Neural Processes Underlying Perceptual Learning of a Difficult Second Language Phonetic Contrast

Daniel Callan*, †, Keiichi Tajima†, Akiko Callan*, Reiko Akahane-Yamada†, and Shinobu Masaki*

*Brain Activity Imaging Center, Information Sciences Division, ATR International, Japan
†Spoken Language Acquisition Project, Information Sciences Division, ATR International, Japan
dcallan@isd.atr.co.jp

Abstract

Neural processes underlying the perceptual learning of the English /r-l/ phonetic contrast by native Japanese speakers before and after extensive perceptual identification training using feedback was investigated using fMRI. Relative to control conditions (English /b-v/ and /b-g/ contrasts), the /r-l/ contrast showed greater brain activity as well as functional connectivity (reflecting underlying global mappings) postrelative to pre-training bilaterally in frontal and temporal brain areas involved with speech processing as well as the cerebellum and the putamen.

1. Introduction

Adult speakers have great difficulty learning to identify certain second-language (L2) phonetic contrasts that their native language (L1) does not use. A classic example of this is the English /r-l/ contrast for native Japanese speakers. The Japanese language does not phonetically distinguish between /r/ and /l/. It has been found that native Japanese adults, even after years of exposure to English, cannot reliably identify and discriminate minimal pairs of words contrasting in /r/ and /l/ (e.g. rake and lake) [15]. Recent studies that carried out perceptual identification training with feedback using a number of different speakers as stimuli have shown an increase in /r-l/ minimal-pair identification performance for adult native Japanese speakers (approximately 20%) that generalizes to novel stimuli and is long lasting [1, 5].

The primary objective of the current study is to investigate the neural processes associated with perceptual learning of difficult L2 phonetic contrasts using MRI. Studies investigating L1 processing suggest that the region of the left—and in some studies right—superior temporal gyrus STG (Brodmann area BA 21 and 22) anterior to the primary auditory area (BA 42) is involved with phonological processing [2,3,4,8,18]. Other brain areas that are said to be involved with phonological processing include the medial temporal gyri [4], the left temporal pole [16], and the left hemisphere inferior temporal regions including Broca’s area (BA 44-45) [8,18], and the premotor cortex BA 6 [18]. Brain areas implicated in verbal working memory include bilaterally dorsolateral prefrontal cortex DLPFC BA 46, anterior cingulate, posterior parietal area BA 39, inferior frontal area BA 44-45, premotor area BA 6, and the cerebellum (mainly right cerebellum) [9,11].

Imaging studies concerned with bilingual language processing have focused mainly on determining differences in brain areas implicated in L1 and L2 processing [6]. These studies implicate greater left temporal lobe activity for L1 than for L2 processing. Additional studies have shown that L2 processing activates different areas for the two languages including non overlapping subregions of Broca’s area [12] as well as right hemisphere temporal and frontal regions [7]. However, the neural processes underlying the acquisition of L2 have not been extensively investigated.

The present study investigates perceptual learning of difficult L2 phonetic contrasts using fMRI. Relative to pre-training bilaterally in frontal and temporal brain areas involved with speech processing as well as functional connectivity (defining global mappings) in brain regions thought to be involved with speech processing (STG anterior to auditory cortex and Broca’s area) as well as brain areas involved with learning, memory and integrative functions (hippocampus, basal ganglia, and cerebellum). In this study, native Japanese monolinguals were tested and scanned using fMRI both before and after training for the English /r-l/ contrast (difficult for native Japanese speakers). The English /b-v/ contrast (also difficult for native Japanese speakers) and the /b-g/ contrast (easy for native Japanese speakers) were also tested and served as a control to ensure that changes in neural activity are a result of learning induced plasticity rather than possible confounds that could arise because of task difficulty or because of the long separation between pre- and post-training scanning sessions.

2. Methods

2.1. Subjects

Nine right-handed monolingual Japanese speakers participated in this study. Subjects were between 21 and 30 years of age. Two subjects were female and seven were male. All of the subjects received some degree of English education in junior and senior high school, in which reading and writing were emphasized rather than conversational skills. None of the subjects had lived in an English speaking community. Subjects were paid for their participation, and gave written informed consent.

2.2. Training

Subjects were trained to identify the English /r-l/ contrast by undergoing perceptual training with feedback [5]. Stimuli used during training were 68 pairs of English words minimally contrasting in /r/ and /l/ spoken by five native American English speakers. On each trial, subjects saw one minimal pair on the computer screen, and heard one of the two words presented through headphones. Subjects responded by clicking the word that they thought they heard. The program then gave immediate feedback on whether the response was correct or
not; if incorrect, the trial was repeated with the same stimulus until the subject responded correctly. The training lasted approximately one month, and consisted of 45 sessions, each session containing 272 trials. In each session, the 68 minimal pairs spoken by one of the native speakers were presented twice each in a randomized order.

Before and after the training, subjects took a pre-test and post-test in which the same 2-alternative forced-choice task was used. However, subjects did not receive feedback about their responses during these tests. Each test consisted of one session with about 160 English words containing /r/ and /l/, none of which appeared during the training phase.

