



Directing Attention during Perceptual Training: A Preliminary Study of Phonetic Learning in Southern Min by Mandarin Speakers

Ying Chen¹, Eric Pederson²

¹School of Foreign Studies, Nanjing University of Science and Technology, China

²Department of Linguistics, University of Oregon, USA

ychen@njjust.edu.cn, epederso@uoregon.edu

Abstract

Previous studies have shown that directing learners' attention during perceptual training facilitates detection and learning of unfamiliar consonant categories [1, 2]. The current study asks whether this attentional directing can also facilitate other types of phonetic learning. Monolingual Mandarin speakers were divided into two groups directed to learn either 1) the consonants or 2) the tones in an identification training task with the same set of Southern Min monosyllabic words containing the consonants /p^h, p, b, k^h, k, g, t^h, t, e/ and the tones (55, 33, 22, 24, 41). All subjects were also tested with an AXB discrimination task (with a distinct set of Southern Min words) before and after the training. Unsurprisingly, both groups improved accuracy in the sound type to which they attended. However, the consonant-attending group did not improve in discriminating tones after training and neither did the tone-attending group in discriminating consonants -- despite both groups having equal exposure to the same training stimuli. When combined with previous results for consonant and vowel training, these results suggest that explicitly directing learners' attention has a broadly facilitative effect on phonetic learning including of tonal contrasts.

Index Terms: attention, perceptual training, phonetic learning, tone, Southern Min

1. Introduction

Attention can be oriented exogenously or endogenously [3, 4]. Exogenous orienting is the preferential focus of one's attention driven by particular external stimuli and is essentially bottom-up and stimulus-driven attentional control. Endogenous orienting is the directing of one's own attention according to one's own goals or expectations. It is top-down and goal-directed attentional control. Endogenous attention can be readily manipulated via experimental direction. Controlling visual attention in experiments was found to play a crucial role in learning the structure and meaning of languages [5, 6, 7, 8]. Controlling auditory attention also positively affects the results of learning speech sounds in a novel or second language [1, 2, 9, 10, 11].

However, previous studies found no effect of manipulated attention *during testing* in the discrimination tasks of non-native phonetic contrasts [12, 13]. Previous studies of different training conditions also found that learners did not pay equal attention to speech-like phonetic cues that were emphasized [9]. Participants have been found to shift their attention from one acoustic cue to another in training which used category-level identification as feedback. Further, these trained acoustic cues weighed more heavily in post-training tests [10, 11]. Unlike the

presumed implicit orienting of participants' attention of these studies, [1] directly manipulated endogenous orienting of attention via training instructions to learn either consonant categories or lexical meaning in an unfamiliar language. [2] also explored this effect of attention during training by directing participants' attention to different new phonetic categories. They found a positive effect of attention for American English speakers learning Hindi consonants. However, the effect was not found with vowels presumably because the accuracy of vowel discrimination in the pre-training test was already too high to allow an improvement in the post-training test.

The current study selects Quanzhou Southern Min as the target language and its consonants and tones as target categories. In Quanzhou Southern Min, there is a three-way voicing contrast in initial consonants (voiced, voiceless unaspirated, and voiceless aspirated) and five tones in unchecked syllables—high-level (55), mid-level (33), low-level (22), falling (41) and rising (24). In contrast, Beijing Mandarin has a two-way voicing contrast (voiceless unaspirated and voiceless aspirated) and only one tone is level (55). According to the Native Language Magnet model, native-language experience warps the perceptual space of non-native sounds and deviates it from the acoustic space as it should be [14, 15]. Learners need to employ attentional resources to detect the difference between native and non-native sounds and select the most reliable acoustic cues for perceptual learning [16, 17]. Therefore, this study asks two research questions: 1) whether in addition to facilitating the learning of consonants (as has been earlier demonstrated) the directing of attention during perceptual training also facilitates learning tones; and 2) do the error patterns in discriminating phonetic contrasts reflect the perceptual difficulties in learning Southern Min aspiration and pitch for Mandarin speakers?

