Neural plasticity revealed in perceptual training of a Japanese adult listener to learn American /l-r/ contrast: a whole-head magnetoencephalography study

Yang Zhang[1], Patricia K. Kuhl[1], Toshiaki Imada[2&3], Paul Iverson[4], John Pruitt[5], Makoto Kotani[6], Erica Stevens[1]

[1]Department of Speech and Hearing Sciences, University of Washington, Seattle, Washington 98195, USA; [2]NTT Communication Science Laboratories Laboratory, [3]Real World Computing Partnership, Nippon Telegraph and Telephone Corporation, Atsugi-shi, Kanagawa 243-0198, Japan; [4]Department of Phonetics & Linguistics, University College London, London NW1 2HE, England; [5]Microsoft Corporation, Redmond, WA 98052, USA; [6]Tokyo Denki University, Tokyo, 101-8457, Japan.

ABSTRACT

In this study, behavioral and brain measures were taken to assess the effects of training a Japanese adult subject to perceptually distinguish English /l/ and /ɹ/. Behavioral data showed significant improvement in identifying both trained and untrained speech stimuli. Correspondingly, neuromagnetic results showed enhanced mismatch field responses in the left hemisphere and reduced activities in the right hemisphere. This pattern of neural plasticity was not observed for truncated non-speech stimuli.

1. Introduction

Language experience has a dramatic impact on speech perception and production. One classic example is that of Japanese listeners' poor performance on the English /l-r/ distinction. Early work on developmental speech perception has demonstrated that at a young age infants are capable of detecting phonetic differences regardless of the tested language [1,2]. Evidence of linguistic experience in mapping the sounds of what will become the native language begins to show up as early as at six months of age [3], and by the end of the first year of life, infants have become adult-like in their perception of speech sounds [4]. Japanese infants were found to be no exception in the /l/-/ɹ/ case [5,6].

The study of development is one way of exploring how language experience alters our ability to perceive and produce speech**.** A different method is to study the effects of training adult listeners to perceive non-native speech sounds. The successes and failures of various training methods may provide us with a better understanding of the underlying perceptual mechanisms and the nature of neural plasticity for learning in the formation of new phonetic categories. Many training studies using synthetic speech found that despite substantial stimulus-specific improvement**,** subjects' ability to generalize this training to natural listening situations may remain relatively poor [7,8]. Some recent training studies using a high-variability natural-token procedure, however, did show long-term retention of generalizable training effects in perception as well as production [9,10]. How to integrate the training methods and optimize the interactions of stimulus variables, task variables, and subject characteristics for successful perceptual learning remains a challenge to researchers [11].

Modern brain imaging and neurophysiological tools provide good temporal and spatial resolutions suitable for a direct noninvasive assessment of the neural structures and brain mechanisms that are responsible for cognitive processes. Recent works using event related potentials (ERP) and magnetoencephalography (MEG) indicate that certain components of neural activity such as mismatch negativity (MMN) and mismatch field (MMF) reflect not only pre-attentive sensory detection of small acoustic changes in auditory stimuli but also a higher level of processing for speech sounds that involves language-specific representations [12,13,14]. These studies consistently showed that given equalized amount of acoustic difference for the native and nonnative phonetic contrasts, the neural mismatch responses for nonnative pair were significantly diminished. However, it is unclear how a nonnative phonetic category can be learned and how the internal structure of the learned phonetic category may influence speech perception. Recent studies on brain plasticity showed a promising line of research to address these questions [15].

In this report we describe preliminary results from an ongoing cross-language project using Functional Magnetic Resonance Imaging (fMRI) and MEG to investigate brain plasticity in perceptual training. Given the accuracy of MEG data that is meaningful on a single-subject level [16], this report looked at one Japanese listener's training data and his MMF responses.

2. Methods

2.1 Features of the Training Software Program

A training software program was developed on the basis of Pruitt's original work [17]. The program utilizes the following training methods that are considered to be conducive to speech and language learning:

1. <u>Use of an identification task</u>. Discrimination task only focuses on differences between stimuli, which may not facilitate phonetic categorization.
2. <u>Incremental levels of difficulty</u>. Difficulty is implemented in the variability of talker, vowel context, syllabic context, and the amount of acoustic exaggeration. The use of exaggerated speech is to mimic the listening experience of infants who are exposed in great numbers to the exaggerated acoustic events contained in infant-directed speech (known as "motherese"). This speaking style consists of greater acoustic exaggeration and variety than adult-directed speech and may facilitate the formation of prototypical representation of a phonetic category [18].
3. <u>Bimodal speech cues</u>. A static photographic image of each talker articulating /r/ or /l/ was provided simultaneously for each acoustic presentation.
4. <u>Self-directed, adaptive, motivational training with immediate feedback</u>. Each correct answer is registered onscreen and above-chance performance is recognized with small monetary reward. Incorrect answers are indicated and prompted with playback.

