EFFECTS OF TRAINING MODALITY ON AUDIO-VISUAL PERCEPTION OF NONNATIVE SPEECH CONTRASTS

Yue Wang[1], Dawn Behne[2], and Haisheng Jiang[1]
[1]Dept. of Linguistics, Simon Fraser University, 8888 University Drive, Burnaby, Canada, V5A 1S6
[2]Dept. of Psychology, Norwegian University of Science and Technology, Dragvoll, 7491, Trondheim, Norway

1. INTRODUCTION

Speech perception often involves integrated auditory and visual modalities [1,2]. While nonnative perceivers can be facilitated by visual information when perceiving L2 sounds just as the natives, they may also be impeded in correct use of L2 visual cues, as they are not sensitive to the visual cues non-existent in their L1 [3,4,5,6]. Further research indicates that sensitivity to the visual information in L2 sounds can be enhanced through audio-visual (AV) training [3,7,8,9]. Findings also raise the question of the relative effectiveness of AV and V-only training. On the one hand, AV training may be preferred as learners presumably need to extract information from as many input channels as possible to perceive the challenging nonnative sounds [7]. On the other hand, V-only training may force the learners to focus more on the visual speech cues, and therefore reduce the cognitive load required to attend to both auditory and visual input [3]. However, the results from the previous separate AV and V-only training studies cannot be directly compared as they involve different training materials and sessions [9].

On the basis of these previous findings, the current study explored the effects of training on the perception of nonnative speech sounds using A, V, and AV training input modalities. Specifically, we examined how Mandarin learners of English could learn to perceive the A and V cues to the English interdental fricatives non-existent in Mandarin, compared to their familiar labiodentals and alveolars. A study with Mandarin learners is of particular interest. Since Mandarin perceivers demonstrate a lesser degree of attentiveness to visual speech information in their L1 compared to perceivers of other languages [5,10], questions arise as to how this lack of visual attentiveness affect their perception of L2.

2. METHOD

Forty-four young adult Mandarin Chinese natives with less than five years' residency in Canada participated in the study. They were randomly assigned to a control group (n=11) and three training groups (n=11 per group), with each training group receiving training with a different speech input modality: A, V, and AV. All participants took the same pre- and posttest during which they were presented with stimuli from three modalities: A, V, and AV.

Pre/posttest stimuli were recorded of an adult male speaker of Canadian English. The stimuli were based on 18 English CV syllables having a fricative followed by a vowel: [fi, fa, fu, vi, va, vu, θi, θa, θu, ði, ða, ðu, si, sa, su, zi, za, zu]. The fricatives differed in place of articulation, (POA: labiodental, interdental, alveolar). The participants' task was to identify the fricatives while listening to the sounds over a headset, or viewing the speaker mouth movements on the screen, or both.

The perceptual training program followed the high variability procedure demonstrated to be highly effective in auditory perception training [11,12]. Learners were trained to identify the target fricative contrasts appearing in a variety of phonetic contexts, in both natural and nonsense words, and produced by four native speakers of Canadian English (2 males, 2 females). For each training trial, the trainees' task was identification followed by feedback. The stimuli were presented as A-only for the A-train group, V-only for the V-train group, and AV for the AV-train group. The training was completed in two weeks, including six sessions of 45 minutes each.

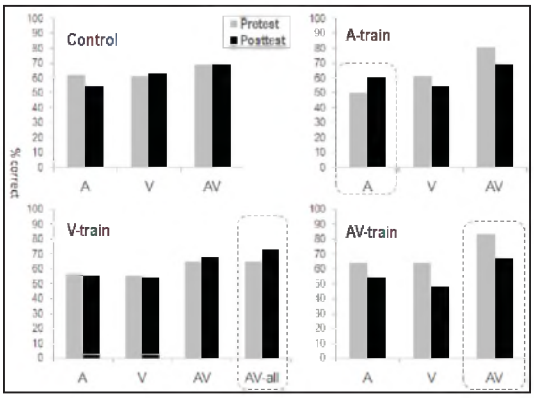
3. RESULTS

Percent correct identification of the fricatives at pretest and posttest was analyzed using a 4-way mixed analysis of variance (ANOVA), with Group (Control, A-train, V-train, AV-train) as a between-subject factor, and Test (pretest, posttest), POA, (labialdental, interdental, alveolar), and Modality as repeated measures. The dependent variable was perceivers' correct identification for POA regardless of voicing since POA was the focus of interest.

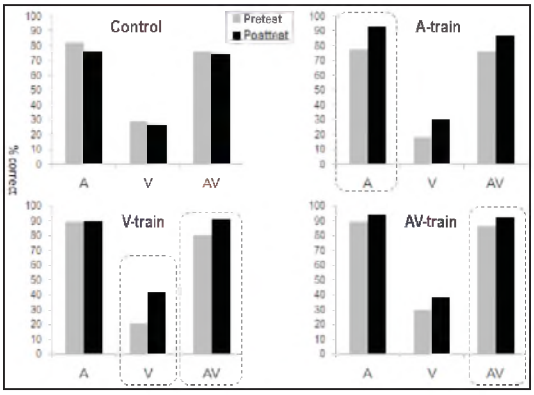
The results showed significant interactions of Group x Test [$F(3, 40)=3.15$, $p<.035$], Group x Test x Modality [$F(6, 80)= 2.30$, $p<.042$], and Group x Test x Modality x POA [$F(12, 160)=22.92$, $p<.001$]. Therefore, data were further analyzed with sets of one-way ANOVAs for each Group to compare the pre- and posttest performance in each Modality and POA. The results of these comparisons are displayed in Figure 1. Since no reliable differences were observed for labiodental perception across groups and tests, the results were excluded.

