Rapid Tuning of Auditory "What" and "Where" Pathways by Training

Yi Du[1,2], Yu He[1], Stephen R. Arnott[1], Bernhard Ross[1], Xihong Wu[2], Liang Li[2] and Claude Alain[1,3]

[1]Rotman Research Institute, Baycrest Centre for Geriatric Care, Toronto, Ontario, Canada M6A 2E1, [2]Department of Psychology, Speech and Hearing Research Center, Key Laboratory on Machine Perception (Ministry of Education), PKU-IDG/McGovern Institute for Brain Research, Peking University, Beijing 100871, China and [3]Department of Psychology, University of Toronto, Ontario, Canada M8V 2S4

Address correspondence to Claude Alain, Rotman Research Institute, Baycrest Centre for Geriatric Care, 3560 Bathurst Street, Toronto, Ontario, Canada M6A 2E1. Email: calain@research.baycrest.org

Behavioral improvement within the first hour of training is commonly explained as procedural learning (i.e., strategy changes resulting from task familiarization). However, it may additionally reflect a rapid adjustment of the perceptual and/or attentional system in a goal-directed task. In support of this latter hypothesis, we show feature-specific gains in performance for groups of participants briefly trained to use either a spectral or spatial difference between 2 vowels presented simultaneously during a vowel identification task. In both groups, the neuromagnetic activity measured during the vowel identification task following training revealed source activity in auditory cortices, prefrontal, inferior parietal, and motor areas. More importantly, the contrast between the 2 groups revealed a striking double dissociation in which listeners trained on spectral or spatial cues showed higher source activity in ventral ("what") and dorsal ("where") brain areas, respectively. These feature-specific effects indicate that brief training can implicitly bias top-down processing to a trained acoustic cue and induce a rapid recalibration of the ventral and dorsal auditory streams during speech segregation and identification.

Keywords: attention, hearing, MEG, plasticity, speech

Introduction

Learning often involves a rapid improvement in performance that occurs within the first hour. These rapid changes in behavioral performance were previously thought to solely re ect procedural learning, that is, changes in participants' strategies that occurred during task familiarization (Karni and Bertini 1997; Karni et al. 1998; Wright and Fitzgerald 2001). However, evidence from behavioral studies suggests that increased accuracy during the first hour may also involve increased perceptual sensitivity for auditory (Hawkey et al. 2004) or visual (Hussain et al. 2012) stimuli. That is, learning within the first hour shows specificity to the trained stimulus or feature.

The hypothesis that the first hour of training can increase perceptual sensitivity has received support from animal studies investigating rapid learning-induced plasticity in auditory processing. For instance, changes in the receptive fields of ferret auditory cortex can occur within minutes of a tone discrimination task that was previously learned (Fritz et al. 2003, 2005a, 2005b). However, these neuroplastic changes were smaller or absent if the animal listened passively to the same sounds (i.e., when the sounds are task-irrelevant). Such rapid frequency-specific changes in the receptive fields of auditory neurons have been observed in a wide variety of situations including classical conditioning (Bakin and Weinberger 1990; Edeline et al. 1993), instrumental avoidance conditioning (Bakin et al. 1996), and discrimination learning (Fritz et al. 2003). Training-induced plasticity in auditory localization has been observed in mammals (Kacelnik et al. 2006; Bajo et al. 2010), but it remains to be determined whether these changes can occur quickly within the auditory system.

Using the scalp recording of auditory event-related potentials (ERPs), Alain et al. (2007) found that behavioral improvement during the first hour of a concurrent-vowel identification task correlated with enhancements in early (~130 ms) and late (~340 ms) ERPs originating from the right superior temporal gyrus (STG) and inferior prefrontal cortex, respectively. These ERP changes re

have provided evidence for rapid adjustment of the perceptual system during training (for a review see Fritz et al. 2005a), it remains to be determined whether such feature-specific effects take place in humans. In human studies, the nonspecific behavioral improvement observed in the early stages of learning (i.e., lack of specificity) is often taken as evidence for procedural learning with feature-specific effects emerging only after multiple daily training sessions (Watson 1980; Karni and Bertini 1997). Although there is some behavioral evidence for specificity within the first training session (Hawkey et al. 2004; Hussain et al. 2012), neurophysiological evidence supporting such feature-specific effects is lacking.

In the present study, we investigated whether a brief 45-min training program can yield a rapid feature-specific modulation on both behavior and neuromagnetic activity during subsequent measurement while participants identified 2 vowels presented simultaneously. Two groups of participants were trained to utilize differences either in the fundamental frequency (Δf_0) or in the spatial location (Δlocation) of the 2 vowels presented simultaneously, acoustic features well known to stimulate distinct ventral ("what") and dorsal ("where") neural pathways in humans (Alain et al. 2001; Arnott et al. 2004). Shortly after training, we measured neuromagnetic brain activity using magnetoencephalography (MEG) while participants from both groups performed the same task using an identical set of double-vowel stimuli that shared the same f_0 and location or differed in either f_0 or location or both. This procedure enabled us to investigate the behavioral and neuroplastic changes induced by training listeners to use either spectral or spatial cues in speech separation and identification while holding the bottom-up sensory input constant. If the rapid improvement re ects primarily procedural learning that is independent of trained features, one would anticipate that the training effects on behavioral and neuromagnetic data would be comparable between the differentially trained groups. On the other hand, if learning does involve a rapid adjustment in perceptual sensitivity to the task-relevant attribute, then one would predict feature-specific gains in performance. That is, participants that receive spectral training should perform better when a Δf_0 cue is available relative to when a Δlocation cue is present, and vice versa. We anticipated that feature-specific gains in performance will be paralleled by changes in neuromagnetic brain activity, which will be illustrated by greater source activity in ventral and dorsal brain regions in groups trained on Δf_0 or Δlocation, respectively.

Materials and Methods

Participants

Twenty-four participants who provided written informed consent according to the University of Toronto and Baycrest Hospital Human Subject Review Committee guidelines were randomly assigned into 2 training groups: the Frequency group (7 women; aged 20–33 years; mean: 24 years) and the Location group (8 women; aged 20–31 years; mean: 24 years). All participants were right-handed, native English speakers, and had normal pure-tone thresholds at both ears (<25 dB HL for 250–8000 Hz).