2.2 Behavioral Experiments

For a baseline measure, ten native speakers of Japanese (3 females, 7 males, age range: 21-24, mean 22.3) and ten native speakers of American English (5 females, 5 males, age range: 20-30, mean 23.6) were recruited. Japanese subjects were all college students in Japan who received English instruction in the mid- and high- school level as well as in college. The American subjects were monolingual undergraduate students at University of Washington. All subjects are right-handed with no speech and hearing disorders in medical history.

A /ra-la/ continuum was created using SenSyn program. Figure 1 shows spectrograms for the endpoints in the continuum. There were eleven syllables of 400 ms in duration. All acoustic parameters were kept the same except the F3 transition slope whose starting frequencies varied in the range of 1325 to 3649 Hz at eleven levels equally spaced on the mel scale. The initial 155 ms of stimuli were composed of steady formant structure. The F3 transition had 100 ms duration that ended in a steady third formant at 2973 Hz for /a/. Specific parameters were adopted from a previous study [19]. The syllables were resampled to 48 kHz 16-bit using SoundEdit1.6 to accommodate the stimulator for MEG experiments.

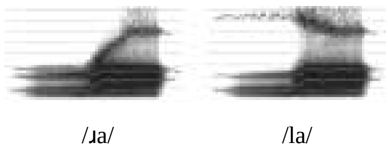

Figure1. Schematic spectrograms for the synthetic stimuli.

Subjects completed an identification test in an acoustically treated booth. The test began with a familiarization session of 11 trials followed by a testing session of 40 trials for each stimulus. The stimuli were randomly presented to the right-ear headphone at 80dB SPL. After the test, one Japanese subject was chosen to complete the training.

2.3 Training Protocol

A pretest-intervention-posttest design was implemented to assess the listener's initial capability and the training effects. A Mac G3 computer was used as the platform for the program. The subject listened to stimuli via a headset at a comfortable level. Stimuli were prepared first by recording natural tokens of /r/ and /l/ from eight native speakers of American English for five vowels in CV and VCV contexts. These tokens were submitted to an LPC analysis-resynthesis procedure to exaggerate the formant frequency differences between pairs of /r-l/ tokens, and to reduce the bandwidth of F3. Temporal exaggeration of the /r-l/ stimuli was made to the stimuli using a time warping technique (pitch synchronous overlap and add). These acoustically modified stimuli and the digitized versions of the naturally-produced tokens were used for the training phase of the experiment, while only the natural tokens were presented in the pretest and posttest. The pretest consisted of 4 blocks with 320 tokens by 8 speakers in 80 contexts. The posttest was identical. Training consisted of twelve sessions of approximately 50~60 minutes a session. Each session had a total of 400 listening trials arranged in 10 blocks with short intermittent tests of 10 trials that assessed progress. The series of training sessions commenced with presentation of tokens that were highly exaggerated but progressed over the course of the training to less exaggerated versions of the tokens. To address generalizability of training effects to novel /l/ and /r/ sounds, five talkers were used for training, but tokens from all talkers were presented in the pre- and posttests. For both pre- and post-tests, MEG experiments preceded behavioral experiments.

2.4 MEG Experiments

MEG experiments were conducted using the oddball paradigm. The subject was instructed to read a self-chosen book and ignore the auditory stimuli. Four conditions were designed to examine neural correlates of discrimination and categorization.

1. <u>Single condition</u>. The endpoint stimuli in the continuum were used for standard and deviant. This pair maximizes the acoustic difference between /l/ and /r/.
2. <u>Multiple condition</u>. Three stimuli from each category in the continuum were used for standard and deviant.
3. <u>Truncated 155ms condition</u>. The initial 155ms segments of the stimuli in Single condition were used. The purpose was to examine MMF characteristics elicited by different portions of acoustic difference in /l-r/.
4. <u>Truncated 100ms condition</u>. The middle 100ms of F3 transition of the Single condition stimuli were used.

Stimuli were monaurally delivered to the right ear via a plastic tube at 80 dB SPL. Deviant occurrence was at 0.15 probability with at least two intervening standards. Interstimulus intervals were randomized between 800 and 1200 ms. There were two blocks of stimuli with the standard and the deviant reversed in the second block. A ten-minute break was inserted between the two blocks in one experiment to reduce effects of habituation and fatigue. After the break, head positioning data were fitted again on four coils pasted on the scalp with 98% accuracy or above. Head origin deviations were adjusted in the range of 0~3.0 mm before proceeding to the second block.

