View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by KU ScholarWorks

Training American listeners to perceive Mandarin tones

Yue Wang[a] and Michelle M. Spence
Cornell Phonetics Laboratory, Cornell University, Ithaca, New York 14853

Allard Jongman and Joan A. Sereno
Linguistics Department, University of Kansas, Lawrence, Kansas 66045

(Received 20 April 1999; revised 19 July 1999; accepted 20 July 1999)

Auditory training has been shown to be effective in the identification of non-native segmental distinctions. In this study, it was investigated whether such training is applicable to the acquisition of non-native suprasegmental contrasts, i.e., Mandarin tones. Using the high-variability paradigm, eight American learners of Mandarin were trained in eight sessions during the course of two weeks to identify the four tones in natural words produced by native Mandarin talkers. The trainees' identification accuracy revealed an average 21% increase from the pretest to the post-test, and the improvement gained in training was generalized to new stimuli (18% increase) and to new talkers and stimuli (25% increase). Moreover, the six-month retention test showed that the improvement was retained long after training by an average 21% increase from the pretest. The results are discussed in terms of non-native suprasegmental perceptual modification, and the analogies between L2 acquisition processes at the segmental and suprasegmental levels. © *1999 Acoustical Society of America.* [S0001-4966(99)04811-0]

PACS numbers: 43.71.Hw, 43.71.Es [JMH]

INTRODUCTION

It is commonly stated that Mandarin tones are difficult for American learners to acquire (e.g., Kiriloff, 1969; Bluhme and Burr, 1971; Shen, 1989), since English and Mandarin differ in their pitch patterns, distributions, and functions (Chen, 1974; White, 1981). In the present study, American listeners were trained to identify the four Mandarin tones, using an auditory training procedure which has been shown to be effective in helping learners acquire non-native segmental contrasts in a comparatively short period of time.

A. Auditory training

Research in the domain of second language (L2) acquisition has generally found that adults are inferior to children in the ability to perceive and produce foreign speech sounds, manifested by the commonly known ''adult foreign accent.'' The belief in the possibility that children enjoy an innate ability to acquire languages more easily and accurately than adults leads to the Critical Period Hypothesis (CPH), stating that cerebral lateralization occurs after puberty, accompanied by the loss of neurological plasticity of the brain, resulting in a reduction in language learning ability (Lenneberg, 1967).

An alternative account of foreign accent is the phonologically based argument that foreign accent is not caused by the completion of cerebral lateralization, but is rather the result of the interaction between L2 learners' two phonetic systems (e.g., Flege, 1995; Best, 1995). In this view, adult L2 learners differ from children acquiring their first language (L1) in that the former perceive and produce L2 sounds with reference to the linguistic categories of their existing native language system. Thus the influence of the adults' firmly established L1 phonetic system is believed to be responsible for ''foreign accent.'' However, unlike the CPH statement of a complete diminution of speech learning ability at puberty, the phonologically based argument is that the decline in human vocal learning ability with age does not apply to all L2 sounds. It is assumed that the degree of approximation to L2 sounds depends on learners' ''perceived phonetic similarity'' of L2 sounds to their L1 phonetic categories. Empirical research has revealed that, with sufficient experience and exposure, adult L2 learners can authentically perceive or produce novel L2 phones which are judged to have no L1 phonemic counterparts, although it is still difficult for them to form separate phonetic categories for those L2 sounds that are similar to L1 counterparts but realized in a phonetically different manner (Flege, 1987; Best *et al.*, 1988).

The evidence that learners can improve their L2 pronunciation at least for some target language sounds suggests adult perceptual mechanisms have more plasticity than was previously recognized. Therefore, researchers have attempted to train listeners to perceive non-native sounds in a linguistically meaningful manner, based on the assumption that the perceptual system of mature adults can be 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 approach was to train American listeners to perceive three-way (i.e., voiced, voiceless unaspirated, voiceless aspirated) voice onset time (VOT) distinctions (e.g., Pisoni *et al.*, 1982; McClaskey *et al.*, 1983), since English does not phonemically distinguish voiced and voiceless unaspirated stops. There were also experiments that trained French listeners to identify the English /ɵ-ð/ contrast, which is absent in French (e.g., Jamieson and Morosan, 1986, 1989). Most recent training studies have concentrated on training Japanese listeners to identify English /r/ and /l/

[a]Electronic mail: yw36@cornell.edu

(e.g., Strange and Dittmann, 1984; Logan *et al.*, 1991; Lively *et al.*, 1993; Lively *et al.*, 1994; Bradlow *et al.*, 1997).

Summing up the results of these training studies, first and most importantly, the identification of non-native 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 /ɵ/ or /ð/) improved from the pretest (68% correct responses) to the post-test (79% correct responses) by 11%. Logan *et al.*'s (1991) study on training Japanese listeners to perceive English /r/ and /l/ also showed a significant increase of 8% from pretest (78%) to post-test (86%). Similarly, there was a 16% increase (from 65% to 81%) 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 to six 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 is to facilitate the development of a new phonemic category that is usable among a variety of sources of variability (Logan and Pruitt, 1995). 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. While Strange and Dittmann (1984) report no significant effect of discrimination training using synthetic stimuli in only one phonetic environment, Logan *et al.* (1991) demonstrated that a high-variability training paradigm (i.e., identification of natural stimuli in various phonetic contexts and spoken by various talkers) encouraged long-term modification of listeners' phonetic perception.

B. Mandarin tones

Mandarin phonemically distinguishes four tones, with Tone 1 having high-level pitch, Tone 2 high-rising pitch, Tone 3 low-dipping pitch, and Tone 4 high-falling pitch (Chao, 1948). The prosodic features of tones are manifested physically by different fundamental frequency ($F0$) values, as shown in Fig. 1. Moreover, the $F0$ pattern for particular tones varies as a function of vowel (Howie, 1976). In addition, the intrinsic duration differs for the four tones, the longest being Tone 3, and the shortest being Tone 4 (Lin, 1965). Intrinsic amplitude has been found to vary among the four tones as well, with Tone 3 having the lowest, and Tone 4 the highest amplitude (Chuang *et al.*, 1972).

