDs that arose from the overlapping channel configurations (Supplemental Fig. 2).

To implement ECD clustering spatially restricted by ROIs instead of blind clustering using the spatial coordinates of x, y and z and the 20 mm separation distance parameter, anatomical areas were specified for ECD localization based on the individual subjects' MRIs. The superior temporal cortex (ST) was defined as the region below the lateral sulcus (inclusive) and above the superior temporal sulcus (not inclusive). The middle temporal cortex (MT) was defined as the region below the superior temporal sulcus (inclusive) and above the inferior temporal sulcus (not inclusive). The inferior parietal region (IP) was defined as the region posterior to the post-central sulcus (not inclusive), below the intraparietal sulcus (inclusive), above the line connecting the point where the lateral sulcus starts ascending and the anterior end point of the lateral occipital sulcus. The inferior frontal (IF) cortex was defined as the region below the inferior frontal sulcus (inclusive), anterior to the inferior precentral sulcus (inclusive), and posterior to the lateral orbital sulcus (not inclusive). Every ECD that passed the selection criteria was first plotted and visually inspected on

the individual subject's MRIs. The anatomical locations for each selected ECD were identified and verified by two experienced researchers. Each ECD was then manually entered on the spreadsheet for each individual subject based on their anatomical locations for clustering and further statistical analysis.

MCE analysis

To cross-validate the MEG activation patterns shown by our ECD modelling and clustering approach, we performed minimum current estimates (MCE, Neuromag MCE Version 1.4) using the minimum L1-norm (Uutela et al., 1999). Unlike the ECD approach with over-specified assumptions about the point-like source activity under selected channel configurations, the MCE algorithm searched for the best estimate of a distributed primary current and selected the solution with the smallest norm from all current distributions to explain the measured magnetic field. The MCE approach required no a priori assumptions about the nature of the source current distribution, which is considered to be more appropriate when the activity distribution is poorly known (Hari et al., 2000).

Specifically, the MCE analysis projected the solutions on the triangularized gray matter surface of a standard brain for visualization and statistical analysis (Zhang et al., 2005). Our MCE solutions assumed a minimum separation of electric current locations at 10 mm apart from each other. Thirty singular values were employed for regularization. The procedure included the following steps: (1) Realistic Boundary Element Models (BEMs) for individual subjects were constructed from their MRIs using Neuromag software. (2) The individual BEMs and their head origin parameters were used for the conductor model in the forward calculation. Estimates of individual signals were calculated at every sampling point for the original averaged waveform (not the subtracted waveforms of MMF) of each syllable stimulus. (3) The BEM of a standard brain provided by the MCE software was used as the standard source space for all subjects. The individual estimates were aligned on the standard brain space by applying a 12-parameter affine transformation and a refinement with a non-linear transformation based on the comparison of gray-scale values of the individual's MRIs and the standard brain. (4) Aligned estimates were averaged across subjects for visualization of MCE solutions at the group level. Integration of MCEs over the selected time window (20–700 ms) for group averages was performed for the entire brain. (5) Unlike our ROI definitions for ECD clustering, which used anatomical boundaries on each subject's MRIs, the spatial extent of each ROI in MCE analysis was operationally defined in an ellipsoidal shape superposed on the BEM of the standard brain. The spatial parameters for the center points of the ROIs were based on the x, y, and z coordinates for the selected regions (superior temporal, middle temporal, inferior parietal and inferior frontal) in our extended ECD analysis. The coordinates of the ROI center points were transformed to Talairach coordinates for the standard brain. (6) A weighting function was implemented to calculate the sum amplitude from the MCE estimates in the ellipsoidal space. The maximum weight was in the center of the ellipsoid, and the edges of the ROI were set at 60% weight of the maximum. (7) The analyzed MCE data for individual subjects were exported to Matlab for time-point-by-time-point statistical comparison between the standard and deviant stimuli as well as between pre-test and post-test for all the stimuli. (8) The MCEs for the four selected ROIs included current direction specification (either positive or negative for each time point). Because the negative estimates could not reflect increase vs. decrease in absolute amounts in the same way as positive estimates did, the MCEs for the ROIs were fully rectified before the statistical comparison.

Statistical analysis

Statistical analyses for both behavioral and MEG data used paired Student's *t*-test (two-tailed) and repeated-measures ANOVA. Brain-behavior correlates were performed using Pearson Correlation analysis for paired data. Where the MEG data exceeded the statistical assumption of normal distribution for performing the repeated-measures ANOVA tests, a multiplicative correction procedure was adopted using a Matlab program developed by colleagues at Research Center for Advanced Technology, Tokyo Denki University (Nemoto et al., 2008). Wherever applicable, the reported *p*-values were either *Bonferroni-adjusted* or Greenhouse–Geisser corrected.

Results

Behavioral effects on trained and untrained natural speech stimuli

The 12 training sessions produced a rapid and highly significant improvement of 21.6% in the trainees, an increase from 60.1% to 81.7% [two-tailed *t*-test, $p < 0.0001$] (Fig. 2a). Neither of the two subjects who did not receive training showed comparable changes (-1.3% for KO and $+2.8\%$ for KI) (Fig. 2b), and they were both statistical outliers against the trainee group [one-tailed *Z*-test, $p < 0.00001$]. The average improvements were 19.5% for the CV tokens and 23.7% for the VCV tokens (Fig. 2d), confirming the previous observation that the syllable-initial context of the /r–l/ contrast was harder for Japanese adults to learn (Logan et al., 1991).

The trainees showed steady progress as a function of the training sessions [$R = 0.973$ in linear regression analysis, $p < 0.0001$] (Fig. 2c), which attested the effectiveness of adaptive training. In contrast, there was no significant effect of stimulus block for the four blocks of stimuli in either the pre-test or the post-test, suggesting that there was no significant "spontaneous learning" due to increased familiarity with the stimuli during the pre- and post-tests.

Significant transfer of learning was found in the trainees (Figs. 2e, f). The /r–l/ identification improvement was 18.1% for the two untrained vowels [two-tailed *t*-test, $p < 0.001$, for each vowel] as compared to 23.9% for the three trained ones [two-tailed *t*-test, $p < 0.0001$ for each vowel]. Similar results were obtained with respect to the talker factor. There was a 19.9% improvement for the three untrained talkers in comparison with 22.6% improvement for the five trained ones. ANOVA results revealed significant between-subject differences [$(F(6,48) = 12.01, p < 0.00001$] and talker differences for the stimuli [$F(4,24) = 7.98, p < 0.01$ for the five trained ones; $F(2,12) = 9.88, p < 0.01$ for the three untrained ones], suggesting that Japanese listeners did not improve uniformly for the /r–l/ tokens recorded from different native English speakers. Nevertheless, the overall training success, including transfer of learning to novel stimuli, was not affected as the talker factor had no significant interaction with the pre–post measures.

Behavioral effects on untrained synthetic speech stimuli

Transfer of learning was examined in detail by manipulating two acoustic dimensions of the untrained synthetic stimuli to determine how Japanese trainees utilized F2 and F3 cues for /r–l/ identification before and after training. Repeated measures ANOVA on trainees' identification scores for the untrained synthetic stimuli (Figs. 3a–c) showed significant main effects on both acoustic dimensions in the stimuli [$F(2,12) = 93.78, p < 0.000001$ for F2 with three levels of difference; $F(5,30) = 13.42, p < 0.01$ for F3 with six levels of difference]. There were significant two-way interactions of F3 with training [$F(5,30) = 4.42, p < 0.05$], F2 with F3 [$F(10,60) = 4.55, p < 0.05$], and a marginally significant interaction of F2 with training [$F(2,12) = 3.32, p = 0.07$].

We hypothesized that specific stimuli on the stimulus grid, including those near the phonetic boundary on each F3 synthetic continuum, would indicate the effects of training. As expected, significant pre–post changes in stimulus identification [two-tailed *t*-test, $p < 0.05$] were found for stimuli 5 and 7 from the first continuum (C1 in Fig. 1a), and stimuli 5, 7 and 11 from the second continuum (C2
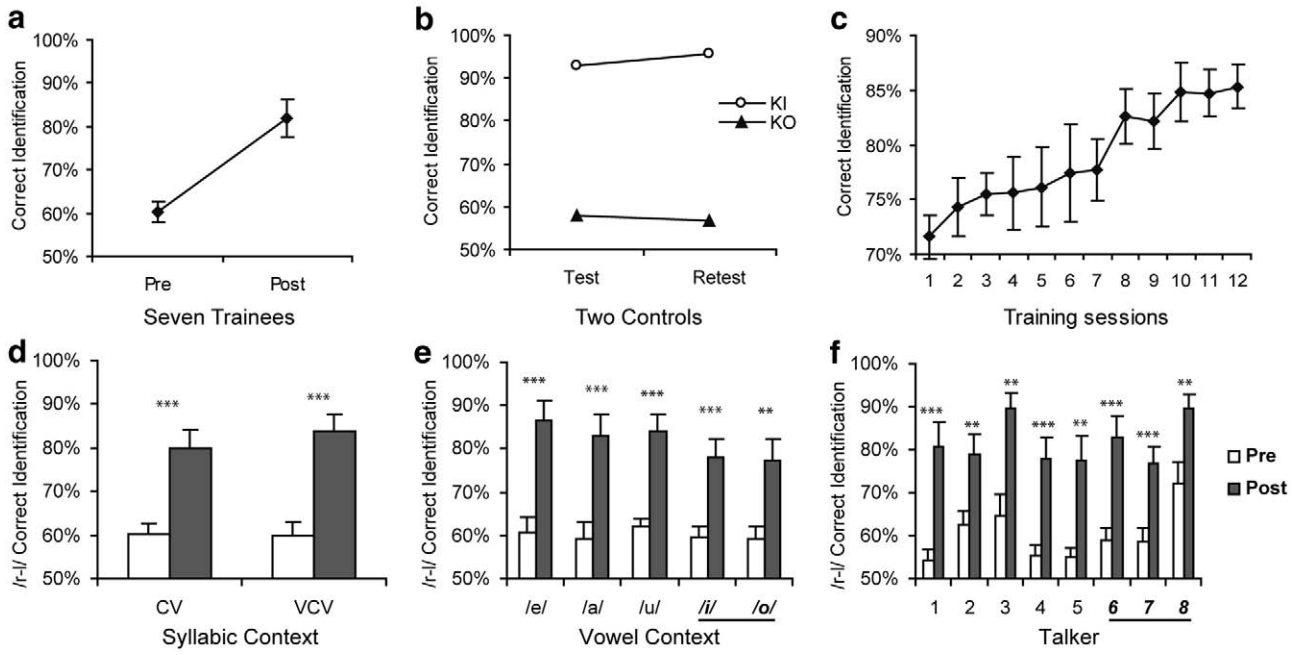
Fig. 2. Percent correct identification of the natural speech stimuli. (a) Overall pre–post improvement in the seven trainees. (b) Little pre–post changes in the two controls. (c) Quiz scores in the 12 training sessions. (d) Pre–post scores sorted by syllabic contexts for trainees. (e) Pre–post scores sorted by vowel contexts for trainees. (f) Pre–post scores sorted by talkers for trainees. Untrained tokens are marked in underlined bold italics on the x-axis. Standard error bars are marked where applicable [*** stands for $p<0.001$; ** $p<0.01$.].

in Fig. 1a). Marginally significant changes [one-tailed t-test, $p<0.05$] were found for stimuli 3 and 5 from the third continuum (C3 in Fig. 1a).