Stimuli and Task

Stimuli were 4 synthetic steady-state American English vowels: /a:/ (as in *fa e*), /3/ (as in *e*), /i:/ (as in *ee*), and /u:/ (as in *e*), henceforth referred to as "ah," "er," "ee," and "oo," respectively (Assmann and Summerfield 1994). Each vowel was 200 ms in duration (2442 samples at a 12.21-kHz sample rate, 16-bit quantization), low-pass filtered at 5 kHz, with f_0 (100–126 Hz, see later for details) and formant frequencies held constant for the entire duration (see Supplementary Fig. 1). Formant frequencies were patterned after a male talker from the North Texas region. The source signal was the same in all the 4 vowels, simulating "equal vocal effort." Onsets and offsets were shaped by 2 halves of an 8-ms Kaiser window. Double-vowel stimuli were created by adding together the digital waveforms of 2 different vowels and then dividing the sum by 2. Each vowel was paired with every other vowel. Stimuli were examined using an oscilloscope to ensure that there was no "clipping." The vowels were added in phase and this resulted in smaller amplitude when the 2 vowels differed in f_0.

Stimuli were converted to analog forms (TDT RP-2 real-time processor, Tucker Davis Technologies, Alachua, FL, USA), fed into a headphone driver (TDT HB-7), and presented binaurally at 75 dB sound pressure level (SPL) through Etymotic ER-3A inserted earphones (Etymotic Research, Elk Grove, IL, USA) connected with a 1.5-m re ection-less plastic tube. The intensity of the stimuli was measured using a Larson-Davis SPL meter (Model 824, Provo, UT, USA). The plastic tubes from the ER-3A transducers were attached to a 2-cc coupler on an artificial ear (Model AEC100l) connected to the SPL meter. Separate measurements were taken for both left and right ear channels. Perceived sound locations were induced by applying a head-related transfer function (HRTF) coefficient from the TDT library to the vowels prior to sending them to the headphone driver (for a detailed description and behavioral validation of the HRTF coefficient, see Wightman and Kistler 1989a, 1989b; Wenzel et al. 1993). The HRTF coefficients were individually determined by a brief sound localization task at the beginning of the experiment. On each trial, 1 of the 4 vowels was presented at 1 of the 5 azimuth locations (i.e., −90°, −45°, 0°, 45°, 90°) using a variety of HRTF coefficients selected from the TDT library that best suited the participant's head size. Participants were asked to point toward the sound source location. The HRTF coefficient that resulted in the most accurate localization responses was then determined and used for the remainder of the experiment for each participant.

Before the training task, participants were provided with written instruction, as well as exemplars of the various stimuli. Each vowel was presented individually (16 trials, 4 vowels by 2 f_0 levels, 100 and 106 Hz), and participants identified the vowel by pressing 1 of 4 keys on the keyboard, marked "AH," "ER," "EE," and "OO." All participants achieved single vowel accuracy of 95% or better.

Following the familiarization with the stimuli and task, participants underwent a 45-min training session. For the Frequency group, each vowel pair contained 1 vowel with f_0 at 100 Hz and the other f_0 at 100, 103, 106, 112, or 126 Hz, resulting in 5 levels of Δf_0: 0, 0.5, 1, 2, or 4 semitones. Both vowels were presented from the midline. For the Location group, the 2 vowels in each pair had equal f_0 (100 Hz) and were presented both from the midline (0°) or from 15°, 30°, 45°, or 60° away from the midline (i.e., one from left of the midline, and the other from right of the midline), resulting in 5 levels of Δlocation: 0°, 30°, 60°, 90°, or 120°. All vowel pairs were presented in the horizontal plane, randomized, and balanced in 4 blocks of 120 trials. Participants were told that 2 different vowels will always be presented on each trial and the 2 vowels might have the same or different pitch (for the Frequency group) or come from the same or different locations (for the Location group). Their task was to identify both vowels by sequentially pressing corresponding keys on the keyboard. In other words, participants were implicitly trained to utilize the spectral or spatial difference between the 2 vowels to facilitate their segregation and identification. This is different from an explicit frequency or spatial discrimination task, which would have required participants to indicate whether the 2 sounds had the same or different pitch or spatial location. Five milliseconds after participant's second response, a visual feedback occurred on the screen in front of the participant for 1 s, showing the stimuli and response for the last trial. The next trial started 2 s after participant's second response.

The MEG session started 15 min after training. Participants were presented with 4 trial types, which were created by the orthogonal combination of Δf_0 and Δlocation. That is, the 2 vowels could have

either the same (100 or 106 Hz) or different f_0 (one at 100 Hz and the other at 106 Hz, i.e., 1-semitone Δf_0), and they could come from either the same (midline) or different azimuth locations (one from 45° to the left and the other from 45° to the right, i.e., 90° Δlocation). These 4 trial types were labeled as follows: same f_0 same location (SFSL), same f_0 different location (SFDL), different f_0 same location (DFSL), and different f_0 different location (DFDL), and were randomized and balanced in 4 blocks of 144 trials. Participants performed the same task as during training without feedback, and they were told that the 2 vowels could have the same or different pitch and come from the same or different locations. The next trial started 1.5 s after participant's second response.

MEG Acquisition and Analysis

MEG data were recorded in a magnetically shielded room using a 151-channel whole-head neuromagnetometer (VSM Medtech, Port Coquitlam, BC, Canada). Participants were in the upright seating position with their head resting in the helmet-shaped sensor array. Head localization coils were placed on the nasion, left and right preauricular points for coregistration of the MEG data with anatomical magnetic resonance images (MRIs) and/or realistic estimates of the participant' head shape by a 3-dimensional digitization system (Fastrak, Polhemus, Colchester, VT, USA) obtained prior to MEG recording. The neuromagnetic activity was sampled at 625 Hz and low-pass filtered at 200 Hz, and 4 blocks were collected with each one lasting about 9 min.

The synthetic aperture magnetometry (SAM), a minimum-variance beamformer algorithm (Van Veen et al. 1997), was used as a spatial filter to estimate the time course of source activity on a lattice of 5-mm spacing across the whole-brain volume in the 0.3 to 20 Hz frequency range. A multiple-sphere head model was used for the beamformer analysis in which a single sphere was fit to the digitized head shape for each MEG sensor. Waveforms of averaged source activity for each trial type were calculated following the event-related SAM approach (ER-SAM, Robinson 2004; Cheyne et al. 2006). The time course of source activity at each node/voxel was estimated as a weighted linear combination of the magnetic field measured at all MEG sensors and represented as a normalized pseudo-*Z* measure (Robinson and Vrba 1998). For data reduction, the time series were down-sampled by the factor of 5 (i.e., one sample point every 8 ms). Time series of volumetric maps of group mean pseudo-*Z* values for each trial type were normalized to the Talairach stereotaxic space, spatially smoothed using a Gaussian filter with a full width at half maximum value of 4.0 mm, overlaid on the anatomical image of a template brain (colin27, Montreal Neurological Institute, Holmes et al. 1998), and visualized with the Analysis of Functional Neuroimages software (AFNI version 2.56a, Cox 1996). As the aim of this study was to examine the learning effect on speech segregation rather than response processing, all trials were included regardless of accuracy.

