Available online at www.sciencedirect.com
ScienceDirect
Journal of Phonetics 36 (2008) 268–294
Journal of Phonetics
www.elsevier.com/locate/phonetics

Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers

Alexander L. Francis[a,b,*], Valter Ciocca[a], Lian Ma[a,c], Kimberly Fenn[d]

[a]*Division of Speech and Hearing Sciences, Prince Philip Dental Hospital, University of Hong Kong, Hong Kong SAR, China*
[b]*Department of Speech, Language & Hearing Sciences, Purdue University, Heavilon Hall, 500 Oval Drive, West Lafayette, IN 47907, USA*
[c]*CLP Centre, School of Stomatology, Beijing Medical University, Hai Dian, wei Gong Cun, Beijing 10081, People's Republic of China*
[d]*Department of Psychology, University of Chicago, 5848 S. University Avenue, Chicago, IL 60637, USA*

Received 8 November 2005; received in revised form 6 June 2007; accepted 8 June 2007

Abstract

Two groups of listeners, one of native speakers of a tone language (Mandarin Chinese) and one of native speakers of a non-tone language (English) were trained to recognize Cantonese lexical tones. Performance before and after training was measured using closed response-set identification and pairwise difference rating tasks. Difference ratings were submitted to multidimensional scaling (MDS) analyses to investigate training-related changes in listeners' perceptual space. Both groups showed comparable initial performance and significant improvement in tone identification following training. However, the two groups differed in terms of the tones they found most difficult to identify, and in terms of the tones that were learned best. Differences between the two groups' training-induced changes in identification (confusions) and perceptual spaces demonstrated that listeners' native language experience with intonational as well as tone categories affects the perception and acquisition of non-native suprasegmental categories.

1. Introduction

When learning a foreign language, listeners must learn to redistribute attention, directing more attention to previously ignored acoustic patterns (Francis & Nusbaum, 2002; Guion & Pederson, 2007) and/or ignoring properties that function in the native language but not in the new one (Yamada & Tohkura, 1992). Previous research has shown that both linguistic experience and perceptual training can change the way listeners attend to the acoustic patterns of speech (Francis, Baldwin, & Nusbaum, 2000; Francis & Nusbaum, 2002; Guion & Pederson, 2007; McCandliss, Fiez, Protopapas, Conway, & McClelland, 2002; Pisoni, Lively, & Logan, 1994). However, until recently, the vast majority of research on phonetic learning has been carried out on features of segmental (vowel and consonant) categories (though cf. Mennen (2004) for a cross-linguistic study of learning intonation). In the present article we extend this research into the domain of lexical tones.

*Corresponding author. Department of Speech, Language & Hearing Sciences, Purdue University, Heavilon Hall, 500 Oval Drive, West Lafayette, IN 47906, USA. Tel.: +1 765 494 3815.
E-mail addresses: francisa@purdue.edu (A.L. Francis), vciocca@hkusua.hku.hk (V. Ciocca), lmaa@hkucc.hku.hk (L. Ma), kmfenn@uchicago.edu (K. Fenn).

0095-4470/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.wocn.2007.06.005

In a tone language, meaningful lexical differences can be indicated simply by changing the fundamental frequency (f0) pattern of a given syllable. For example, in Cantonese the syllable [ji] means "cure" when produced with a 55 (High Level) tone, but when produced with a 21 (Low Falling) tone it means "son".[1] Although other acoustic properties can also function as perceptual cues to tone identity, it has been shown that, in Mandarin, non-frequency-related features, including amplitude envelope contour and duration and other vocal quality properties such as glottalization, typically function as secondary cues and listeners generally base their tone category decisions primarily on f0 contours when these are available (Fu, Zeng, Shannon, & Soli, 1998; Liu & Samuel, 2004; Whalen & Xu, 1992). In Cantonese, no non-f0-related properties have yet been shown to correlate consistently with tone identity, or to be used consistently by listeners even in the absence of f0 information (Ciocca, Francis, Aisha, & Wong, 2002; Fok Chan, 1974; Vance, 1976).

Recently, a number of studies have begun to examine tone perception by both tone and non-tone language speakers (e.g. Gandour, Dzemidzic, et al., 2003; Hallé, Chang, & Best, 2004; Krishnan, Xu, Gandour, & Cariani, 2005; Wang, Behne, Jongman, & Sereno, 2004; Wong, 2002; Xu, Gandour, & Francis, 2006). In general, results suggest that speakers of non-tone languages process speech f0 differently, both behaviorally and neurologically, than do speakers of tone languages. Such differences may give rise to the well-known difficulty experienced by adult speakers of non-tone languages when attempting to learn an unfamiliar tone language (cf. Wang, Spence, Jongman, & Sereno, 1999).

Two general kinds of theories have been proposed to account for this observed difficulty. According to what might be termed a "levels of representation" account, speakers of non-tone languages are simply unable to relate lexical tones to familiar (native) linguistic categories because there is nothing in their native grammar that prepares them for using prosodic properties such as f0 in a lexically contrastive manner. Thus, although speakers of a language such as English have categories defined by f0 characteristics, these categories are intonational, rather than lexical, and it is this difference that creates the difficulty. On the other hand, a "category assimilation" account might attribute this difficulty to difficulty in mapping foreign tone categories onto native ones, whether intonational or lexical. According to this second kind of theory, which is compatible with, if not explicitly derived from, current models of prosodic phonology that treat intonational categories as comparable to segmental ones (see e.g. Ladd, 1992, 1996), non-tone language speakers will process foreign lexical tones with reference to their native intonational categories, just as tone language speakers will process foreign tones with reference to their native tone categories. The increased difficulty experienced by non-tone language speakers derives either (or both) from a greater degree of mismatch between the native intonational and foreign tone categories as compared to that between native and foreign tone categories, and (perhaps) from a weaker (less categorical) mental representation of native intonational categories as compared to that of native tone categories.