Discrimination scores for stimulus pairs on C2 further illustrated the impact of the training program designed to direct the trainee's attention to the critical F3 dimension for /r–l/ categorization (Figs. 3d, e). For stimulus pairs with small acoustic differences (1–3, 3–5, 5–7, 7–9, 9–11), the d′ scores were all below 0.9 before and after training, and there were no significant changes towards native-like categorical perception in the second language learners (Figs. 1c and 3d). However,

when the stimulus pairs involved larger acoustic separations, percent correct measures showed significant improvements for the best exemplars of r and l, the cross-category stimulus pair 1–11 [two-tailed t-test, $p<0.01$], and the within-category pair 7–11 [two-tailed t-test, $p<0.05$], but not for the cross-category pair 3–7 (Fig. 3e). Average d′ increased from 1.69 to 2.24 for the 1–11 pair and from 1.29 to 1.98 for the within-category pair, but the pre–post d′ values for the cross-category pair were both below 1.0. Although the short-term training program successfully led to increased sensitivity to F3
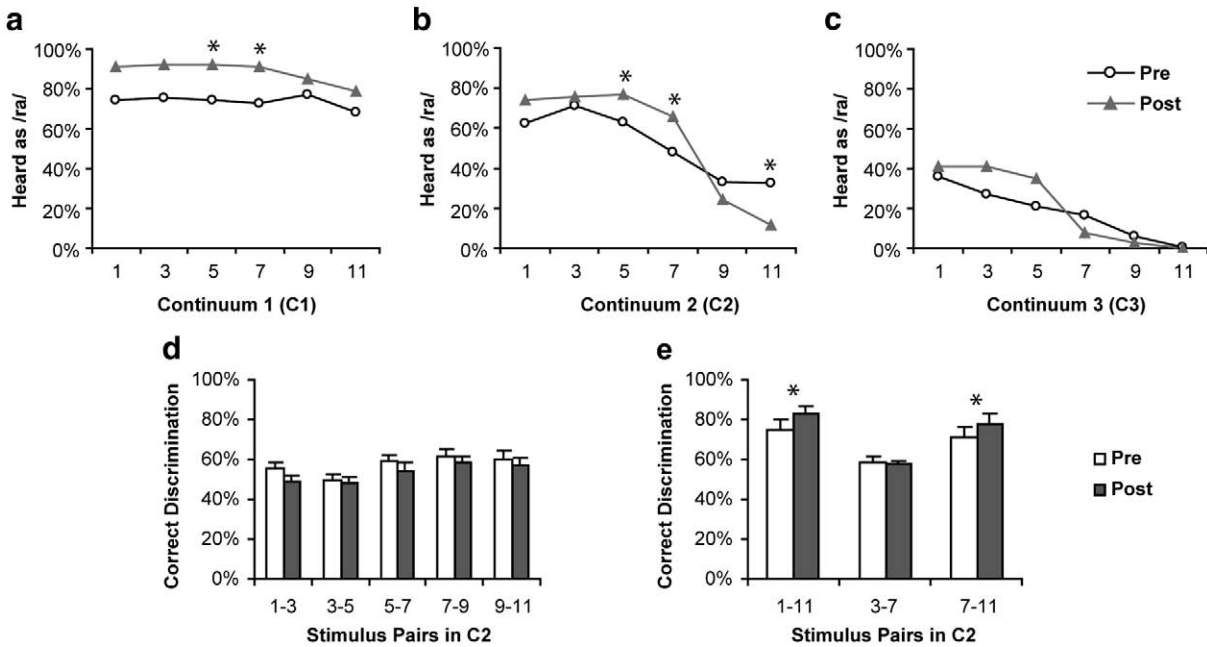

Fig. 3. Behavioral results of trainees' pre–post tests of the synthetic stimuli. (a) Identification scores for the C1 continuum stimuli. (b) Identification scores for C2 stimuli. (c) Identification scores for C3 stimuli. (d) Discrimination scores for stimulus pairs with small equated intervals in C2 continuum. (e) Discrimination scores for three stimulus pairs with larger intervals, prototype pair (1–11), cross-category pair (3–7) and within-category pair (7–11). The naming of prototype within- and cross-category pairs here was based on previous American data (Fig. 1c) [*$p<0.05$].

differences in the prototypical /ra/ and /la/ syllables, it did not induce a native-like "phonetic boundary" as seen in Fig. 1c.

Consistent with the behavioral results for the natural stimuli, the two subjects without going through the training sessions did not show changes of comparable size for the synthetic stimuli. The pre–post d' changes for stimuli in C2 for these two subjects were both statistical outliers in the discrimination task for the stimulus pairs 1–11 and 7–11 [one-tailed Z-test, $p<0.01$]. Unlike the low-performing control (KO), the high-performing control (KI) showed a perceptual pattern that resembled the American listeners' categorical perception of the synthetic /r–l/ continuum (Fig. 1 and Supplemental Fig. 3). Thus it was not impossible for Japanese adults to have native-like phonetic boundary for the /r–l/ continuum. This was also consistent with KI's pre-test score of above 90% accuracy for the natural speech stimuli (Fig. 2b).

Training effects in MMF responses for untrained synthetic stimuli

Training resulted in enhancement of the mismatch field responses for the untrained synthetic stimulus pair 1–11 [$F(1,6)=10.64$, $p<0.05$] (Figs. 4a, b). Head position changes between pre- and post-recording sessions were found to be statistically negligible. Standardized head positions in grand averaging further eliminated differences that could arise simply from head positioning. There was a significant interaction of hemisphere with stimuli [$F(1,6)=6.36$, $p<0.05$]. MMFs increased from 35.25 to 49.38 fT/cm in left hemisphere (LH) and 40.67 to 42.23 fT/cm in right hemisphere (RH), indicating that the training effect for this "prototypical" stimulus pair was primarily shown in the left hemisphere [post-hoc two-tailed t-test, $p<0.05$]. The trainees' average laterality index as expressed by (LH − RH) / (LH + RH) changed from −0.07 before training to 0.08 after training, indicating a group-level shift towards left hemisphere processing.

Two other pairs of untrained synthetic stimuli (within- and cross-category pairs) were tested to assess the training program's success in producing a native-like phonetic boundary (Figs. 4c, d). Consistent with the behavioral results (Fig. 3e), significant MMF enhancement

was found for the within-category pair 7–11 [$F(1,6)=9.15$, $p<0.05$] but not for the cross-category pair 3–7. Unlike the left-dominant MMF changes observed for stimulus pair 1–11, the pre–post changes in the MMF for 7–11 were dominant the right hemisphere [post-hoc two-tailed t-test, $p<0.05$]. These results confirmed transfer of learning to untrained synthetic stimuli, primarily a left-hemisphere effect for the cross-category stimulus pair 1–11 and a right-hemisphere effect for within-category stimulus pair 7–11. Consistent with the behavioral results, the MMF data did not reveal a native-like phonetic boundary as a result of training. Instead of increased amplitude, a slight reduction in MMF was found in the two control subjects who did not go through training. Again, both subjects were statistical outliers in their MMF amplitude changes when pooled against the trainee group for the 1–11 pair and 7–11 pair [one-tailed Z-test, $p<0.001$].

Training effects in ECD cluster and duration measures

For the seven trainees, phonetic training resulted in overall bilateral decreases in ECD cluster and duration across all four regions of interest in the neural representation of the prototypical stimuli 1 and 11 (Fig. 5) but not for the perceptually more ambiguous stimuli, 3 and 7, in the /ra–la/ continuum C2. The results confirmed our hypothesis that training would reduce the focal extent and duration of cortical activation for coding the prototypical /ra/ and /la/ sounds (Zhang et al., 2005). Significant main effects of region were observed for both ECD measures for the prototypical stimuli [$F(3,18)=5.83$, $p<0.01$ for ECD clusters; $F(3,18)=4.63$, $p<0.05$ for ECD duration]. Specifically, the reduction in ECD duration was significant in the inferior parietal (IP) region bilaterally [$F(1,6)=8.15$, $p<0.05$] (Fig. 5c). Due to large inter-subject variability, effects were non-significant in the medial temporal (MT), superior temporal (ST) and inferior frontal (IF) regions. No significant hemisphere effect was found in any of the regions. It was striking that six out of seven trainees showed ECD reduction in various degrees. The one subject who showed ECD increase retrospectively reported a violation of the experimental condition by frequently attending to the occurrence of the oddball
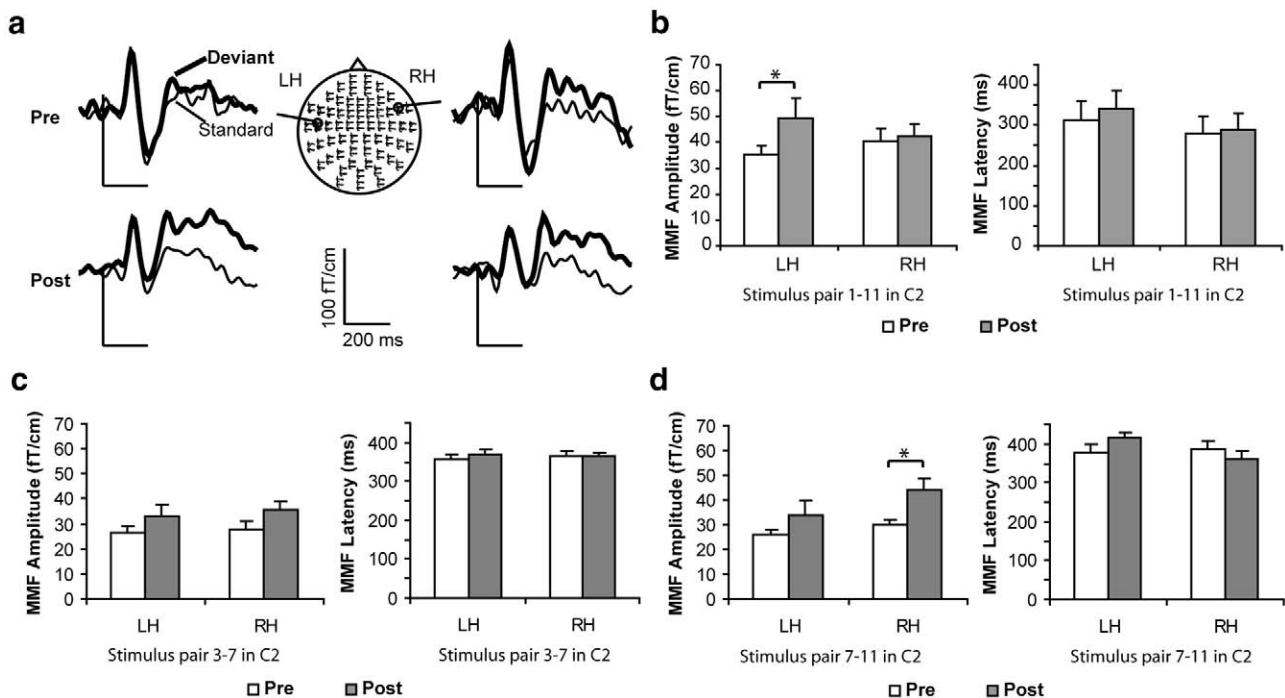