Statistical Analysis

Repeated-measures analysis of variance (ANOVA) followed by 1-way ANOVA, Bonferroni post hoc tests, and -tests were conducted for behavioral data with the null-hypothesis rejection level set at 0.05.

For MEG data, group analysis was conducted on the grand mean pseudo-*Z* values across stimulus types of each 40 ms epoch centered at 50, 100, 150, 200, 250, 300, and 350 ms after stimulus onset. The time windows were chosen to encompass the transient auditory evoked fields elicited by double-vowel stimuli (i.e., P1m, N1m, and P2m), which peaked, respectively, at about 45, 110, and 205 ms after sound onset. Moreover, these windows cover the time periods that have shown rapid neuroplastic changes related to spectral (Alain et al. 2007) or spatial training (Spierer et al. 2007). The use of the mean value of a 40-ms epoch also smoothed the individual difference in temporal processing of double-vowel stimuli, took into account the learning-related modulation of brain activity during certain period, and increased the statistical sensitivity and power. First, a voxel-wise, mixed-effect 2-factor ANOVA with group as the fixed factor and with participant as the random factor was computed for each epoch using the 3dANOVA2 function in AFNI. To correct for multiple comparisons, a spatial cluster extent threshold was applied by using AlphaSim with 4096 (2^{12}) Monte Carlo simulations. Using an uncorrected *P*-value threshold of 0.05, the minimum cluster size with a family-wise, false-positive probability of $P < 0.05$ was 2048 μL (32 voxels) for 30–70 ms epoch, 1792 μL (28 voxels) for 80–120 ms epoch, 1408 μL (22 voxels) for 130–170 ms epoch, 1152 μL (18 voxels) for 180–220 ms epoch, 1344 μL (21 voxels) for 230–270 ms epoch, 1088 μL (17 voxels) for 280–320 ms epoch, and 1280 μL (20 voxels) for 330–370 ms epoch. Thus, only significant activations with the cluster size reached a specific cluster extent threshold listed above were reported for each contrast during each epoch.

Results

Behaviors

Figure 1*A* shows the group mean proportion of trials in which both vowels were correctly identified during the training phase as a function of Δf_0 or Δlocation. Both the Frequency and Location groups achieved about 40% accuracy (the chance level is 25%) under the baseline condition (when the 2 vowels had 0-semitone difference in Δf_0 and 0° spatial separation) in the first block of the training session, indicating no remarkable group difference before training. In both groups, a repeated-measures ANOVA revealed a significant main effect of

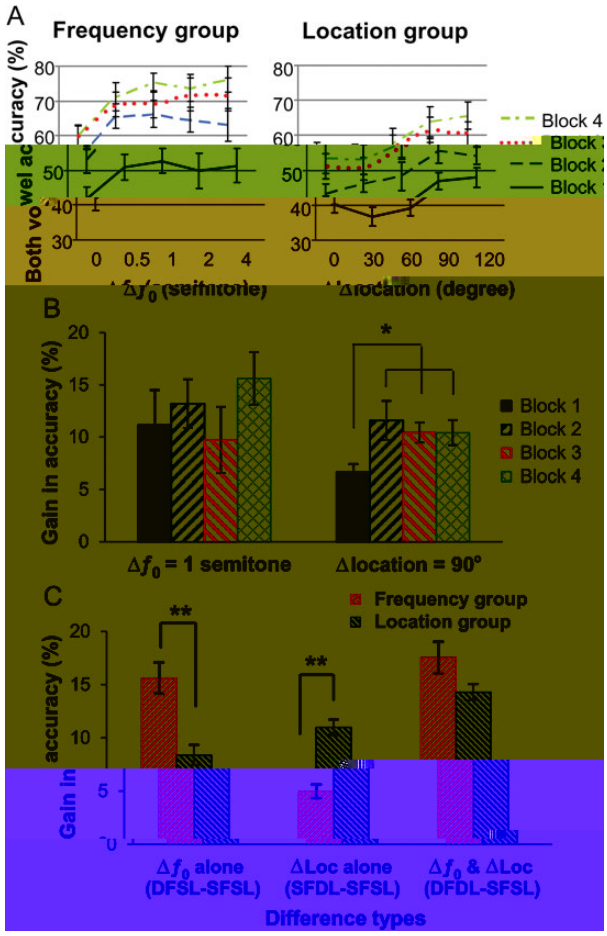

Figure 1. Behavioral performance during training and MEG recording. (*A*) Group mean accuracy for identifying both vowels during training is plotted as a function of Δf_0 or Δlocation. (*B*) Group mean gain in accuracy when Δf_0 was 1 semitone or Δlocation was 90° during training. $^*P < 0.05$ by paired *t*-tests. (*C*) Group mean gain in accuracy for Δf_0 alone, Δlocation alone, and both Δf_0 and Δlocation during MEG recording. $^{**}P < 0.001$ by independent-sample *t*-tests. The error bars represent the standard error of the mean.

block ($F_{3,33} = 37.30$ and 15.02, respectively, $P < 0.001$ in both cases), indicating improved vowel accuracy with practice.

In the Frequency group, pair-wise comparisons showed that accuracy improved significantly from the first to the second block of trials ($P < 0.001$). Accuracy measures in all subsequent blocks were also significantly higher than that in the first block of trials ($P < 0.001$ in all cases). After the second block of trials, the gain in accuracy was smaller, with participants showing neither a significant improvement between the second and third blocks of trials nor between the third and fourth blocks of trials ($P = 0.072$ and 1.00, respectively). Nonetheless, participants were more accurate in the fourth than in the second block of trials ($P < 0.01$). Finally, accuracy increased with increasing f_0 separation between the 2 vowels ($F_{3,33} = 12.38$, $P < 0.001$ in both cases). Pair-wise comparison revealed that performance improved with increasing f_0 separation up to 1 semitone (all $P < 0.05$), and plateaued thereafter from 1 to 4 semitones.