2. Methods

2.1. Participants

Learner participants were 36 native speakers of Beijing Mandarin that do not speak other Chinese languages. They had no exposure to Southern Min before participating in this study. Learners were randomly assigned to two groups in the experiments: consonant-attending and tone-attending. Both groups were presented with the same stimuli sets during both training and testing, but received different instructions.

2.2. Stimuli

Recorded stimuli were created from native speakers of Quanzhou Southern Min (four males and four females).

2.2.1. Discrimination tests

Nine pairs of consonant contrast /b-p, p-p^h, b-p^h, g-k, k-k^h, g-k^h, e-te, te-te^h, e-te^h/ and ten pairs of tone contrast (22-33, 33-55, 22-55, 22-24, 22-41, 33-24, 33-41, 55-24, 55-41, 24-41) in Southern Min were tested in an AXB discrimination task [1, 2] for the pre-training and post-training tests. Each pair uses three different vowels from the Southern Min vowel list /i, e, a, ə, u, u, o, ɔ/. The AXB test therefore involves 27 pairs of syllables. Each pair was designed in a paradigm of AAB, ABB, BBA, and BAA with each of the three tokens in any AXB trial physically distinct and produced by different speakers. Two female and two male speakers of native Southern Min produced the 228 pairs of syllables. The speaker gender alternated in every trial so that two successive tokens were always produced by different genders.

2.2.2. Identification training

Each target sound was mapped to a figure displayed on the computer screen during the training session. Consonants with same place of articulation were mapped to the same shape and those with same aspiration manner were mapped to the same color (see Figure 1, which was presented to the consonant-attending group). Tones were mapped to images indicating pitch height and slope (see Figure 2, which was presented to the tone-attending group). (The IPA and number characters of consonants and tones are only shown in Figures 1 and 2 for reader clarity were not shown in the training instructions to the participants.) Thirty syllable-tone sequences were selected for training of both consonants and tones. Each syllable-tone sequence was produced two times by four speakers. Therefore, there were 240 tokens used in the training sessions.

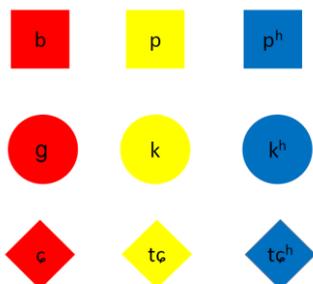


Figure 1: Consonant training paradigm.

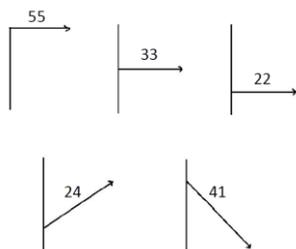


Figure 2: Tone training paradigm.

2.3. Procedure

All the syllable-tone sequences in the stimuli are real words in Quanzhou Southern Min. They were listed in a random order, embedded in a carrier phrase of /X, tɛit24 li41 ei22 X li41/ ('X, this character is X character') and recorded three times. Only

the first token of the target sounds was extracted from the carrier phrase. All the target tokens were RMS normalized to a 70dB intensity.

An AXB discrimination task with these target tokens was designed with E-prime 2.0 and used in both pre-training and post-training tests. Reaction time was recorded but not reported in the current paper. Before these 228 trials (as described above), there were five practice trials with different sounds to familiarize the participants with the procedure. The same discrimination test was used for both participant groups and for both pre- and post-training testing.

After the initial discrimination test, participants went through their first training session. First, participants were exposed to the sound-image mapping paradigm (as in Figures 1 and 2) using a PowerPoint presentation. Then, in the actual training trials (also designed with E-prime 2.0), after the participant heard a sound and pressed a keypad key corresponding to one of the images, feedback was immediately shown on the screen: for the correct response, showing "Correct"; for the incorrect response, showing "Incorrect" and the correct image. The trial that received an incorrect response would be repeated until there was a correct response before the procedure would move to the next trial. Only the first-time responses were recorded as the identification results. Both groups were trained with the same 240 syllables produced by four speakers different from the speakers who were used for the discrimination test stimuli.

Participants then returned on a different day within three days for a second session. This consisted of the same identification task as the first training session, followed by the post-training test, which was the same AXB discrimination test as in the pre-training test.

