Working Papers of the Cornell Phonetics Laboratory
1998, vol. 12, pp. 243-254

Training American Listeners to Perceive Mandarin Tones: A Pilot Study*

Yue Wang, Allard Jongman and Joan A. Sereno

Auditory training of nonnative speech contrasts is based on the assumption that the adult perceptual system can be modified. Previous research has shown substantial improvement in the identification of nonnative segmental distinctions after a simple phonetic laboratory training procedure. This study attempted to investigate whether such a procedure is applicable to the acquisition of nonnative suprasegmental contrasts, i.e., Mandarin tones. In four sessions during the course of a week, four American learners of Mandarin were trained to identify the four tones in natural words produced by both male and female native Mandarin talkers. The trainees' identification accuracy in the pretest and posttest were compared, showing an average 11% increase after training. This indicates that tone contrasts can be obtained and improved by training. The results are discussed in terms of nonnative perceptual modification at the suprasegmental level, as well as some methodological implications for further studies.

1 Introduction

It is commonly stated that the speech sounds of Mandarin do not present any particular difficulty for American learners of Mandarin; rather, its tones are difficult for them to acquire (Kiriloff 1969; Bluhme and Burr 1971; Shen 1989), since English and Mandarin differ in inextricably subtle ways in their pitch patterns, and in their distributions and functions (Chen 1974; White 1981). The present study attempted to train American listeners to identify the four Mandarin tones, using an auditory training procedure which has been shown to be highly effective to help learners acquire nonnative segmental contrasts in a comparatively short period of time.

1.1 Background

It is believed that, as compared to infants, adults depend more on linguistic experience than on auditory mechanisms in the perception of speech sounds. That is, they identify and discriminate speech sounds with reference to the linguistic categories of their native language. They are inferior to infants or children in the ability to distinguish novel foreign sounds. However, recent research suggests that adult perceptual mechanisms have more plasticity than was previously recognized. Therefore, a number of researchers have attempted to train listeners to perceive nonnative sounds in a linguistically meaningful manner, based on the assumption that the perceptual system of mature adults can be

*We would like to acknowledge the valuable comments of the participants of the Fall 1997 Empirical Research Workshop, Steve Bonta, Jenna Dalious, Meredith Doran, Pat Genung, Michael McCarthy, Yas Shirai, Traci Suiter, Hongyin Tao, and of the two reviewers, Niken Adisasmito-Smith and Abby Cohn.

modified. The goal of these auditory training studies is, by using relatively simple laboratory procedures, to help listeners create a new phonetic category that is usable in various phonetic contexts and can be retained in long-term memory.

An early attempt of this type of study was to train American listeners to perceive 3-way (i.e., voiced, voiceless unaspirated, voiceless aspirated) VOT distinctions (e.g., Pisoni, Aslin, Perey and Hennessy 1982; McClaskey, Pisoni and Carrell 1983), since English does not phonemically distinguish voiced and voiceless unaspirated stops. There were also experiments that trained French listeners to identify the English /θ - ð/ contrast, which is absent in French (e.g., Jamieson and Morosan 1986, 1989). Most training studies have concentrated on training Japanese listeners to identify English /r/ and /l/ (e.g., Strange and Dittman 1984; Logan, Lively and Pisoni 1991; Lively, Logan and Pisoni 1993; Lively, Pisoni, Yamada, Tohkura and Yamada 1994; Bradlow, Pisoni, Yamada and Tohkura 1997).

Summing up the results of these training studies, first and importantly, the identification of nonnative speech contrasts generally improved after training. For instance, Jamieson and Morosan (1986) reported that the French trainees' average percentage of correct identification for natural stimuli (containing /θ/ or /ð/) improved from the pretest (68% correct responses) to the posttest (79% correct responses) by 11%. Logan et al. (1991)'s study on training Japanese listeners to perceive English /r/ and /l/ also showed a significant increase in the percentage of correct responses from the pretest (78%) to the posttest (86%). Similarly, there was a 15% increase in the Japanese trainees' /r-l/ identification accuracy in Bradlow et al. (1997).