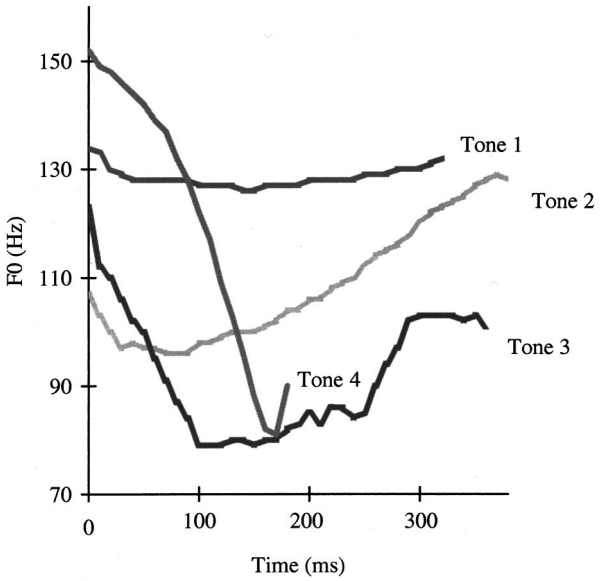

FIG. 1. $F0$ contours for the four Mandarin tones, each combined with the syllable *fa*, produced by a male native speaker of Mandarin.

Studies in the perceptual domain have shown that the above acoustic cues are functionally integrated in the identification of Mandarin tones by native listeners. For example, perception tests using synthetic $F0$ contours and multidimensional scaling studies have demonstrated the two dimensions of $F0$ height and contour as fundamental perceptual cues of Mandarin tones, of which listeners seem to attach more importance to the ''contour'' than ''height'' dimensions (Gandour, 1984; Massaro *et al.*, 1985). $F0$ contour as a perceptual cue has been further investigated in terms of $F0$ turning point, i.e., the point at which the direction of the $F0$ contour changes from falling to rising, the results of which showed that the timing of $F0$ turning point constitutes a salient perceptual cue for differentiating Tone 2 from Tone 3 (Shen and Lin, 1991; Moore and Jongman, 1997), and Tone 3 from Tone 4 (Gårding *et al.*, 1986). In addition, duration has also been shown to affect tonal perception. For instance, Blicher *et al.* (1990) reported that systematic lengthening of the vowel shifted the labeling boundary in the direction of the Tone 2 exemplar, thus producing more Tone 3 responses. Moreover, native Mandarin listeners have been found to refer to extrinsic $F0$ (corresponding to speaker identity) as a frame of reference for tone perception; that is, they perceive tones by normalizing for speaker $F0$ range (Moore and Jongman, 1997).

Perception studies on Mandarin tones have also been conducted cross-linguistically to examine if and how nontonal listeners distinguish themselves from the Mandarin listeners by their patterns of perceptual processing of the dimensions of $F0$. For example, by comparing tone perceptual patterns of native English and Mandarin (as well as Cantonese, Taiwanese, and Thai) listeners, Gandour (1983) found that native English listeners attached more importance to the height, and less to the direction dimension, than did listeners from most of the tone languages. He argued that since English has no contrastive tones, contour or otherwise, English listeners directed their attention almost exclusively

to the $F0$ height of the stimuli. Addressing the same question, Leather (1987) examined the identification of Mandarin Tone 1 and Tone 2 (in a synthetic Tone 1-2 continuum) by native listeners of English and Dutch (both nontonal), as compared to that by Mandarin listeners. The result of a greater spread in location of the category crossover among the Dutch and English, as opposed to the Chinese, reflects linguistically inappropriate perceptual weighting of the parameters of $F0$ contour by the phonetically unskilled nonnatives. Stagray and Downs (1993) examined the differential sensitivity for frequency among Mandarin and English listeners from a psychoacoustic perspective. They found that Mandarin listeners had poorer differential sensitivity than English listeners because the former had learned to categorize sounds of similar frequency together to facilitate their perception of tones. Taken together, these cross-linguistic studies suggest that linguistic experience plays an important role in tone perception.

For adult nontonal speakers learning Mandarin as an L2, tones have presented great difficulty (e.g., Kiriloff, 1969; Bluhme and Burr, 1971; Shen, 1989). For native speakers acquiring Mandarin as L1, tonal pattern is an integral part of each word they learn, but such functional association between segmental structure and $F0$ contour is nonexistent, for example, in American learners' linguistic behavior. Therefore, the source of difficulty in learning tones has generally been attributed to interference from English suprasegmental features. Knowledge of the function of pitch in the English stress and intonation systems was found to highly influence American listeners' perception of Mandarin tones (White, 1981; Broselow *et al.*, 1987; Chen, 1997). For example, White (1981) claimed that English listeners will perceive the Mandarin high tones as stressed and the low Tone 3 as unstressed, despite the fact that in Mandarin, the stress on a syllable is mainly realized by duration and amplitude rather than $F0$. Given her observations that Tones 1 and 4 are more difficult to acquire, Shen (1989) argued that these two tones are more likely to be receptive to L1 interference since they are prosodically less marked than Tones 2 and 3. It should be noted that, although Tones 2 and 3 have been observed to be easier to learn than Tones 1 and 4, this tone pair is still the most confusing pair for English learners of Mandarin (Kiriloff, 1969).

C. The present study

As reviewed previously, research has shown substantial improvements (8%–16%), after simple phonetic laboratory training procedures, in the identification of segmental distinctions which are absent in the listeners' native language. However, little research has reported the application of such training procedures to the acquisition of non-native speech contrasts at the suprasegmental level. Since the acquisition of Mandarin tones has been found to be difficult for native nontonal learners, it provides an ideal case for the study of suprasegmental training. By training American listeners to perceive Mandarin tones, the goal of the present study was to examine whether auditory training, which has been shown to be effective at the segmental level, can be applied to the acquisition of non-native suprasegmental contrasts.

