PLASTICITY OF NON-NATIVE PHONETIC PERCEPTION AND PRODUCTION: A TRAINING STUDY

Satoshi Imaizumi[1], Hidemi Itoh[2], Yuji Tamekawa[1], Toshisada Deguchi[3], Koichi Mori[4]

[1]Department of Speech and Cognitive Sciences, Graduate School of Medicine, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-0033 Japan, [2] Tohoku University, School of Dentistry, Sendai, 980-8575 Japan, [3]Tokyo Gakugei University, Nukui-kita-Machi, Koganei, 184-8501 Tokyo, Japan, [4]Research Institute of National Rehabilitation Center for the Disabled, 4-1 Namiki, Tokorozawa, Saitama 359-8555, Japan

ABSTRACT

Audiovisual perceptual training of non-native phonetic contrasts was conducted for 10 naive Japanese adults using audiovisual recordings of 13 native English speakers articulating 90 rl minimal word pairs, and analyzed changes in perceptual and articulatory representations of non-native phonetic contrasts. The speech identification score drastically improved during the training. The improvement in non-native rl perceptual distinction was clearly associated with the changes in the perceptual and articulatory representations, which represents perceptual/articulatory dissimilarities between the non-native and native phonemes as maps created using a multi-dimensional scaling analysis (MDS). Results suggested that the new non-native phonetic categories can be acquired through proper training even in adulthood so that distances among exemplars within each of the acquired categories shrunk and distances between the categories stretch considerably compared to those of pretraining stage in the perceptual and articulatory representations.

1. INTRODUCTION

Language-specific speech perception develops within the first year of life[1] and becomes less plastic in adulthood[2], as is the case in other sensory and motor domains[3]. Perception of non-native speech contrasts, however, improves in the course of learning even in adulthood[4-9]. Monolingual Japanese adults have difficulty in distinguishing the English /r/ and /l/ (such as in "right" vs. "light"), which do not occur contrastively in Japanese[10-13]. The brain mechanisms responsible for difficulty and possibility in acquiring the non-native phonetic contrasts are still not fully understood.

We conducted audiovisual training of non-native speech discrimination for 10 naive Japanese adults using audiovisual recordings of 13 native English speakers articulating 90 rl and 32 sT (T=voiceless /th/) minimal word pairs, and analyzed changes in perceptual, articulatory, and cortical representations of native and non-native phonetic contrasts. The focus of this study was on the relationship between articulatory and perceptual representations of L1 and L2 phonetic categories, which may cause difficulties in L2 acquisition, and possible benefits of visual cues in resolving the difficulties.

2. METHOD

We analyzed changes in articulatory, perceptual and neuronal representations of Japanese and English syllables /ra/, /la/ and /wa/ during audio-visual training for English spoken-word identification.

2.1. Audiovisual Training

We videotaped the faces of 13 native English speakers (three Americans and 10 Canadians, 5 males and 8 females) articulating 90 rl and 32 sT minimal pair words. The utterances were recorded on a computer-controlled laser video-disk. One training session consisted of identification practice with correct answer rewarded for 122 x 6 words spoken by six speakers. At the very beginning of training, each trainees were instructed to attend to picture, and then to gradually shift attention to sound. When the discrimination score for rl word identification became stable in a range higher than 90%, they were instructed to turn off the picture and attend to sound.

2.2. Articulatory Representation

Palato-lingual contact patterns of utterances /right/, /light/, /fright/, and /flight/ were measured by dynamic palatography from the trainees. These words were selected since word-initial rl and consonant-cluster rl were the hardest for Japanese trainees to perceptually discriminate. Three Japanese words, /raito/, /furaito/, /waito/, were included in the dynamic palatographic measurement. The Japanese words were presented in Japanese kana characters. Palato-lingual contact patterns (Pijk, i=electrode number, j=token number, k= subject number) of middle parts of /r/ or /l/ segments were extracted and dissimilarity matrices between all pairs of three tokens of each word were calculated for each trainees. Dissimilarity D(J1, J2) between two palato-lingual contact patterns, Pijk, j=J1, J2, was defined as the number of electrodes which have a common value between PiJ1k and PiJ2k, i=1,2,,,64. The obtained dissimilarity matrix, D(J1, J2), was analyzed using a multidimensional scaling method of SAS to display articulatory relationships between English /r/ and /l/, and Japanese /r/ and /w/.