In addition, researchers have also found an effect of training with regard to generalization and long-term retention. First, experience gained from training on one phonetic category (e.g., VOT contrast for labial stops) can be transferred to another phonetic category (e.g., VOT for alveolar stops) without additional training (McClaskey et al. 1983). Second, generalization can extend to novel words and talkers that are not used in training (Lively et al. 1993). Third, contrasts learned can be maintained long (i.e., three months) after training (Lively et al. 1994). And finally, contrasts gained perceptually can be transferred to production without additional training (Rochet 1995; Bradlow et al. 1997).

Concerning methodological issues, the previous studies have agreed that training should be designed to ensure the formation of a robust phonetic category, since the ultimate goal would be to "facilitate the long-term development of a novel phonemic category that is potentially usable among a variety of phonetic contexts, talkers, and other sources of variability" (Logan and Pruitt 1995, p. 353). Previous studies have made every effort to achieve this goal. For example, Jamieson and Morosan (1986, 1989) designed the fading technique (i.e., training is not only on the prototypical stimuli, but also on a variety of exemplars within the category) in an attempt to extend generalization from synthetic to natural stimuli. Lively et al. (1994) demonstrated that the high-variability training paradigm they adopted (i.e., stimuli were put in various phonetic contexts and spoken by various talkers) encouraged a long-term modification of listeners' phonetic perception.

Previous studies have shown that the general procedure of an auditory training experiment typically involves a pretest, training sessions, a posttest, and additional posttests if generalization and long-term retention are examined. The pretest is usually used to compare with the posttest to determine the effectiveness of training. The training session is the core of the experiment, in which the basic question to consider is stimulus presentation. The typical tasks that have been used include discrimination and identification (with immediate feedback). However, Jamieson and Morosan (1986) compared both types of tasks and concluded that training using an identification task was more likely to result in an improvement in the perception of nonnative phonetic categories than using a discrimination task, because the former forces listeners to incorporate within-category variability in the formation of the novel phonetic categories.

1.2 The present study

As reviewed above, previous research has shown substantial improvements in the identification of segmental distinctions which are absent in the listeners' native language after simple phonetic laboratory training procedures. However, little research has reported the application of such training procedures to the acquisition of nonnative speech contrasts at the suprasegmental level. The goal of the present study was to examine whether auditory training which has been shown to be effective at the segmental level is applicable to the acquisition of nonnative suprasegmental contrasts, taking Mandarin tones as a case study.

In Mandarin, there are four distinctive tones, with Tone 1 having high-level pitch, Tone 2 high-rising pitch, Tone 3 low-dipping pitch, and Tone 4 high-falling pitch. Following Chao (1948), the tonal contours are represented schematically by time-pitch graphs attached to the left of a vertical reference line divided into four intervals by five points, as illustrated in Figure 1.

Based on previous findings, this study's hypothesis is that the American listeners' identification of the four Mandarin tones will be improved after training.

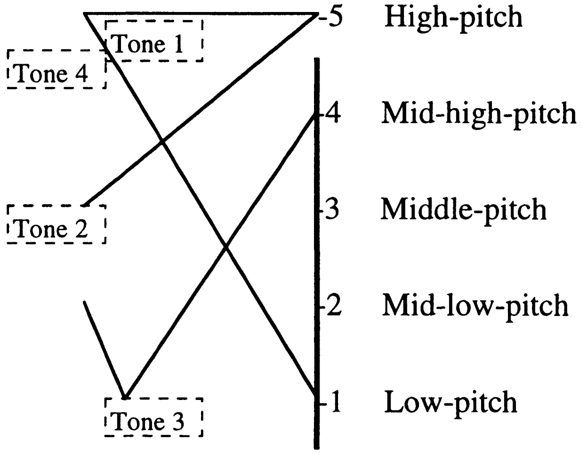

Figure 1. Relative pitch values of the four tones in Mandarin, with the contours going from the left to the right.

2 Method

2.1 Participants

Eight listeners participated in the study voluntarily, with four as trainees, and four as control subjects. All of them are undergraduate students at Cornell University who were taking a second year Mandarin Chinese language course at the time of the study. The subjects are all native speakers of American English, although two bilinguals were involved (one is English-Japanese, the other is English-Korean). None of the subjects claimed to have any experience with a tone language prior to learning Mandarin, and none of them have ever lived in a Mandarin-speaking environment.