3. Results

The percent accuracy of the pre-training and post-training discrimination tests were analyzed separately by attending-group in two-way repeated measures ANOVAs. In this statistical model, subjects are already taken as a random factor and the fixed factors are test (two levels: pre-training and post-training) and sound (two levels: consonant and tone). The results of both groups show no interaction between test and sound but main effects of test in both the consonant-attending group [$F(1, 17)=8.736, p=0.009$] and the tone-attending group [$F(1, 17)=8.945, p=0.008$] and sound in both the consonant-attending group [$F(1, 17)=113.535, p<0.001$] and the tone-attending group [$F(1, 17)=42.286, p<0.001$]. Post hoc t-tests show the accuracy of consonant discrimination in the post-training test is significantly different from that of the pre-training test for the consonant-attending group [$t(17)=3.516, p=0.003$] and the accuracy of tone discrimination in the post-training test is significantly different from that of the pre-training test for the tone-attending group [$t(17)=2.874, p=0.011$]. There is no statistical difference in the accuracy of tone discrimination between the post-training test and the pre-training test for the consonant-attending group nor was there any statistical difference in the accuracy of consonant discrimination between the post-training test and the pre-training test for the tone-attending group. These results are indicated in Figures 3 and 4.

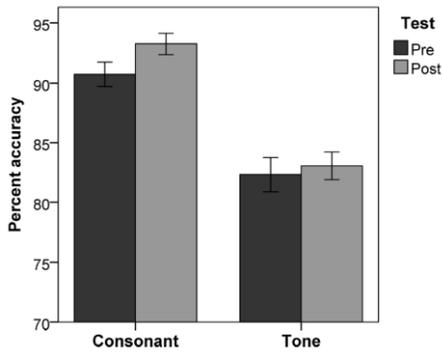


Figure 3: Results of the AXB discrimination task in pre-training and post-training tests of consonant-attending group.

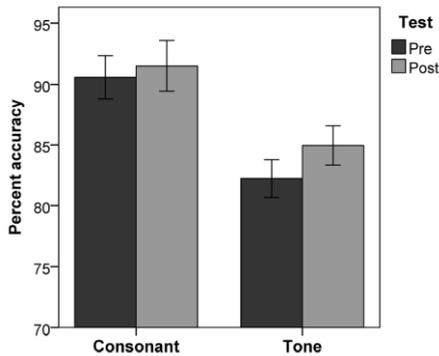


Figure 4: Results of the AXB discrimination task in pre-training and post-training tests of tone-attending group.

T-tests also show significant differences of the accuracy in the identification task between the first session and the second session in the perceptual training of both consonants [$t(17)=5.568, p<0.001$] and tones [$t(17)=5.854, p<0.001$]. The results are combined and shown in Figure 5.

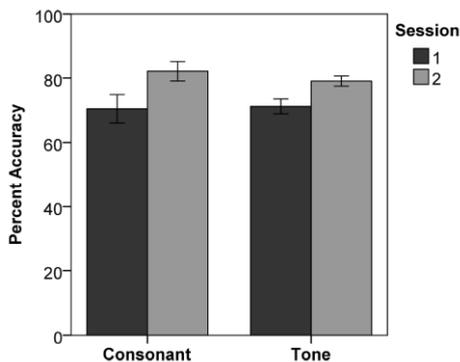


Figure 5: Results of the identification training combining both consonant-attending and tone-attending groups.

Four two-way ANOVAs were conducted to examine error patterns of the AXB discrimination task by target sounds and attending-group. The factors are test (two levels: pre and post) and contrast (nine levels for consonant: b-p, b-p^h, p-p^h, g-k, g-k^h, k-k^h, e-te, e-te^h, te-te^h; ten levels for tone: 22-33, 33-55, 22-55, 22-24, 22-41, 33-24, 33-41, 55-24, 55-41, 24-41). None of the ANOVAs show an interaction between test and contrast. The results of consonant discriminations by the consonant-

attending group show main effects of test [$F(1,17)=16.713, p=0.001$] and contrast [$F(8,136)=9.176, p<0.001$] (see Figure 6). The results of consonant discriminations by the tone-attending group show no main effect of test but contrast [$F(8,136)=13.277, p<0.001$] (see Figure 7). The results of tone discriminations by the tone-attending group show main effects of test [$F(1,17)=9.340, p=0.007$] and contrast [$F(9,153)=72.315, p<0.001$] (see Figure 8). The results of tone discriminations by the consonant-attending group show no main effect of test but contrast [$F(9,153)=55.503, p<0.001$] (see Figure 9).