2.3. Perceptual Representation

Four Japanese (two females and two males) articulating /wa/ and /ra/, and four English speakers (two females and two males) articulating /wa/, /ra/ and /la/ in their native languages were videotaped. All possible pairs of those videotaped samples were presented in random orders to 20 naive Japanese trainees in three presentation modes; auditory-only, visual-only, and audiovisual modes. The subjects were instructed to assign a large number if they felt the pair was dissimilar to each other in the phonetic quality, or a small number if the pair was similar. The obtained dissimilarity matrix was analyzed using a multidimensional scaling method of SAS.

2.4. Neuronal Representation

Changes in neuronal representations were measured using magnetoencephalography. Measurements were performed in an acoustically and magneto-electrically shielded room. Blocks of randomly ordered seven synthetic /la/ samples were binaurally presented through plastic tubes and earpieses at a sensation level of 70 dB with a duration of 200ms. 15% of the stimuli were randomly appearing deviant samples of synthetic /ra/. In a different measurement block, the standards and the deviants were altered. The onset-to-onset interstimulus interval was 800ms. The magnetic responses were measured using a helmet-shaped whole-head Neuromag 122 magnetometer. The locus of the neuronal activity was estimated by determining the equivalent current dipole (ECD) at the latency where the ECD magnitude is maximum and the goodness-of-fit is higher than 80%.

3. RESULTS AND DISCUSSION

The speech discrimination score drastically improved during the training confirming previous reports. The improvement in non-native rl distinction was clearly associated with the changes in the mental representations of non-native and native phonemes, which represents articulatory and perceptual dissimilarities between the non-native and native phonemes as a map (Fig. 1 and 2).

In the pretraining articulatory representation as shown in Fig. 1a, the tokens of English /ra/ and /la/ and Japanese /ra/ scattered widely and partially overlapped with each other before training. Japanese /ra/ located closer to English /ra/ rather than English /la/. After training, these three phonemes separated into three groups (Fig. 1b). For English /la/, the tongue tip made contact with the alveolar ridge only, while English /ra/ was produced with a constriction occurring in the area of the velar consonants. The palato-lingual contact patterns during flapping movement of the tongue for Japanese /ra/ were variable among the trainees.

In the perceptual representations before training, particularly when only an auditory signal was presented, the tokens of the five phonemes could be classified into only two groups /w/ and the others (Fig. 2a). The tokens of English /ra/ and /la/ and Japanese /ra/ scattered widely and overlapped with each other. The trainees had significant difficulty in distinction between /ra/ and /la/. While, after the training,

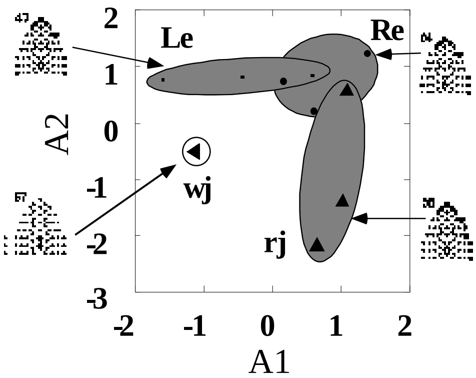

Figure 1a: The pretraining articulatory representation of non-native and native phonemes, English /r/ (Re) and /l/ (Le) and Japanese /w/ (wj) and /r/ (rj), derived from the multidimensional scaling applied to the dissimilarity matrix measured from palato-lingual contact patterns of a subject. The tokens of English /right/ and /light/ and Japanese /raito/ scattered widely and partially overlapped with each other before training. A1 and A2 are the articulatory factors extracted by MDS.

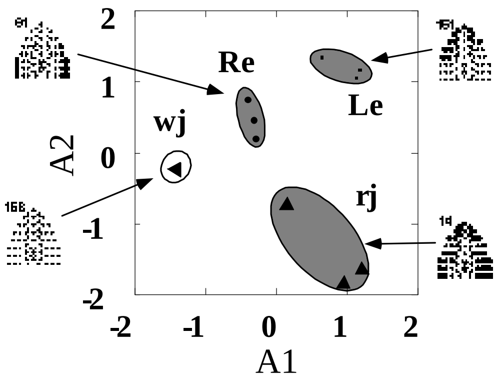

Figure 1b: The articulatory representation after training of non-native and native phonemes (Japanese /w/ and /r/, and English /r/ and /l/) which separated into three groups. For English /la/, the tongue tip made contact with the alveolar ridge only, while English /ra/ was produced with a constriction occurring in the area of the velar consonants.

the five phonemes separated into three groups: /wa/ of both languages, English /ra/, and overlapping Japanese /ra/ and English /la/ (Fig. 2b). Each groups were more tightly clustered than the pretraining representation, suggesting that the trainees acquired sensitivity for the non-native /r/ and /l/ contrast.