The discussion presented by Wayland and Guion (2004) represents a good example of the first (levels of representation) account. They found that native Chinese (Mandarin and Taiwanese) speakers were better at discriminating a Thai tone contrast (mid- vs. low-tone) than were native English speakers, both before and after training, although both groups still performed worse than did native Thai speakers (who also showed improvement at discriminating this difficult contrast after training). Moreover, the native English group showed no significant improvement even after a week (five 30-min sessions) of training. Wayland and Guion (2004) argue that these results suggest either that native Chinese listeners were able to transfer or extended a native-language (L1)-based propensity for tracking f0 to the perception of a new language while English

[1]For the purposes of comparing tonal patterns across languages, in this article we will refer to individual tones using a tone number system based on that of Chao (1947). In this system tones are assigned two or three numerical identifiers indicating the relative pitch height on a scale from 1 (bottom of the talker's canonical pitch range) to 5 (top of the talker's canonical pitch range) at the beginning (first number) and end (last number) of the syllable carrying the tone. In cases such as the Mandarin Tone 3 (dipping tone) a third number indicating the relative pitch around the middle of the syllable is also included (e.g. 214). For ease of interpretation we will also refer to each tone using the descriptive terminology of linguistic research in the language from which the particular tone is drawn (e.g. we will say "the Mandarin 25 tone (Tone 2) …" or "the Cantonese 23 (Low Rising) tone …" Note that these numerical descriptors, while originally based on impressionistic perceptual analyses of tone productions, accord very well with instrumental acoustic analyses of syllables produced in isolation (Bauer & Benedict, 1997). Thus, they can be used as a terminological shorthand referring to the *phonetic* properties of each syllable (equivalent to the IPA 'tone letters', IPA 1999) without necessarily requiring the adoption of any specific phonological model of lexical tones (i.e. their use here is neutral with respect to any debate over the precise number of tone levels that must be represented in a phonology).

speakers could not, or that native Chinese listeners were able to map the non-native tones onto (different) native Chinese (phonological) tone categories while English speakers could not. In either case, the assumption is that English speakers are lacking some fundamental capability or representation that Chinese speakers possess and that facilitates the perception and acquisition of non-native lexical tone categories.

In contrast, the work of Hallé et al. (2004), insofar as it adopts Best's (1995) perceptual assimilation model (PAM), appears to represent the second (category assimilation) approach, although their specific conclusions in fact argue for little or no assimilatory effects in the specific case examined. Hallé et al. (2004, experiments 1 and 3) tested French and Taiwanese listeners on discrimination of three Mandarin lexical tone continua, and found that French listeners were uniformly sensitive to f0 differences across the continua, while Taiwanese listeners were more sensitive to differences between tokens belonging to different tone categories than they were to similar acoustic differences between same-category pairs. They argue that these results suggest that Taiwanese listeners' perception was influenced by their native tone categories, while French listeners were unable to treat the Mandarin tones as "basic prosodic units bearing contrastive linguistic significance" (i.e. phonological categories) despite being able to process them as "prosodic aspects of speech" (p. 417). While this characterization evokes a "levels of representation" model (i.e. French speakers do not perceive Mandarin tones correctly because they have no native tone categories to assimilate them to), Hallé et al. (2004) also state that "Tone contours thus are not completely irrelevant to a French ear with respect to their putative linguistic value" because "the acoustic correlates of tones, f0 and intensity contour, are used in French, just as in any language, at the sentential intonation level." Thus, French listeners must have been primarily influenced by "the perceived salience of the phonetic (i.e., intonational) differences involved" which "might be non-language-specific in this case, or they could be evocative of language-specific intonation patterns" (pp. 417–418). In other words, according to Hallé et al. (2004), French speakers were unable to perceive Mandarin tones in a categorical manner not because they were unable to process *tone* per se, but rather merely because they were unable to clearly map the Mandarin tones onto any particular native French phonological (intonational) categories.

A possible third approach may be found in the work of Wang et al. (2004), who compared hemispheric lateralization of Mandarin lexical tones by Mandarin, English and Norwegian listeners. They found that, even though Norwegian listeners were familiar with lexical tones from their native language, these listeners, like the non-tone language speaking English listeners, showed no hemispheric asymmetry for Mandarin tones, although Mandarin and Mandarin–English bilingual subjects did. They interpret these results as suggesting that hemispheric lateralization depends on familiarity with the specific acoustic properties of the stimuli. Norwegian listeners, although familiar with the use of pitch as a cue to lexical tone, and although they did possess native tone category representations, were not familiar with the specific acoustic features that distinguish Mandarin tones, and therefore processed them bilaterally just as English listeners did. This argument suggests that it is not the native tone or intonational categories that play the pivotal role in influencing cross-language perception of tones, but rather that what may matter most in cross-tone-language perception of tone categories is the degree to which the acoustic *features* used to define tones in the native language correspond to those used to define tones in the foreign language.

Such a feature-based perspective is supported by research on the cross-language perception of f0 patterns by both tone and non-tone language speakers. Gandour and colleagues have shown that the basic perceptual dimensions of lexical tone space tend to be the same across tone and non-tone languages, but that the relative weighting of each dimension varies across languages (Gandour, 1983; Gandour & Harshman, 1978). These findings support the idea that native intonational as well as tone representations influence the perception of the same acoustic features: f0 height and direction of change. For example, Gandour (1983; see also Guion & Pederson, 2007) showed that Mandarin, Taiwanese, Thai, and English speakers' perception of f0 contours can be characterized adequately within the same two-dimensional space defined by height (average f0) and direction of change of f0 (level, rising or falling) across the syllable. However, English and Cantonese speakers give more weight to the height dimension than do speakers from the other three languages, while Cantonese and Mandarin speakers give more weight to the direction dimension than do English speakers. These patterns (at least with respect to Mandarin, English and Cantonese listeners) are presumably a consequence of the need to optimize perception of f0 patterns for distinguishing frequency-based categories irrespective of whether these categories are tone (Cantonese, Mandarin) or intonational (English).

1.1. Tone inventories of Mandarin, Cantonese and English

Standard (Beijing) Mandarin is typically described as having four tone categories: a High (55) tone (Tone 1), a Rising (25) tone (Tone 2), a Dipping (214) tone (Tone 3), and a Falling (51) tone (Tone 4), as shown in Fig. 1 (Li & Thompson, 1989; Norman, 1988). These f0 patterns are generally maintained in fluent speech, with a few exceptions known as tone sandhi in which the f0 contour of one syllable will vary depending on the tones following (and possibly preceding) it. In most cases, Mandarin tone sandhi changes one of these f0 contours into another one in the list (a phonological change). For example, when a syllable with a Dipping tone (214) is directly followed by another syllable with the same tone, the first syllable will be produced as a rising tone (25) instead. However, one sandhi rule does result in the production of a fifth f0 contour: When a syllable with a Dipping tone is directly followed by any tone other than another Dipping tone, the first syllable will be produced with a *low falling* (21) tone, similar to the Low Falling tone of Cantonese (Li & Thompson, 1989). Finally, some syllables receive a lexically determined "weak stress" in connected speech, resulting in a reduced f0 contour strongly influenced by the tone of the preceding (fully stressed) syllable but most plausibly described as having a mid-level f0 target (Chen & Xu, 2006).