Three native speakers of Mandarin Chinese participated as talkers. One female talker read the pretest and the posttest stimuli, and two others (one male, and one female) provided the training stimuli.

2.2 Stimuli

The stimuli are real monosyllabic Mandarin words presented in isolation. In order to ensure context variability, the stimuli were chosen to have combinations of different initial consonants and final vowels, and different syllabic structures (i.e., Vowel, ConsonantV, VNasal, CVN, CGlideV, CGVN). A total of 100 items were used in the pre/posttest (25 for each tone), and 160 were used in training (40 for each tone). The stimuli used in training were tape-recorded in a sound-proof booth, using a cardioid microphone (Electrovoice RE 20) and a cassette recorder (Carver TD-1700).

2.3 Procedure

The perceptual training program followed the high-variability procedure developed by Logan et al. (1991). This procedure consisted of a pretest phase, a training phase, and a posttest phase. Both the trained and the control subjects took the pretest, in which they were presented with the 100 randomized stimuli (written in *pinyin* romanization) on a sheet of paper. When the stimuli were presented, subjects were to mark (using tonal diacritics), for each item, which of the four tones they had heard. The pretest took place in a language classroom, and lasted about ten minutes.

Only the four trainees participated in the one-week training program, which consisted of four sessions of forty minutes each. During the training sessions, the trainees were presented with tape-recorded stimuli produced by either of two talkers. In each session, stimuli from only one talker were presented. The four tones were trained pairwise (i.e., Tones 1 and 2, Tones 1 and 3, Tones 1 and 4, Tones 2 and 3, Tones 2 and 4, and Tones 3 and 4). The trainees' task was two-alternative forced choice identification. They were to mark which tone they had heard on a sheet of paper with the stimuli written in *pinyin*. Immediate feedback was given after each trial, with the talker first indicating the correct response, and then repeating both tones in the corresponding tone pair. Each session ended with a test of twenty selected trained stimuli in that session without feedback.

Immediately after the training program, both the trained and the control subjects took the posttest, which was otherwise identical to the pretest, except that the stimuli were re-randomized.

3 Results

Figure 2 shows the overall results of training. It displays the percentage of correct identification for the trained (left bars) and the control (right bars) groups at the pretest and the posttest. As shown in the left bars, the trained subjects showed an improvement in their identification scores from pretest (77% correct identification) to posttest (88% correct identification), which indicates the trainees' substantial gains in tone identification accuracy (11% increase). In contrast, although on average, the control subjects started at approximately the same level as the trained group in the pretest (78% correct identification), they exhibited little improvement in the posttest (80% correct responses).

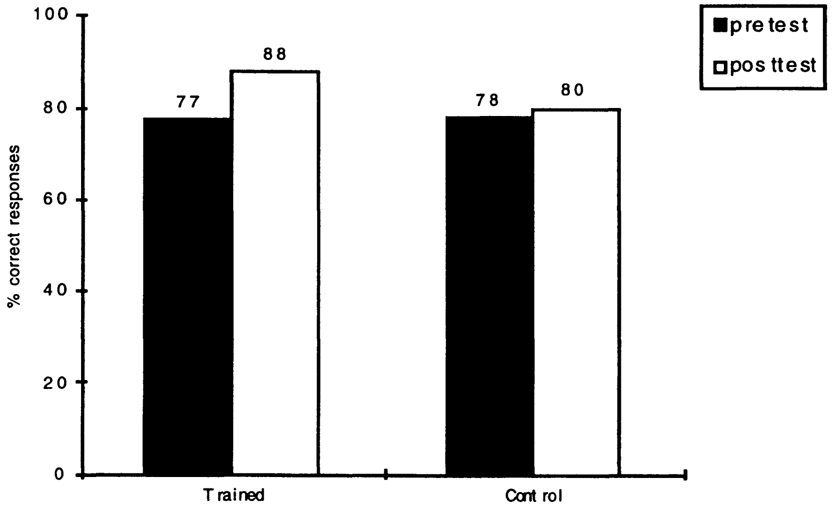

Figure 2. Percent correct identification of the four Mandarin tones for trained (n=4) and control (n=4) subjects at pretest and posttest

