ELSEVIER

Physiology & Behavior 77 (2002) 657–662

PHYSIOLOGY & BEHAVIOR

Teaching the /r/–/l/ discrimination to Japanese adults: behavioral and neural aspects

James L. McClelland[a,b,c,*], Julie A. Fiez[a,b,d], Bruce D. McCandliss[e]

[a]Center for the Neural Basis of Cognition, 115 Mellon Institute, 4400 Fifth Avenue, Pittsburgh, PA 15213, USA
[b]Department of Neuroscience, University of Pittsburgh, Pittsburgh, PA 15260, USA
[c]Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213, USA
[d]Department of Psychology and Learning Research and Development Center, University of Pittsburgh, Pittsburgh, PA 15260, USA
[e]Sackler Institute for Developmental Psychobiology, Weill Medical College, Cornell University, Box 140, 1300 York Avenue, New York, NY 10021, USA

Received 28 August 2002; accepted 21 September 2002

Abstract

Several studies have been conducted to address the learning of a nonnative speech contrast in adulthood, using native speakers of Japanese and the English /r/–/l/ contrast. Japanese adults were asked to identify contrasting /r/–/l/ stimuli (e.g., "rock–lock"). An adaptive training regime starting with initially easy stimuli was contrasted with a fixed training regime using difficult stimuli, with some subjects receiving feedback on the correctness of their responses and others receiving no feedback in both conditions. After three and six sessions of training, subjects received tests assessing identification and discrimination of /r/–/l/ stimuli as well as generalization. In all cases except fixed training without feedback, subjects showed clear evidence of learning, and several indicators suggested that training affects speech perception, rather than simply auditory processes. Neuroimaging studies currently underway are examining the neural basis of these findings.

Keywords: Plasticity; Learning; Speech discrimination training; Hebbian learning; Language learning in adulthood

1. Introduction

One of the mainstays of support for the notion of a critical period in language learning is the finding that people often have great difficulty acquiring aspects of a foreign language to which they are first exposed in adulthood. A well-known case in point is the difficulty Japanese adults have acquiring the distinction between English /r/ and /l/. Not only do Japanese adults have difficulty saying /r/ and /l/ correctly; they also have great difficulty perceiving the difference, and this difficulty can persist despite many years of exposure to English [1].

We and several collaborators have been interested in understanding whether this difficulty might be overcome by specialized training methods and what aspects of such methods are most effective. We have also been interested in determining whether any learning actually reflects changes in perception of the sounds as speech sounds and whether the learning affects the same neural structures that are ordinarily used in speech perception. Here, we consider these three issues in turn.

2. Methods for teaching nonnative speech discriminations

A key idea that motivated the initial stages of our work was the idea that the difficulties adults may have in learning new speech contrasts in adulthood might reflect an undesirable characteristic of Hebbian synaptic modification [2]. Hebb's well-known postulate states that when one neuron participates in firing another, the strength of the connection from the first to the second will be increased [3]. Consider the implications of this in a situation where a stimulus producing a given pattern of activity in sensory input neurons elicits a particular percept in neurons further downstream. The strengths of the connections from the sensory input neurons to the activated downstream neurons (and perhaps also connections among the activated downstream

* Corresponding author. Center for the Neural Basis of Cognition, 115 Mellon Institute, 4400 Fifth Avenue, Pittsburgh, PA 15213, USA. Tel.: +1-412-268-3157; fax: +1-412-268-5060.

E-mail address: jlm@cnbc.cmu.edu (J.L. McClelland).

0031-9384/02/$ – see front matter © 2002 Published by Elsevier Science Inc.
PII: S0031-9384(02)00916-2

neurons) should thereby be increased. This would be expected to strengthen the tendency for a given input to elicit the same resulting percept on a later occasion. Thus, we conjectured, extensive experience in Japanese, where the range of sounds we treat as /r/ and /l/ are all mapped to the same (apparently /l/-like) percept, might lead to a situation where both /r/ or /l/ inputs would elicit the same percept. Once this was established, further presentations might elicit Hebbian learning, but this would simply strengthen the tendency of each sound to elicit the same percept, which would be counterproductive to discrimination.