Fig. 4. Trainee's mismatch field (MMF) responses of pre- and post-tests for the synthetic prototype stimuli. The naming of prototype stimuli here was based on previous American data (Zhang et al., 2005). (a) Waveform responses from one representative subject for the stimulus /la/, showing increased MMF due to training. (b) MMF amplitudes in group averages pooled for the /ra/ and /la/ stimuli. (c) MMF latencies in group averages pooled for the /ra/ and /la/ stimuli. LH = left hemisphere, RH = right hemisphere [*$p<0.05$].
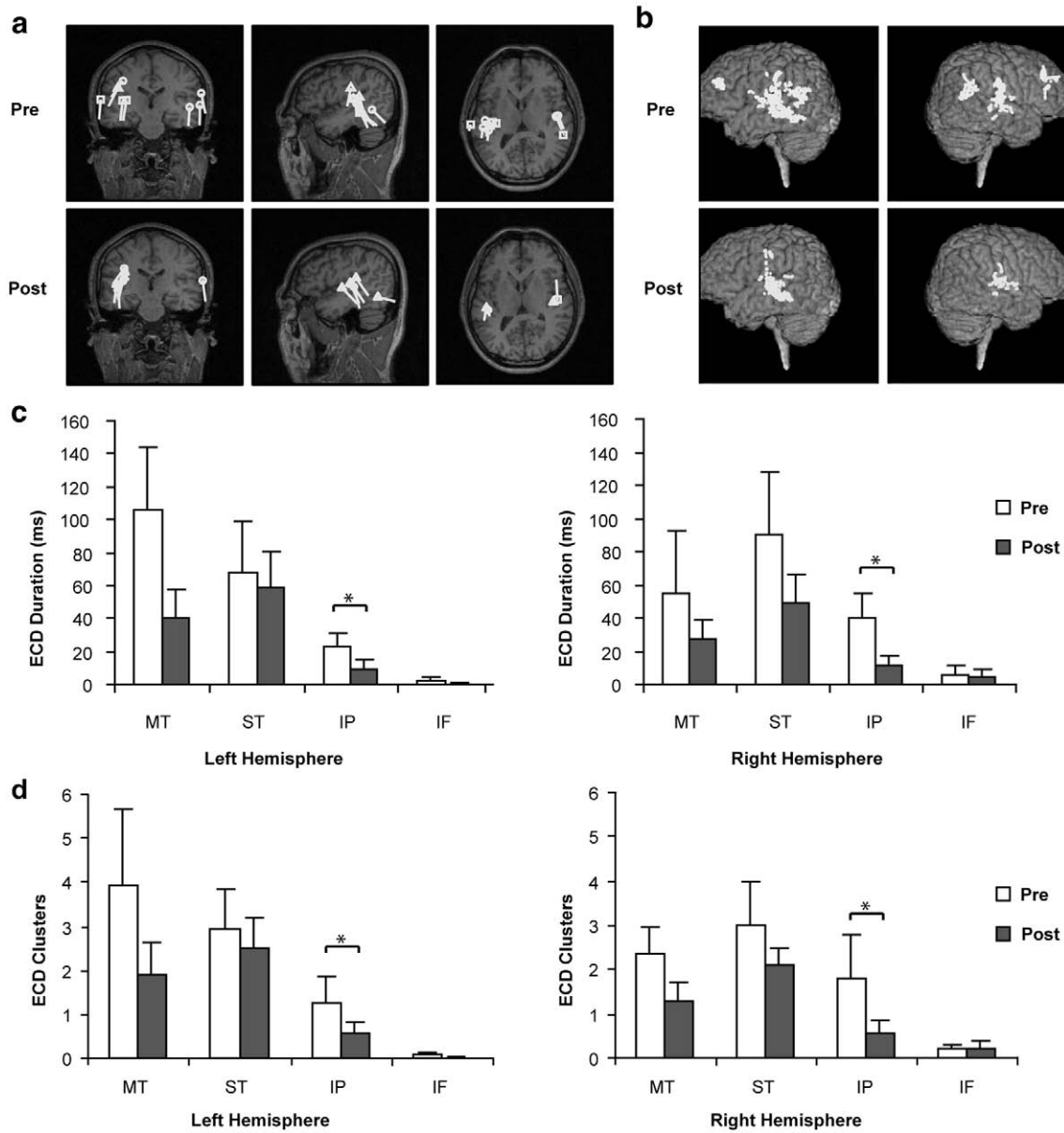
Fig. 5. Neural efficiency measures of pre- and post-tests for the synthetic prototype stimuli (stimuli 1 and 11 on the C2 continuum in Fig. 1). (a) ECDs of one spatial configuration on the coronal, sagittal, and axial planes from one representative subject. (b) Cortical surface renditions of all the ECDs from the same subject. The plots in (b) could not properly convey the ECD depth information as seen in (a). (c) Number of ECD clusters in the middle temporal (MT), superior temporal (ST), inferior parietal (IP) and inferior frontal (IF) regions. (d) ECD durations in the four cortical regions [*$p < 0.05$].

stimulus in the post-training test for the 1–11 pair. Nevertheless, inclusion of this aberrant data point did not change the overall significant reduction in both ECD cluster and duration measures in the IP region at the group level. Neither of the two control subjects who did not go through training showed similar reductions in the ECD measures in the IP region [one-tailed Z-test, $p < 0.01$] (Supplemental Fig. 4).

Training effects in brain-behavior correlations

Individual trainees were highly variable in tests using the untrained synthetic stimuli (Fig. 1a). The existence of the large variability in a small subject pool allowed us to examine brain-behavioral correlations. All seven trainees showed increased neural sensitivity to the synthetic 1–11 pair, and a significant positive correlation was observed between training-induced behavioral d' changes and overall MMF amplitude changes

(averaged across the two hemispheres) for the best synthetic /ra–la/ exemplars [Pearson's $r = 0.78$, $p < 0.05$] (Fig. 6a). Similarly, despite the existence of one aberrant data point in terms of training-induced changes in the ECD cluster measures, there were significant negative correlations between pre–post behavioral d' changes and ECD cluster changes (Fig. 6b) and between pre–post behavioral d' changes and ECD duration changes (Fig. 6c). The ECD cluster and ECD duration measures exhibited striking similarity (Figs. 5c, d and 6b, c) with highly significant correlations [Pearson's $r = 0.98$, $p < 0.0001$].

Training effects in MCE analysis

Overall, the MCE results for the trainee group showed brain activation patterns consistent with the MMF results and ECD clustering analysis (Figs. 7 and 8). First, significant mismatch field responses in total MCE activity for the deviant and standard stimuli
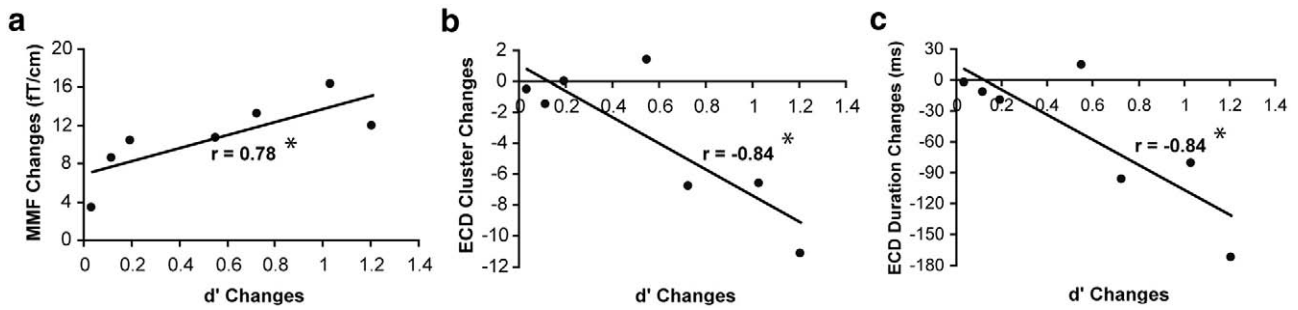
Fig. 6. Brain-behavior correlations in training-induced plasticity. (a) Positive correlation between *d′* changes and pooled mismatch field responses changes. (b) Negative correlation between *d′* changes and pooled ECD cluster changes. (c) Negative correlation between *d′* changes and pooled ECD duration changes [*$p < 0.05$].

were observed for both the pre-test and post-test, and the MMF was enhanced and extended over a longer time window after training (Fig. 8a). Second, despite the existence of enhanced MMF, significant reduction in total MCE activity was observed for both the deviant and standard stimuli (Fig. 8b). Third, ROI results revealed dominant MCE

activities in the superior temporal regions of both hemispheres (Figs. 7 and 8c). There was pre–post reduction of MCE activities over relatively long time windows in the ST, MT and IP regions (Figs. 8c, d, e).