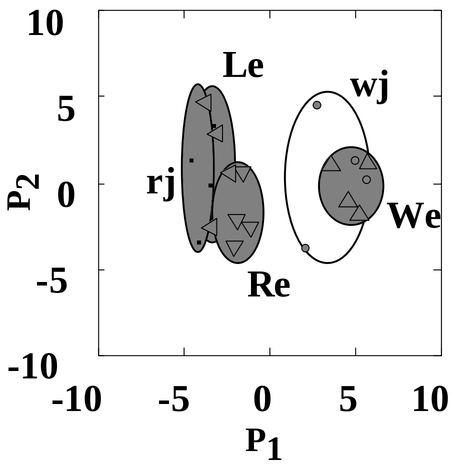

Figure 2a: The perceptual representation before training of non-native and native phonemes (Japanese /wa/ and /ra/, and English /ra/, /la/ and /wa/), derived from the multidimensional scaling applied to the dissimilarity matrix measured under auditory-only presentation. The overlapped English /r/ and /l/ groups suggest that the trainees had no sensitivity to the /r/ and /l/ contrast. P1 and P2 are the perceptual factors extracted by MDS.

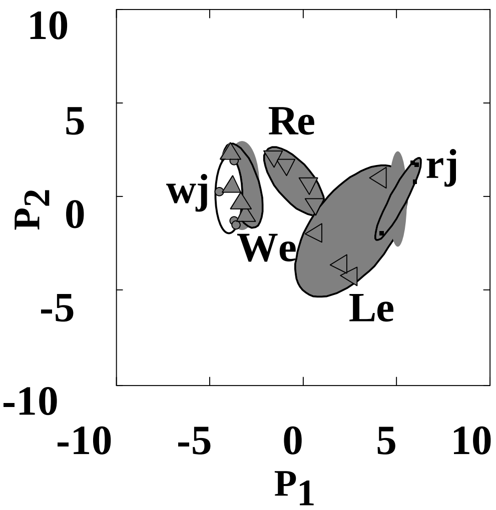

Figure 2b: The perceptual representation after the training. The English /r/ and /l/ separation suggests that the trainees acquired sensitivity to the non-native /r/ and /l/ contrast.

The benefits of audiovisual training is evident by comparing the auditory (Fig. 2a) and audiovisual representations (Fig. 2c). Visually presented lip protrusion for English /ra/ and /wa/ differentiates them from other phonemes, and constriction of the tongue tip at the alveolar ridge for English /la/ and flapping movement of the tongue for Japanese /ra/ differentiates them from other phonemes, when

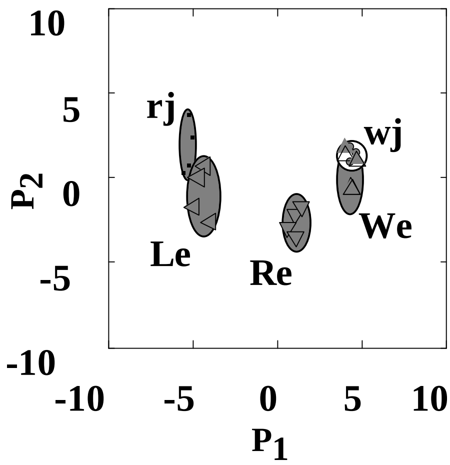

Figure 2c: The audiovisual representation after the training. Audiovisual cues further enhance the inter-phoneme separation and condense intra-phoneme scattering.

the tongue is visible. Because of these visual cues, the five phonemes were separated more distinctively in the audiovisual representation than in the auditory representation. Visual cues are known to affect[15] and facilitate speech perception [16, 17]. Young infants can imitate facial expressions[18] and sounds [19], and can detect auditory-visual correspondences of speech[20], suggesting that visually presented articulatory cues facilitate speech acquisition. The present results suggest that those cues accelerate forming proper representations of non-native phoneme categories in adulthood.

To ascertain this possibility, we measured the mismatch magnetic fields (MMF), i.e., the auditory-evoked brain magnetic responses generated by an attention-independent change-detection neural process. The magnitude of MMF elicited by English /ra/ and /la/ changed significantly by training. Before training, the MMF magnitude was small without any significant differences between the right and left hemispheres. After training, the MMF magnitude increased with a significant dominance of the left hemisphere. The source of the learning-dependent changes in the MMF magnitude was located in the auditory cortex of the left hemisphere. The left-hemisphere dominance of the MMF enhancement due to training suggests that the neural phoneme traces were created for the non-native phonemes in the left auditory cortex. Typical exemplars of native phonemes generate a larger mismatch negativity (MMN) in electroencephalography and a larger mismatch magnetic field (MMF) than an atypical or non-native phoneme in the human cortex, especially in the left hemisphere[21]. MMN and MMF enhancement is associated with learning-related improvement in sensitivity of slight changes in stimuli[22].