Other, non-f0, features have also been shown to function as cues to Mandarin tones in the absence of identifiable f0 cues, including the shape of the amplitude envelope and syllable duration (Fu et al., 1998; Liu & Samuel, 2004; Whalen & Xu, 1992). Furthermore, although the Mandarin fourth (High Falling) tone is frequently reported to be accompanied by a glottalized (vocal fry) voice quality (Liu & Samuel, 2004), we are not aware of any studies demonstrating that this property functions as an acoustic cue to this lexical tone in the same way that f0, duration and amplitude envelope have been shown to function as cues in the absence of fundamental frequency information. Even still, all of these properties are clearly subordinate to f0 cues in the perception of natural, unmodified speech, and, individuals with poor pitch perception (e.g. cochlear implant users) typically exhibit poor perception of Mandarin tones (Wei, Cao, & Zeng, 2004) though some individuals may perform quite well (Peng, Tomblin, Cheung, Lin, & Wang, 2004).

In Hong Kong Cantonese there are six contrastive tones (Bauer & Benedict, 1997; Fok Chan, 1974): High Level (HL): 55; High Rising (HR): 25; Mid-Level (ML): 33; Low Rising (LR): 23; Low Level (LL): 22; Low Falling (LF): 21. Representative f0 contours from an adult male native speaker are shown in Fig. 2. With respect to non-f0 cues, the Low Falling tone is commonly produced with some degree of glottalization, but it

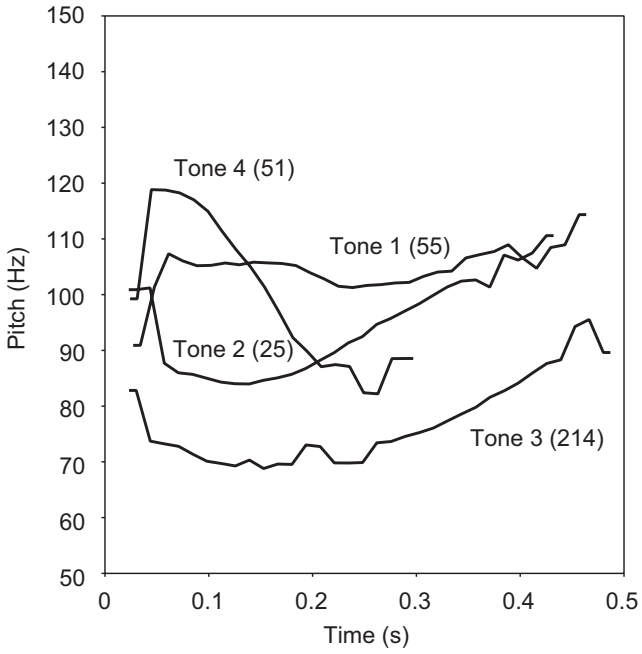

Fig. 1. Fundamental frequency contours for representative Mandarin tones produced on the syllable [da] by an adult male native speaker of the Beijing dialect.

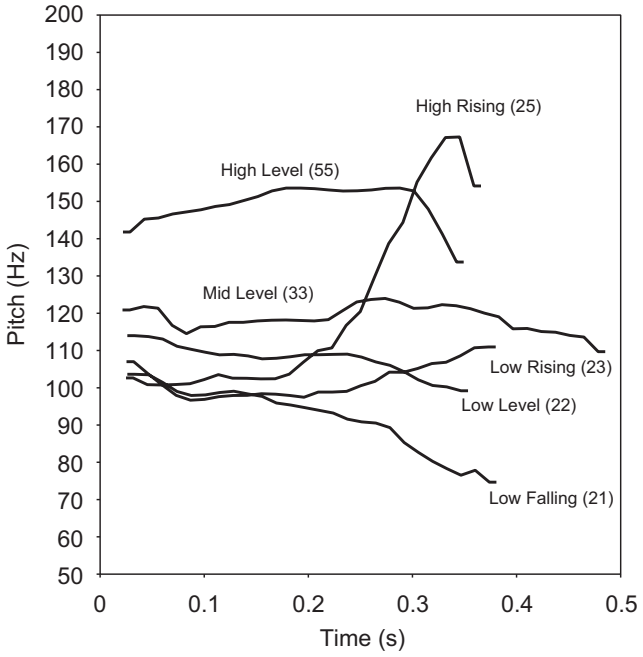

Fig. 2. Fundamental frequency contours for representative Cantonese tones produced on the syllable [ji] by an adult male native speaker from Hong Kong.

has been shown that this property does *not* function as a consistent cue for Cantonese listeners (Vance, 1976). No other non-f0 cues have been proposed for Cantonese, and, as with Mandarin, Cantonese-speaking cochlear implant users have considerable difficulty identifying Cantonese tones, suggesting that f0 patterns serve as the primary, and perhaps sole, cues to lexical tones in Cantonese (Ciocca et al., 2002; Lee, van Hasselt, Chiu, & Cheung, 2002).

Characterizing the English tone inventory is less straightforward, and there are, historically, a number of different characterizations. There is general agreement concerning the existence of tone patterns specified on prominent syllables and other aspects of tone patterns specified at final edges (Cruttenden, 1997; Ladd, 1996). For the sake of clarity, we adopt the Tones and Breaks Indices (ToBI) transcription system (Beckman & Hirschberg, 1994; Beckman, Hirschberg, & Shattuck-Hufnagel, 2005; Pierrehumbert, 1980) because of its familiarity, though our purpose is not to argue for or against this particular system. In this spirit, it may be observed that research on the interpretation of ToBI-defined intonational patterns (Pierrehumbert & Hirschberg, 1990) suggests that the intonational system of (American) English should be able to distinguish between about 22 different simple contours. This number is derived by identifying all possible combinations of the six pitch accents, two phrase accents, and two boundary tones defined in the ToBI system, and eliminating combinations that result in acoustically indistinguishable contours (see Ladd, 1992, pp. 81–82).[2] While Pierrehumbert and Hirschberg (1990) do not seem to take a strong position on whether every one of these possible contours can actually serve to indicate categorically distinct meanings as a phoneme should, their discussion of the compositionality of tune meaning suggests that they assume that this will prove to be the case. Moreover, Ladd (1996, Table 3.1) shows that these 22 legal intonational patterns of American English correspond surprisingly well with patterns proposed in other (e.g. British) intonational frameworks. Note also that even researchers that explicitly reject this approach still tend to focus on a relatively small number of tones with relatively distinct meanings (e.g. Cruttenden's (1997, pp. 50–54) derivation of seven nuclear tones that "suffice for the usual level of delicacy that is required"). Indeed, for our purposes, the precise number of intonational categories, or even their specific characterization in terms of an autosegmental vs.