Figure 3 compares, in more detail, each of the trainees' correct responses at the pretest and the posttest, which reveals that each subject's performance has improved to a certain extent after training (26%, 12%, 5%, and 1% improvement for subject 1 to 4, respectively). It is also noted that there is a large degree of variation among the four trainees' initial levels. Thus, while both Subject 1 and Subject 2 showed substantial improvement from the pretest to the posttest, training effects were much smaller for Subject 3 and Subject 4, perhaps since they started at ceiling level.

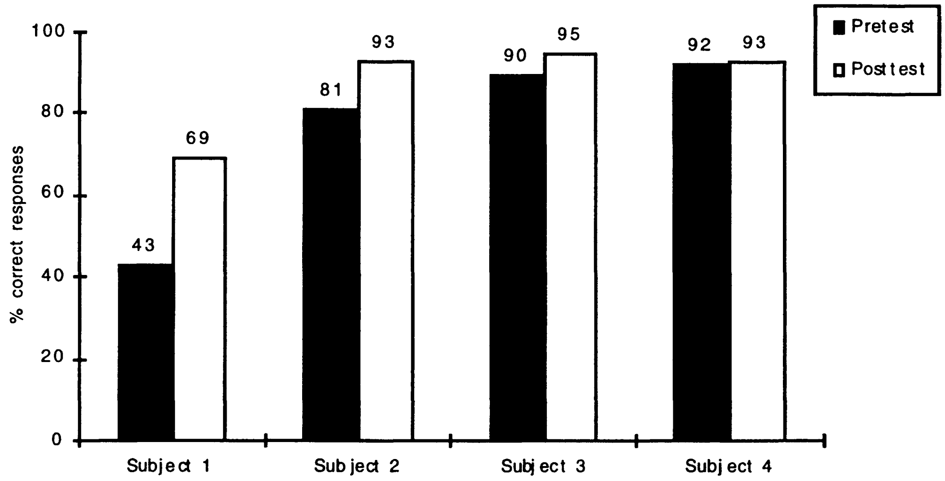

Figure 3. Percent correct identification of the four tones for each of the trained subjects at pretest and posttest

The trained group's performance in each individual tone is illustrated in Figure 4. It reveals that the trainees' identification of each tone improved from the pretest to the posttest (*i.e.*, 11% improvement for Tone 1, 7% for Tone 2, 20% for Tone 3, and 6% for Tone 4). In addition, the figure indicates that the trainees' identification of Tone 2 and 3 was comparatively poor at the pretest (70% and 68% correct responses, respectively). However, there was a substantially greater improvement for Tone 3 than for Tone 2.

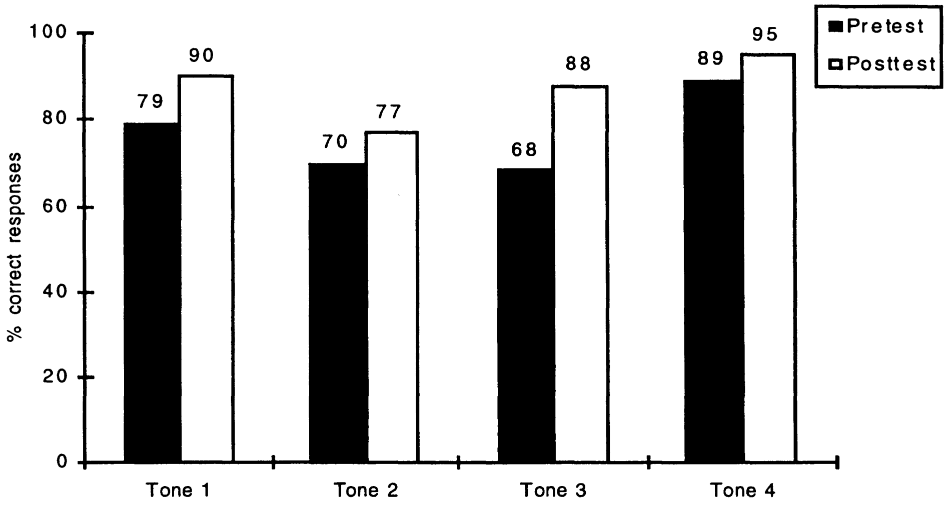

Figure 4. Trained subjects' (n=4) percent correct identification for each tone at pretest and posttest