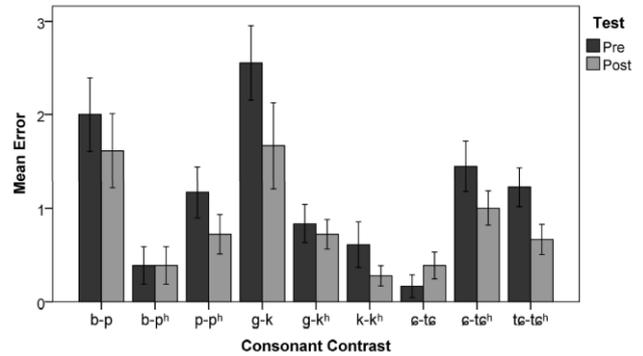


Figure 6: Mean errors of consonant discriminations by consonant-attending group (n=12).

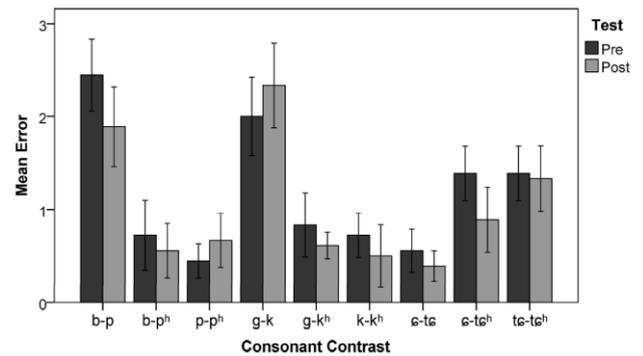


Figure 7: Mean errors of consonant discriminations by tone-attending group (n=12).

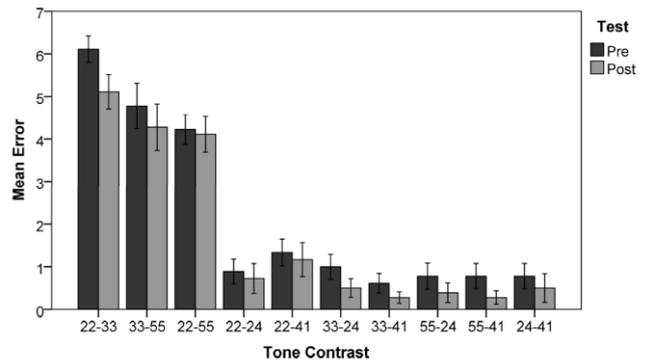


Figure 8: Mean errors of tone discriminations by tone-attending group (n=12).

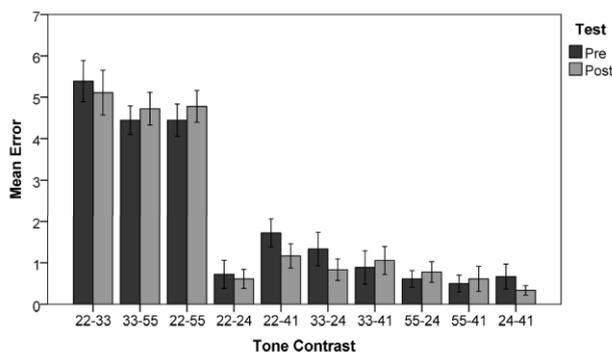


Figure 9: Mean errors of tone discriminations by consonant-attending group ($n=12$).