[2]The specific numbers of accents and tones are debated. For example, Herman and McGory (2002) explore the perceptual similarity of nine types of pitch accent. However, the validity of the basic observation remains: The set of possible intonational contours is finite and relatively small, comparable in order to the number of segmental phonemes in many languages.

non-autosegmental model, is not an issue. For present purposes, the important conclusion to draw from this research is what Ladd (1996) calls the "linguist's theory of intonational meaning" (pp. 39–40), namely that intonational patterns, however they are defined, are associated with specific meanings (see discussions of this approach in Ladd, 1996, Chap. 3, especially 3.4, and a similar discussion from a variety of theoretical perspectives in Cruttenden, 1997, Section 4.4), and thus may be represented as part of the listener's phonological system in a more or less categorical manner comparable to that of segmental phonemes.

Research by Ward and Hirschberg (1985) strongly supports this hypothesis. They showed that, in American English, the fall–rise contour clearly expresses talker uncertainty, and subsequent research suggests that this intonational pattern, at least, may be considered to be perceived categorically (Pierrehumbert & Steele, 1989). However, the categorical status of other contours is more heavily debated (see Ladd, 1996, Chap. 3 for discussion). A recent investigation of cross-language perception of intonation found evidence supporting a distinction only between two main contours, one falling and one rising, with some degree of (complex) variation within the two classes (Grabe, Rosner, García-Albea, & Zhou, 2003).

Despite the current uncertainty regarding the categorical status of specific English intonational contours, consideration of the set of 22 possible contours described by Pierrehumbert and Hirschberg (1990) (or their British-style equivalents provided by Ladd, 1996, Table 3.1) and their putative meanings provides a plausible basis for characterizing English listeners' native language experience with f0-based linguistic distinctions. For example, it seems quite reasonable to assume that English speakers should be familiar with making intonational distinctions between a f0 rise (e.g. standard yes–no question contour) and fall (e.g. neutral declarative intonation). They may also be adept at distinguishing the relative endpoint of rising contours: Compare the high-rise (implied) question inviting the listener to agree (H*HH%), as in "*I thought it was good…*" (implying, *what do you think*?) with standard yes-know question intonation (L*HH%) as in "*Did you think it was good?*" (example from Pierrehumbert and Hirschberg (1990), see pp. 290–291 for discussion). Finally, the basic structure of the ToBI system, and indeed, all autosegmental-metrical approaches to American English intonation, in which intonational contours are considered to be derived from sequences of fundamentally low (L) and high (H) tones (cf. Cruttenden, 1997, Chap. 3; Ladd, 1996, Chap. 3), also suggests that English speakers may be able to distinguish more or less categorically between a relatively low tone (L) and a relatively higher one (H).

1.2. Dimensions of perceptual contrast

1.2.1. Initial (cross-language) perception

Having established the existence of prosodic categories in both Mandarin and English, it is possible to use the similarity of L2 tokens to L1 categories to make some tentative predictions regarding the perception of Cantonese by Mandarin and English speakers from the perspectives of a model of category assimilation (e.g. PAM, Best, 1995). Mandarin listeners should be good at identifying the Cantonese 55 (High Level) and 25 (High Rising) tones because they are acoustically virtually indistinguishable from, and thus would assimilate quite well to, the Mandarin 55 (Tone 1) and 25 (Tone 2) tones, respectively. It is also possible that the Cantonese 23 (Low Rising) tone might assimilate to the Mandarin 214 (Tone 3) tone, but here the mapping is not quite as straight forward and the predictions of different category-based theories might differ. For example, Flege's (1995) speech learning model (SLM) considers that a native category that is similar but not entirely identical to a non-native one might actually interfere with the perception (and certainly the acquisition) of a native one. Thus, the SLM might predict great difficulty for Mandarin speakers' perception of the 23 tone due to interference from the similar, but not identical, native 214 tone while PAM might lead one to expect this to be treated as a kind of *category goodness* assimilation, with the 23 tone being perceived as a poor (but identifiable) realization of the native 214 tone. On the other hand, depending on the degree of perceived difference between the 23 and 214 tones, PAM could also predict that the 23 tone might be treated as an *uncategorizable* phoneme, and thus one that would be easy to distinguish but very difficult to learn to identify. Performance on the other three tones would also be expected to depend on the relative degree of perceived similarity, with the 33 (Mid-Level) tone possibly also mapping moderately well onto the 214 tone but the 22 (Low Level) and 21 (Low Falling) tone mapping poorly if at all to any of the canonical Mandarin tone categories. Ultimately, because these categories are defined in terms of multiple acoustic features (at least,

average f0 *height* as well as *direction* of f0 change), the issue of how close the foreign category might be to the native one will necessarily depend on the relative weighting of each dimension in the native as compared to the target language, already suggesting that a feature weighting-based model of cross-language perception may have more utility in this domain.

Following a category assimilation model, English listeners should be able to identify the 25 (High Rising) and 23 (Low Rising) tones based on their similarity to English question intonation patterns and final continuation rises (H* HH% and L*LH%), although they might also confuse the two based on their auditory similarity, if the acoustic realization of the Cantonese tone contrast does not correspond sufficiently to that of the English one. They might also be able to identify the 21 (Low Falling) tone on the basis of its similarity to a number of English falling intonational contours.

From a feature-weighting perspective, the results presented by Gandour and colleagues (Gandour, 1983; Gandour & Harshman, 1978) and Guion and Pederson (2007) suggest that Mandarin listeners should be sensitive to both the direction of f0 change and relative (average) f0 height of a syllable, with more weight being given to direction than to height. Thus, with respect to confusions, we might expect Mandarin listeners to be prone to confusing the 23 and 25 (Low and High Rising) tones because these contrast more in terms of height than direction. On the other hand, the 21 and 23 tones (Low Falling and Low Rising) may be less easily confused because they differ both in terms of direction of f0 change and in average f0.