TABLE I. Characteristics of the trainees and the controls in terms of language background.

	Gender	Age	Mode of learning	Length of learning[a]	Class when training[b]	L2 experience
Trainee						
1	F	20	class[c]	7 months	yes	none
2	M	25	class	4 months	no	Spanish
3	F	19	class	7 months	yes	French
4	F	29	class	4 months	no	Cantonese
5	F	20	class	7 months	yes	French
6	M	24	intensive[d]	7 months	yes	none
7	F	19	class	7 months	yes	Cantonese
8	F	24	intensive	7 months	yes	none
Control						
1	M	21	class	7 months	yes	none
2	F	20	class	4 months	no	Cantonese
3	M	25	class	7 months	yes	Japanese
4	M	22	class	10 months	yes	Cantonese
5	M	23	class	7 months	yes	Spanish
6	M	21	class	7 months	yes	French
7	F	20	class	7 months	yes	Spanish
8	M	22	class	10 months	yes	none

[a]Length of learning Mandarin as a foreign language.
[b]Whether taking Mandarin course during the training period.
[c]A first-year Chinese course (5 hours/week).
[d]An intensive Mandarin program (20 hours/week).

I. METHOD

The perceptual training program followed the high-variability procedure developed by Logan *et al.* (1991). That is, American listeners were trained to identify the four Mandarin tones appearing in a variety of phonetic contexts in natural words, produced by a variety of talkers. In order to assess the trainees' improvements, the program included a pretest before training, a post-test, two generalization tests, and a long-term retention test. Listeners' performance in the pretest and the post-test was compared to determine to what extent tone identification could be improved due to training. The two generalization tests were designed to examine if any improvement gained in training could be extended to novel stimuli (Generalization Test 1), and to novel talkers and stimuli (Generalization Test 2). The retention test was conducted six months after the training program to determine the long-term training effects.

A. Participants

Sixteen native speakers of American English without speech and hearing impairments participated in the study, with eight as trainees and eight as controls. All were paid for their participation. The trainees and controls are all students at Cornell University who have taken one or two semesters of Mandarin Chinese language courses. None of the trainees or controls has ever lived in a Mandarin-speaking environment, and most of them (except for the four who speak limited Cantonese) have no experience with a tone language prior to learning Mandarin. The characteristics of the trainees and controls are described in Table I.

Six native speakers of Mandarin Chinese participated voluntarily as talkers. One male speaker read the pretest and post-test stimuli, while four others (two males and two females) served as talkers during training. One of these male

speakers also read the novel stimuli for Generalization Test 1 (henceforth Gen 1). The sixth speaker was a female who provided the novel stimuli for Generalization Test 2 (henceforth Gen 2).

B. 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 various initial consonants and final vowels, and different syllabic structures (i.e., V, CV, CVNasal, VN, CGlideV, CGVN). A total of 400 different stimuli were selected: 100 items (25 for each tone) were used in the pre/post-test, 180 (45 for each tone) in training, 60 (15 for each tone) in Gen 1, and an additional 60 (15 for each tone) in Gen 2. The stimuli used in the retention test were the same as those in the post-test.

The stimuli were tape-recorded in a soundproof booth in the Cornell Phonetics Laboratory, using a cardioid microphone (Electrovoice RE20) and a cassette recorder (Carver TD-1700). They were then digitized at 11 kHz and low-pass filtered at 5 kHz using WAVES+/ESPS speech analysis software running on a SUN Sparc Station, after which they were transferred to a Swan 386/25 PC for the perceptual tests and training, using the BLISS software (Mertus, 1989).

Before the training program started, the intelligibility of the stimuli provided by the six talkers was assessed by one male and one female native speaker of Mandarin Chinese. Listeners indicated which tone they heard by pressing one of four response buttons. For both listeners, identification accuracy was 100% for all stimuli and all talkers.

C. Procedure

The training program consisted of a pretest phase, a training phase, and a post-test phase. Both the tests and training were conducted at the Cornell Phonetics Laboratory, where listeners were tested or trained in a sound-treated cubicle. Stimuli were presented binaurally at a comfortable sound level over Sony MDR-V6 headphones. Listeners were instructed to indicate their responses by pushing corresponding buttons representing each of the four tones. The four buttons were labeled from left to right by the numbers 1 to 4, as well as by the tonal diacritics (stylized pitch contours).

1. Pretest

Both the trainees and the controls took the pretest, in which they were presented with 100 randomized stimuli, with an inter-trial-interval of 3 s. The listeners were told to respond after each stimulus. They were encouraged to guess if unsure. No feedback was given at any time. The pretest lasted about 10 min, with no more than four listeners tested at any one time. All listeners were tested within a one-week period.

2. Training sessions

Immediately after the pretest, only the eight trainees participated in the two-week training program, consisting of eight sessions of 40 min each, during which the trainees were trained auditorily with the stimuli produced by four talkers.

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). Pairwise presentation during training allowed for a systematic increase in difficulty of tone contrasts. The order of tone pair presentation was from easiest to most difficult, in accordance with the error analysis obtained from the trainees' pretest. That is, for each training set, the first session always started with Tones 1 and 3, followed by Tones 3 and 4, and Tones 1 and 4; the second session had Tones 1 and 2, Tones 2 and 4, and Tones 2 and 3 presented in succession. Three tone pairs were trained in each session, such that it took a training set of two successive sessions to complete one talker for a total of 180 stimuli. The order of presentation in terms of talker was counterbalanced for the eight trainees, but male and female talkers were always presented alternately.

During each session, the trainees' task was two-alternative forced-choice identification. They were to indicate (within 2 s) which tone of a certain tone pair they had heard by pressing the corresponding button. Immediate feedback was given after each stimulus, with a neutral voice indicating the correct response in English, and the talker repeating both tones in the tone pair. For example, for target stimulus *bei 3* (bearing Tone 3) in tone pair 3 and 4 training, stimulus presentation and feedback went as follows:

Talker: *bei 3*.
Trainee's response.
Neutral English voice: *That was Tone 3*.
Talker's repetition: *bei 3*.
Neutral English voice: *Tone 4 is*:
Talker: *bei 4*.