4. Discussion

The results (Figures 3 and 4) indicate that the consonant-attending group improved in discriminating consonant contrasts and the tone-attending group improved in discriminating tone contrasts after the identification training. However, even though both groups were equally exposed to the same training stimuli, the consonant-attending group did not improve in discriminating tones after training and neither did the tone-attending group in discriminating consonants. These results are consistent with the findings in [2] of Hindi by American English speakers in which the group instructed to attend to consonants during training demonstrated learning of consonants while the group instructed to attend to vowels instead did not demonstrate learning of consonant contrasts. The current study reconfirms that explicitly directing learners' attention during perceptual training facilitates learning new phonetic categories whereas the unattended phonetic categories show no improvement. This study expands those earlier results in that both target categories (consonants and tones) were learned during training and in this study one attended category was tone which had not previously been tested.

The error patterns in the consonant discrimination tests show more errors in the /b/-/p/ contrast and the /g/-/k/ contrast than any other consonant contrasts (Figures 6 and 7). This presumably is because of the lack of voiced stops and thus no /b/ and /g/ in Mandarin. The VOT (voice onset time) difference between voiced stops and voiceless unaspirated stops is less robust than that between voiceless unaspirated stops and voiceless aspirated stops. Accordingly, the discriminating accuracy of /b/-/p^h/ and /g/-/k^h/ was significantly higher than that of /b/-/p/ and /g/-/k/.

In the tone discrimination tests, there were more errors of level tone contrasts (22-33, 33-55, 22-55) than any other tone contrasts (Figures 8 and 9). This is not surprising since Mandarin has only one level tone (55) and height contrasts between level tones are difficult to perceive for Mandarin speakers. That the 22-33 contrast had the highest error rate of all the tone discrimination tests is predictable too since there is a lack of mid-level and low-level tones in Mandarin. Other tone contrasts were much better discriminated because there was always contrast between a contour tone and a level tone or a contrast between falling and rising tones. These results reconfirm the previous finding that Mandarin speakers rely more on pitch slope than pitch height to differentiate tones [1, 18, 19].

Despite these difficulties in discriminating certain phonetic categories in Southern Min, Mandarin learners' accuracy in discriminating aspiration contrasts and pitch contrasts increased after they were directed to attend the target categories during training. Increased accuracy in the identification task from the first training session to the second training session (Figure 5) suggests a process of perceptual learning, i.e., detecting the new phonetic categories, selecting the reliable perceptual cues, processing in working memory, and transferring to long-term memory [16, 17, 20, 21]. Importantly, the effect of learning carries over from the first session to the second session on a later day indicating a reasonably lasting effect.

The speaker variability in the current study's testing and training stimuli increased the difficulties of discriminating and identifying the target sounds (cf. [22, 23]). The use of varied speakers during training likely enhanced participants' discrimination ability in the post-training discrimination test with different speaker stimuli indicating a broadly applicable representation of these phonetic categories independent of particular speaker productions.

5. Conclusions

Endogenous orienting of attention to specific phonetic categories during perceptual training has clearly enhanced the learning of these phonetic categories. Mere exposure to unattended phonetic categories during training appears to have had little effect. The categories each group attended to were presented not only within the same stimuli, but within the same simple syllable. This suggests that attention is specifically to particular acoustic properties rather than there merely being a timing of attentional resources to particular times during stimuli presentation. The role of attention in the learning of relatively conscious categories (such as grammar or meaning) has been well established, but it appears to be vital for the acquisition of a broad range of phonetic categories as well. Future work will involve comparing the effect of orienting attention on learning novel tones for speakers of a different tone language (Mandarin) with that for speakers of a language without lexical tone (English).

6. Acknowledgements

This work was supported by the National Science Foundation of China 61573187. We thank Susan Guion-Anderson (1966-2011) for her suggestions to our initial design of this study.

7. References

- [1] S. G. Guion and E. Pederson, "Investigating the role of attention in phonetic learning," in O.-S. Bohn and M. Munro (eds.) *Language Experience in Second Language Speech Learning*, Amsterdam: John Benjamins, pp. 57-77, 2007.
- [2] E. Pederson and S. Guion-Anderson, "Orienting attention during phonetic training facilitates learning," *Journal of the Acoustical Society of America*, vol. 127, no. 2, pp. EL54-EL59, 2010.
- [3] M. I. Posner, "Orienting of attention," *Quarterly journal of experimental psychology*, vol. 32, pp. 2-25, 1980.
- [4] M. M. Chun, J. D. Golomb, and N. B. Turk-Browne, "A taxonomy of external and internal attention," *Annual review of psychology*, vol. 62, pp. 73-101, 2011.
- [5] R. S. Tomlin and V. Villa, "Attention in cognitive science and second language acquisition," *Studies in Second Language Acquisition*, vol. 16, pp. 183-203, 1994.
- [6] L. B. Forrest, *Discourse goals and attentional processes in sentence production: The dynamic construal of events*.