The present results suggest that the new non-native phoneme categories can be acquired through proper training even in adulthood so that MDS distances among exemplars within each of the acquired categories shrunk and distances between the categories stretch considerably compared to those of pretraining stage in both the perceptual and articulatory representations.

4. CONCLUSION

This study of audiovisual training of non-native phonemes shows that multimodal exposure to speech effectively alters articulatory, perceptual and neuronal representations of non-native phoneme categories. Whole-head magnetic recordings located the source of phoneme-related plasticity in the left auditory cortex.

5. REFERENCES

1. Kuhl, P., Williams, K.A., Lacerda, F. *et al*., “Linguistic experience alters phonetic perception in infants by 6 moonths of age,” Science 255(5044): 606-608, 1992.

2. Best, C.T., McRoberts, G.W. and Sithole, N.N. “Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants,” J. Exp. Psychol.: Human Perception & Performance 14(3): 345-360, 1988.

3. Singer, W. “Development and plasticity of cortical processing architectures,” Science 270(5237): 758-64, 1995.

4. Pisoni, D.B., Aslin, R.N., Perey, A.J., *et al*., “Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants,” J. Exp. Psychol.: Human Perception & Performance 8(2): 297-314, 1982.

5. Jamieson, D.G. and Morosan, D.E. “Training non-native speech contrasts in adults: acquisition of the English /delta/-/theta/ contrast by francophones,” Perception & Psychophysics 40(4): 205-215, 1986.

6. Logan, J.S. Lively, S.E. and Pisoni, D.B. “Training Japanese listeners to identify English /r/ and /l/: a first report,” J. Acoust. Soc. Am. 89(2): 874-886, 1991.

7. Best, C.T. and Strange, W. “Effects of phonological and phonetic factors on cross-language perception of approximants,” J. Phonetics 20: 305-330, 1992.

8. Yamada R. and Tohkura, Y. “The effects of experimental variables on the perception of American English /r/ and /l/ by Japanese listeners,” Perception & Psychophysics 52(4): 376-392, 1992.

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.,” J. Acoust. Soc. Am. 94(3): 1242-1255, 1993.

10. Goto, H. “Auditory perception by normal Japanese adults of the sounds "L" and "R",” Neuropsychologia 9: 317-323, 1971.

11. Miyawaki, K., Strange, W., Verbrugge, R. *et al*., “An effect of linguistic experience: the discrimination of [r] and [l] by native speakers of Japanese and English,” Perception and Psychophysics 18: 331-340, 1975.

12. Mochizuki, M. “The identification of /r/ and /l/ in natural and synthetized speech,” J. Phonetics 9: 283-303, 1981.

13. Flege, J.E., Takagi, N. and Mann, V. “Lexical familiarity and English-language experience affect Japanese adults' perception of /r/ and /l/,” J. Acoust. Soc. Am. 99(2): 1161-73, 1996.

14. Best, C.T., McRoberts, G.W. and Sithole, N.N. “The phonological basis of perceptual loss for non-native contrasts: maintenance of discrimination among Zulu clicks by English-speaking adults and infants,” J. Experimental Psychology: Human Perception and Performance 14: 345-360, 1988.

15. McGurk, H., and MacDonald, J.W. “Hearing lips and seeing voices,” Nature 264: 126-130, 1976.

16. O'Neil, J.J. “Contribution of the visual components of oral symbols to speech comprehension,” J. Speech Hear. Dis. 19: 429-439, 1954.

17. Calvert, G.A., Bullmore, E.T., Brammer M.J. *et al*., “Activation of auditory cortex during silent lipreading,” Science 276(5312): 593-596, 1997.

18. Meltzoff, A.N., and Moore, M.K. “ Imitation of facial and manual gestures by human neonates,” Science 198 (4312): 74-78, 1977.

19. Kuhl, P.K., and Meltzoff, A.N. “Infant vocalizations in response to speech: vocal imitation and developmental change,” J. Acoust. Soc. Am. 100 (4 Pt 1): 2425-38, 1996.

20. Kuhl, P.K., and Meltzoff, A.N. “The bimodal perception of speech in Infancy,” Science 218 (4577): 1138-1141, 1982.

21. Naatanen, R., Lehtokoski, A., Lennes, M. et al., “Language-specific phoneme representations revealed by electric and magnetic brain responses,” Nature 385 (6615): 432-434, 1997.

22. Naatanen, R., Schroger, E., Karakas, S. et al., “Development of a memory trace for a complex sound in the human brain.,” Neuroreport 4(5): 503-506, 1993.