Similarly, in the Location group, pair-wise comparisons showed significant gains in accuracy between the first and second blocks of trials ($P < 0.02$) and between the second and third blocks of trials ($P < 0.05$). Accuracy in all the subsequent blocks of trials was also significantly higher than in the first block of trials ($P < 0.01$ in all cases). There was no significant increase in accuracy between the third and fourth blocks of trials ($P = 1.00$). Accuracy also increased with increasing spatial separation between the 2 vowels ($F_{3,33} = 56.27$, $P < 0.001$ in both cases). Participants performed better when the 2 vowels were separated by 9(v)1 trrreas1243260hin cythetafter-23415.7(92.7(also)193327.ncr)233(trrr)23(eassm11Tf6(er)1(343260h)4345(trrr)43.7

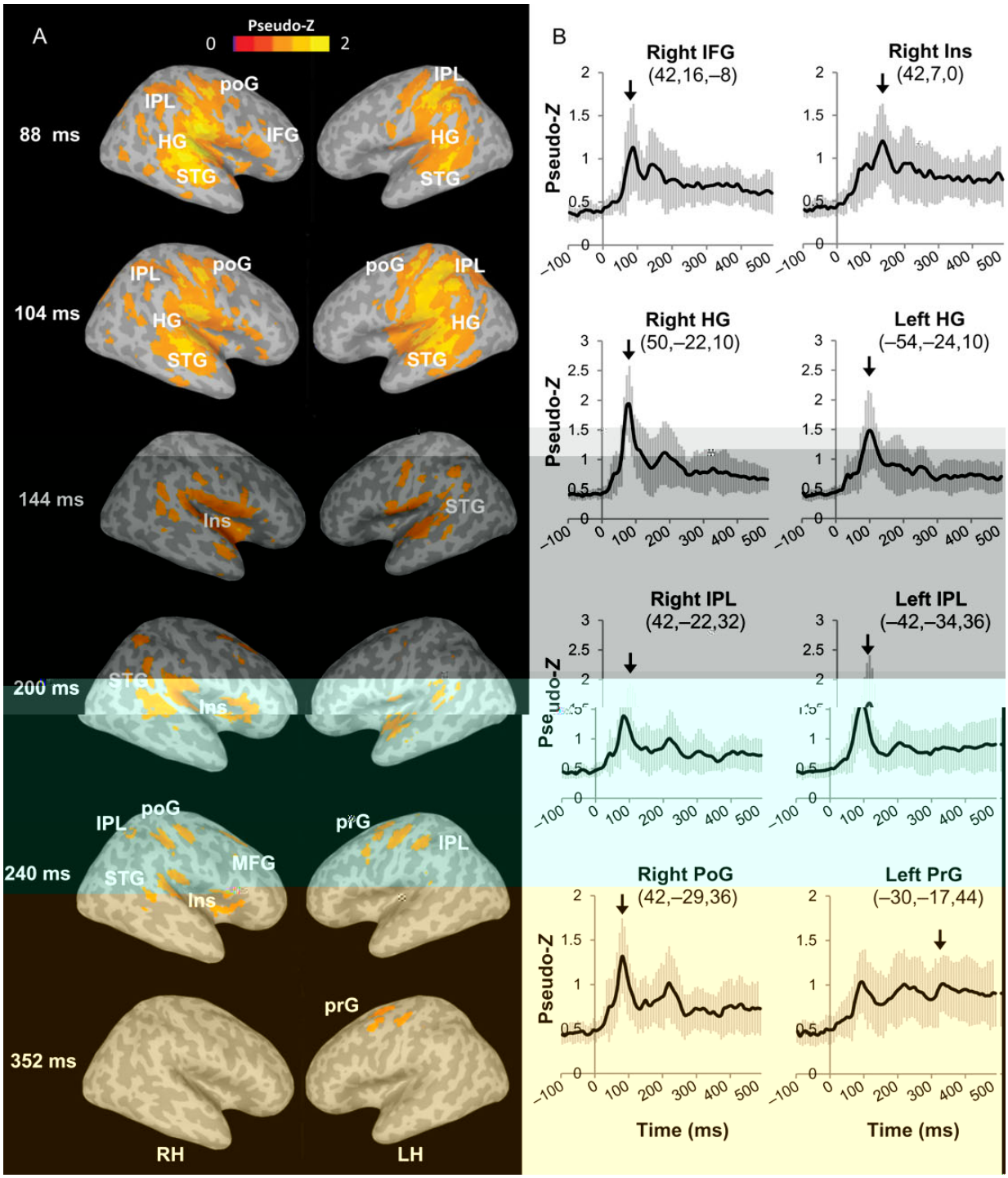

Figure 2. Grand mean ER-SAM maps and source waveforms during the double-vowel task. (*A*) ER-SAM maps averaged across all participants and all stimulus types at selected latencies are thresholded at pseudo-$Z > 0.9$ and overlaid onto a template brain. (*B*) MEG source waveforms averaged across all participants and all stimulus types at selected peak voxel in ER-SAM maps. The error bars represent the standard error of the mean. The arrows indicate the latency when the chosen voxel was defined as local maxima. HG, Heschl's gyrus; IFG, inferior frontal gyrus; Ins, insula; MFG, middle frontal gyrus; IPL, inferior parietal lobule; poG, postcentral gyrus; prG, precentral gyrus; STG, superior temporal gyrus.

supramarginal gyrus (SMG, 30–70, 230–270 and 330–370 ms intervals), the bilateral IPL (right IPL, 80–120 and 280–320 ms intervals; left IPL, 130–170 and 230–270 ms intervals), the left angular gyrus (AG, 130–170 ms interval), the left MTG (180–220 ms interval), the posterior area of the left STG (230–270 ms interval), the right prG (230–270 ms interval), the right poG (280–320 ms interval), and the right middle frontal gyrus (280–320 ms interval). Thus, although both groups were processing the same set of stimuli, distinct brain networks were revealed depending on the trained features, with the Frequency group recruiting more anterior and ventral temporo-frontal areas ("what" pathway) while the Location group activating more posterior and dorsal temporo-parietal areas ("where" pathway).