With this hypothesis as background, there were two predictions. The first prediction was that learning might still be possible in adulthood if we used sounds that so strongly exaggerated the contrast between /r/ and /l/ that our subjects would actually hear them as distinct. The second prediction was that feedback would not be necessary to induce learning and might even be irrelevant. Here, we review a published behavioral study by our group that tested these predictions [4].

Native speakers of Japanese living in the United States were recruited for participation in a speech discrimination learning experiment. The participants were trained in four different regimes that differed in whether initially exaggerated stimuli were used and whether feedback was given. Half of the subjects in each regime were trained to discriminate the minimal pair "rock" vs. "lock", and the other half were trained on "road" vs. "load". The stimuli used were derived from natural spoken examples of the stimuli in question, produced by a male speaker. These examples were used to synthesize two continua of sounds, both interpolating between the natural examples and extrapolating by exaggerating the differences (See Fig. 1). All subjects used in the study performed below 70% correct in discriminating the fixed training stimuli, which were examples of /r/ and /l/ that native English speakers reliably discriminate. Stimuli were presented one at a time over headphones, and subjects were instructed to indicate whether they thought the stimulus began with /r/ or /l/ by pressing corresponding response keys. There were three training sessions, each lasting about 20 min, consisting of 480 training trials plus 50 interleaved probe trials using the fixed training stimuli.

In the crucial adaptive/no feedback regime, subjects initially received training stimuli that exaggerated those properties of the stimuli distinguishing /r/ and /l/. Furthermore, if the subject responded incorrectly to a particular stimulus, the task was made easier by exaggerating the difference one step further, alternating whether the /r/ or the /l/ was further exaggerated until the end of the continuum was reached. If the subject made eight correct responses in succession, the task was made one step harder, also alternating so that the stimuli remained roughly equidistant from