In addition, identification responses were collected during the fMRI scanning before and after training for three phonetic contrasts, /r-l/, /b-v/, and /b-g/. Since these identification tests were conducted during scanning, the signal-to-noise ratio during these tests was much lower than in the above pre- and post-tests.

2.3. Imaging

Brain imaging was performed using a 1.5 Tesla Marconi Magnex Eclipse scanner. First, high-resolution anatomical T2 weighted images were acquired using a fast spin echo sequence. These scans consisted of 50-52 contiguous axial slices with a .75x.75x3mm voxel resolution (matrix size 256x256 pixels) covering the cortex and cerebellum. Second, functional T2* weighted images were acquired using a gradient echo-planar imaging sequence (echo time, 55ms; repetition time, 6000ms; flip angle, 90°). Depending on head size, a total of 50-52 contiguous axial slices were acquired with a 3x3x3mm voxel resolution (matrix size 64x64x64 pixels). The field of view included the cortex and cerebellum.

2.4. Stimuli

The speech stimuli were 24 minimal pairs of monosyllabic English words differing between /r/ and /l/. Analogous sets were also made for /b-v/ (e.g., boat and vote) and for /b-g/ (e.g., bun and gun). The speech stimuli were spoken by a male native English speaker and were recorded digitally in an anechoic chamber. The reference control stimuli consisted of signal correlated noise constructed for each of the word stimuli by randomly multiplying each point by 1 or -1. Signal correlated noise stimuli preserve the temporal envelope of the original word stimuli but do not contain the spectral information.

2.5. Procedure

A block design was used in which 12 monosyllabic words were presented (approximately 85-90 dB SPL) at a rate of one every two seconds via MR-compatible headphones (Hitachi ceramic transducer headphones; frequency range 100-30000 Hz, approximately 20dB SPL passive attenuation) followed by 12 control stimuli (signal correlated noise). This pattern was repeated four times until all 48 monosyllabic words were presented. Separate experimental sessions were conducted for the three contrasts; /b-g/, /b-v/, and /r-l/. The task for the subject was to determine the initial phoneme of each word by button press in the word blocks and to randomly press one of the two buttons in the control blocks. Although this response paradigm attempts to control for brain activity related to the act of pressing a button, it is understood that random versus directed button press in the control versus experimental blocks is likely to involve somewhat different neural processes. Brain imaging was conducted before and after approximately one month of perceptual identification training.

2.6. Data Analysis

2.6.1. Preprocessing of Images

Images were preprocessed using programs within SPM99b (Welcome Department of Cognitive Neurology, London UK). Differences in acquisition time between slices were accounted for. Movement artifact was removed by realigning the images acquired within a single session using a 6 parameter ridged body spatial transform and a least squares approach. The images were spatially normalized to a standard space using a template EPI image (Bounding Box, –90 to 91 mm, -126 to 91 mm, -72 to 109 mm; voxel size, 3x3x3 mm). For the statistical parametric mapping analysis SPM, images were smoothed by a 6-mm FWHM Gaussian kernel. For the independent component analysis ICA, the common voxels across all subjects corresponding to the segmented gray matter were extracted (28642 voxels total for each image). The extracted gray matter images were then smoothed using a 6-mm isotropic FWHM Gaussian kernel.

2.6.2. Statistical Parametric Mapping Analysis

Regional brain activity for the various conditions was assessed on a voxel-by-voxel basis using SPM. A mixed effect model was employed. At the first level (fixed effect within subject), the experimental (word stimuli) versus the control (signal correlated noise stimuli) conditions were assessed separately for each subject and each phonetic contrast. At this level global normalization and grand mean scaling is carried out. In addition the data was modeled using a fixed box car response as well as taking into account the hemodynamic response function. At the second level (random effect between subject) the contrast image of the parameter estimates of the first level analysis for each subject are used as input for a SPM model employing a basic one-sample t-test. Differences in regional activity post- relative to pre- training are determined by masking the post- training resultant SPM (assessed at T=4.5, p < 0.001, df=8, spatial extent threshold 5 voxels) by the pre-training resultant SPM (assessed at T=1.86, p < 0.05, df=8, spatial extent threshold 5 voxels).

2.6.3. Independent Component Analysis

Functional connectivity was assessed using independent component analysis ICA. It is maintained that the fMRI data is composed of a great deal of artifacts as well as various types of task and non task-related physiological activity. ICA is able to maximally separate the fMRI data into component maps (spatial distribution of voxel values) and their associated time course of activation [14]. This is accomplished by using an unsupervised neural network that iteratively determines the unknown unmixing matrix from the entire set of input fMRI data by maximizing the joint entropy (thus minimizing mutual information) between components [14]. Functional connectivity can be assessed because the activity within a independent component is thought to result from the same underlying source (the source may be task-related for some independent components or it may be related to an artifact
such as head movement, machine noise, non-task-related physiological processes, etc…).