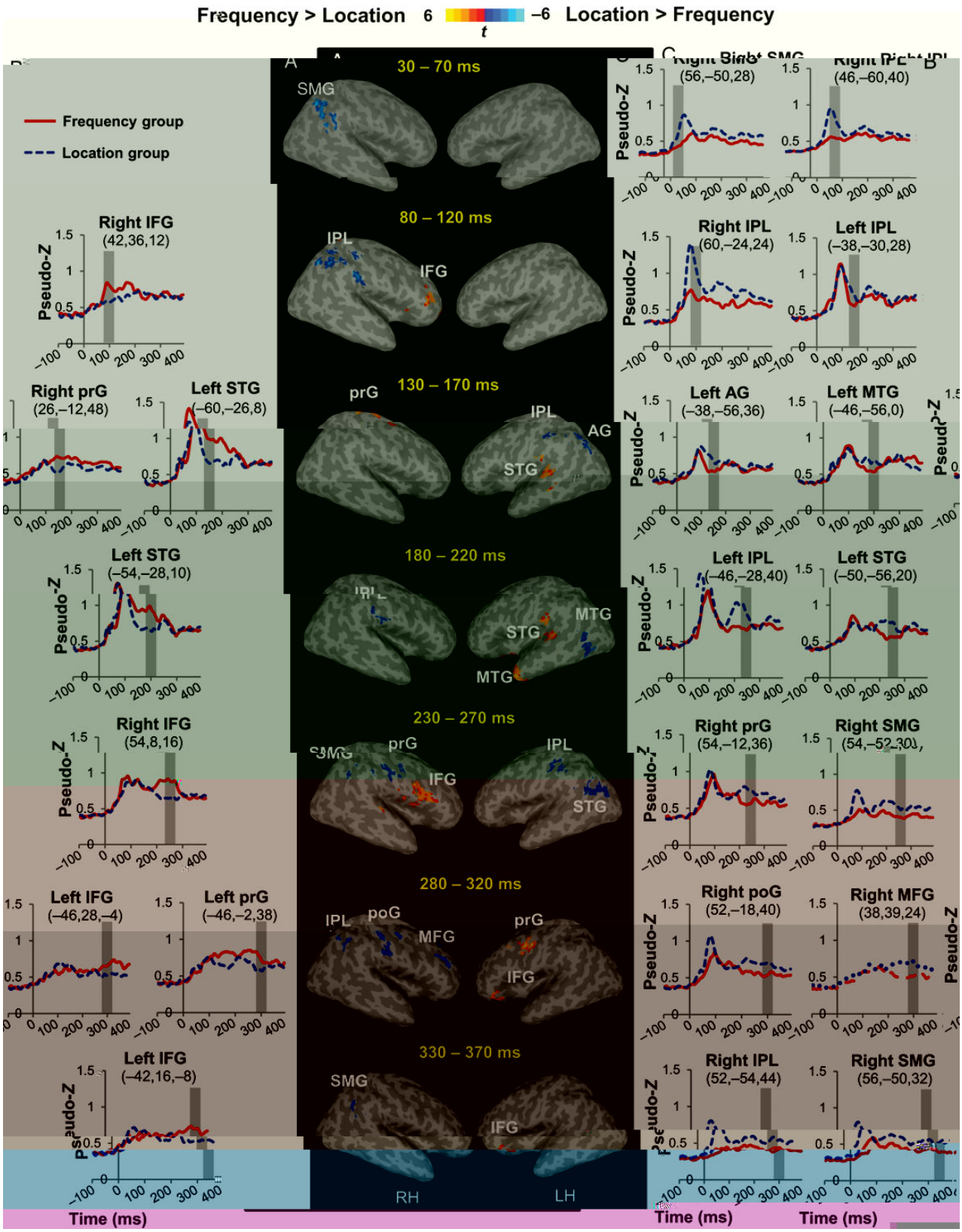

Figure 3. Activation maps and source waveforms showing the feature-specific training effect. (*A*) Contrast maps of the MEG source activity between the 2 groups across stimulus types for six 40-ms intervals are overlaid on a template brain. All activations are significant at corrected $P < 0.05$ and cluster size >1088 μL. (*B* and *C*) Group mean MEG source waveforms at selected clusters exhibiting remarkable group differences, (*B*) Frequency group > Location group; (*C*) Location group > Frequency group. The numbers below each cluster label show the Talairach coordinates of the peak voxel. The gray bars indicate the 40-ms interval showing a significant group difference. AG, angular gyrus; IFG, inferior frontal gyrus; IPL, inferior parietal lobule; MFG, middle frontal gyrus; MTG, middle temporal gyrus; poG, postcentral gyrus; prG, precentral gyrus; SMG, supramarginal gyrus; STG, superior temporal gyrus.

Table 1.
Feature-specific training effect on neuromagnetic activity

Latency	Brain regions	BA	Peak Talairach coordinate			*t*-value	No. of voxels
			x (mm)	*y* (mm)	z (mm)		
Frequency group > Location group							
80–120 ms	R inferior frontral gyrus	46	42	36	12	3.503	37
130–170 ms	R precentral gyrus	4	26	−12	48	2.663	117
	L superior temporal gyrus	42	−60	−26	8	3.508	29
180–220 ms	L middle temporal gyrus	21	−46	7	−32	6.077	303
	L superior temporal gyrus	41	−54	−28	10	3.511	32
230–270 ms	R inferior frontal gyrus	44	54	8	16	2.825	94
280–320 ms	L precentral gyrus	6	−46	−2	38	5.559	54
	L inferior frontal gyrus	47	−46	28	−4	3.276	37
330–370 ms	L inferior frontal gyrus	47	−42	16	−8	2.443	21
Location group > Frequency group							
30–70 ms	R supramarginal gyrus	40	56	−50	28	4.177	83
80–120 ms	R inferior parietal lobule	40	46	−60	40	5.335	115
		40	60	−24	24	4.087	36
130–170 ms	L inferior parietal lobule	40	−38	−30	28	2.531	43
	L angular gyrus	39	−38	−56	36	3.415	27
180–220 ms	L middle temporal gyrus	19	−46	−56	0	4.159	29
	R inferior parietal lobule	40	65	−26	24	3.651	27
230–270 ms	L superior temporal gyrus	22	−50	−56	20	3.037	78
	L inferior parietal lobule	40	−44	−30	42	2.954	25
	R precentral gyrus	4	54	−12	36	3.568	26
	R supramarginal gyrus	40	54	−52	30	3.414	25
280–320 ms	R postcentral gyrus	3	52	−18	40	4.056	46
	R middle frontal gyrus	10	38	39	24	3.348	36
	R inferior parietal lobule	40	52	−54	44	4.272	30
330–370 ms	R supramarginal gyrus	40	56	−50	32	3.183	22

Note: All activations are significant at $P < 0.05$ and survive family-wise correction for multiple comparisons. BA, Brodmann's area.