Thus the above block was considered a training trial, with an inter-trial-interval of 5 s. In addition, to focus the trainees' attention, each trial started with a 500-Hz pure tone. Each tone pair training (i.e., 30 trials) ended with a short break.

After each two consecutive sessions (i.e., a single talker), trainees were given a test of 60 selected trained stimuli produced by the same talker. No feedback was given. Since there were four different talkers for training, four assessments (training set 1–4) were administered.

3. Post-tests

Immediately after the training program, both the trained and the control listeners took the post-test, which was otherwise identical to the pretest, except that the stimuli were re-randomized. The listeners then took Gen 1, with 60 novel stimuli produced by one of the male talkers from training, and Gen 2, with an additional 60 novel stimuli produced by a new female talker; the procedures of both were comparable with the pretest. The post-tests were completed within a week's period.

4. Retention test

Six months after training, four trainees (Trainees 1, 3, 4, and 7 in Table I) and four controls (Controls 1, 2, 4, and 6 in Table I) were available for the long-term retention test, which involved the same stimuli and procedure as the post-

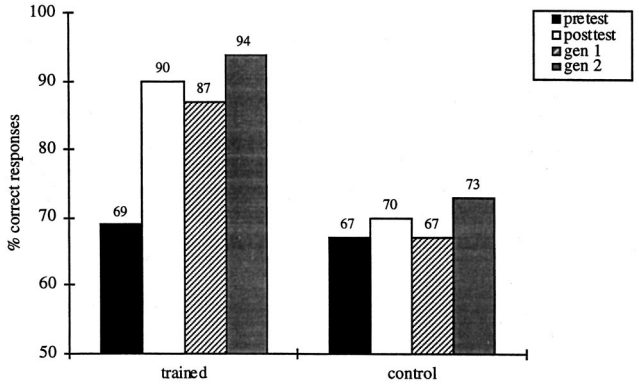

FIG. 2. Mean percent correct identification of the four Mandarin tones for trained ($n=8$) and control ($n=8$) subjects at pretest, post-test, generalization test 1 (Gen 1: old talker, new stimuli), and generalization test 2 (Gen 2: new talker, new stimuli).

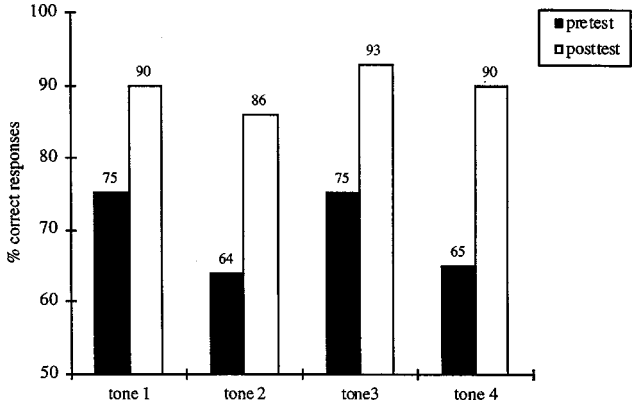

FIG. 3. Trained subjects' mean percent correct identification for each tone at pretest and post-test.

test. All four trainees and two of the controls (Controls 2 and 4) had not been exposed to Mandarin for at least three months (summer break) before the retention test. The other two controls (Controls 1 and 6), however, had been in Taiwan for three months taking an intensive Mandarin course.

II. RESULTS

A. Overall improvement and generalization

Correct identification scores for the trained and the control groups at the pretest, post-test, generalization test 1, and generalization test 2 are displayed in Fig. 2. As shown in the left-hand bars, the trainees showed an improvement in their identification scores from pretest (69% correct identification) to post-test (90% correct identification), a substantial 21% increase in tone identification accuracy. Moreover, this increase in performance was also revealed in the two generalization tests (87% correct identification in Gen 1; and 94% correct identification in Gen 2), indicating tone contrasts gained in training were extended to novel talkers and stimuli.

In contrast, as the right-hand bars show, although the control listeners started at approximately the same level as the trainees in the pretest (67% correct identification), they exhibited little improvement in the three post-tests (70% in the post-test, 67% in Gen 1, and 73% in Gen 2).

The overall results were analyzed using a two-way ANOVA of Test (pretest, post-test, Gen 1, Gen 2) and Group (trained, control), with Test as the repeated measure. There was a significant main effect of Test [$F(1,14)=25.10$, $p<.0001$], Group [$F(1,14)=7.65$, $p<.015$], and a significant Group x Test interaction [$F(3,42)=11.61$, $p<.0001$]. To further investigate these effects, two one-way ANOVAs were conducted. First, a one-way ANOVA was calculated for each test, with Group as factor. As expected, no reliable difference was obtained between the trained and control group at pretest [$F(1,14)=0.15$, $p>.703$]. However, the two groups were significantly different at the post-test [$F(1,14)=10.33$, $p<.006$], Gen 1 [$F(1,14)=10.59$, $p<.006$], and Gen 2 [$F(1,14)=12.25$, $p<.003$]. This indicates that the trained and control subjects' tone identification accuracy was comparable to start with, but their performance was different after training. Second, a one-way ANOVA with Test as factor showed, for the trained group, a significant difference among the four tests [$F(3,28)=13.73$, $p<.0001$]. *Post hoc* comparison (Tukey-HSD) showed that the pretest score was significantly lower than that of either the post-test or Gen 1, or Gen 2. Moreover, there were no significant differences among post-test, Gen 1, and Gen 2. Conversely, for the control group, no reliable difference was found among the four tests [$F(3,28)=0.32$, $p>.812$]. Since no difference was found among the post-test, Gen 1, and Gen 2 for either the trained or the control group, subsequent analyses were conducted using the post-test as the representative of the three tests.