The MEG data were collected using the Neuromag 122-channel whole-head SQUID gradiometer housed in a four-layered magnetically shielded room at NTT Communications Science Laboratories in Japan. The analog filter was 0.01~100 Hz, and the sampling frequency was 497 Hz. Epochs with MEG ≥ 3000 fT/cm or EOG ≥ 150 µV indicative of artefacts were rejected online. At least 100 epochs were averaged for the deviant and the standard immediately before the deviant. The data were digitally filtered at 0.8 ~ 40 Hz offline. The analysis time was -100 ~ 800 ms. For each experiment, N1m was determined at post-stimulus 80~160 ms using a subset of 44 channels from both hemispheres. The MMF peak was determined from subtracted waves using a time window of 200 ms after N1m.

3. Results

3.1 Behavioral Identification Functions

Figure 2 shows the group average identification functions for Japanese and American subjects. Overall, Japanese listeners were more biased in labeling more stimuli as /ra/. Nonparametric two-tailed Kolmogorov-Smirnov tests on the percent-correct identification indicated significant difference between the two groups of subjects on every stimulus on the continuum except No.5 ($p < .01$).

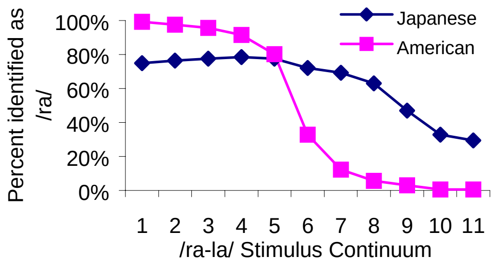

Figure 2. Average identification functions for 10 Japanese and 10 American subjects.

3.2 Behavioral Pretest and Posttest

Table 1 summarizes the pretest and posttest data for 320 identification trials. On average, training resulted in improvement of correct identification from 57% to 79%. Binomial tests indicated that before training above chance level performance was observed only in the subcategory of Talker 2, and after training performance in correct identification became significant for all sub-categories. Training effects were found to be transferable to the novel untrained stimuli at a sizable improvement of 26.7%.

(a)

Syllabic context	CV	VCV
Pre	.53	.60
Post	.79	.78

(b)

Vowel context	/i/	/e/	/a/	/o/	/u/
Pre	.55	.59	.55	.56	.58
Post	.83	.89	.73	.73	.75

(c)

Talker	1	2	3	4	5	**6**	**7**	**8**
Pre	.55	.75	.50	.55	.48	**.55**	**.63**	**.53**
Post	.83	.70	.88	.73	.68	**.88**	**.75**	**.88**

Table 1. Pre- and post- test correct identifications according to (a) syllabic contexts, (b) vowel contexts, and (c) talkers. Numbers in bold face were for stimuli not used in training.

Comparable training effects were observed for all stimuli in the /ra-la/ continuum except No. 1 and No. 4 (Figure 3). However, the post-training Japanese subject's phonetic boundary for /ra-la/ (between stimuli No. 8 and No. 9) shifted towards /la/ from the American subjects' boundary location (between No. 5 and No. 6) by three steps of the equalized acoustic change in the continuum.

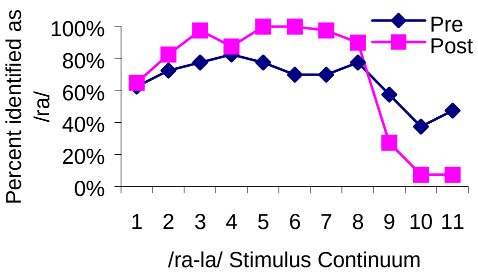

Figure 3. Pretest and Posttest identification functions.

3.2 Neuromagnetic data

Tables 2 & 3 show pretest and posttest mismatch field results for all four MEG conditions. Compared to the noise level in baseline, the MMF values were significant. For the speech stimuli, training results in enhanced MMF peak responses in the left hemisphere (163.0~269.6 ms) coupled with reduced activities in the right hemisphere (169.1~295.8 ms). Figures 4 & 5 illustrate this pretest and posttest change in waveforms and MMF dipole localization. This pattern of neural plasticity was not observed for truncated stimuli, which the subject perceived as non-speech sounds. Prior to training, the right hemisphere appeared to be heavily involved in detecting the acoustic differences in stimuli for all four conditions. After training, MMF data exhibited left-hemisphere dominance in the Single condition, bilaterally equalized cortical involvement in the Multiple condition, and bilaterally increased mismatch activity for truncated stimuli with one exception in the 155ms condition.