Brain-Behavior Correlations

Figure 4 shows the correlations between brain neuromagnetic activity and cue-induced gain in performance in 2 regions showing significant (corrected $P < 0.05$) feature-specific training effect in source activity. As shown by Figure 4*A*, individuals' mean source activity in the right IFG during the 80–120 ms interval significantly correlated with listeners' gain in accuracy for Δf_0 cue alone ($R^2 = 0.58$, $P < 0.01$, Pearson correlation). Participants from the Frequency and Location groups nicely formed 2 clusters with the former showing a larger gain in performance by Δf_0 cue and stronger source activity in the right IFG than the latter. In contrast, individuals' mean source activity in the right IPL during the 80–120 ms interval remarkably correlated with listeners' gain in accuracy for Δlocation cue alone (Fig. 4*B*, $R^2 = 0.63$, $P < 0.01$). The 2 groups clearly separated from each other with participants from the Location group achieving a higher gain in accuracy by Δlocation cue and a stronger source activity in the right IPL than the Frequency group.

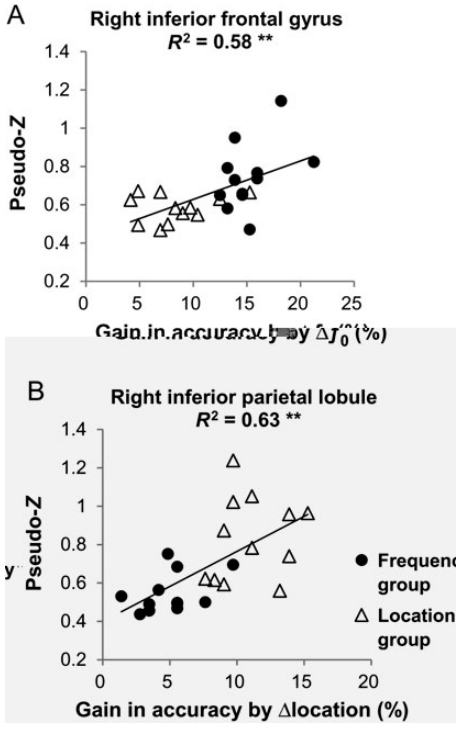

Figure 4. Brain-behavior correlation. Individuals' mean source activity in the right inferior frontal gyrus (*A*) and in the right inferior parietal lobule (*B*) during the 80–120 ms interval from both the Frequency (filled circles) and Location groups (open triangles) are plotted against the listeners' gain in accuracy for Δf_0 cue alone and Δlocation cue alone, respectively. The 2 regions were chosen as showing a significant (corrected $P < 0.05$) group difference in source activity. $^{**}P < 0.01$ by Pearson correlation.

Discussion

The present study was designed to investigate whether a brief training program could yield feature-specific gains in performance that would coincide with changes in neuromagnetic brain activity. As expected, participants' accuracy in identifying 2 vowels presented simultaneously improved during a 45-min training session. More importantly, the training effects transferred to the MEG recording session in a feature-specific fashion, such that participants who learned to use f_0 separation between the 2 vowels showed greater accuracy when this

cue was present than when the location difference was available. Conversely, participants who learned to use the spatial separation between vowels showed greater benefit when that information was present compared with trials where vowels differing in f_0 were presented at the same location. In addition, in both trained groups, the gain in accuracy for the trained feature was greater than that observed in participants trained on the other feature or a control group (Du et al. 2011) who was not exposed to nor trained on the same stimuli prior to the MEG recording session. These results are consistent with prior behavioral studies (Hawkey et al. 2004; Hussain et al. 2012) and provide further evidence that brief training can enhance perceptual sensitivity.

In the present study, the training phase may act as a priming task that biased attention toward the trained acoustic feature even when, in the subsequent task, participants were not explicitly told to focus attention on the voice or location. That is, the training may implicitly bias participants to focus their attention on the trained spectral or location difference between the 2 vowels, which would appear more salient than the untrained cue during MEG measurement that followed the brief training session. The observed pattern of behavioral results and source brain activity are consistent with a rapid adjustment in perceptual sensitivity such that top-down instructions and task parameters during the training phase confer the selectivity, which is necessary to modify a single feature representation (f_0 or location) without affecting other spatially organized feature representations embedded within the same neural circuitry. This training-induced selectivity appears to be fairly long-lasting and resilient, as it lingers throughout the delay between training and MEG recording and remains even when participants did not receive feedback on their performance during the MEG session.

Although the pattern of behavioral and neuromagnetic data appears to be consistent with an attention bias and/or changes in perceptual sensitivity, it remains possible that changes in performance and brain activity can be partially due to a repetition effect. Evidence from a repetition priming paradigm has shown superior performance in identifying auditory stimuli when the same stimuli were used in a prior task (Stuart and Jones 1995). Consequently, in the Frequency and Location groups, the enhanced performance for Δf_0 and Δlocation, respectively, could be attributed to the fact that the participants were more familiar with the stimuli. However, this explanation is unlikely for 2 reasons. First, the SFSL condition was also presented during both training and MEG recording, but there was little evidence of repetition priming for this trial type. More importantly, the feature-specific effects observed in the present study re ect a relative rather than an absolute difference between the training and the MEG recording. That is, the analyses focused on the gain related to having frequency or spatial separation, and this approach controlled for having the same stimuli present in the training and the MEG recording session.

Using the spatial filtering technique of ER-SAM for imaging cortical source activity, we have shown that concurrent-vowel segregation and identification engaged a widely distributed neural network that comprised the primary and associative auditory cortices as well as prefrontal, inferior parietal, and response-related sensorimotor areas. Many of the regions showing significant source activity have been implicated in spatial and nonspatial auditory selective attention tasks and are part of a brain network that controls the focus of attention to task-relevant feature or stimuli (Maeder et al. 2001; Rämä et al. 2004; Degerman et al. 2006, 2008; Krumbholz et al. 2007; Alain et al. 2008; Paltoglou et al. 2011). For instance, enhanced activity in the IPL has been consistently reported during auditory localization tasks (Arnott et al. 2004) and is thought to play an important role in auditory spatial working memory (Alain et al. 2008; Alain, Shen, et al. 2010) and/or transforming auditory location into visuo-spatial coordinates that can guide the ocular system toward the sound sources (Arnott and Alain 2011). In the present study, the increased activity in the IPL could also index auditory source separation using spatial cues. Such an account would be consistent with functional MRI (fMRI) studies showing enhanced activation in the IPL with an increasing number of spatially distinct sound sources presented simultaneously (Zatorre et al. 2002; Smith et al. 2010).