B. Individual tones and tone pairs

The trainees' performance for each individual tone is illustrated in Fig. 3, revealing that identification of each tone improved significantly from the pretest to the post-test: 15% improvement for Tone 1 ([$F(1,14)=5.15$, $p<.006$]); 22% for Tone 2 ([$F(1,14)=7.12$, $p<.001$]); 18% for Tone 3 ([$F(1,14)=2.87$, $p<.05$]); and 25% for Tone 4 ([$F(1,14)=6.20$, $p<.002$]). Interestingly, there was no significant difference among the four tones at either pretest [$F(1,30)=0.73$, $p>.545$], or post-test [$F(1,30)=0.62$, $p>.607$],

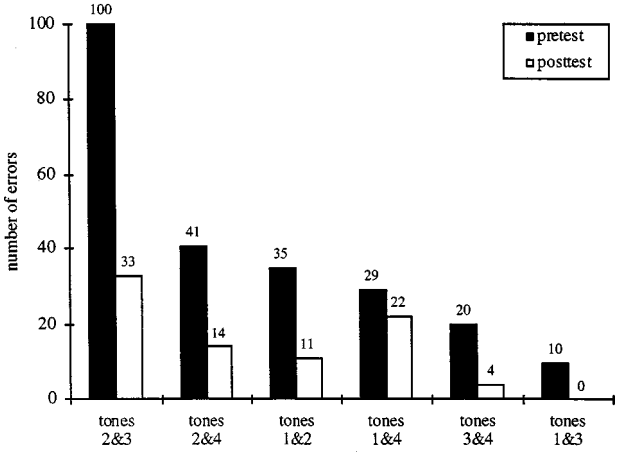

FIG. 4. Tone pair confusions for trained subjects at pretest and post-test. The number of errors (out of 400) for each tone pair refers to misperception of one tone as the other in the corresponding pair.

even though, at pretest, the trainees' identification of Tone 2 and Tone 4 appears poorer as compared to that of Tone 1 and Tone 3.

An analysis of tone confusions is shown in Fig. 4, which compares, for the pretest and the post-test, the number of errors the trainees made for each tone pair out of a total of 400 (25syllables×2tones×8trainees) (see Appendix for complete pretest and post-test confusion matrices). 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. In agreement with the overall data, a comparison of the errors made at the pretest and post-test shows a decrease of errors for each tone pair.

The tone pair confusion analysis demonstrated significant differences among the tone pairs for both tests (pretest: $[F(1,46)=9.70,\ p<.0001]$; post-test: $[F(1,46)=3.81,\ p<.006]$). *Post hoc* analyses reveal that at pretest, the most difficult tone pair was Tones 2 and 3, followed by Tones 2 and 4, Tones 1 and 2, Tones 1 and 4, Tones 3 and 4, and Tones 1 and 3 (as mentioned previously, this provided the rationale for the reversed order of tone pair presentation during training). However, at post-test, tone pair 1 and 4 became the second most confusing pair next to tone pair 2 and 3. Analysis of variance revealed a significant interaction of tone pair and test (pretest, post-test) $[F(3,92)=9.70,\ p<.0001]$. More specifically, while all other tone pairs showed a reliable decrease in errors from the pretest to the post-test, the difference between the two tests for Tones 1 and 4 was not significant $[F(1,14)=0.32,\ p>.577]$. Thus it appears that tone pair 1 and 4 was most resistant to improvement. Nonetheless, the rank order of the tone pairs at pretest and post-test was still highly correlated (Spearman $r=0.83$, $p<.04$), which indicates that the pattern of tone confusion before and after training is to a large extent comparable.

C. Performance during training

The results from the four assessments during training were analyzed as a function of training set and as a function of talker. Trainees' performance from training set 1 to training set 4 was not significantly different $[F(1,30)=0.61,\ p>.617]$. The trainees' scores were already very high after the first training set (88% correct identification), and were maintained in the following three assessments (92%, 88%, and 92%, respectively), revealing little progressive improvement as training went along. The high identification accuracy of the four assessments during training might be attributed to the fact that subjects were only tested on the stimuli that were just used in that training session. In addition, since each test represents a different talker, a progressive improvement may not necessarily be expected.

No reliable difference as a function of talker was observed $[F(1,30)=0.38,\ p>.770]$, nor was there any significant difference between the male and female talkers $[F(1,30)=0.88,\ p>.355]$. Identification scores were 93% and 89% for the two female talkers, and 90% and 88% for the two male talkers.

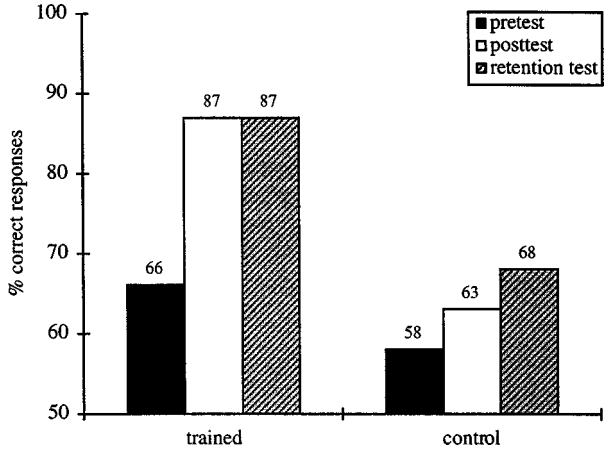

FIG. 5. Mean percent correct identification of the four Mandarin tones for the trained ($n=4$) and the control ($n=4$) subjects at pretest, post-test, and the retention test six months after training.