MMF (fT/cm)	Single				Multiple			
	Left		Right		Left		Right	
	la-ra	ra-la	la-ra	ra-la	la-ra	ra-la	la-ra	ra-la
Pre	22.35	23.60	28.91	27.78	21.22	23.23	31.00	27.73
Post	30.29	30.57	23.61	Null	23.51	24.28	23.22	24.29

Table 2. Pre- and post- test MMF peak magnitude in Single and Multiple conditions. "la-ra" indicates subtraction of deviant /la/ from standard /ra/. Vice versa for "ra-la".

MMF (fT/cm)	155ms				100ms			
	Left		Right		Left		Right	
	s1-s2	s2-s1	s1-s2	s2-s1	s3-s4	s4-s3	s3-s4	s4-s3
Pre	19.50	Null	32.52	27.98	16.93	22.30	28.06	23.72
Post	25.10	24.24	28.94	37.20	21.88	26.28	36.83	25.35

Table 3. Pre- and post- test MMF peak magnitude in 155ms and 100ms conditions. "s1" and "s2" indicate the 155ms portions from /la/ and /ra/. "s3" and "s4" are for the 100ms portions.

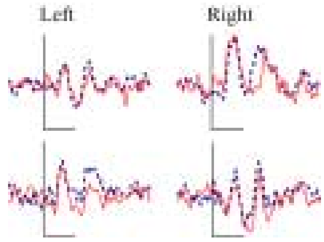

Figure 4. Waveforms of two individual channels respectively from the left and right hemispheres in the Single condition. The upper panel is for pretest and lower panel for posttest. The solid wave is for standard and dotted wave for deviant. MEG scales: vertical = 100 fT/cm, horizontal = 200 ms.

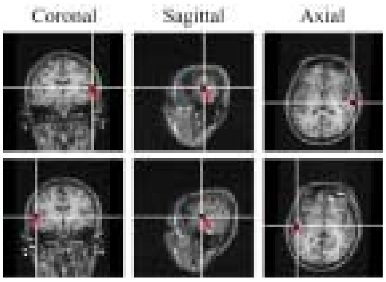

Figure 5. Dipole source localization for MMFs corresponding to the waveforms in Figure 4.

4. Discussion

The success of the behavioral training program is evidenced by substantial improvement in /ɹ/ and /l/ identification of the Japanese listener, and this ability was transferable to untrained stimuli. However, the Japanese subject's enhanced sensitivity to the crucial F3 transition appeared to be driven by the transitional direction and the initial steady F3 frequency locale in relation to F2. Among the 11 synthetic syllables in the continuum, No. 1-8 have a rising F3 transition and No. 9-11 a falling F3 transition. We speculate that the Japanese subjects including the trained subject primarily used rising F3 transition as an acoustic cue for /ra/. The fact that Americans treated No. 7 and 8 with shallow rising F3 as /la/ could be due to the perceptual magnet effect that the prototype /la/ assimilates its neighboring stimuli into the category [19]. In this perspective, the /ra/ prototype in Japanese could interfere with the training process as a limit on plasticity. Work in our laboratory showed that unlike the Americans Japanese listeners tend to separate /l/ and /r/ stimuli acoustically on the F2 dimension. This factor could also interfere with the subject's performance, especially when the initial portion of F3 in the consonant is very close to F2 as in Stimuli No. 1 and No. 2.

Neural plasticity in the Japanese subject was found to show a right-to-left hemispheric shift of cortical MMF activities in the establishment of linguistic /l/ and /r/ categories. Earlier we reported that an American control's MMFs showed a left hemisphere dominance for /l/ and /r/ whereas the Japanese subjects showed bilateral involvement [20]. Training appeared to lead to more linguistic analysis of the speech stimuli in the left hemisphere. The patterns in pre-attentive MMF activation underlying stimulus discriminability and categorization confirmed that the MMF is a sensitive measure of neural activities in acoustic and phonetic processing.

Little is known about how linguistic experience causes people to attend to different dimensions of the stimuli and how language training could be designed to accurately map linguistic representations for nonnative speech contrasts. It is reported that even early and extensive exposure to a second language is not sufficient to attain the ultimate phonological competence of native speakers [21]. The formation of a new phonetic category in mental representation as well as its influence on speech perception in terms of behavioral and neurophysiologic measures merits further empirical work.

Acknowledgements

This work has been supported by NIH (HD 37954) and Human Frontiers Science Program (HFSP 159) to Dr. Patricia K. Kuhl and by NTT traineeship to Yang Zhang. MRI facilities were provided by Tokyo Denki University. The authors would like to thank Dr. Yoh'ichi Tohkura for his support for this project.