An analysis of tone confusions is shown in Figure 5, which compares, for the pretest and the posttest, the number of errors the trainees made for each tone pair. For example, the number of errors for tone pair 1 and 2 is the sum of misperceptions of both Tone 1 as Tone 2, and Tone 2 as Tone 1. Consistent with the data shown in Figure 4 that identification was poor for Tone 2 and Tone 3, Figure 5 reveals that the trainees made more errors for Tones 2 and 3 than for other tone pairs, which indicates that these two tones were most easily confused. Also in agreement with the overall data, a comparison of the errors made at the pre- and the posttest shows a decrease of errors for each tone pair. In sum, Figures 2 to 5 indicate unanimously that the trainees' identification of the four tones improved after training.

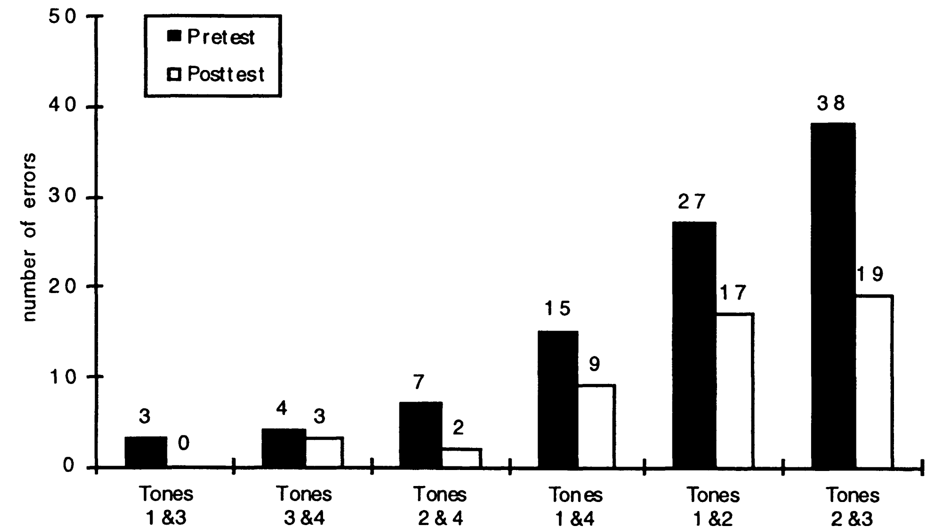

Figure 5. Tone pair confusion for trained subjects (n=4) at pretest and posttest. The number of errors for each tone pair refers to misperceptions of one tone as the other in the corresponding tone pair.

The overall results were analyzed using a 2-way ANOVA with test (pre, post) and group (trained, control) as factors. Both the main effect of group [F(1, 14)=0.19, p>.670], and the group x test interaction [F(3,14)=0.32, p>.585] failed to reach significance. More specifically, for the trained group, another 2-way ANOVA was calculated, with test (pre, post) and tone (tones 1 to 4) as factors. The results showed neither a significant effect of test [F(1, 30)=2.14, p>.157], nor a test x tone interaction [F(3, 30)=0.15, p>.931]. Given the substantial improvement in tone identification shown in Figures 2 to 5, it is a bit surprising that the statistical analyses did not reach significance. Two reasons might account for this result, the first being that the scope of this study is relatively small (*i.e.*, only four trainees and four control subjects). With more subjects, a statistically significant difference may be expected. Second, as was shown in Figure 3, the trainees were not at the same level of tone identification accuracy. This variability among subjects might make it difficult to reach statistical significance.

4 Discussion and conclusion

The results of the present experiment suggest that the perception of Mandarin tones can be improved using a simple training task. This indicates that the procedure which has been adopted in training the acquisition of nonnative segmental contrasts can also be applied at the suprasegmental level. As a pilot study, these preliminary results obtained have some implications for future studies, a number of which will be discussed in this section.