For English listeners, Gandour (1983), Gandour and Harshman (1978) and Guion and Pederson (2007) showed that relative f0 height was considerably more important than direction of f0 change. Therefore, we might expect English listeners to perform well at distinguishing the three Cantonese level tones (22, 33, and 55). Indeed, English listeners' performance on Cantonese level tone perception might even be comparable to that of native Cantonese listeners, because direction of f0 change also plays a relatively weak role in Cantonese listeners' tone perception compared to f0 height (Gandour and Harshman (1978), though see Gandour (1981) for somewhat different results for identification of specifically Cantonese tones). On the other hand, Gandour and Harshman's (1978) observation that English listeners gave less weight to the dimension of direction of f0 change than even Cantonese listeners might suggest that English listeners should have considerable difficulty distinguishing between, e.g. the 21 and 23 (Low Falling and Low Rising) tones which are relatively similar in average f0 but differ significantly in terms of direction of f0 change. Of course, this distinction could in principle still be accomplished on the basis of average f0 across the syllable, in which case English listeners should once again have little trouble with it.

1.2.2. Perceptual learning

A major advantage of a feature-weighting perspective on cross-language perception is that it allows more specific predictions to be made regarding acquisition. In particular, the relative weight given to both dimensions of contrast, height and direction, by successful learners in both the English and Mandarin groups should become more similar to the relative weighting shown by native Cantonese listeners. Based on the differences between the three language groups' *a priori* dimensional weights as identified by Gandour (1983), English listeners should learn to give more weight to the direction dimension (improving their recognition of contour tones) while Mandarin listeners should learn to give more weight to the height dimension (improving their recognition of level tones). In support of this hypothesis, Guion and Pederson (2007) found that native English speakers with significant (adult) Mandarin experience showed a more Mandarin-like pattern of perceptual weighting of f0 slope as compared to native English speakers with no tone language experience.

1.3. Summary

The present experiment was designed to compare the perceptual learning of non-native tone categories by native speakers of a tone language (Mandarin Chinese) and a non-tone language (American English). The goal was to explore the degree to which patterns of cross-language tone perception and perceptual learning of tones may be understood in terms of the influence of native prosodic categories and/or biases in perceptual weighting of specific prosodic acoustic features resulting from listeners' prior experience with f0-based linguistic contrasts (both tone and intonational). Thus, we focus not on differences in overall performance, as might be the case in a study of training methods, but rather on differences in the *pattern* of performance

between groups demonstrating comparable degrees of overall learning. That is, instead of simply evaluating the two groups' proportion correct tone identification across all trials, we compared each groups' performance on individual tones as well as the pattern of their *incorrect* responses, in an effort to determine how they were treating the separate acoustic properties that define tone (and intonational) categories before and after training.

2. Methods

2.1. Subjects

Ten native speakers of Mandarin Chinese (five men and five women), 10 native speakers of American English (five men and five women), and 12 native speakers of Cantonese (four men and eight women) participated in this experiment. None of the Mandarin speakers had any familiarity with Cantonese or other tone language (by self-report), and none of the English speakers had any familiarity or exposure to Cantonese or any other tone language (also by self-report). Mandarin participants were all students at the Beijing Medical University (now Beijing University Health Science Center), English participants were undergraduate and graduate students at the University of Chicago and Cantonese participants were students at the University of Hong Kong. The posttest data from one Mandarin-speaking participant (male) were lost due to a computer error, so only results from the remaining nine Mandarin participants were analyzed.[3]

2.2. Stimuli

Stimuli consisted of seven semantically neutral Cantonese sentences (e.g. "I will say—for you to hear.", six target syllables ([fan], [fu], [jau], [ji], [se], [si]) each carrying each of the six Cantonese tones: 21 (Low Falling), 22 (Low Level), 23 (Low Rising), 33 (Mid-Level), 25 (High Rising), and 55 (High Level) (see Appendix A for stimuli). Thus, there were a total of 252 sentences (seven contexts times six syllables times six tones), each recorded twice by six college-aged talkers (three men and three women). For testing a total of 12 tokens were used, consisting of one frame sentence (#3) combined with one syllable ([ji]) carrying each of the six possible tones syllables and produced twice by the third male talker.

For training, sentences 1, 2, 4, 5 and 6 were used, each with all of the other syllables ([fan], [fu], [jau], [se] and [si]) carrying each of the six possible tones and spoken twice each by all the talkers except for the test talker (1500 utterances total). Listeners did not hear the test sentence or target syllable in training, and none of the training stimuli were produced by the (male) test talker. Thus, any change observed from pretest to posttest would be indicative of generalized learning.

A set of synthetic stimuli were also generated for examining changes in listeners' perceptual space as a result of training. These consisted of 22 Klatt-synthesized [ji] syllables with varying f0 contours. Six of the contours were based on canonical (citation form) recordings of the test talker's productions of the six Cantonese tones. An additional set of eight variants of two of these canonical tones (23 (Low Rising) and 33 (Mid-Level)) were also generated. However, results from these 16 tokens will not be discussed here. Listeners heard each pairwise combination of these 22 stimuli, for a total of 484 pairs.

2.3. Procedure

Mandarin and English speaking listeners completed a pretest, training phase and posttest. Cantonese listeners completed only the pretest. On the pretest and posttest, listeners completed both an identification and a difference rating task. In the identification task listeners heard a total of 66 trials made up of 11 blocks containing each of the six test stimuli in random order. On each trial listeners heard a single sentence and were then asked to identify the target word by clicking a mouse button on the appropriate on-screen symbol.

[3]One of the remaining Mandarin speakers was also a speaker of Uyghur (a Turkic, non-tonal language of Central Asia), but reported using Mandarin nearly 100% of the time as an adult and was judged as speaking and understanding Mandarin like a native by report of her peers.

Presentation of the sentence was accompanied by an on-screen presentation of the traditional Chinese characters for the sentence, an English gloss and Romanization/transcription of the sentence and an English translation of the sentence (see Fig. 3a). Response choices were indicated in traditional Chinese characters as well as their English gloss, traditional tone number (1–6) and a schematic diagram of their canonical f0 contour (see Fig. 3b). Listeners were asked to respond quickly and accurately, but there was no upper time limit on their response and response times were not recorded.