REFERENCES

1. Eimas, P.D., Siqueland, E.R., Jusczyk, P., and Vigorito, J. "Speech perception in infants." *Science, 171*, 303-306, 1971.
2. Werker, J. F., and Tees, R. C. "Cross-language speech perception: Evidence for perceptual reorganization during the first year of life." *Infant Behavior and Development*, 7, 49-63, 1984.
3. Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., and Lindblom, B. "Linguistic experience alters phonetic perception in infants by 6 month of age." *Science, 255*, 606-608, 1992.
4. Werker, J. F., and Tees, R. C. "Influences on infant speech processing: toward a new synthesis." *Ann. Rev Psychol.* 50:509-35, 1999.
5. Tsushima, T., Takizawa, O., Sasaki, M., Shiraki, S., Nishi, K., Kohno, M., Menyuk, P., and Best, C. "Discrimination of English /r-l/ and /w-y/ by Japanese infants at 6-12 months: language-specific developmental changes in speech perception abilities." *Proceedings of ICSLP, 4*, 1695-1698, 1994.
6. Kuhl, P. K., Kiritani, S., Deguchi, T., Hayashi, A., Stevens, E. B., Dugger, C. D., and Iverson, P. "Effects of language experience on speech perception: American and Japanese infants' perception of /ra/ and /la/." *JASA, 102*, 3135, 1997.
7. Strange, W., and Dittmann, S. "Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English." *Perception & Psychophysics, 36*, 131-145, 1984.
8. McCandliss, B.D., Fiez, J.A., Conway, M., Protopapas, A., and McClelland, J.L. "Eliciting adult plasticity: Both adaptive and non-adaptive training improves Japanese adults identification of English /r/ and /l/." *Society of Neuroscience Abstracts, 24*, 1898, 1998.
9. Lively, S.E., Logan, J.S., and Pisoni, D.B. "Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories." *JASA, 94*, 1242-1255, 1993.
10. Bradlow, A.R., Akahane-Yamada R., Pisoni, D.B., and Tohkura, Y. "Training Japanese listeners to identify English /r/ and /l/: long-term retention of learning in perception and production." *Perception & Psychophysics,* 61(5):977-985, 1999.
11. Logan, J.S., and Pruitt, J.S. "Methodological issues in training listeners to perceive non-native phonemes." In Strange, W. (Ed.) *Speech Perception and Linguistic Experience: Issues in Cross-language Research*, pp 351-377. Baltimore: York Press, 1995.
12. 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., and Alho, K. "Language-specific phoneme representations revealt by electric and magnetic brain responses." *Nature, 385(6615)*, 432-434, 1997.
13. Winkler, I., Kujala, T., Tiitinen, H., Sivonen, P., Alku, P., Lehtokoski, A., Czigler, I., Csepe, V., Ilmoniemi, R.J., and Näätänen, R. "Brain responses reveal the learning of foreign language phonemes." *Psychophysiology,* 36(5), 638-642, 1999.
14. Sharma, A., and Dorman, M.F. "Neurophysiologic correlates of cross-language phonetic perception." *JASA*, 107(5):2697-2703, 2000.
15. Kraus, N., McGee, T.J., and Koch, D.B. "Speech sound representation, perception, and plasticity: a neurophysiologic perceptive." *Audiol. Neurootol.*, 3:168-82, 1998.
16. Lounasmaa, O.V., Hämäläinen, M., Hari, R., and Samelin, R. "Information processing in the human brain: Magnetoencephalographic approach." *PNAS*, 93, 8809-8815, 1996.
17. Pruitt, J.S. *The perception of Hindi dental and retroflex stop consonants by native speakers of Japanese and American English*. Doctoral Dissertation, University of South Florida, 1995.
18. Kuhl, P. K., Andruski, J. E., Chistovich, I. A., and Chistovich, L. A. "Cross-language analysis of phonetic units in language addressed to infants." *Science, 277,* 684-686, 1997.
19. Iverson, P. and Kuhl, P.K. "Influences of phonetic identification and category goodness on American listeners' perception of /r/ and /l/." *JASA, 99(2)*, 1130-1140, 1996.
20. Zhang, Y., Kuhl, P.K., Imada, T., Kotani, M., Stevens, E., and Coffey-Corina, S. *Poster at CNS2000* , 2000.
21. Pallier C, Bosch L, Sebastian-Galles N. "A limit on behavioral plasticity in speech perception." *Cognition*. 64(3):B9-17, 1997.