D. Long-term retention

Figure 5 illustrates the four trained and four control listeners' performance in the retention test as compared to that in the pretest and post-test, revealing that the trainees' improvement was maintained six months after training. The mean identification accuracy for the trainees in the retention test (87%) retains the post-test level (87%), both of which are higher than in the pretest (66%). By contrast, for the controls, the progression from the pretest (58%) to the post-test (63%) and retention test (68%) is much smaller. Detailed analysis of individual listeners revealed that the controls' mean retention score was boosted by the two listeners with three-month Mandarin exposure in Taiwan (32% and 7% improvement from the pretest). Omitting the data from these two listeners would result in a retention test score of 58% for the remaining control subjects, identical to their pretest scores.

A two-way ANOVA of Test (pretest, post-test, retention) and Group (trained, control), with Test as repeated measure, revealed a significant difference in both Test $[F(2,21)=12.44,\ p<.001]$ and Group $[F(1,22)=5.79,\ p<.05]$, but there was no significant Test x Group interaction $[F(2,21)=3.02,\ p>.087]$. More specifically, a one-way ANOVA was conducted for each group with Test as factor. For the trained group, an expected difference was observed for the three tests $[F(2,9)=9.89,\ p<.005]$, with the pretest score significantly lower than the post-test and retention test (Tukey-HSD). Although the controls show a slight progression of the mean scores from the pretest to the post-test and retention test, there was no significant difference among these tests $[F(2,9)=0.50,\ p>.619]$.

E. Individual trainees

Individual trainee and control performance at pretest, post-test, and retention test is summarized in Table II. Each trainee's identification accuracy improved after training (ranging from 6% to 33%), and the improvement was retained. It should also be noted that there is a large degree of variability among the eight trainees' initial levels, which seems to be reflected the extent of the training effects. Thus

TABLE II. Individual listeners' tone identification accuracy (%) at pretest, post-test, and retention test.

	Pretest	Post-test	Improvement	Retention
Trainee				
1	59	88	+29	81
2	62	83	+21	···
3	63	80	+17	76
4	67	87	+20	92
5	67	100	+33	···
6	73	90	+17	···
7	75	93	+18	98
8	89	95	+6	···
Control				
1	42	60	+18	74
2	55	48	−7	58
3	58	58	0	···
4	60	58	−2	58
5	75	82	+7	···
6	75	86	+11	82
7	85	90	+5	···
8	85	77	−8	···

while the listener with a lower initial score (e.g., listener 1: 59%) showed substantial improvement (29%) in the post-test, training effects were much smaller (6%) for the one who started high (e.g., listener 8: 89% at pretest). It appears a bit surprising that listener 5 reached 100% correct identification at post-test, given that her pretest score was comparatively low (67%). However, a closer inspection of her data showed that in her pretest, 90% of the errors was due to misperception of Tone 3 as Tone 2. Since her problem was limited to one tone pair, improvement may have been easier. The retention test shows that for each of the four trainees, the improvement gained from training was maintained after six months. In particular, the training effect does indeed appear robust, given that listeners were not exposed to Mandarin for as long as three months prior to the retention test.

The trainees' self-evaluation of their performance (obtained from debriefing) is summarized in Table III. Consistent with their actual performance, all listeners recognized some degree of improvement after training. Given that many of them did not claim to have other sources of input that specifically influenced their tone perception, their improvement could largely be attributed to the training. However, although some trainees reported a progressive improvement during training, and many of them considered female talkers more intelligible, neither of these assessments was mirrored in the data.

Finally, in connection with the language background information of the trainees (cf. Table I), two other minor observations could be made based on the above individual analyses. First, neither trainee 2 nor trainee 4 was taking a Mandarin course during the time of training, yet their improvement (21% and 20%, respectively) was at the average level (21%), which further demonstrates the robustness of training. Second, two listeners with some experience with another tone language (Cantonese) were involved in the training program (trainee 4 and trainee 7). However, an examination of their overall improvement and tone confusion patterns shows that their performance was comparable with the other ''nontonal'' listeners.

III. DISCUSSION

The present study demonstrated that the perception of Mandarin tones can be improved using a simple training task, indicating that the procedure which has been adopted in training the acquisition of non-native segmental contrasts can also be applied at the suprasegmental level.

The results showed a robust effect of training by a substantial 21% increase in the trainees' overall tone perception accuracy, a significant improvement which also holds true for each of the four tones, and for each individual trainee. More importantly, the improvement gained in training was generalized to new stimuli (18% increase) and new talkers and stimuli (25% increase), and was retained by listeners six months after training (21% increase). These results are comparable to those obtained in the segmental training studies described previously (e.g., Jamieson and Morosan, 1986; Logan *et al.*, 1991; Lively *et al.*, 1994; Bradlow *et al.*, 1997).

Several aspects of tone training warrant discussion related to the general L2 acquisition domain. First, as discussed above, one of the ultimate goals of the acquisition of L2 is the construction of new phonetic categories of the target language. Logan *et al.* (1991) pointed out that the high-variability training procedure facilitates the formation of novel phonetic categories in that stimulus variability exposes learners to the full range of acoustic phonetic cues that char-

TABLE III. Trainees' self-reported performance in the training program.

Trainee	Degree of improvement after training	Progression[a]	Degree of attentiveness in training	More intelligible talker-voice	Other source of tone input
1	moderate	yes	attentive	female	no
2	great	yes	attentive	female	no
3	moderate	yes	occasionally not attentive	female	no
4	great	yes	very attentive	female	no
5	moderate	no	attentive	female	no
6	moderate	not known	attentive	higher voice	self-practice
7	moderate	no	occasionally not attentive	no difference	self-practice
8	moderate	not known	attentive	female	no

[a]Was identification progressively easier from sessions 1 to 8?

acterize those categories, while talker variability enables listeners to overcome idiosyncrasies due to differences in talkers' vocal tract size, glottal source function, and speaking rate.