Unlike the ECDs with sparse time points for significant dipole activities as a result of the stringent ECD selection criteria, MCE
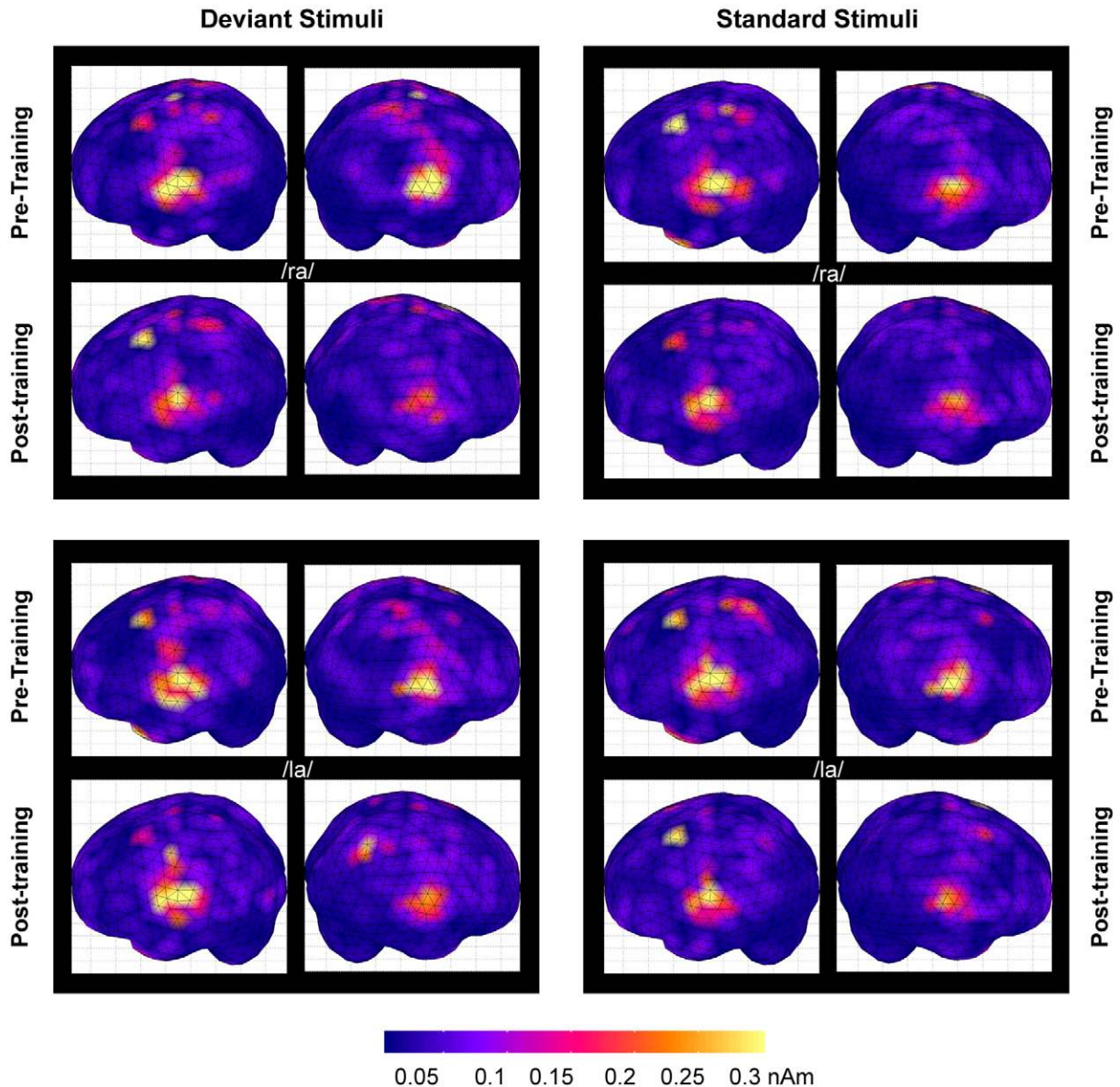

Fig. 7. Minimum current estimates (MCEs) plotted on the standard brain surface for each stimulus in the pre-test and post-test MEG experiments for the prototypical /ra/ and /la/ stimuli. Samples between 20 and 700 ms were integrated and averaged for the trainee group.
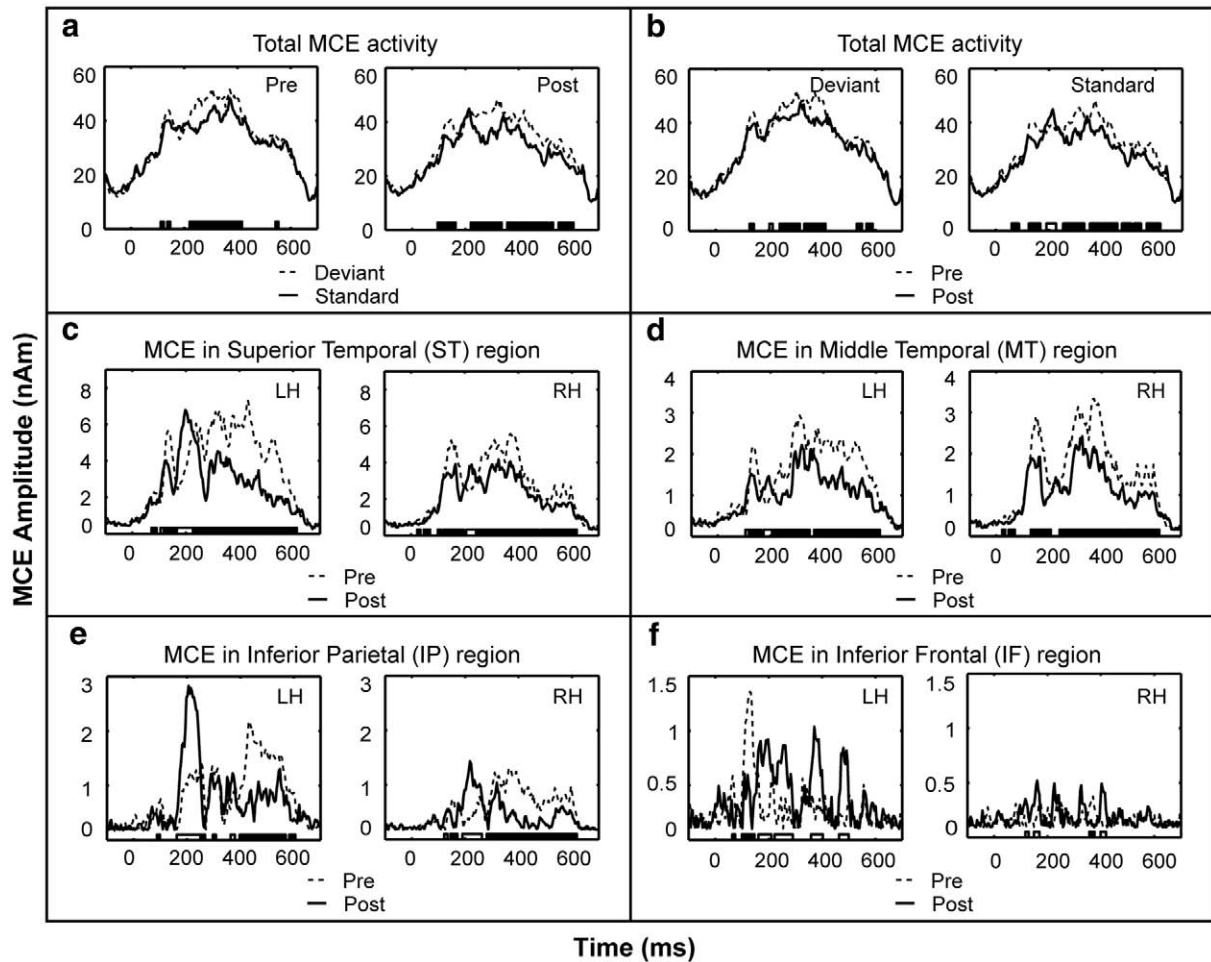
Fig. 8. Combined minimum current estimates plotted at each time point for the window of −100–700 ms. (a) Total MCE activity for the deviant vs. standard contrast in the pre-test and the post-test. Training-induced MMF enhancement was observed in the same latency window (200–400 ms) as in Fig. 4b. (b) Total MCE activity for the pre–post comparison respectively for the deviant stimuli and standard stimuli. (c) Pre–post MCEs for the Superior Temporal (ST) regions in the left hemisphere (LH) and right hemisphere (RH). (d) Pre–post MCEs for the Middle Temporal (MT) regions in the left and right hemispheres. (e) Pre–post MCEs for the left and right Inferior Parietal (IP) regions. (f) Pre–post MCEs for the left and right Inferior frontal (IF) regions. Significance for time-point-by-time-point comparisons was shown by the black and white bars on the x axis. The black bars on the x axis showed time windows of significant MCE reduction (solid curve relative to the dotted curve), and the white bars showed time windows of significant increase [$p < 0.01$]. Only windows equal to or longer than 10 ms were shown in the significance bars.

calculation preserved the entire temporal scale for all time points of interest, including the baseline of −100–0 ms. Unlike the ECD results, pre–post time-point-by-time-point differences revealed significant increases of total MCE activity at around 200 ms, approximately 12 ms of increased activities for the deviant stimuli and approximately 26 ms for the standard stimuli (Fig. 8b). This enhanced MCE activity in the post-training test was also seen in the ROI data in the left hemisphere (Figs. 8c, d, e, f) as well as in the right hemisphere (Figs. 8c, e). Nevertheless, the overall MCE activities in ST, MT, and IP showed training-induced reduction in both hemispheres except the IF region (Fig. 8f). Relatively speaking, the MCE amplitude in the IF region was the smallest in scale among the four ROIs; therefore, the significantly increased pre–post activities at latencies after 234 ms in the left and right IF regions (Fig. 8f) were not shown in the total MCE activity (Fig. 8b).

Discussion

Successes and limitations of the IDS-motivated training program

The behavioral data provided strong evidence of success for our training program. Japanese listeners achieved a 21.6% improvement in identifying naturally spoken /r–l/ syllables in 12 training sessions,

which generalized to the untrained voices, vowel contexts and synthetic prototype stimuli (Fig. 2). This is remarkable given that many Japanese adults continue to misperceive and mispronounce the English /r–l/ sounds after months of laboratory training or years of residence in the US (Takagi, 2002). It is also noteworthy that our /r–l/ training program shortened the training period for the Japanese listeners by more than 70% for equivalent effects reported in other studies (Bradlow et al., 1999; Callan et al., 2003). Our behavioral data did not show "spontaneous learning" from mere exposure to the stimuli in the test and retest procedure; rather, behavioral improvement (Fig. 2c) illustrated a steady progressive trajectory in the adaptive training sessions. Although a much faster rate of improvement was previously reported in adaptive training based on the Hebbian learning model (McCandliss et al., 2002; Tricomi et al., 2006; Vallabha and McClelland, 2007), their observed effects were of smaller size in terms of transfer of learning as well as of limited scale in comparison to natural speech stimuli with the five vowel contexts, two syllabic contexts, and eight talkers and synthetic stimuli with systematic variations on the F2 and F3 dimensions used in our study.