In the difference rating task, listeners were presented with a single pair of sounds separated by a 250 ms inter-stimulus interval (ISI). Pairs were presented in random order, and each pair, including same-token pairs (e.g. 22–22, Low Level followed by Low Level) was presented once, for a total of 484 difference rating trials on the pretest, and 484 on the posttest. For the different-token pairs there was one presentation of each order (e.g. pair 21–55, Low Falling followed by High Level, was presented once, as was pair 55–21). After hearing a given pair, listeners were asked to rate the degree of similarity of the two sounds on a scale of 1–10, where 1 indicated perfectly similar (identical) and 10 indicated "as different as possible."

The pretest was carried out over 2 days with the difference rating task on day 1 and the identification task on day 2. The posttest followed the reverse format, with the identification task on day 1 and the difference rating task on day 2.

The training phase of the experiment took a total of approximately 10 h over the course of 10 days (1 h per day on consecutive days) for the Mandarin listeners, and over the course of between 16 and 30 days for the English listeners (also 1 h per day, on consecutive days whenever possible, except for the intervention of weekends and, for some participants, temporary emergency closure of the university). Listeners always completed at least one training session on the day of or immediately preceding the first day of the posttest. The structure of each trial in the training task was identical to that of the identification task on the pretest and posttest, except that listeners were provided with feedback and auditory reinforcement (repetition) of the stimulus sentence. Thus, whether or not a listener identified the target syllable correctly, they were told whether or not they were correct, the correct response was indicted by an on-screen arrow, and they heard the

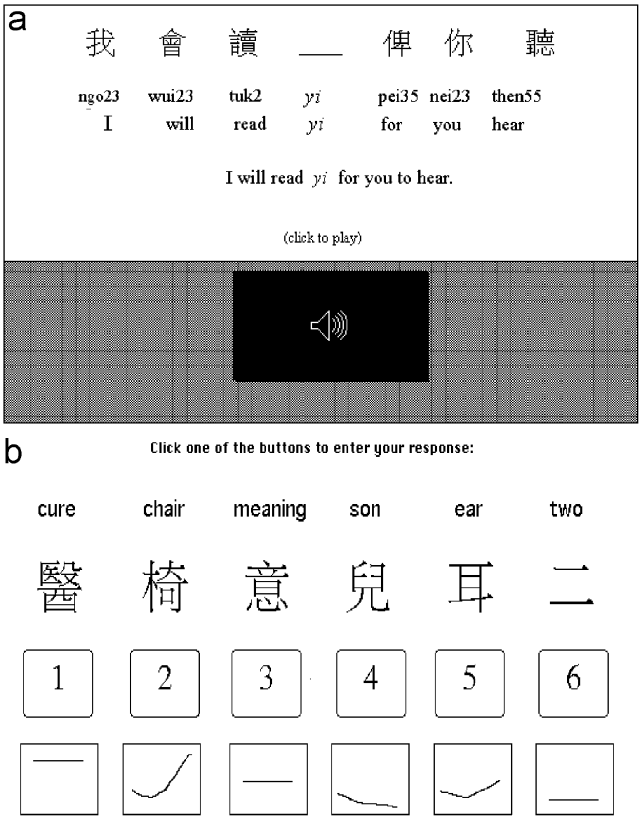

Fig. 3. Stimulus presentation (a) and response collection (b) screens.

stimulus again. Presentation of training stimuli was blocked by target syllable within sentences, and presented in a quasi-randomized order (all trainees heard the same order of presentation). For example, in the first training session listeners heard context sentence 2 containing the target syllable [fan], sentence 4 with the target syllable [se] and sentence 5 with the target syllable [jau], each with all six possible tones as produced by all five talkers (two repetitions per talker), for a total of 180 trials. Mandarin subjects were able to complete 180 trials per session over the course of either 9 or 10 1 h-long training sessions. Pilot testing revealed that English subjects, by contrast, were able to complete about twice as many trials per session as Mandarin subjects. In order to maintain approximately equal duration of training for each group, English subjects repeated each day's training session twice while Mandarin listeners completed each training block only once.

Training *time*, rather than *number of tokens* was equalized because it initially appeared as if the English listeners' considerably faster performance might reflect a more casual attitude toward the task, possibly manifesting itself in less attention being paid on any given trial. By doubling the number of trials we hoped to ensure that the English listeners would end up directing about the same amount of attention to the training stimuli across all training trials as did the Mandarin listeners (cf. Guion & Pederson (2007), on the importance of attentional direction in phonetic learning). In other words, since the English participants appeared to be listening to each trial for about half the time as the Mandarin participants, by doubling the number of trials we hoped to equalize the amount of time the two groups spent attending to each training token. Some possible reasons for the observed difference in trial completion time, and possible consequences, are discussed in the General Discussion, below.

3. Results

3.1. Tone identification

3.1.1. Training performance

Despite differences in the spacing of training sessions and the number of tokens presented in each training session, the Mandarin and English groups showed nearly identical patterns of performance across the training sessions, as shown in Fig. 4. A mixed factorial ANOVA with the between groups factor of Group (English and Mandarin) and the within subjects factor of training Session (1–6) and subsequent post hoc (Tukey HSD) analysis showed no significant effect of Group, $F(1,17) = 0.14$, $p = 0.71$, but a significant effect of Session, $F(9,153) = 37.22$, $p < 0.001$, and a significant interaction between the two, $F(9,153) = 2.24$, $p = 0.02$. Training was successful for both groups, as reflected by the results of a post hoc (Tukey HSD) analysis, showing significant ($p < 0.05$) differences between the first and last training sessions for both groups. However, post hoc analyses showed no significant differences between the two groups at any training session, suggesting that the effects of training were comparable across groups, despite differences in training schedules.

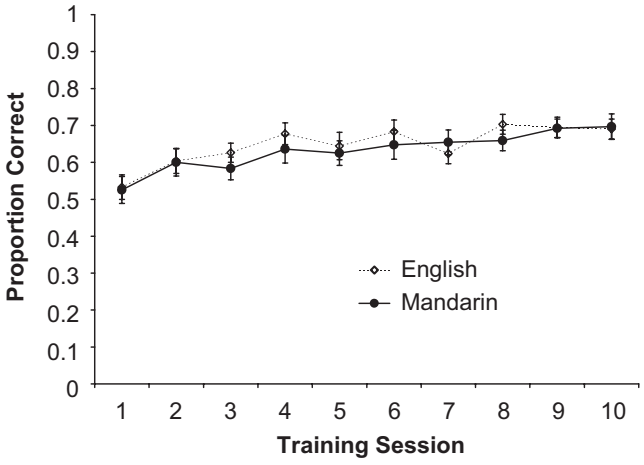

Fig. 4. Proportion correct identification for each training session. Error bars indicate standard error.