For the interdentals (Figure 1a), significant improvement from the pretest (50% correct identification) to the posttest (60% correct) was observed for the A-train group in the A modality when only the audio stimuli were presented [$F(1,10)=5.2$, $p<.046$]. For the V-train group, there revealed no reliable test difference across modalities. However, a more detailed analysis (taking voicing into consideration) revealed an increase post-training in perceiving the AV modality for both voiceless (from 65% to 73%) and voiced (from 47% to 59%) interdentals. For the AV-train group, a significant decrease (83% to 67%) was unexpectedly observed in the perception of the AV modality [$F(1,10)=7.6$, $p<.020$].

For the alveolars (Figure 1b), significant improvements were observed for the following groups and modalities: (1) A-train group with A modality [from 77% to 93%; $F(1,10)=7.9$, $p<.019$]; (2) V-train group with V modality [from 21% to 42%; $F(1,10)=31.9$, $p<.001$], and AV modality [from 80%, to 91%; $F(1,10)=7.9$, $p<.019$]; and (3) AV-train group with AV modality [Pretest: 86%, Posttest: 92%; $F(1,10)=5.7$, $p<.038$].



(a) Interdentals



(b) Alveolars

Fig. 1. % correct responses for (a) interdental and (b) alveolar fricative perception at pretest and posttest by perceivers in Control, A-train, V-train, and AV-train groups. (Significant pre- and posttest differences are circled. AV-all: % correct for both POA and voicing.).

4. DISCUSSION AND CONCLUSIONS

The results revealed a noticeable effect of training modality, where the extent of post-training improvement was consistent with the type of training; that is, the A-train group improved most in the perception of the A modality, whereas the V-train group improved most in the perceiving the V or AV modality. An exception to this general pattern of improvement was the decreased correct AV perception of the interdentals after AV training. The result appears to support the view that training with both auditory and visual input may add cognitive load to the learners, resulting in poorer performance [3]. This implicates that at the initial stage (especially with less advanced L2 perceivers learning a difficult L2 contrast), focused training with a single modality may be more effective than with multi-modalities which may be adopted at a later stage.

Regarding POA, the results showed a greater degree of training effect on the perception of the less visually distinct alveolars, as compared to that of the interdentals. Indeed, it has been speculated that learners (such as Mandarin) whose L1 possesses relatively low visual influence may not easily attune to the visual speech cues and therefore need to be trained to attend more to the visual information [3]. The results support this speculation, that despite the alveolars are familiar to the Mandarin perceivers, their perception can still benefit from training.

In sum, the current findings are consistent with the previous research showing the effectiveness of training on the AV perception of L2 speech [3,7,8], with the degree of training effect corresponding to factors such as perceivers' L1 experience, training input modality, and the visual salience of the speech segments.

REFERENCES

[1] Massaro, D. W. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry*. Erlbaum: Hillsdale, NJ.

[2] Jongman, A., Wang, Y., and Kim, B. (2003). Contributions of semantic and facial information to perception of nonsibilant fricatives. *JSLHR* 46, 1367-1377.

[3] Hazan, A., Sennema, A., Iba, M., & Faulkner, A. (2005). Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech Commun.* 47, 360-378.

[4] Hazan, V., Sennema, A., Faulkner, A., & Ortega-Llebaria, M. (2006). The use of visual cues in the perception of non-native consonant contrasts. *JASA* 119, 1740-1751.

[5] Sekiyama, K. (1997). Cultural and linguistic factors in audiovisual speech processing: The McGurk effect in Chinese subjects. *Percept. Psychophys.* 59, 73–80.

[6] Werker, J. F., Frost, P. E., & McGurk, H. (1992). Cross-language influences on bimodal speech perception. *Can. J. Psychol.* 46, 551-568.

[7] Hardison, D. M. (2003). Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Appl. Psycholing.*, 24, 495-522.

[8] Hardison, D. M. (2005). Second-language spoken word identification: Effects of perceptual training, visual cues, and phonetic environment. *Appl. Psycholing.* 26, 579-596.

[9] Chen, Y., & Hazan, V. (2007). Language effects on the degree of visual influence in audiovisual speech perception. *Proceedings of the 16th ICPhS*, Saarbrueken, 2177-2180.

[10] Burnham, D., Lau, S., Tam, H., & Schoknecht, C. (2001). Visual discrimination of Cantonese tone by tonal but non-Cantonese speakers, and by non-tonal language speakers. *Proceedings of AVSP 2001*, 155-160.

[11] Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *JASA* 89, 874–886.

[12] Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *JASA* 106, 3649–3658.

ACKNOWLEDGEMENTS

We thank Angela Cooper, Nina Leung, Rebecca Simms, and Jung-yueh Tu for their assistance. This project was funded by SSHRC.