Conceptual structure, discourse and language. Stanford, CA: CSLI Publications, 1996.

- [7] R. Schmidt, "Attention," in P. Robinson (ed.) *Cognition and Second Language Instruction*, Cambridge: Cambridge University Press, pp. 3-32, 2001.
- [8] A. Myachykov, R. S. Tomlin, and M. I. Posner. "Attention and empirical studies of grammar," *The Linguistic Review*, vol. 22, no. 4, pp. 347-364, 2005.
- [9] L. A. Christensen and L. E. Humes, "Identification of multidimensional stimuli containing speech cues and the effects of training," *Journal of the Acoustical Society of America*, vol. 102, pp. 2297-2310, 1997.
- [10] A. L. Francis, K. Baldwin, and H. C. Nusbaum, "Effects of training on attention to acoustic cues," *Attention, Perception, & Psychophysics*, vol. 62, pp. 1668-1680, 2000.
- [11] A. L. Francis and H. C. Nusbaum, "Selective attention and the acquisition of new phonetic categories," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 28, pp. 349-366, 2002.
- [12] J. F. Werker and R. C. Tees, "Phonemic and phonetic factors in adult cross-language speech perception," *Journal of the Acoustical Society of America*, vol. 75, pp. 1866-1878, 1984.
- [13] L. Polka, "Characterizing the influence of native language experience on adult speech perception," *Attention, Perception, & Psychophysics*, vol. 52, pp. 37-52, 1992.
- [14] P. K. Kuhl and P. Iverson, "Linguistic experience and the perceptual magnet effect," in W. Strange (ed.) *Speech perception and linguistic experience: Issues in cross-language research*, Timonium, MD: York Press, pp. 121-154, 1995.
- [15] P. Iverson, P. K. Kuhl, R. Akahane-Yamada, E. Diesch, Y. Tohkura, A. Kettermann, and C. Siebert, "A perceptual interference account of acquisition difficulties for non-native phonemes," *Cognition*, vol. 87, pp. B47-B57, 2003.
- [16] W. Strange and V. L. Shafer, "Speech perception in second language learners: The re-education of selective perception," in J. G. Hansen Edwards and M. L. Zampini (ed.) *Phonology and second language acquisition*, Amsterdam: John Benjamins, pp. 153-191, 2008.
- [17] W. Strange, "Automatic selective perception (ASP) of first and second language speech: a working model," *Journal of phonetics*, vol. 39, pp. 456-466, 2011.
- [18] J. Gandour, "Tone perception in far eastern-languages," *Journal of Phonetics*, vol. 11, no. 2, pp. 149-175, 1983
- [19] J. Gandour, "Tone dissimilarity judgments by Chinese listeners," *Journal of Chinese Linguistics*, vol. 12, pp. 235-261, 1984.
- [20] M. I. Posner and S. E. Peterson, "The attention system of the human brain," *Annual Review of Neuroscience*, vol. 13, pp. 25-42, 1990.
- [21] J. S. Logan, S. E. Lively, and D. B. Pisoni, "Training Japanese listeners to identify English /r/ and /l/: A first report," *Journal of the Acoustical Society of America*, vol. 89, pp. 874-886, 1991.
- [22] D. B. Pisoni, "Long-term memory in speech perception: Some new findings on talker variability, speaking rate and perceptual learning," *Speech communication*, vol. 13, no. 1, pp. 109-125, 1993.
- [23] Y. Wang, M. M. Spence, A. Jongman, and J. A. Sereno, "Training American listeners to perceive Mandarin tones," *Journal of the Acoustical Society of America*, vol. 106, pp. 3649-3658, 1999.