This training procedure was also adopted in the current study; that is, training stimuli were chosen to represent a variety of phonetic environments, and were produced by a number of talkers of both genders. Acoustic analysis has shown that the $F0$ pattern for a particular tone is subject to change in different vowels (Howie, 1976). Therefore, it is important that various vowels are used in order for the physical stimuli to be mapped onto more abstract phonemic representations. Talker variability is particularly crucial in tone training, since different talkers (especially males and females) have different fundamental frequencies. It has been reported that native Mandarin speakers use changes in $F0$ contours more than height to distinguish among tones, whereas native English listeners tend to attach more importance to height (Gandour, 1983, 1984). Thus by using different talkers, learners are trained to focus on detecting the pitch contour differences of the tones, and to normalize the differences in $F0$ height of various talkers. In addition, since intrinsic duration also differs for the four tones (Lin, 1965), talker variability would enable listeners to normalize differences in speaking rate.

All these measures were taken to enhance the tonal category distinctions for the American trainees. As reviewed previously, English listeners' discrimination and identification of Mandarin tones tend to be less ''categorical'' as compared to Mandarin listeners (Leather, 1987; Stagray and Downs, 1993). Therefore, if training emphasizes those perceptual cues employed by native Chinese to categorize the four tones, the formation of these tonal categories by English learners should be expected. In the present study, the fact that the trainees' identification accuracy increased to a large extent for all the four tones independent of stimuli and talkers, and that the increase had been retained in the trainees' long-term memory, suggests that a separate category for each tone may have been formed and maintained after training.

These results strongly support the previous claim in the segmental domain that adult L2 learners can establish separate phonetic categories for those L2 sounds that are nonexistent in their L1 sound systems (e.g., Flege, 1992). While for native Mandarin speakers tonal pattern is an integral part of the lexicon, such functional association between segmental structure and $F0$ contour does not exist in American learners' phonetic systems. In this sense, forming tonal categories is comparable with forming new segmental categories, which may be effortful, but attainable, for adult L2 learners.

However, since, for American listeners, acquiring the Mandarin tone system involves the integration of $F0$ information at the lexical and sentential level, their knowledge of the function of pitch in the stress and intonation systems of English may be evident as well. In the present study, although the trainees exhibited an increase in the identification of all the four tones, their tone pair confusion patterns showed that these four tones were indeed differentially acquired. Tone pair 1 and 4 was most resistant to improvement and was reported by many trainees as ''confusing.'' These two tones were also found difficult for Americans to acquire by Shen (1989), who proposed that Tone 1 and Tone 4 are prosodically less marked for English listeners than Tone 2 and Tone 3. Similarly, White (1981) found that English listeners perceive Mandarin high tones as stressed, and the low Tone 3 as unstressed. Given these findings, it might be that, in this study, Tone 1 and Tone 4 are most resistant to improvement since they are both comparable to the English unmarked or stressed condition, while the other tone pairs each involve at least one tone that is novel or ''unnatural'' in English. While the initial difficulty in distinguishing Tones 2 and 3 has been attributed to their acoustic similarities (Chen, 1997; Moore and Jongman, 1997), Tones 2 and 3 were improved greatly after training. It might be speculated that since these two tones are so novel to the English listeners, these listeners are more attentive to their distinctions in training. That training can fine-tune distinctions as subtle as tones 2 and 3 may well be due to the novel nature of these two tones to the American listeners.

These findings are consistent with those in the studies of L2 segmental acquisition. For instance, in their study of English vowel acquisition, Bohn and Flege (1992) hypothesized that phonetic learning for similar sounds does not progress much along with L2 experience, whereas new sounds benefit from learning. Likewise, learners are more likely to perceive or produce new, rather than similar, L2 phones authentically (Flege, 1987). Taken together, the present results provide a piece of evidence that the pattern of L2 suprasegmental acquisition might be analogous to that of segmental acquisition, with respect to L1 interference. Although more studies on the comparison of English and Mandarin prosodic patterns are needed to provide a more definite interpretation for the present results, the potential mapping of the patterns of L2 acquisition at segmental and suprasegmental levels is indeed interesting.

IV. CONCLUDING REMARKS

In this study, auditory training at the suprasegmental level was demonstrated to be effective. That is, the perception of Mandarin tones by American learners can be improved with training. The contrasts can be generalized to novel words and talkers, and maintained in long-term memory.

These results raise the question of 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 found that learning gained perceptually can benefit production (Rochet, 1995; Bradlow, 1997), it is worthwhile to test if such transfer will 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 quantitatively judge the trainees' improvement after training. Finally, this study only presented training stimuli in isolation. Given that, more often than not, tones are to be perceived and produced in context, training at the

phrasal or sentential levels should also be involved in future studies. These future studies will allow further investigations of the acquisition of Mandarin tones as well as the interaction of L1 and L2 at a suprasegmental level.

ACKNOWLEDGMENTS

The authors wish to thank Qinhong Anderson and Xiaoxin Sun for recruiting listeners, Yufen Lee Mehta, Ningkang Jiang, Yonghong Mao, Li Tang, Qiuyun Teng, and Dongming Zhang for providing the training stimuli, Eric Evans and Doug Stauffer for technical support, the 16 listeners for their persistent participation, and editor James M. Hillenbrand and three anonymous reviewers for valuable comments and suggestions. Portions of this study were reported at the 135th meeting of the Acoustical Society of America [J. Acoust. Soc. Am. **103**, 3090 (1998)].

APPENDIX

Confusion matrices for the trained group at (a) pretest and (b) post-test (25 stimuli×8 trainees=200 responses for each tone).

(a) Pretest

Perceived as	*Stimulus* Tone 1	Tone 2	Tone 3	Tone 4
Tone 1	**152**	7	4	15
Tone 2	28	**126**	37	37
Tone 3	6	63	**152**	13
Tone 4	14	4	7	**135**

(b) Posttest

Perceived as	*Stimulus* Tone 1	Tone 2	Tone 3	Tone 4
Tone 1	**180**	4	0	9
Tone 2	7	**172**	11	11
Tone 3	0	21	**185**	0
Tone 4	13	3	4	**180**

Best, C. T. (**1995**). ''A direct realist view of cross-language speech perception,'' in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Baltimore), pp. 171–204.