The tests using synthetic stimuli testified the importance of analyzing contributions of different acoustic dimensions in understanding the nature of second-language phonetic learning in adulthood. While impressive gains were obtained in the trainees, our results also indicated that 12 h of training did not produce a native-

like phonetic boundary in nonnative listeners as shown in both behavioral results (Figs. 1 and 3) and MMF results (Fig. 4).

Training succeeded in improving listeners' sensitivity to the critical F3 dimension as expected. However, signal treatment (formant frequency separation between /r/ and /l/, formant bandwidth, and duration) in the F3 dimension alone did not prevent native-language interference from the F2 dimension (Fig. 3). More specifically, Japanese listeners appeared to judge the /r–l/ distinction analogous to their native /r–w/ contrast based on the F2 onset frequency and the amount of frequency separation between F2 and F3 (Iverson et al., 2005; Lotto et al., 2004; Sharf and Ohde, 1984; Zhang et al., 2000). This result should not be surprising given the following facts: (a) behavioral data show that nonnative listeners find it difficult to attend to new acoustic parameters for phonetic categorization (Best and Strange, 1992; Francis and Nusbaum, 2002; Underbakke et al., 1988), (b) models and theories predict its difficulty (Best et al., 2001; Kuhl et al., 2008; Vallabha and McClelland, 2007), and (c) even highly proficient second language learners do not attain the same level of perceptual competence as native listeners (Bosch et al., 2000; Gottfried, 1984; Guion et al., 2000; Pallier et al., 1997; Sebastián-Gallés and Soto-Faraco, 1999; Takagi and Mann, 1995).

Of particular interest to theory of language acquisition, the infant-directed speaking style, providing greater acoustic exaggeration and variety than adult-directed speech, may facilitate the formation of prototypical representations of a phonetic category, which can be predictive of high-order linguistic skills (Kuhl et al., 2008). The essential aspects of infant-directed speech (IDS, commonly referred to as "motherese") are found across cultures: (1) exaggeration of spectral and temporal cues for phonetic categories as well as for pitch variations in terms of range and contour shape (Burnham et al., 2002; Fernald and Kuhl, 1987; Kuhl et al., 1997; Liu et al., 2003), (2) high stimulus variability in multiple speakers and phonetic contexts (Davis and Lindblom, 2001; Katrin and Steven, 2005; Werker et al., 2007), (3) visual exposure to the face of the talker and articulatory motion, which encourages cross-modal sensory learning and the binding of perception and action through imitation (Kuhl and Meltzoff, 1982, 1996), (4) naturalistic (or unsupervised) listening of a large number of tokens statistically separable in the distribution of acoustic cues without demanding overt identification or discrimination responses (de Boer and Kuhl, 2003; Katrin and Steven, 2005; Maye et al., 2002; Vallabha et al., 2007), and (5) listener-oriented adaptive delivery that helps focus on and hold attention to the critical speech features (Englund and Behne, 2006; Fernald and Kuhl, 1987; Kuhl et al., 2003; Uther et al., 2007). The idea of IDS-based input manipulation in aiding second language acquisition is consistent with other models of speech learning although not all of the models incorporate a developmental perspective (Best et al., 2001; Escudero and Boersma, 2004; Flege, 1995; McClelland, 2001; Pisoni and Lively, 1995; Vallabha and McClelland, 2007). In fact, one or more IDS features were incorporated in previous training studies regardless of the theoretical perspectives (Akahane-Yamada et al., 1997; Bradlow et al., 1999; Hazan et al., 2006; Iverson et al., 2005; Jamieson and Morosan, 1986; Logan et al., 1991; McCandliss et al., 2002; Pisoni et al., 1982; Pruitt et al., 2006; Wang et al., 2003; Zhang et al., 2000).

But not all aspects of IDS necessarily facilitate phonetic learning. For example, although heightened pitch may be helpful in attention arousal, it could impede infants' vowel discrimination (Trainor and Desjardins, 2002). Given the premise that phonetic learning is a cornerstone for both first and second language acquisition (Zhang and Wang, 2007), one practical challenge for future studies is to determine what aspects of IDS might be helpful for language learning in childhood and in adulthood. Another would be how to integrate useful features for optimal learning, which has important implications for language intervention and neurological rehabilitation. Our training program, notwithstanding its limitations, provided the initial thrust of research efforts in this regard.

Neural markers of phonetic learning

The MEG data confirmed our hypothesis that training-induced changes could be simultaneously reflected in the measures of neural sensitivity and efficiency. First, consistent with our preliminary data and our previous study comparing American and Japanese subjects' MMF responses to the same /ra–la/ stimuli (Zhang et al., 2000; Zhang et al., 2005), there were bilateral contributions to the MMF response, and phonetic training led to significant MMF enhancement in the left hemisphere (Figs. 4a, b). Our neural sensitivity interpretation is in line with Näätänen's model for the mismatch response (Näätänen et al., 1997; see Näätänen et al., 2007 for a review; Shestakova et al., 2002). In this model, two processes contribute to the mismatch response of speech discrimination, a bilateral subcomponent for acoustic processing and a left-dominant subcomponent for linguistic processing based on existing long-term memory traces for syllables created in the course of learning to extract categorical information from variable phonetic stimuli. In addition to the amplitude and laterality changes in MMF, the trainees also demonstrated an overall trend of MMF peak delay in the left hemisphere (Figs. 4b, c, d). In learning a new phonetic distinction where the critical acoustic cue, the F3 transition, started 155 ms after the stimulus onset, the left-hemispheric contribution to the phonetic subcomponent of MMF developed to capture the F3 transition. This could result in a little longer latency than the acoustic subcomponent of MMF. This interpretation was consistent with our previous finding that American listeners showed a later MMF peak response than Japanese listeners for the /ra–la/ contrast (Zhang et al., 2005). Furthermore, our MMF data showed earlier right-hemisphere MMF enhancement for the within-category pair (Fig. 4d) in contrast with the left-hemisphere MMF enhancement for the cross-category pair for which the trainees showed significant behavioral improvement. This could also be nicely explained by Näätänen's model in line with existing literature (Kasai et al., 2001; Tervaniemi and Hugdahl, 2003). Detecting changes in phonetic stimuli without a relatively clear category distinction would primarily depend on acoustic processing with more right hemisphere involvement.

Second, the pre–post ECD cluster and duration data showed an overall bilateral reduction of cortical activation in both spatial and temporal domains (Fig. 5), which was consistent with the MCE results (Figs. 7 and 8). We interpret the reduced activation as an indicator of increased neural efficiency in processing the speech information, and the lack of a laterality effect here is consistent with our previous study (Zhang et al., 2005). Parallel to increased sensitivity for phonetic categorization, the learning process can be conceived as an increased differentiation of the activation pattern, so that when performance is more specialized and highly efficient, only mechanisms absolutely necessary for the occurrence of performance are activated (Näätänen, 1992). Strikingly similar pattern in MEG data was reported earlier: more focal MEG activations and a shift in MMF laterality to the learnt patterns following intensive training were observed in subjects learning to use and discriminate Morse code (Kujala et al., 2003). Learning-induced decreases in brain activation have been reported in many fMRI and PET studies involving a variety of linguistic or nonlinguistic stimuli and tasks such as sound-category learning (Guenther et al., 2004), visual priming (Squire et al., 1992), orientation discrimination (Schiltz et al., 1999), verbal delayed recognition (Jansma et al., 2001), and procedural learning of nonmotor skills (Kassubek et al., 2001). Structural analysis of the brain regions responsible for phonetic processing indicated increased white matter density to support efficient learning (Golestani et al., 2007; Golestani et al., 2002). Practice-related decreases in activation may reflect the optimization of brain resource allocation associated with decreased neural computational demands within and across information processing modalities (Gaab et al., 2005; Poldrack, 2000; Smith and Lewicki, 2006; Zhang et al., 2005). Together, the brain imaging data are consistent with the observation that language

learning itself enables better processing efficiency by allowing automatic focus at a more abstract (linguistic) level and thus freeing attentional resources (Jusczyk, 1997; Mayberry and Eichen, 1991).

The correlations between behavioral and neuromagnetic measures illustrate two important points (Fig. 6). First, as many studies have shown (Näätänen et al., 2007), the MMF response faithfully reflects training-induced changes in perceptual sensitivity. Second, as previously shown in fMRI data that the degree of behavioral improvement can predict learning efficiency in temporal-parietal and inferior frontal activation (Golestani and Zatorre, 2004), the ECD cluster and duration measures appear to be good predictors of perceptual learning.

Neural substrate and activation patterns for phonetic learning

The ECD distribution patterns (Fig. 5 and Supplemental Fig. 4) were consistent with our previous study (Zhang et al., 2005) and the overall MCE results (Figs. 7 and 8), indicating the bilateral support for acoustical and phonetic processing in the superior temporal, middle temporal, inferior parietal and inferior frontal regions. The reduction of ECD activities was most conspicuous in the inferior parietal region, an area affected by language experience (Callan et al., 2004; Zhang et al., 2005). The reduced ECDs in IP may reflect more efficient phonological encoding after training (Hickok and Poeppel, 2000). Similarly, an overall trend of bilateral reduction in ECD activity was also observed in ST and MT, but large inter-subject variability existed in the temporal lobe, which could reflect different degrees of acoustic vs. linguistic coding for the synthetic stimuli (Liebenthal et al., 2005). The large inter-subject variability found in ECD measures was reduced in the MCE solutions. As a result, the MCE results consistently showed significant overall reductions in the IP, ST and MT regions except for a small number of time points that fell within the MMF window (Fig. 8). The differences in our ECD measures of neural efficiency and the MCE data for the four ROIs could additionally result from the differences in terms of the operational definitions of the ROIs themselves. The ROIs for ECDs were defined using anatomical boundaries based on each individual subject's 3-D MRIs. The ROIs for MCEs were implemented as weighted ellipsoids. The center points of the ROIs corresponded well with each other in the two approaches, but the boundaries did not.