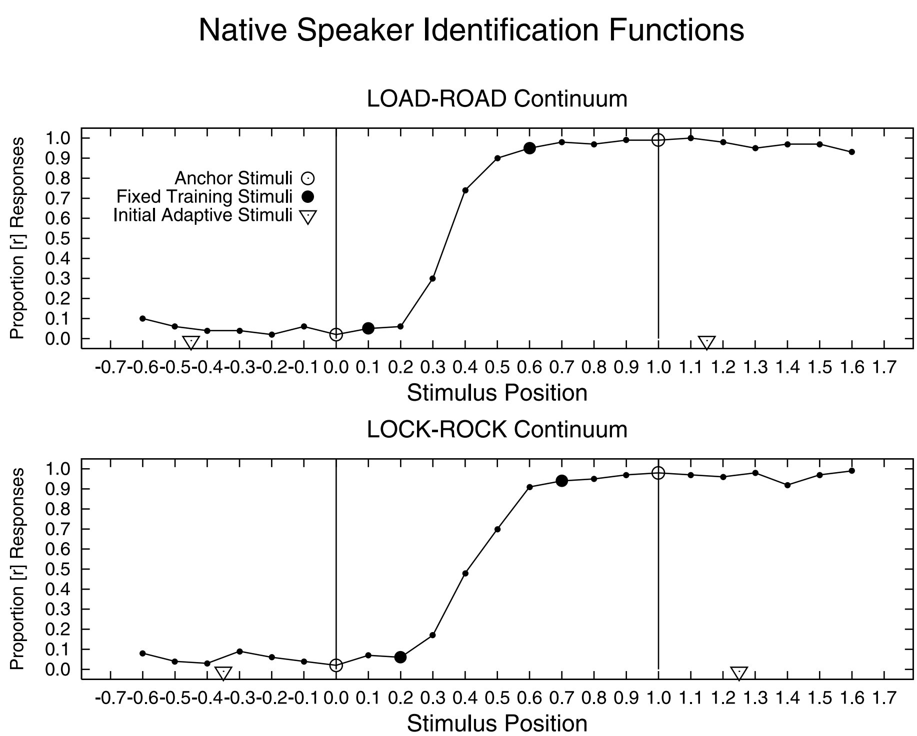

Fig. 1. Mean categorization functions of 12 native English speakers for synthesized speech stimuli from each of the two continua used in the experiments. The *X*-axis represents the position on the stimulus in relation to the anchor stimuli, which are resynthesized versions of naturally spoken stimuli without exaggeration or interpolation. Percentages of trials eliciting *R* responses are plotted on the *Y*-axis for each stimulus. Large empty circles represent the anchor stimuli resynthesized from the recorded base stimuli. Data points between the anchor stimuli are responses to stimuli interpolated between these anchors, and data points in the peripheral regions represent responses to extrapolated speech stimuli. Stimuli used for the fixed training condition are indicated with large filled circles. Triangles point to the positions of the initial stimuli used in the adaptive training condition. Reprinted from Fig. 1, p. 92 in Ref. [4].

the /r/–/l/ boundary for native English speakers. In this condition, there was simply a slight delay between stimuli, and subjects received no feedback on the accuracy of their responses. Conditions were the same in the fixed/no feedback regime, except that two fixed training stimuli were used on all trials, so that subjects received stimuli they could not initially discriminate throughout the training period. The two remaining conditions, adaptive/with feedback and fixed/with feedback, were identical to the corresponding no-feedback conditions, except that a row of green check marks or red × 's was shown to signal correct or incorrect responses, immediately following the subject's response.