In this study the infomax version of ICA was implemented [13]. The segmented gray matter voxels common across all subjects were convolved with the hemodynamic response function. In order to control for differences between experimental sessions and for individual differences, the data was normalized by dividing each voxel value by the average of all voxel values within a scan. For each of the experimental conditions and for each of the subjects an ANOVA was carried out for each voxel between the experimental condition (word stimuli) and the control condition (signal correlated noise stimuli) and a measure of the effect size (omega squared) was calculated. The omega squared values for each voxel for each subject and condition were used as input for the ICA. The input matrix dimensions were 54 (9 subjects, 6 conditions) by 28642 (number of voxels). After training the independent component weight matrix was assessed using ANOVA to determine the task related components that showed a significant difference only for the /r-l/ condition post- relative to pre- training, but not for the /b-v/ and /b-g/ conditions.

3. Results

3.1. Behavioral Performance

Subjects’ ability to identify English /r/ and /l/ improved significantly, from a mean of 62.5% in the pre-test to 80.4% in the post-test [F(1,8) = 78.3; p < .001]. This suggests that the training was indeed effective in enhancing Japanese listeners’ ability to perceive the /r-l/ contrast. Subjects’ performance collected during fMRI scanning on the /b-g/ contrast was, not surprisingly, very high (above 97% on average) both before and after training. For the /r-l/ contrast, performance again improved significantly, from 62.7% before training to 73.2% after training [F(1,6) = 10.5; p < .05]. Finally, the /b-v/ contrast also showed a small but significant improvement, from 60.1% to 68.4% [F(1,7) = 8.5; p < .05]. Even though subjects were not explicitly trained on the /b/-/v/ contrast, it is conceivable that training on the /r-l/ contrast was sufficient to enhance subjects’ sensitivity to other nonnative phonetic categories.

3.2. Statistical Parametric Mapping Analysis

Figs. 1-3 show regional brain activity determined by SPM on a voxel by voxel bases for the various conditions (/r-l/, /b-v/, /b-g/) pre- and post- training (as well as post- training masked by thresholded pre- training) rendered on the surface of the brain. For all of the conditions, pre- and post- training, there is a fair amount of activity in left and right hemisphere superior temporal areas (BA 21 and 22), and the right cerebellum. For the /r-l/ and /b-v/ conditions there is also activity, the left cerebellum, and the left and right hemisphere areas including BA 44-45 Broca’s area, BA 46, BA 47, the insula, the premotor cortex and supplementary motor area SMA BA 6.

The results indicate that there is far more brain activity unique to post- training after masking out pre- training activity for the /r-l/ condition than for the /b-v/ and the /b-g/ conditions. This is clearly shown in Figs. 2-4. For the /r-l/ condition this activity is localized to the left and right superior temporal areas (BA 21 and 22), the right hemisphere analogue of Broca’s area (BA 44-45), the DLPFC (BA 9, BA 46), BA 47, the premotor and SMA BA 6, the putamen bilaterally, the left medial genulate nucleus MGN of the thalamus, and the left, and to a lesser degree the right cerebellum. For the /b-v/ condition there is a small amount of differential activity in the right hemisphere BA 44 and the left cerebellum. For the /b-g/ condition there is a very small amount of differential activity in the left hemisphere BA 47, the left insula, the left MGN, the left putamen, right BA 22, and the cerebellum.
3.3. Independent Component Analysis

Analysis of the ICA weight matrix revealed a task related component (IC48 out of 54) that showed a significant difference only for the /r-l/ condition post- relative to pre-training \( [F(1,8) = 1.02; p = .34] \) and \( /b-g/ [F(1,8) = .13; p = .73] \) conditions. The task-related component map is plotted using a threshold of \( Z = 1.96 \) and a spatial extent threshold of 5 voxels. Fig. 4 shows the task-related component map reflecting the functional connectivity between various regions of the brain. Bilateral activity occurred within the STG (BA 21-22), the temporal pole (BA 38), the DLPFC (BA 46, BA 9), as well as both temporal and frontal areas. Lateralized activity was found for the left fusiform gyrus (BA 37), the left MGN, and the right putamen.

4. Discussion

The results of the SPM analysis revealed increased activity for the /r-l/ condition post- relative to post- training in areas of the brain involved with speech processing, including the STG (anterior to auditory cortex) and Broca’s area. Consistent with other studies [7,12] investigating L2 processing in low fluency bilinguals, brain activity for English words relative to signal correlated noise was found to be localized bilaterally in superior temporal regions as well as inferior frontal regions. A comparison of brain regions activated for the various conditions suggests that difficult contrasts have more frontal activity than easy contrasts. This is consistent with the idea that articulatory maps may be utilized to facilitate perception (perhaps in the form of global mappings [10]). If this is the case, it is necessary for the various regions of the brain to functionally interact. Although integrative processes in the brain can not be determined by an investigation of regional activity on a voxel by voxel basis they can be determined to some degree by measures of functional connectivity.

The results of the ICA revealed functional connectivity between temporal and frontal areas as well as the cerebellum and putamen for only the /r-l/ condition post- relative to pre-training. Activity in the cerebellum and putamen (also revealed by SPM analysis) are interesting in that both of these structures are known to be involved with integrative functions in the brain as well as in feedback based learning.

Further research will explore neural processes involved with speech production and perception for L2 phonetic processing, as well as the use of different training techniques.

5. References