Unlike the fMRI studies which showed significant changes in the inferior frontal area as a result of phonetic training (Callan et al., 2003; Golestani and Zatorre, 2004; Wang et al., 2003), we found very small ECD activity in this region in both pre–post tests with no significant change. Only three subjects consistently showed significant ECDs for all the stimuli in the IF region. This could arise partially from the stringent ECD selection criteria and partially from the passive listening task since all the fMRI studies used active listening tasks. The low ECD activity in IF was consistent with MCE amplitude scale in this region (Fig. 8f). The relatively small IF activity was also consistent with our previous study on Japanese and American subjects using the same stimulus pair 1–11 and three source estimation solutions — ECD, MCE, and MNE (Zhang et al., 2005). While some MEG studies on phonetic processing also showed little IF activation (Breier et al., 2000; Maestu et al., 2002; Papanicolaou et al., 2003), others reported significant IF activities for speech as well as nonspeech stimuli (Imada et al., 2006; Maess et al., 2001; Pulvermüller et al., 2003; Shtyrov and Pulvermuller, 2007). Thus the difference between our MEG results and the previous fMRI findings regarding the role of the IF region in phonetic learning cannot be simply explained as a result of instrumental capacity.

What makes phonetic learning complicated is the fact that speech perception involves brain regions for acoustic-phonetic as well as auditory–articulatory mappings. Infant MEG data, for instance, suggests an early experience-dependent perceptual–motor link in the left hemisphere that gets strengthened during development (Imada et al., 2006). The long-term memory traces for the phonetic categories may be specifically formed not just on the basis of the auditory exposure but as a result of auditory–motor integration in the process of learning. This interpretation is consistent with the MCE data but not the ECD data in the IF region in our training study (Fig. 8f). Despite the small amplitudes, there was significant pre–post reduction of IF activity in early time window (50–180 ms) and significantly increased IF activity in the later window (186–548 ms) in the left hemisphere after training. Interestingly, the right IF region also showed an overall increase in MCE activity with a different temporal pattern: there was an early pre–post increase in MCE activities (146–156 ms and 164–180 ms) and a decrease (386–394 ms) followed by another increase in MCE activities (402–416 ms). The increases in the left IF in terms of MCE activities were also consistent with the fMRI results (Callan et al., 2003, Wang et al., 2003), suggesting that phonetic learning may strengthen the perceptual–motor link by recruiting the Broca's area. Further research is needed to determine the role of IF in supporting phonetic processing, in linking perceptual and motor systems during language acquisition, and how the IF activities may vary as a function of age and learning experience.

Relationships between brain plasticity, neural sensitivity and efficiency

Although learning improves sensitivity and efficiency of the neural system, brain plasticity does not progress in a monotonic linear fashion. In theory, learning involves continuously updated cognitive and attentional strategies, which could conceivably be supported by *decreases*, *increases*, and *shifts* in brain activation as well as additional *recruitment* of specific brain regions with nonlinear trajectories (Zhang and Wang, 2007). For instance, the Hebbian learning model allows behavioral improvement to be supported by stronger, longer, and expanded brain activities (McClelland, 2001; Vallabha and McClelland, 2007). The MMF, ECD and MCE data in this MEG study and our previous study (Zhang et al., 2005) have shown that this is not necessarily the case.

The seemingly contradictory phenomenon with larger MMFs versus smaller values for the ECDs or MCEs had been reported in our previous cross-language study (Zhang et al., 2005), where MEG responses to the same /ra–la/ contrast were recorded from American and Japanese subjects in two separate tests. Increased MMF activity was computed directly from the deviant minus standard subtraction, and it could occur with reduced activation level for automatic processing of the linguistic stimuli with more focal activation in a passive listening task. We had shown that the American subjects consistently showed larger MMFs but an overall reduced activation level than the Japanese subjects. One plausible interpretation is that the responses in Japanese trainees started to move towards more native-like brain activation patterns for the prototypical /r-l/ stimuli as a result of training.

Several factors may affect the strength of brain activation in the course of learning because the relative attentional demands depend on the difficulty of the stimuli, the experimental task, the learning outcome in individual subjects, and the subjects' ability to follow the instructions consistently. In fMRI studies involving active phonetic identification/discrimination tests, many subjects showed increased activities in the temporal and frontal regions in the post-training measures (Callan et al., 2003; Golestani and Zatorre, 2004; Wang et al., 2003). MEG studies also indicated significant attention-modulated effects on regional and hemispheric levels of activation for speech and nonspeech auditory stimuli (Gootjes et al., 2006; Hari et al., 1989; Obleser et al., 2004). How the brain reallocates its resources as a result of learning is also directly related to the task — discrimination training can result in increased activation whereas categorization training can result in decreased activation (Guenther et al., 2004). Our training study used the identification task, and our MEG tests used the passive

listening condition with a distracting reading task to examine the automatic neural processing of phonetic learning outside the focus of attention. Except for one subject who, self-reportedly, did not follow the instruction in the post-training test, all the other six trainees showed an overall reduction in ECD activities to various degrees. Despite the differences in terms of neural efficiency in MEG and fMRI studies of second language phonetic learning in adulthood, the high achievers showed more efficient processing with less ECD clusters and shorter cumulative activation duration, which was consistent with the fMRI findings about good vs. not-as-good learners (Golestani and Zatorre, 2004; Wang et al., 2003).

The MCE data nicely demonstrate another aspect of brain plasticity previously unseen in the MMF, ECD or the fMRI data: the co-existence of increasing and decreasing activities in the pre–post comparison depends on not only the region of interest but also the temporal point or window of analysis. There are different amounts of contributions from different ROIs to the total MCE activity, and some of the neuronal currents at different ROIs may cancel each other out at the same time point in the total MCE activity. Increased MMFs could take place in the presence of the overall reduced activation level for stimulus coding in terms of total MCE activity for both the deviant and the standard stimuli in the passive listening oddball paradigm. Future studies can be designed to explore the brain activation changes in different ROIs in the course of learning, and how the learning outcome could be explained using the relative ROI contributions and their integration under conditions with different attentional demands.

A cautionary note is necessary on how to interpret neural processing efficiency in connection with neural sensitivity (Zhang et al., 2005). Neural sensitivity in our study was a one-time-point "pre-attentive" measure for the MMF peak of the subtracted waveform in the so-called "ignore" condition (Näätänen et al., 2007). By contrast, our definition for neural efficiency in terms of the number of ECD clusters and their cumulative duration reflected the composite neural coding process of the stimuli, including multiple event-related components such as P1, N1, and P2 in a relatively long time window. Conceivably, increased neural sensitivity to detect a change could result from more focused attention, which might lead to an overall increase of brain activities, equivalent to a reduction of neural efficiency according to the current mathematical definition for the ECD cluster and duration measures. An alternative account is that language learning fundamentally alters the way people use cognitive strategies and brain resources in processing speech.

Admittedly, although our extended ECD and MCE analyses were largely consistent at finding out the center locations of distributed cortical activations, both the extended ECD modelling approach and the MCE approach as applied in our study fall short of addressing the exact spread or focal extent of activation in a specific brain region. The MCE approach clearly has its advantages for statistical comparison on a time-point-by-time-point basis on the millisecond scale (Fig. 5 vs. Fig. 8), but the MCE solutions, as it stands today, do not have an exact definition for the spatial extent of activation. The current amplitude estimates in the predefined locations in the standard brain have been implemented using a weighting function whose parameter selection is based on previous publications. At its best, our estimates of spatial dispersion in terms of ECD clusters are very crude approximations to overcome the limitation of the MCE approach. But the number of ECD clusters is clearly not an exact measure of spatial extent of activation in terms of voxel-wise calculations. Our extended ECD analysis only took advantage of a limited number of channels in determining localization (Supplemental Fig. 2). It is generally thought that using more channels would give better localization provided that there was an exact analysis method for the multiple-source analysis. Because the multiple sources were unknown to characterize the dispersion of

ECD activities in stimulus coding in our case, we attempted to make use of the advantage of short baseline of our multi-channel gradiometer by assuming that each channel configuration could more correctly detect the activity just beneath its selected area. If we used, in our extended single ECD analysis, more channels than necessary to localize a single ECD among the assumed multiple-source activities, the channels far from the current target ECD would include the stronger magnetic field interferences derived from activity other than our current target. In this regard, the moderate number of channels (Supplemental Fig. 2) might lead us to the more appropriate estimate in our analysis.

Our results are consistent with several other studies, showing that both the extended ECD modelling and the MCE/MNE approaches can give very good approximations of multiple sources of brain activities if the separation between the single sources is sufficiently large, exceeding 20 mm (Hari et al., 2000; Jensen and Vanni, 2002; Komssi et al., 2004; Stenbacka et al., 2002). Due to the inherent limitations of the current source modelling techniques, more research efforts are needed to model the focal extent of MEG activities. Multimodal brain imaging by combining fMRI with MEG can be helpful in this regard (Billingsley-Marshall et al., 2007; Fujimaki et al., 2002; Grummich et al., 2006).

Conclusion

In sum, there are three main findings in the present study. First, our training software program integrating principles of infant phonetic learning can provide the necessary enriched exposure to induce substantial neural plasticity for learning new phonetic categories in adulthood. Second, parallel to the well-known mismatch response for neural sensitivity, neural efficiency as defined in terms of the distribution and duration of neuromagnetic activities appears to be good predictors of phonetic learning. Third, adaptive learning should not only focus on modification of the critical cues for the to-be-learned material but also take into account the neural commitment to prior learning. Native-like speech perception for second language learners may be possible if new methods are developed to help the neurally committed system overcome prior learning. Further training studies and technical improvements for MEG source modelling on the spatial extent of activation are necessary to answer important questions regarding the relative contributions of training features to the phonetic learning process, the association of neural plasticity for phonetic learning with the acquisition of higher-order linguistic and cognitive skills, the relationships between neural sensitivity measures and neural efficiency measures as a function of stimulus, task and subject characteristics, and how the neural markers of learning may vary as a function of age.

Acknowledgments

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2009.01.028.

References

Akahane-Yamada, R., Bradlow, A.R., Pisoni, D.B., Tohkura, Y.i., 1997. Effects of audio-visual training on the identification of English /r/ and /l/ by Japanese speakers. J Acoust. Soc. Am. 102, 3137.

Akamatsu, T., 1997. Japanese Phonetics: Theory and Practice. LINCOM EUROPA, München.

Best, C.T., McRoberts, G.W., Goodell, E., 2001. Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listeners' native phonological system. J. Acoust. Soc. Am. 109, 775–794.

Best, C.T., Strange, W., 1992. Effects of language-specific phonological and phonetic factors on cross-language perception of approximants. J. Phon. 20, 305–330.

Billingsley-Marshall, R.L., Clear, T., Mencl, W.E., Simos, P.G., Swank, P.R., Men, D., Sarkari, S., Castillo, E.M., Papanicolaou, A.C., 2007. A comparison of functional MRI and magnetoencephalography for receptive language mapping. J. Neurosci. Methods 161, 306–313.