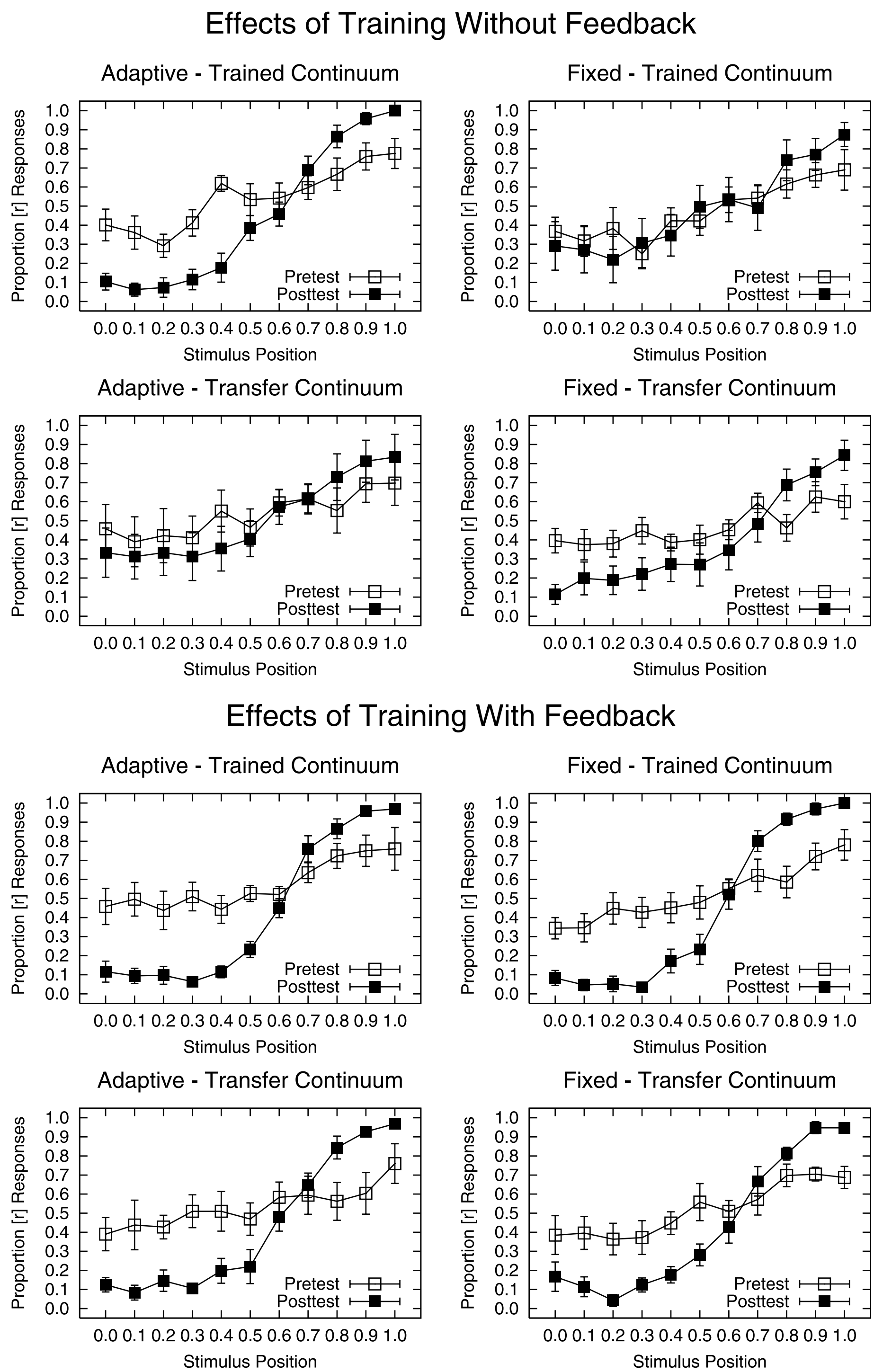

Fig. 2. Mean categorization functions (with standard error bars) for four groups of Japanese subjects ($n = 8$) before and after three 20-min training sessions in the four training conditions of the experiment. Pre- and post-test results are shown on the continuum used in training and on the other continuum used to assess transfer. Adapted from Fig. 2, p. 95, and Fig. 6, p. 99, in Ref. [4].

In accordance with predictions from the Hebbian account, adaptive training without feedback resulted in robust learning, whereas fixed training without feedback resulted in a failure to learn. Extensive exposure to stimuli that the subjects could not initially discriminate thus produced little or no benefit, but exposure to stimuli that were exaggerated so that the subjects could perceive them as distinct led to fairly rapid and robust learning. The main measure of learning (Fig. 2) is based on identification of stimuli ranging between the spoken examples of /r/ and /l/. Before training, identification functions were fairly flat, showing little differentiation of the stimuli. After 3 days of adaptive training without feedback, however, all subjects showed improved identification, and the group average curve showed the S-shape characteristic of normal function, albeit with a slightly shallower slope and some errors in the tails. There was a slight change from the pre- to the post-test in the fixed, no-feedback condition, but a similar change was seen in a control group that received no training. The control results suggest that the pretest (which contained some exaggerated stimuli) may have produced a slight benefit, but that fixed training without feedback had no additional effect.

While results from the no-feedback conditions matched the predictions of the Hebbian account, results from the conditions where feedback was used indicated that, in fact, feedback made quite a bit of difference. Specifically, subjects in both the adaptive and the fixed training regimes showed very good learning when feedback was used. Apparently, Hebb's postulate is not sufficient in and of itself to account for the conditions under which learning occurs; rather, it is clear that feedback can have a substantial effect on the outcome of learning.