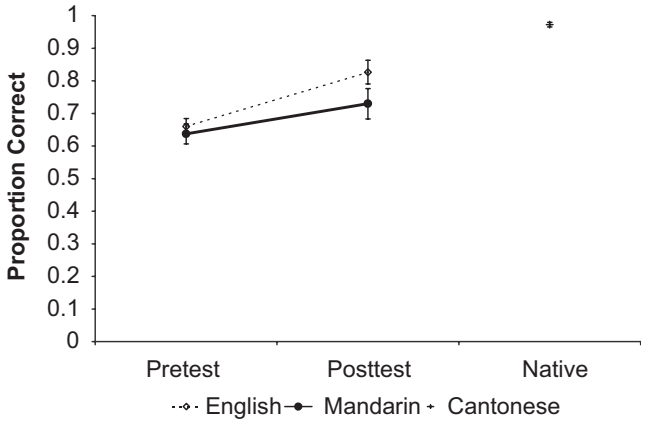

Fig. 5. Proportion correct identification of all Cantonese lexical tones by native speakers of English (open symbols, dashed line), Mandarin (closed symbols, solid line) and Cantonese (cross) before and after training (English and Mandarin speakers only). Error bars indicated standard error.

3.1.2. Pretest and posttest performance

Identification performance on the pretest and posttest was calculated in terms of proportion of Cantonese tones identified correctly. As shown in Fig. 4, both the English and Mandarin groups showed clear improvement in proportion correct tone identification from pretest to posttest. The Mandarin speakers improved from 63.7% to 73.0% correct, English speakers improved from 66.0% to 82.7% correct, and Cantonese speakers showed an average of 97% correct on these stimuli (Fig. 5).

In order to better understand the learning processes underlying the two non-native groups' improvements in tone identification, responses for Mandarin and English listeners were broken down by tone for statistical analysis (shown in Fig. 6). A mixed factorial ANOVA was computed with the between-groups factor of Group (English and Mandarin), and the within-subjects factors of Test (pretest and posttest) and Tone (21 Low Falling, 22 Low Level, 23 Low Rising, 33 Mid-Level, 25 High Rising, and 55 High Level). The main effect of test was significant, $F(1,17) = 47.50$, $p < 0.001$, as was that of Tone, $F(5,85) = 40.55$, $p < 0.001$, but there was no main effect of Group, $F(1,17) = 1.71$, $p = 0.21$. The interaction between Test and Group approached significance, $F(1,17) = 3.88$, $p = 0.07$; post hoc (Tukey HSD) analysis revealed no significant difference between Mandarin and English listeners' performance on either the pretest or posttest. The same tests also showed that the nearly significant effect was mainly a result of the effect of the factor Test, such that each groups' proportion of correct responses improved significantly from pretest to posttest ($p < 0.05$).[4]

There was a significant interaction between Test and Tone, $F(5,85) = 3.66$, $p = 0.005$, and post hoc analysis showed that this was mainly due to significant improvement on tone 21 (Low Falling) (from 52% to 77% correct), and 22 (Low Level) (from 22% to 48% correct). The lack of apparent improvement on the 23 (Low Rising) tone (changing from 49% to 66% correct) was due entirely to the Mandarin listeners' lack of change from 52% to 54% correct. The English listeners were apparently successful at learning to identify this tone, progressing from 45% to 77% correct, and this change was significant at the $p = 0.05$ level.[5] Performance on the other three tones did not improve significantly because both groups of listeners performed well on these tones on the pretest. Performance on the Mid-Level tone changed from 82% to 84% correct, the High Rising tone changed from 96% to 99% correct, and the High Level tone actually decreased non-significantly from 96% to 94% correct. This may be compared with the performance of native Cantonese listeners on the same

[4]Results using arcsine-transformed proportions were comparable to those using untransformed percentages for all main effects, Group, $F(1,17) = 2.23$, $p = 0.15$, Test, $F(1,17) = 36.07$, $p < 0.001$, Tone, $F(5,85) = 47.72$, $p < 0.001$, and for the interactions, Test × Group, $F(1,17) = 3.16$, $p = 0.09$, Tone × Group, $F(5,85) = 1.98$, $p = 0.09$, Test × Tone, $F(5,85) = 3.67$, $p = 0.005$, and Test × Tone × Group, $F(5,85) = 2.58$, $p = 0.03$, as well as all post hoc tests.

[5]This is the only contrast that was significant by post hoc (Tukey HSD) analysis when using raw proportions but not significant when using arcsine-transformed proportions. Thus, the reliability of this particular statement is somewhat more than usually open to interpretation.