Bosch, L., Costa, A., Sebastián-Gallés, N., 2000. First and second language vowel perception in early bilinguals. Eur. J. Cogn. Psychol. 12, 189–221.

Bradlow, A.R., Akahane-Yamada, R., Pisoni, D.B., Tohkura, Y., 1999. Training Japanese listeners to identify English /r/ and /l/: long-term retention of learning in perception and production. Percept. Psychophys. 61, 977–985.

Breier, J.I., Simos, P.G., Zouridakis, G., Papanicolaou, A.C., 2000. Lateralization of activity associated with language function using magnetoencephalography: a reliability study. J. Clin. Neurophysiol. 17, 503–510.

Burnham, D., Kitamura, C., Vollmer-Conna, U., 2002. What's new pussycat: on talking to animals and babies. Science 296, 1435.

Callan, D.E., Jones, J.A., Callan, A.M., Akahane-Yamada, R., 2004. Phonetic perceptual identification by native- and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulatory–auditory/orosensory internal models. Neuroimage 22, 1182–1194.

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.

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

Davis, B.L., Lindblom, B., 2001. Phonetic variability in baby talk and development of vowel categories. In: Claes von Hofsten, F.L., Heimann, M. (Eds.), Emerging Cognitive Abilities in Early Infancy. Lawrence Erlbaum, Mahwah, NJ, pp. 135–171.

de Boer, B., Kuhl, P.K., 2003. Investigating the role of infant-directed speech with a computer model. Acoust. Res. Lett. Online 4, 129–134.

Englund, K., Behne, D., 2006. Changes in infant directed speech in the first six months. Infant Child Dev. 15, 139–160.

Escudero, P., Boersma, P., 2004. Bridging the gap between L2 speech perception research and phonological theory. Stud. Second Lang. Acquis. 26, 551–585.

Fernald, A., Kuhl, P., 1987. Acoustic determinants of infant preference for motherese speech. Infant Behav. Dev. 10, 279–293.

Flege, J., 1995. Second language speech learning: theory, findings, and problems. In: Strange, W. (Ed.), Speech Perception and Linguistic Experience: Theoretical and Methodological Issues. York Press, Timonium, MD, pp. 229–273.

Flege, J.E., Yeni-Komshian, G.H., Liu, S., 1999. Age constraints on second-language acquisition. J. Mem. Lang. 41, 78–104.

Francis, A.L., Nusbaum, H.C., 2002. Selective attention and the acquisition of new phonetic categories. J. Exp. Psychol. Hum. Percept. Perform. 28, 349–366.

Fujimaki, N., Hayakawa, T., Nielsen, M., Knosche, T.R., Miyauchi, S., 2002. An fMRI-constrained MEG source analysis with procedures for dividing and grouping activation. Neuroimage 17, 324–343.

Gaab, N., Tallal, P., Kim, H., Lakshminarayanan, K., Archie, J.J., Glover, G.H., Gabrieli, J.D.E., 2005. Neural correlates of rapid spectrotemporal processing in musicians and nonmusicians. Ann. N.Y. Acad. Sci. 1060, 82–88.

Golestani, N., Molko, N., Dehaene, S., LeBihan, D., Pallier, C., 2007. Brain structure predicts the learning of foreign speech sounds. Cereb. Cortex 17, 575–582.

Golestani, N., Paus, T., Zatorre, R.J., 2002. Anatomical correlates of learning novel speech sounds. Neuron 35, 997–1010.

Golestani, N., Zatorre, R.J., 2004. Learning new sounds of speech: reallocation of neural substrates. Neuroimage 21, 494–506.

Gootjes, L., Bouma, A., Van Strien, J.W., Scheltens, P., Stam, C.J., 2006. Attention modulates hemispheric differences in functional connectivity: evidence from MEG recordings. Neuroimage 30, 245–253.

Gottfried, T.L., 1984. Effects of consonant context on the perception of French vowels. J. Phon. 12, 91–114.

Grummich, P., Nimsky, C., Pauli, E., Buchfelder, M., Ganslandt, O., 2006. Combining fMRI and MEG increases the reliability of presurgical language localization: a clinical study on the difference between and congruence of both modalities. Neuroimage 32, 1793–1803.

Guenther, F.H., Nieto-Castanon, A., Ghosh, S.S., Tourville, J.A., 2004. Representation of sound categories in auditory cortical maps. J. Speech Lang. Hear. Res. 47, 46–57.

Guion, S.G., Flege, J.E., Akahane-Yamada, R., Pruitt, J.C., 2000. An investigation of current models of second language speech perception: the case of Japanese adults' perception of English consonants. J. Acoust. Soc. Am. 107, 2711–2724.

Hari, R., Hamalainen, M., Kaukoranta, E., Makela, J., Joutsiniemi, S.L., Tiihonen, J., 1989. Selective listening modifies activity of the human auditory cortex. Exp. Brain Res. 74, 463–470.

Hari, R., Levanen, S., Raij, T., 2000. Timing of human cortical functions during cognition: role of MEG. Trends Cogn. Sci. 4, 455–462.

Hazan, V., Sennema, A., Faulkner, A., Ortega-Llebaria, M., Iba, M., Chung, H., 2006. The use of visual cues in the perception of non-native consonant contrasts. J. Acoust. Soc. Am. 119, 1740–1751.

Hernandez, A., Li, P., 2007. Age of acquisition: its neural and computational mechanisms. Psychological Bulletin 133, 638–650.

Hickok, G., Poeppel, D., 2000. Towards a functional neuroanatomy of speech perception. Trends Cogn. Sci. 4, 131–138.

Imada, T., Mashiko, T., Sekine, H., 2001. Estimation errors of single dipole model applied to twin dipole activity: computer simulation study. In: Nenonen, J., Ilmoniemi, R.J., Katila, T. (Eds.), Proceedings of the 12th International Conference on Biomagnetism. Helsinki University of Technology, Espoo, Finland, pp. 733–737.

Imada, T., Zhang, Y., Cheour, M., Taulu, S., Ahonen, A., Kuhl, P.K., 2006. Infant speech perception activates Broca's area: a developmental magnetoencephalography study. Neuroreport 17, 957–962.

Imaizumi, S., Tamekawa, Y., Itoh, H., Deguchi, T., Mori, K., 1999. Effects of L1 phonotactic constraints on L2 speech perception and production. J. Acoust. Soc. Am. 105, 1094.

Iverson, P., Hazan, V., Bannister, K., 2005. Phonetic training with acoustic cue manipulations: a comparison of methods for teaching English /r/-/l/ to Japanese adults. J. Acoust. Soc. Am. 118, 3267–3278.

Iverson, P., Kuhl, P.K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., Siebert, C., 2003. A perceptual interference account of acquisition difficulties for non-native phonemes. Cognition 87, B47–B57.

Jamieson, D.G., Morosan, D.E., 1986. Training non-native speech contrasts in adults: acquisition of the English /delta/-/theta/ contrast by francophones. Percept. Psychophys. 40, 205–215.

Jansma, J.M., Ramsey, N.F., Slagter, H.A., Kahn, R.S., 2001. Functional anatomical correlates of controlled and automatic processing. J. Cogn. Neurosci. 13, 730–743.

Jensen, O., Vanni, S., 2002. A new method to identify multiple sources of oscillatory activity from magnetoencephalographic data. Neuroimage 15, 568–574.

Johnson, J., Newport, E., 1989. Critical period effects in second language learning: the influence of maturation state on the acquisition of English as a second language. Cogn. Psychol. 21, 60–99.

Jusczyk, P.W., 1997. The Discovery of Spoken Language. MIT Press, Cambridge, MA.

Kasai, K., Yamada, H., Kamio, S., Nakagome, K., Iwanami, A., Fukuda, M., Itoh, K., Koshida, I., Yumoto, M., Iramina, K., Kato, N., Ueno, S., 2001. Brain lateralization for mismatch response to across- and within-category change of vowels. Neuroreport 12, 2467–2471.

Kassubek, J., Schmidtke, K., Kimmig, H., Lucking, C.H., Greenlee, M.W., 2001. Changes in cortical activation during mirror reading before and after training: an fMRI study of procedural learning. Cogn. Brain Res. 10, 207–217.

Katrin, N., Steven, S., 2005. Statistical properties of infant-directed versus adult-directed speech: insights from speech recognition. J. Acoust. Soc. Am. 117, 2238–2246.

Komssi, S., Huttunen, J., Aronen, H.J., Ilmoniemi, R.J., 2004. EEG minimum-norm estimation compared with MEG dipole fitting in the localization of somatosensory sources at S1. Clin. Neurophysiol. 115, 534–542.

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

Kuhl, P.K., Andruski, J.E., Chistovich, I.A., Chistovich, L.A., Kozhevnikova, E.V., Ryskina, V.L., Stolyarova, E.I., Sundberg, U., Lacerda, F., 1997. Cross-language analysis of phonetic units in language addressed to infants. Science 277, 684–686.

Kuhl, P.K., Conboy, B.T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., Nelson, T., 2008. Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). Philos. Trans. R. Soc. B Biol. Sci. 363, 979–1000.

Kuhl, P.K., Meltzoff, A.N., 1982. The bimodal perception of speech in infancy. Science 218, 1138–1141.

Kuhl, P.K., Meltzoff, A.N., 1996. Infant vocalizations in response to speech: vocal imitation and developmental change. J. Acoust. Soc. Am. 100, 2425–2438.

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. Proc. Natl. Acad. Sci. U. S. A. 100, 9096–9101.

Kuhl, P.K., Tsao, F.M., Liu, H.M., Zhang, Y., de Boer, B., 2001. Language/culture/mind/brain: progress at the margins between disciplines. Annals of the New York Academy of Sciences 935, 136–174.

Kujala, A., Huotilainen, M., Uther, M., Shtyrov, Y., Monto, S., Ilmoniemi, R.J., Näätänen, R., 2003. Plastic cortical changes induced by learning to communicate with non-speech sounds. Neuroreport 14, 1683–1687.

Lenneberg, E.H., 1967. Biological foundations of language. Wiley, New York.

Liebenthal, E., Binder, J.R., Spitzer, S.M., Possing, E.T., Medler, D.A., 2005. Neural substrates of phonemic perception. Cereb. Cortex 15, 1621–1631.

Lin, F.H., Witzel, T., Ahlfors, S.P., Stufflebeam, S.M., Belliveau, J.W., Hämäläinen, M.S., 2006. Assessing and improving the spatial accuracy in MEG source localization by depth-weighted minimum-norm estimates. Neuroimage 31, 160–171.