3. What did the Japanese adults learn?

Given our training regime, it is possible that the subjects did not really treat the stimuli we used as speech, but instead simply treated them as auditory stimuli with particular distinguishing acoustic properties. To address this issue, we considered two measures not yet discussed. First, we considered whether the training resulted in the formation of new perceptual categories characteristic of speech, as evidenced by the presence of a discrimination peak at the point where their identification functions indicated a crossover from /r/ to /l/ percepts. This part of the investigation is based on an extensive scientific literature that interprets the presence of a discrimination peak at the category boundary as indicating that perception is categorical in nature. The idea is that stimuli in the same category are perceived as the same, but those in different categories are perceived as different, so that discrimination of stimuli in the same category will be poor, but discrimination of stimuli in two different categories will be much better. The data from a discrimination test conducted after training (Fig. 3) indicated that in those conditions where training was effective, a

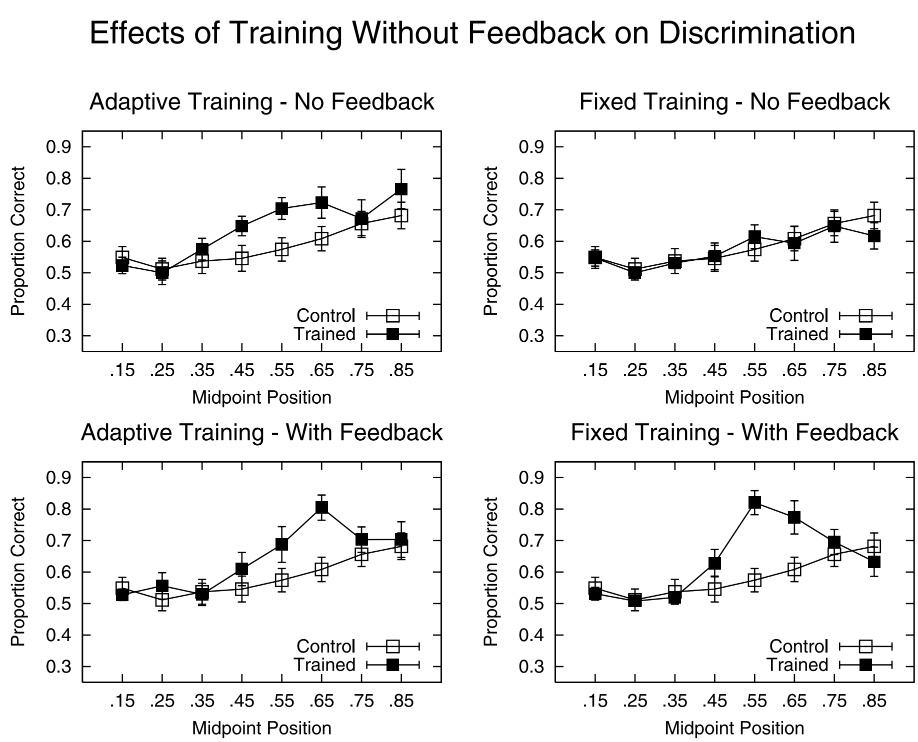

Fig. 3. Effects of training on discrimination of the /r/–/l/ contrast in the four training conditions used in the experiments. All data come from performance on the continuum used in training, based on discrimination performance for stimuli separated by 0.3 units, plotted as a function of the position of their midpoint on the continuum between the two anchor stimuli. The panels contrast performance after training in the indicated condition with performance by control subjects on the post-test. Adapted from Fig. 5, p. 98, and Fig. 7, p. 100, in Ref. [4].

peak in discrimination performance did indeed appear after training.

Second, we considered whether the training generalized from the continuum on which the subjects were trained. The phonetic category of /r/ vs. /l/ is of course not restricted to a specific pair of stimuli such as "rock" vs. "lock" or "road" vs. "load" but extends widely to other cases. Our test for generalization consisted of testing the subjects trained with "rock–lock" on the stimuli from the "road–load" condition, and vice versa. Although this is a fairly limited generalization test and does not tap generalization to new speakers, new positions within a word, or even to many of the possible vowel contexts, we found that those conditions that led to robust learning also resulted in good generalization. These findings suggest that what is learned is at least not hyper-specific and are consistent with the idea that success in learning even with a single pair of stimuli can lead to the acquisition of a speech-like category discrimination.