The STG and IFG have also been repeatedly mentioned in tasks that require identifying sound objects (Clarke et al. 2002; Adriani et al. 2003; Arnott et al. 2004) and are often considered being part of the ventral (what) pathway. These areas may play an important role in speech segregation and identification as evidence by a prior fMRI study showing enhance activity in the left thalamus as well as primary and associative auditory cortices when concurrent vowels differing in f_0 were successfully identified (Alain et al. 2005). Moreover, there is evidence that the activity in primary and associative auditory cortices is modulated by the perception of concurrent sound objects associated with increasing inharmonicity between a lower harmonic and its fundamental (Arnott et al. 2011), and these areas are also part of a network involved during speech in noise perception (Wong et al. 2008; Bishop and Miller 2009; Dos Santos Sequeira et al. 2010).

In the present study, neuromagnetic source activity associated with segregating and identifying 2 vowels presented simultaneously was observed as early as 100 ms after sound onset in sensory-specific as well as multimodal areas such as the IFG and IPL. The time course of source activity observed in auditory, parietal, and prefrontal cortices is consistent with findings from single- and multi-unit recordings in non-human primates (e.g., Vaadia et al. 1986; Mazzoni et al. 1996) as well as intracerebral recording in epileptic patients (e.g., Richer et al. 1989; Molholm et al. 2006), which have revealed time-locked neural activity to auditory stimuli as early as 100 ms after sound onset. Notably, at longer latencies (i.e., 200–400 ms poststimulus), the source activity was predominantly observed in multimodal areas including the parietal and prefrontal cortices as well as motor areas related to response preparation and execution, consistent with hierarchically organized attention-related increased activity in sensory and attention networks (Ross et al. 2010).

Further, the neuroplastic changes associated with the feature-specific gain in performance were revealed by contrasting source activity between the 2 training groups, which yielded a double dissociation with participants trained on spectral and spatial cues showing higher source activity in ventral ("what") and dorsal ("where") brain areas, respectively (Rauschecker and Tian 2000; Alain et al. 2001; Maeder et al. 2001; Arnott et al. 2004; Arnott and Alain 2011). To our knowledge, this is the first demonstration of a feature-specific gain in human brain activity following a brief training designed to enhance perceptual sensitivity to differences in voice pitch and voice location. Notably, this group difference was not all or none, but rather appeared to re ect a bias in recruiting ventral or dorsal brain regions while performing the double-vowel identification task. This is

consistent with prior fMRI studies that have revealed relative differences in activation in auditory "what" and "where" processing streams as a function of task instruction/demand and selective attention effect rather than absolute differences (e.g., Alain et al. 2001; Ahveninen et al. 2006; Degerman et al. 2006; Alain et al. 2008; Paltoglou et al. 2011). The group difference may also indicate changes in the tuning properties of the multimodal neurons engaged in sound identification and localization. These rapid feature-specific changes are consistent with animal studies showing task-relevant changes in the receptive fields and synchronized neuronal firing of auditory neurons within minutes of training (Bakin and Weinberger 1990; Edeline et al. 1993; Bakin et al. 1996; Fritz et al. 2003; Du et al. 2012).

Our findings provide further evidence for rapid changes in cortical evoked responses after less than an hour of auditory training on sound spectro-temporal (Alain et al. 2007; Alain, Campeanu, et al. 2010; Ben-David et al. 2011) or spatial (Spierer et al. 2007) features, and offer the first neuroimaging evidence for rapid perceptual (i.e., feature-specific) learning in humans. The neuromagnetic source analyses nicely complement prior fMRI research and provide unique chronometric information regarding the sequence of neural events associated with rapid learning during speech separation and identification. Enhanced source activity in the left STG around N1m–P2m latency (130–220 ms) and in the bilateral IFG at early (80–120 ms) and late (230–370 ms) latency in participants trained on frequency rather than location cue is consistent with previous reports showing spectral-training-related changes in early (~130 ms) and late (~340 ms) ERPs localized in the right auditory cortex and inferior prefrontal cortex, respectively (Alain et al. 2007). This is also in accordance with rapid changes in sensory evoked responses (N1 and P2 amplitude) and a later ERP (~320 ms) over the left frontal site that differed from changes related to procedural learning during stimulus exposure and task repetition in participants trained on speech content like voice onset time (Alain, Campeanu, et al. 2010; Ben-David et al. 2011). On the other hand, compared with the Frequency group, larger source activity in the left posterior STG (230–270 ms) and multiple bilateral parietal regions including the IPL, SMG, and AG throughout the observed period (30–370 ms) in the Location group provides support for spatial-training-related changes in auditory evoked responses at 195–250 ms originating from the left inferior parietal cortex (Spierer et al. 2007). Notably, the group difference in the ventral and dorsal stream activity began as early as 100 ms after sound onset, suggesting the preset (before the presentation of stimuli) attentional bias on trained attributes and differential "warm-up" of corresponding pathways. This early differential activation in the ventral (e.g., IFG) and dorsal regions (e.g., IPL) correlated with individuals' behavioral improvement from spectral and spatial cues, respectively, indicating the critical role of rapid feature-specific tuning of auditory processing streams in speech segregation and identification. The feature-specific effect at later stage (200–400 ms) may re ect learning-induced alteration on task-related processes, including stimulus classification based on different sound attributes in the anterior and posterior associative auditory cortex, nonspatial and spatial working memory in prefrontal and parietal cortices, response selection, preparation, and execution in motor-related areas. Our results are consistent with the temporal sequence of γ-band increases over the left inferior frontal and left posterior parietal cortex during the delayed maintenance phase of an auditory pattern (Kaiser et al. 2003) and spatial working memory (Lutzenberger et al. 2002) tasks, respectively, and over the prefrontal cortex and higher-order executive networks during the responses in both MEG studies. Moreover, taking advantage of the temporal fidelity of the MEG measurement, our findings shed light on the timing of rapid learning-induced modulation of the ventral (nonspatial) and dorsal (spatial) pathways, which complement prior studies using the fMRI approach (Alain et al. 2001; Ahveninen et al. 2006; Degerman et al. 2006; Paltoglou et al. 2011). Our results suggest that neural systems underlying learning and memory are quickly and adaptively adjusted depending on goal-directed behavior. These may re ect top-down attention to task-relevant attributes to optimally process differences in the frequency or location of the stimulus along the hierarchical auditory processing streams (Woods and Alain 1993; Woods et al. 1994, 2001). Further research combining both MEG and fMRI may help clarify the neural interactions underlying such rapid neuroplastic changes, which could help determine whether these rapid changes in source activity are precursors to long-term changes as the training regimen continues.