Liu, H.M., Kuhl, P.K., Tsao, F.M., 2003. An association between mothers' speech clarity and infants' speech discrimination skills. Dev. Sci. 6, F1–F10.

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

Lotto, A.J., Sato, M., Diehl, R.L., 2004. Mapping the task for the second language learner: the case of Japanese acquisition of /r/ and /l/. In: Slifka, J., Manuel, S., Matthies, M. (Eds.), From Sound to Sense: 50+ Years of Discoveries in Speech Communication. Electronic conference proceedings, Cambridge, MA.

Macmillan, N.A., Creelman, C.D., 1991. Detection Theory: A User's Guide. Cambridge UP, Cambridge, UK.

Maess, B., Koelsch, S., Gunter, T.C., Friederici, A.D., 2001. Musical syntax is processed in Broca's area: an MEG study. Nat. Neurosci. 4, 540–545.

Maestu, F., Ortiz, T., Fernandez, A., Amo, C., Martin, P., Fernandez, S., Sola, R.G., 2002. Spanish language mapping using MEG: a validation study. Neuroimage 17, 1579–1586.

Mayberry, R., Eichen, E., 1991. The long-lasting advantage of learning sign language in childhood: another look at the critical period for language acquisition. J. Mem. Lang. 30, 486–512.

Mayberry, R.I., Lock, E., 2003. Age constraints on first versus second language acquisition: evidence for linguistic plasticity and epigenesis. Brain Lang. 87, 369–384.

Maye, J., Werker, J.F., Gerken, L., 2002. Infant sensitivity to distributional information can affect phonetic discrimination. Cognition 82, B101–B111.

McCandliss, B.D., Fiez, J.A., Protopapas, A., Conway, M., McClelland, J.L., 2002. Success and failure in teaching the [r]–[l] contrast to Japanese adults: tests of a Hebbian model of plasticity and stabilization in spoken language perception. Cogn. Affect. Behav. Neurosci. 2, 89–108.

McClelland, J.L., 2001. Failures to learn and their remediation: a Hebbian account. In: McClelland, J.L., Siegler, R.S. (Eds.), Mechanisms of Cognitive Development. Erlbaum, Mahwah, NJ, pp. 97–121.

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. Learn. Mem. 9, 253–267.

Moulines, E., Charpentier, F., 1990. Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. Speech Commun. 9, 453–467.

Näätänen, R., 1992. Attention and Brain Function. Erlbaum, Hillsdale, NJ.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R.J., Luuk, A., Allik, J., Sinkkonen, J., Alho, K., 1997. Language-specific phoneme representations revealed by electric and magnetic brain responses. Nature 385, 432–434.

Näätänen, R., Paavilainen, P., Rinne, T., Alho, K., 2007. The mismatch negativity (MMN) in basic research of central auditory processing: a review. Clin. Neurophysiol. 118, 2544–2590.

Nemoto, I., Abe, M., Kotani, M., 2008. Multiplicative correction of subject effect as preprocessing for analysis of variance. IEEE Trans. Biomed. Eng. 55, 941–948.

Nenonen, S., Shestakova, A., Huotilainen, M., Näätänen, R., 2005. Speech–sound duration processing in a second language is specific to phonetic categories. Brain Lang. 92, 26–32.

Obleser, J., Elbert, T., Eulitz, C., 2004. Attentional influences on functional mapping of speech sounds in human auditory cortex. BMC Neurosci. 5, 24.

Pallier, C., Bosch, L., Sebastián-Gallés, N., 1997. A limit on behavioral plasticity in speech perception. Cognition 64, B9–17.

Papanicolaou, A.C., Castillo, E., Breier, J.I., Davis, R.N., Simos, P.G., Diehl, R.L., 2003. Differential brain activation patterns during perception of voice and tone onset time series: a MEG study. Neuroimage 18, 448–459.

Pisoni, D., Lively, S., 1995. Variability and invariance in speech perception: a new look at some old problems in perceptual learning. In: Strange, W. (Ed.), Speech Perception and Linguistic Experience: Theoretical and Methodological Issues. York Press, Timonium, MD, pp. 433–459.

Pisoni, D.B., Aslin, R.N., Perey, A.J., Hennessy, B.L., 1982. Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. J. Exp. Psychol. Hum. Percept. Perform. 8, 297–314.

Poldrack, R.A., 2000. Imaging brain plasticity: conceptual and methodological issues. Neuroimage 12, 1–13.

Pruitt, J.S., Jenkins, J.J., Strange, W., 2006. Training the perception of Hindi dental and retroflex stops by native speakers of American English and Japanese. J. Acoust. Soc. Am. 119, 1684–1696.

Pulvermüller, F., Shtyrov, Y., Ilmoniemi, R., 2003. Spatiotemporal dynamics of neural language processing: an MEG study using minimum-norm current estimates. Neuroimage 20, 1020–1025.

Rivera-Gaxiola, M., Csibra, G., Johnson, M.H., Karmiloff-Smith, A., 2000. Electrophysiological correlates of cross-linguistic speech perception in native English speakers. Behav. Brain Res. 111, 13–23.

Schiltz, C., Bodart, J.M., Dubois, S., Dejardin, S., Michel, C., Roucoux, A., Crommelinck, M., Orban, G.A., 1999. Neuronal mechanisms of perceptual learning: changes in human brain activity with training in orientation discrimination. Neuroimage 9, 46–62.

Sebastián-Gallés, N., Soto-Faraco, S., 1999. Online processing of native and non-native phonemic contrasts in early bilinguals. Cognition 72, 111–123.

Sharf, D.J., Ohde, R.N., 1984. Effect of formant frequency onset variation on the differentiation of synthesized /w/ and /r/ sounds. J. Speech Hear. Res. 27, 475–479.

Shestakova, A., Brattico, E., Huotilainen, M., Galunov, V., Soloviev, A., Sams, M., Ilmoniemi, R.J., Näätänen, R., 2002. Abstract phoneme representations in the left temporal cortex: magnetic mismatch negativity study. Neuroreport 13, 1813–1816.

Shtyrov, Y., Pulvermuller, F., 2007. Early MEG activation dynamics in the left temporal and inferior frontal cortex reflect semantic context integration. J. Cogn. Neurosci. 19, 1633–1642.

Smith, E.C., Lewicki, M.S., 2006. Efficient auditory coding. Nature 439, 978–982.

Squire, L.R., Ojemann, J.G., Miezin, F.M., Petersen, S.E., Videen, T.O., Raichle, M.E., 1992. Activation of the hippocampus in normal humans: a functional anatomical study of memory. Proc. Natl. Acad. Sci. U. S. A. 89, 1837–1841.

Stenbacka, L., Vanni, S., Uutela, K., Hari, R., 2002. Comparison of minimum current estimate and dipole modeling in the analysis of simulated activity in the human visual cortices. Neuroimage 16, 936–943.

Strange, W., Dittmann, S., 1984. Effects of discrimination training on the perception of /r–l/ by Japanese adults learning English. Percept. Psychophys. 36, 131–145.

Takagi, N., 2002. The limits of training Japanese listeners to identify English /r/ and /l/: eight case studies. J. Acoust. Soc. Am. 111, 2887–2896.

Takagi, N., Mann, V., 1995. The limits of extended naturalistic exposure on the perceptual mastery of English /r/ and /l/ by adult Japanese learners of English. Appl. Psycholinguist. 16, 379–405.

Tervaniemi, M., Hugdahl, K., 2003. Lateralization of auditory-cortex functions. Brains Res. Rev. 43, 231–246.

Trainor, L.J., Desjardins, R.N., 2002. Pitch characteristics of infant-directed speech affect infants' ability to discriminate vowels. Psychon. Bull. Rev. 9, 335–340.

Tremblay, K., Kraus, N., Carrell, T.D., McGee, T., 1997. Central auditory system plasticity: generalization to novel stimuli following listening training. J. Acoust. Soc. Am. 102, 3762–3773.

Tricomi, E., Delgado, M.R., McCandliss, B.D., McClelland, J.L., Fiez, J.A., 2006. Performance feedback drives caudate activation in a phonological learning task. J. Cogn. Neurosci. 18, 1029–1043.

Underbakke, M., Polka, L., Gottfried, T.L., Strange, W., 1988. Trading relations in the perception of /r/–/l/ by Japanese learners of English. J. Acoust. Soc. Am. 84, 90–100.

Uther, M., Knoll, M.A., Burnham, D., 2007. Do you speak E-NG-L-I-SH? A comparison of foreigner- and infant-directed speech. Speech Commun. 49, 2–7.

Uutela, K., Hämäläinen, M., Somersalo, E., 1999. Visualization of magnetoencephalographic data using minimum current estimates. Neuroimage 10, 173–180.

Vallabha, G.K., McClelland, J.L., 2007. Success and failure of new speech category learning in adulthood: consequences of learned Hebbian attractors in topographic maps. Cognitive Affect. Behav. Neurosci. 7, 53–73.

Vallabha, G.K., McClelland, J.L., Pons, F., Werker, J.F., Amano, S., 2007. Unsupervised learning of vowel categories from infant-directed speech. Proc. Natl. Acad. Sci. U. xS. A. 104, 13273–13278.

Wang, Y., Sereno, J.A., Jongman, A., Hirsch, J., 2003. fMRI evidence for cortical modification during learning of mandarin lexical tone. J. Cogn. Neurosci. 15, 1019–1027.

Werker, J.F., Pons, F., Dietrich, C., Kajikawa, S., Fais, L., Amano, S., 2007. Infant-directed speech supports phonetic category learning in English and Japanese. Cognition 103, 147–162.

Werker, J.F., Tees, R.C., 2005. Speech perception as a window for understanding plasticity and commitment in language systems of the brain. Dev. Psychobiol. 46, 233–251.

Winkler, I., 1999. Brain responses reveal the learning of foreign language phonemes. Psychophysiology 36, 638–642.

Zhang, Y., Kuhl, P.K., Imada, T., Iverson, P., Pruitt, J., Stevens, E., Kotani, M., Tohkura, Y., 2000. Neural plasticity revealed in perceptual training of a Japanese adult listener to learn American /l–r/ contrast: a whole-head magnetoencephalography study. Proceedings of the 6th International Conference on Spoken Language Processing 3, 953–956.

Zhang, Y., Kuhl, P.K., Imada, T., Kotani, M., Tohkura, Y., 2005. Effects of language experience: neural commitment to language-specific auditory patterns. Neuroimage 26, 703–720.

Zhang, Y., Wang, Y., 2007. Neural plasticity in speech learning and acquisition. Bilingualism: Language and cognition 10, 147–160.