Taking all the data considered here together, it is apparent that the conditions in which feedback was used resulted in the clearest evidence of the formation of a new speech discrimination. In these conditions, identification functions on the trained contrast became very sharp, learning generalized robustly and there was clear evidence of an acquired discrimination peak at the category boundary. With adaptive training without feedback, the effects were generally weaker, and there was no clear evidence of generalization. However, a subset of the subjects received three further sessions of training and at the end of that time all three effects were clearly evident.

In summary, we have found that it is possible to train Japanese adults to distinguish speech sounds with which they initially have great difficulty. Furthermore, we have found that the training produces several findings indicative of a real change in perception of the sounds as speech, rather than simply in a change in auditory perception. Regarding what it takes to produce these effects, we have found that the Hebbian model provides only an incomplete account, since feedback does affect the outcome of learning.

4. Insights and contributions from neuroimaging

Our behavioral work has provided a springboard for neuroimaging studies focused on the neural basis of speech perception, plasticity and the effects of feedback on learning. Our goal is to identify the brain regions that support the perception of phonemes and that may exhibit plasticity as a result of training. We began with a pilot study in native English speakers and are now in the process of investigating the neural changes associated with our adaptive training regime in Japanese speakers. The work tests the hypothesis that, following training, Japanese subjects are able to discriminate between /r/ and /l/ because of reorganization within the same perceptual system that supports the discrimination of other preexisting phonetic contrasts.

Prior to training, most Japanese subjects cannot reliably discriminate between /r/ and /l/. This posed a problem because we were interested in imaging subjects before and after training, and we wanted to probe the organization of the system without the confound of asking subjects to perform a task on which they may try but routinely fail. As a solution, we turned to the scientific literature on electrical-evoked response literatures, in which the neural organization of speech perception has been investigated extensively using habituation, or "oddball", paradigms [5]. For the typical oddball paradigm, a subject hears a particular stimulus over and over again, and over a period of seconds, the neural response to the stimulus diminishes, or habituates. The repeating (or standard) stimulus is interrupted intermittently by a different, or oddball, stimulus. In some cases, subjects are asked to detect the oddballs (e.g., by pressing a key whenever they perceive an oddball), but in other cases, subjects simply listen to the stimuli or perform a completely unrelated task. In both cases, changes in neural activity associated with the presentation of an oddball can be isolated. Thus, we anticipated that an oddball paradigm could be a very effective way to explore the organization of the speech perceptual system without requiring subjects to perform an overt task.

In a pilot study involving native English speakers, we confirmed that acoustic–phonetic oddballs produce blood-flow changes that can be reliably detected using functional magnetic resonance imaging (fMRI) [6]. We are now extending this work to native Japanese speakers [7]. The study is designed to investigate the response to a phonetic contrast that is native to the Japanese language (/m/ vs. /n/), and a phonetic contrast that is not native to the Japanese language (/r/ vs. /l/), before and after training. The phonemes were presented in the context of a word, in which all of the stimulus after the initial consonant was acoustically identical across the four stimulus types. Specifically, for four runs that were 5.5 min in length, /mode/ was presented once every 250 ms, interrupted every 16–24 s by the oddball /node/. For another four runs, /road/ was presented once every 250 ms, interrupted every 16–24 s by the oddball /load/. Subjects complete this imaging paradigm before and after 3 days of adaptive training on /r/ vs. /l/ identification without feedback.