The results showed an average 11% increase in the trainees' identification accuracy after training, which is comparable to the results of other training studies discussed previously (e.g., Logan et al. 1991; Bradlow et al. 1997). This is indeed a substantial improvement given the short period of training. Since the previous studies have found progressive improvement in the trainees' identification accuracy as training continued (Logan et al. 1991), a greater degree of improvement might be expected if tone training lasted longer.

The present study adopted the high-variability paradigm designed by the previous studies to promote the formation of robust phonetic categories that are independent of context and talker variabilities. The results show that this procedure is also effective in tone training. In fact, talker variability is crucial in tone training, since different talkers (especially males and females) have different fundamental frequencies (F0). It has been reported that native Mandarin speakers use changes in F0 contours more than heights to distinguish among tones (Kratochvil 1968; Howie 1976; Gandour 1983, 1984). Therefore, if various talkers are used (including both males and females), learners will be trained to focus on detecting the pitch contour differences of the tones, and to normalize the differences in F0 height of various talkers.

As discussed previously, the results of this study failed to reach significance statistically, one of the possible reasons being that there was a great amount of variation among the subjects' identification accuracy. While some of them scored very low (e.g., 43% correct responses for Subject 1 at pretest), others started at ceiling level (e.g., 92% for Subject 4 at pretest). The variation among the subjects' initial levels seemed to correlate with the extent of the training effects. Thus, while the listener with a lower initial score showed substantial improvement in the posttest, training effects were much smaller for the one who started high. It appears that, for the latter, the task might be too easy, and there is little room for improvement during training. Therefore, in future studies, learners at a lower level would be preferred.

The results also showed that the most easily confused tone pair was Tones 2 and 3, followed by Tones 1 and 2, Tones 1 and 4, and Tones 2 and 4; whereas the trainees did not make many errors with Tones 1 and 3, and Tones 3 and 4 (cf. Figure 5). This differential effect strongly suggests that for future studies, training should be dedicated more to those difficult tone pairs. Previous studies have also shown that training was more effective if the focus was on those aspects in which identification performance was initially poor. For instance, in Lively et al. (1993) only tokens containing /r/ and /l/ from the difficult environments (i.e., initial singleton, initial consonant clusters, and intervocalic positions) were presented during training.

This study only examined the effect of tone training by comparing the performance at the pretest and the posttest. However, since the goal of auditory training is to help listeners create new phonetic categories effective in terms of generalization and long-term retention, these two aspects are major concerns in training studies. Previous research has demonstrated that new categories gained by training can be transferred to other phonetic categories or linguistic contexts (McClaskey et al. 1983; Logan et al. 1991), and can be retained in long-term memory (Lively et al. 1994). In future tone training studies, generalization and long-term retention will also be examined.

Another concern is whether perceptual training can be transferred to production, so that training efforts could result in a facilitating effect (i.e., positive transfer) from one modality to the other (Leather and James 1991). Since segmental training studies have also found that learning gained perceptually can benefit production (Rochet 1995; Bradlow et al. 1997), it is worthwhile to test if such transfer can also occur in tone training. Moreover, fine acoustic analysis of American listeners' tone production before and after training, as compared to the native norms, may also be beneficial to quantatitively judge the trainees' improvement after training.

In sum, the results of the present experiment suggest that auditory training can be used to modify American listeners' perception of Mandarin tones in isolated words. Although the scope of this study is too small to reach a convincing conclusion, the results obtained here are promising to inform plans for future studies. Based on the present pilot, a large-scale study is currently underway with more subjects and talkers to examine the training effects on the acquisition of nonnative suprasegmental contrasts, and the transfer, generalization, and long-term retention of the contrasts gained from training.

5 References

Bluhme, H. and R. Burr (1971) An audio-visual display of pitch for teaching Chinese tones. *Studies in Linguistics* 22, 51-57.

Bohn, O.S. (1995) Cross-language speech perception in adults: First language transfer doesn't tell it all. In W. Strange (ed.) *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* . Baltimore: York Press, 273-304.