In summary, a 45-min training session aimed to improve participants' abilities to use f_0 or location cues to separate and identify concurrent vowels yielded behavioral benefits specific to the trained attribute. Gains in performance coincided with rapid feature-specific changes in source activity along the ventral "what" and dorsal "where" auditory pathways, respectively. These group differences re ect a rapid recalibration of the perceptual system with training and cannot be easily accounted for by procedural learning, because the stimulus-response requirements were identical in both groups. Taken together, this study provides the first neuromagnetic evidence for rapid perceptual learning in humans and shows that attention can be quickly and adaptively allocated to sound identity and sound location, an effect that is mediated by the differential engagement of brain areas along the cerebral ventral and dorsal streams.

Supplementary Material

Supplementary material can be found at: http://www.cercor.oxfordjournals.org/

Funding

This research was supported by grants from the Canadian Institutes of Health Research (MOP106619), the Natural Sciences and Engineering Research Council of Canada, Chinese State Scholarship Fund, the "973" National Basic Research Program of China (2009CB320900), and the National Natural Science Foundation of China (31170985).

Notes

C fl c fI e e : None declared.

References

Adriani M, Maeder P, Meuli R, Thiran AB, Frischknecht R, Villemure JG, Mayer J, Annoni JM, Bogousslavsky J, Fornari E et al. 2003. Sound recognition and localization in man: specialized cortical networks and effects of acute circumscribed lesions. Exp Brain Res. 153:591–604.

Ahissar M, Hochstein S. 1993. Attentional control of early perceptual learning. Proc Natl Acad Sci USA. 90:5718–5722.

Ahissar M, Nahum M, Nelken I, Hochstein S. 2009. Reverse hierarchies and sensory learning. Philos Trans R Soc Lond B Biol Sci. 364:285–299.

Ahveninen J, Jaaskelainen IP, Raij T, Bonmassar G, Devore S, Hamalainen M, Levanen S, Lin FH, Sams M, Shinn-Cunningham BG et al. 2006. Task-modulated "what" and "where" pathways in human auditory cortex. Proc Natl Acad Sci USA. 103:14608–14613.

Alain C, Arnott SR, Hevenor S, Graham S, Grady CL. 2001. "What" and "where" in the human auditory system. Proc Natl Acad Sci USA. 98:12301–12306.

Alain C, Campeanu S, Tremblay K. 2010. Changes in sensory evoked responses coincide with rapid improvement in speech identification performance. J Cogn Neurosci. 22:392–403.

Alain C, He Y, Grady C. 2008. The contribution of the inferior parietal lobe to auditory spatial working memory. J Cogn Neurosci. 20:285–295.

Alain C, Reinke K, McDonald KL, Chau W, Tam F, Pacurar A, Graham S. 2005. Left thalamo-cortical network -231.eri2Tjsuccessful. (identi)219.9(pa545.3223.1)10.5ogn7T

Robinson SE. 2004. Localization of event-related activity by SAM(erf). Neurol Clin Neurophysiol. 2004:109.
Robinson SE, Vrba J. 1998. Functional neuroimaging by synthetic aperture magnetometry (SAM). In: Yoshimoto T, Kotani M, Kuriki S, Karibe H, Nakasato N, editors. Recent advances in biomagnetism. Sendai: Tohuku University Press. p. 302–305.
Ross B, Hillyard SA, Picton TW. 2010. Temporal dynamics of selective attention during dichotic listening. Cereb Cortex. 20: 1360–1371.
Shtyrov Y, Nikulin VV, Pulvermuller F. 2010. Rapid cortical plasticity underlying novel word learning. J Neurosci. 30:16864–16867.
Smith KR, Hsieh IH, Saberi K, Hickok G. 2010. Auditory spatial and object processing in the human planum temporale: no evidence for selectivity. J Cogn Neurosci. 22:632–639.
Spierer L, Tardif E, Sperdin H, Murray MM, Clarke S. 2007. Learning-induced plasticity in auditory spatial representations revealed by electrical neuroimaging. J Neurosci. 27:5474–5483.
Stuart GP, Jones DM. 1995. Priming the identification of environmental sounds. Q J Exp Psychol A Hum Exp Psychol. 48:741–761.
Vaadia E, Benson DA, Hienz RD, Goldstein MH Jr. 1986. Unit study of monkey frontal cortex: active localization of auditory and of visual stimuli. J Neurophysiol. 56:934–952.
Van Veen BD, van Drongelen W, Yuchtman M, Suzuki A. 1997. Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. IEEE Trans Biomed Eng. 44:867–880.
Watson CS. 1980. Time course of auditory perceptual learning. Ann Otol Rhinol Laryngol Suppl. 89:96–102.
Wenzel EM, Arruda M, Kistler DJ, Wightman FL. 1993. Localization using nonindividualized head-related transfer functions. J Acoust Soc Am. 94:111–123.
Wightman FL, Kistler DJ. 1989a. Headphone simulation of free-field listening. I: stimulus synthesis. J Acoust Soc Am. 85:858–867.
Wightman FL, Kistler DJ. 1989b. Headphone simulation of free-field listening. II: psychophysical validation. J Acoust Soc Am. 85:868–878.
Wong PC, Uppunda AK, Parrish TB, Dhar S. 2008. Cortical mechanisms of speech perception in noise. J Speech Lang Hear Res. 51:1026–1041.
Woods DL, Alain C. 1993. Feature processing during high-rate auditory selective attention. Percept Psychophys. 53:391–402.
Woods DL, Alain C, Diaz R, Rhodes D, Ogawa KH. 2001. Location and frequency cues in auditory selective attention. J Exp Psychol Hum Percept Perform. 27:65–74.
Woods DL, Alho K, Algazi A. 1994. Stages of auditory feature conjunction: an event-related brain potential study. J Exp Psychol Hum Percept Perform. 20:81–94.
Wright BA, Fitzgerald MB. 2001. Different patterns of human discrimination learning for two interaural cues to sound-source location. Proc Natl Acad Sci USA. 98:12307–12312.
Zatorre RJ, Bouffard M, Ahad P, Belin P. 2002. Where is 'where' in the human auditory cortex? Nat Neurosci. 5:905–909.