Prior to training, robust oddball responses should be observed to the /node/ oddball in brain regions that support speech perception at the phonemic level, since /m/ and /n/ are phonemes in the Japanese language and perceived categorically. In contrast, a minimal response to the /load/ oddball should be observed in such regions, since the difference between /road/ and /load/ is not reliably perceived, and thus, /load/ is not a perceptual oddball for native Japanese speakers. Our preliminary results support this prediction. Several regions in the left and right superior temporal gyrus showed a significant response to the /mode/–/node/ pair, but not to the /road/–/load/ pair. This difference was most significant in the left anterior superior temporal region. After training, the oddball response to /road/–/load/ should increase, since subjects are now able to categorically perceive the difference between these tokens. Once again, our preliminary results support this prediction. Regions that showed a significant response only to the /mode/–/node/ pair prior to training now show a significant response to the /road/–/load/ pair. In fact, in many regions (such as the left anterior temporal region) the oddball response to the /road/–/load/ pair is now greater than the oddball response to the /mode/–/node/ pair.

In summary, our imaging approach allowed us to detect similarities and differences in habituation-related brain responses to native versus newly learned phonetic contrasts, in the context of nearly equivalent phoneme identification performance. This work provides support for the claim that our training regime produces reorganization within the same perceptual system that supports the discrimination of other, preexisting, phonetic contrasts. The study indicates that functional imaging methods may be particularly well suited for addressing situations in which performance in some task (perhaps after training) is similar across two groups or conditions, but there may be important research questions about whether this performance reflects the same underlying processes or alternative, compensatory strategies. Results from brain imaging studies could help in distinguishing these possibilities.

Future imaging work will help to further elucidate the link between changes in behavioral performance and the underlying cognitive processes and neural systems that support such plasticity. For instance, behaviorally, we did not observe significant outcome differences between adaptive and set training with feedback. Functional neuroimaging could be used to determine whether the underlying neural changes are similar across these two approaches to training. On the other hand, behaviorally, we did observe significant outcome differences between set training with versus without feedback. Such differences are not predicted by a Hebbian-based account of the learning, and these findings indicate that feedback may modify Hebbian-based learning or recruit additional learning systems. Neuroimaging could be used to identify differences between the brain regions that participate in learning with versus without feedback, and this information might in turn be used to guide further computational modeling efforts.

Acknowledgements

The work reported in this article was supported by grants from the National Science Foundation (LIS-9720348), the National Institute of Mental Health (MH47566), the National Institute on Drug Abuse (DA1410) and the J.S. McDonnell Foundation (96-34).

References

[1] Miyawaki K, Strange W, Verbrugge R, Liberman A, Jenkins JJ, Fujimura O. An effect of linguistic experience: the discrimination of [r] and [l] by native speakers of Japanese and English. Percept Psychophys 1975;18:331–40.

[2] McClelland JL, Thomas A, McCandliss BD, Fiez JA. Understanding failures of learning: Hebbian learning, competition for representational space, and some preliminary experimental data. In: Reggia J, Ruppin E, Glanzman D, editors. Brain, behavioral, and cognitive disorders: the neurocomputational perspective. Oxford, England: Elsevier; 1999. p. 75–80.

[3] Hebb DO. The organization of behavior. New York: Wiley; 1949.

[4] McCandliss BD, Fiez JA, Protopapas A, Conway M, McClelland JL. 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 2002;2:89–108.

[5] Naatanen R, Lehtoski A, Lennes M, Cheour M, Houtllainen AB, IIvonen A, et al. Language-specific phoneme representations revealed by electric and magnetic brain responses. Nature 1997;385:432–4.

[6] Fiez JA, McCandliss BD. Dishabituation of BOLD responses to phonetic oddballs: an event-related fMRI study of magnitude of acoustic change and native language history. Soc Neurosci Abstr 2000;26:464.4.

[7] Tricomi E, McCandliss BD, Conway M, Fiez JA. Imaging training-based changes in Japanese perception of English /r/ and /l/: hemodynamic responses to acoustic and phonemic oddballs in an event-related fMRI design. Soc Neurosci Abstr 2001;27:949.3.