Best, C. T., McRoberts, G. W., and Sithole, N. N. (**1988**). ''Examination of perceptual reorganization for non-native speech contrasts: Zulu click discrimination by English-speaking adults and infants,'' J. Exp. Psychol. **14**, 345–360.

Blicher, D. L., Diehl, R. L., and Cohen, L. B. (**1990**). ''Effects of syllable duration on the perception of the Mandarin tone2/tone3 distinction: Evidence of auditory enhancement,'' J. Phonetics **18**, 37–49.

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

Bohn, O. S., and Flege, J. E. (**1992**). ''The production of new and similar vowels by adult German learners of English,'' Stud. Second Language Acquisition **14**, 131–158.

Bradlow, A. R., Pisoni, D. B., Yamada, R. A., and Tohkura. (**1997**). ''Training Japanese listeners to identify English /r/ and /l/ IV: Some effects of perceptual learning on speech production,'' J. Acoust. Soc. Am. **101**, 2299–2310.

Broselow, E., Hurtig, R. R., and Ringen, C. (**1987**). ''The perception of second language prosody,'' in *Inter-language Phonology, The Acquisition of Second Language Sound System*, edited by G. Ioup and S. H. Weinberger (Newbury House, Cambridge), pp. 350–361.

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

Chen, G. T. (**1974**). ''The pitch range of English and Chinese speakers,'' J. Chinese Linguistics **2**, 159–171.

Chen, Q. (**1997**). ''Toward a sequential approach for tonal error analysis,'' J. Chinese Language Teachers Assoc. **32**, 21–39.

Chuang, C. K., Hiki, S., Sone, T., and Nimura, T. (**1972**). ''The acoustical features and perceptual cues of the four tones of standard colloquial Chinese,'' Proceedings of the Seventh International Congress on Acoustics (Budapest), 297–300.

Flege, J. E. (**1987**). ''The production of 'new' and 'similar' phones in a foreign language: Evidence for the effect of equivalence classification,'' J. Phonetics **15**, 47–65.

Flege, J. E. (**1992**). ''The intelligibility of English vowels spoken by British and Dutch talkers,'' in Intelligibility in Speech Disorders: Theory, Measurement and Management, edited by R. D. Kent (John Benjamins, Amsterdam).

Flege, J. E. (**1995**). ''Second language speech learning: theory, findings, and problems,'' in *Speech Perception and Linguistic Experience*, edited by W. Strange (York, Baltimore), pp. 233–273.

Gandour, J. T. (**1983**). ''Tone perception in Far Eastern languages,'' J. Phonetics **11**, 149–175.

Gandour, J. T. (**1984**). ''Tone dissimilarity judgments by Chinese listeners,'' J. Chinese Linguistics **12**, 235–261.

Gårding, E., Kratochvil, P., Svantesson, J. O., and Zhang, J. (**1986**). ''Tone 4 and Tone 3 discrimination in modern standard Chinese,'' Language and Speech **29**, 281–293.

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

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

Jamieson, D. G., and Morosan, D. E. (**1989**). ''Training new, nonnative speech contrasts: A comparison of the prototype and perceptual fading techniques,'' Can. J. Psychol. **43**, 88–96.

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

Leather, J. (**1987**). ''*F*0 pattern inference in the perceptual acquisition of second language tone,'' in *Sound Patterns in Second Language Acquisition*, edited by A. James and J. Leather (Foris, Dordrecht), pp. 59–81.

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

Lenneberg, E. (**1967**). *Biological Foundations of Language* (Wiley, New York).

Lin, M. C. (**1965**). ''The pitch indicator and the pitch characteristics of tones in Standard Chinese,'' Acta Acust. (China) **2**, 8–15.

Lively, S. E., Logan, J. S., and Pisoni, D. B. (**1993**). ''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**, 1242–1255.

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y., and Yamada, T. (**1994**). ''Training Japanese listeners to identify English /r/ and /l/ III: Long-term retention of new phonetic categories,'' J. Acoust. Soc. Am. **96**, 2076–2087.

Logan, J. S., Lively, S. E., and Pisoni, D. B. (**1991**). ''Training Japanese listeners to identify English /r/ and /l/: A first report,'' J. Acoust. Soc. Am. **89**, 874–886.

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

Massaro, D. W., Cohen, M. M., and Tseng, C. (**1985**). ''The evaluation and integration of pitch height and pitch contour in lexical tone perception in Mandarin Chinese,'' J. Chinese Linguistics **13**, 267–290.

Mertus, J. (**1989**). *BLISS Manual* (Brown University, Providence).

Moore, C. B., and Jongman, A. (**1997**). ''Speaker normalization in the perception of Mandarin Chinese Tones,'' J. Acoust. Soc. Am. **102**, 1864–1877.

McClaskey, C. L., Pisoni, D. B., and Carrell, T. D. (**1983**). ''Transfer of training of a new linguistic contrast in voicing,'' Percept. Psychophys. **34**, 323–330.

Pisoni, D. B., Aslin, R. N., Perey, A. J., and Hennessy, B. L. (**1982**). ''Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants,'' J. Exp. Psychol. **8**, 297–314.

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

Shen, X. S. (**1989**). ''Toward a register approach in teaching Mandarin tones,'' J. Chinese Language Teachers Assoc. **24**, 27–47.

Shen, X. S., and Lin, M. C. (**1991**). ''A perceptual study of Mandarin tones 2 and 3,'' Language and Speech **34**, 145–156.

Stagray, J. R., and Downs, D. (**1993**). ''Differential sensitivity for frequency among speakers of a tone and a nontone language,'' J. Chinese Linguist. **21**, 143–163.

Strange, W., and Dittmann, S. (**1984**). ''Effects of discrimination training on the perception of /r-l/ Japanese adults learning English,'' Percept. Psychophys. **36**, 131–145.

White, C. M. (**1981**). ''Tonal perception errors and interference from English intonation,'' J. Chinese Language Teachers Assoc. **16**, 27–56.