Bohn, O.S. and J. E. Flege (1990) Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics* 11, 303-328.

Bradlow, A.R., D. B. Pisoni, R. A. Yamada and Y. Tohkura (1997) Training Japanese listeners to identify English /r/ and /l/ IV: Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America* 101, 2299-2310.

Chao, Y.R. (1948) *Mandarin Primer*. Cambridge: Harvard University Press.

Chen, G.T. (1974) The pitch range of English and Chinese speakers. *Journal of Chinese Linguistics* 2, 159-171.

Chen, Q. (1997) Toward a sequential approach for tonal error analysis. *Journal of the Chinese Language Teachers Association* 32, 21-39.

Gandour, J.T. (1983) Tone perception in Far Eastern languages. *Journal of Phonetics* 11, 149-175.

Gandour, J.T. (1984) Tone dissimilarity judgments by Chinese listeners. *Journal of Chinese Linguistics* 12, 235-261.

Howie, J.M. (1976) *Acoustical Studies of Mandarin Vowels and Tones*. Cambridge: Cambridge University Press.

Jamieson, D.G. and D. E. Morosan (1986) Training non-native speech contrasts in adults: Acquisition of the English /θ/-/ð/ contrast by francophones. *Perception and Psychophysics* 40, 205-215.

Jamieson, D.G. and D. E. Morosan (1989) Training new, nonnative speech contrasts: A comparison of the prototype and perceptual fading techniques. *Canadian Journal of Psychology* 43, 88-96.

Kiriloff, C. (1969) On the auditory discrimination of tones in Mandarin. *Phonetica* 20, 63-67.

Kratochvil, P. (1968) *The Chinese Language Today*. London: Hutchinson University Library.

Leather, J. and A. James (1991) The acquisition of second language speech. *Studies in Second Language Acquisition* 13, 305-341.

Lively, S.E, J. S. Logan and D. B. Pisoni (1993) Training Japanese listeners to identify English /r/ and /l/ II: The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America* 94, 1242-1255.

Lively, S.E., D. B. Pisoni, R. A. Yamada, Y. Tohkura and T. Yamada (1994). Training Japanese listeners to identify English /r/ and /l/ III: Long-term retention of new phonetic categories. *Journal of the Acoustical Society of America* 96, 2076-2087.

Logan, J.S., S. E. Lively and D. B. Pisoni (1991) Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America* 89, 874-886.

Logan, J.S., S. E. Lively and D. B. Pisoni (1993) Training listeners to perceive novel phonetic categories: How do we know what is learned? *Journal of the Acoustical Society of America* 94, 1148-1151.

Logan, J.S. and J. S. Pruitt (1995) Methodological issues in training listeners to perceive non-native phonemes. In W. Strange (ed.) *Speech Perception and Linguistic Experience: Issues in Cross-Language Research.*. Baltimore: York Press, 351-377.

McClaskey, C.L., D. B. Pisoni and T. D. Carrell (1983) Transfer of training of a new linguistic contrast in voicing. *Perception and Psychophysics* 34, 323-330.

Miracle, W.C. (1989) Tone produciton of American students of Chinese: A preliminary acoustic study. *Journal of Chinese Language Teachers Association* XXIV, 49-65.

Pisoni, D.B., R. N. Aslin, A. J. Perey and B. L. Hennessy (1982) Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human Perception and Performance* 8, 297-314.

Pisoni, D.B., S. E. Lively and J. S. Logan (1994) Perceptual learning of nonnative speech contrasts: Implications for theories of speech perception. In J. Goodman and H. Nusbaum (eds.) *Development of Speech Perception: The Transition from Recognizing Speech Sounds to Spoken Words*. Cambridge: MIT, 121-166.

Rochet, B.L. (1995) Perception and production of second-language speech sounds by adults. In W. Strange (ed.) *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Baltimore: York Press, 379-410.

Shen, X.S. (1989) Toward a register approach in teaching Mandarin tones. *Journal of Chinese Language Teachers Association* XXIV, 27-47.

White, C.M. (1981) Tonal perception errors and interference from English intonation. *Journal of Chinese Language Teachers Association* 16, 27-56.