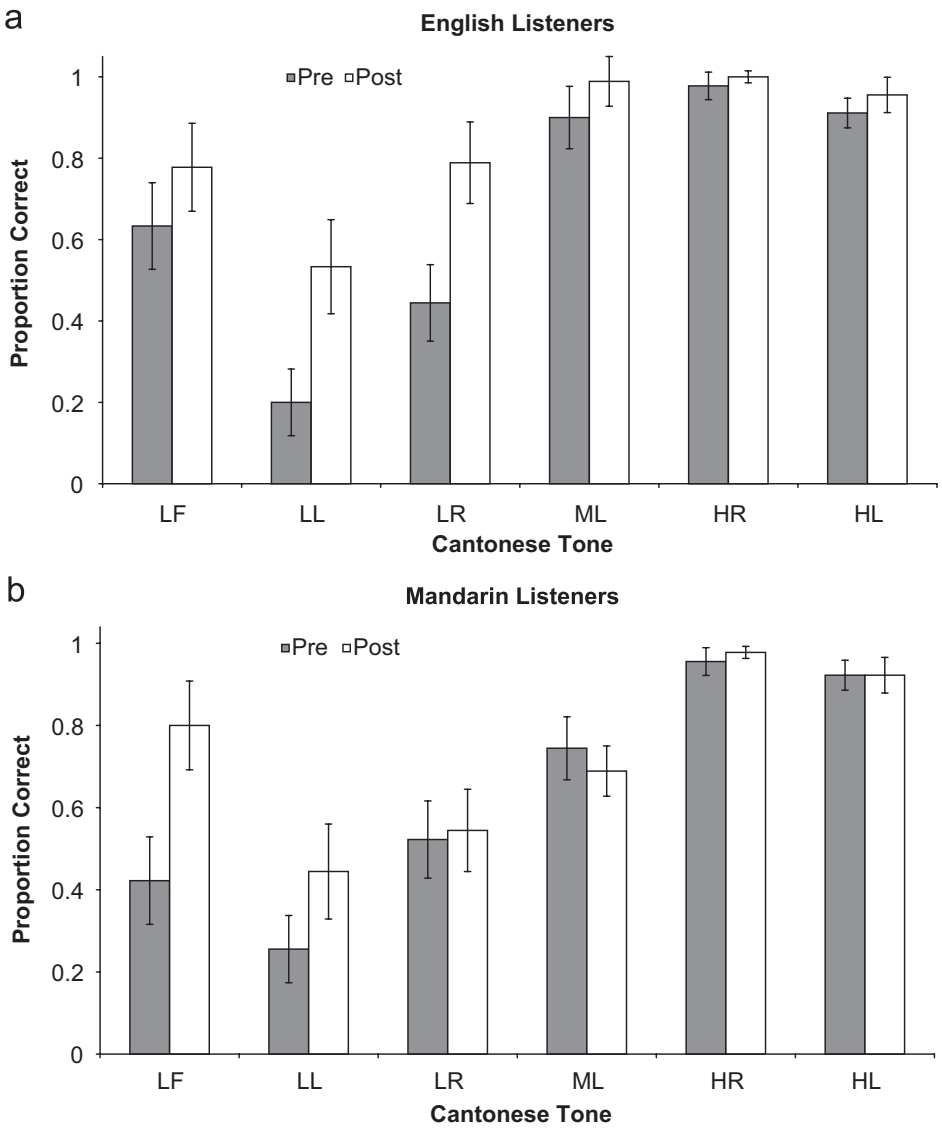

Fig. 6. Proportion correct identification of individual Cantonese lexical tones by native speakers of English (a) and Mandarin (b) before and after training. Error bars indicated standard error.

stimuli: Low Falling, 100%; Low Level, 97%; Low Rising, 94%; Mid-Level, 95%; High Level, 98%; High Rising, 99%.

Finally, results of the three-way ANOVA showed a significant interaction between Test, Tone and Group, $F(5,85) = 2.44$, $p = 0.04$. This interaction was further analyzed using a post hoc (Tukey HSD) analysis with $p \leqslant 0.05$. Results of the post hoc analysis showed no significant differences in performance between the two listener groups for any of the six Cantonese tone categories on either the pretest or posttest. Thus, the primary effects of interest in this interaction involved differences between English and Mandarin listeners in terms of their improvement on particular tones (e.g. the differential pattern of learning for the 23 (Low Rising) tone discussed in the preceding paragraph). Both groups showed no significant differences from pretest to posttest on the Mid-Level, High Rising, and High Level tones. On the Mid-Level tone neither group showed a significant change in performance but the trends for the two groups go in opposite directions (English 89–99%; Mandarin 74–69%). Performance on the High Rising and High Level tones was at or near ceiling for both groups (High Rising: English 97–100%, Mandarin 96–98%; High Level: 86–95%, Mandarin 92% on both tests) and none of these changes was significant.

The two groups showed significantly different patterns of performance on the Low Level, Low Falling, and Low Rising tones. On the 22 (Low Level) tone English listeners' improvement from 18% to 51% correct was

statistically significant, while the Mandarin listeners' change in performance from 26% to 44% correct was not, although it was clearly in the same direction. Although English listeners showed some improvement on identification of the 21 (Low Falling) tone, from 61% to 74% correct, this difference was not significant, but the Mandarin groups' improvement on this tone, from 42% to 80% correct, was statistically significant. In summary, English listeners showed significant improvement on the 23 (Low Rising) and on the 22 (Low Level) tone due to a decrease in the propensity to label both of them as 33 (Mid-Level) tones. Mandarin listeners showed significant improvement only on the 21 (Low Falling) tone due to a decrease in the likelihood of confusing it with the 23 (Low Rising) tone. While patterns of identification performance suggest that native language experience does differentially affect listeners' ability to learn lexical tones, analysis of difference ratings can provide more detailed information on differences in the two groups' perceptual weighting of the acoustic features of tones.

3.2. Difference ratings

Multidimensional scaling (MDS) analysis was applied to the results of the dissimilarity rating task in order to determine the effect of training on listeners' perceptual space as a whole. First, all participants' ratings for each pair were averaged within each group and test, resulting in five six-tone-by-six-tone matrices (two English matrices (one pretest and one posttest) two Mandarin matrices (on pretest and one posttest), and one Cantonese matrix (the pretest) such that each cell in each matrix represented the average difference rating for a given pair of tones on a given test for a particular group of listeners. These difference values were converted to distance measures using the Praat *to Distance* function (Boersma & Weenink, 2004) with standard settings.

A two-dimensional (2D) MDS solution for Cantonese listeners' ratings was calculated first. Two dimensions were chosen because in previous work Gandour (1983) showed that two dimensions suffice to characterize a wide variety of tone languages including Cantonese. Moreover, the individual dimensions of the 2D solution calculated for the Cantonese data were interpretable according to the same principles identified by Gandour and Harshman (1978). This solution accounted for 71% of the variance in the Cantonese data. To analyze the English and Mandarin results, each group's prestest matrices and posttest matrices were submitted to two-way (individual differences) MDS analysis using the Praat INDSCAL analysis option (Boersma & Weenink, 2004). Individual differences scaling differs from standard (one-way) MDS in that it provides not only a single solution space fit to the complete set of inputs (in this case both English and Mandarin listeners' ratings), but also provides a set of ''subject weights'' (or, in this case, ''language weights'') that allow us to compute distinct solution spaces for each group. Separate 2D solutions were computed for the pretest and the posttest in case training resulted in the induction of new or different dimensions (see Francis & Nusbaum, 2002). The Cantonese spatial configuration was used as the starting configuration for both pretest and posttest combined Mandarin/English solutions. The combined pretest solution accounted for 73% of the variance in the two groups' difference ratings, and the posttest solution accounted for 77% of the posttest difference rating variance. These values are comparable to those found by Guion and Pederson (2007). The increase in variance accounted for by the posttest solution suggests that training caused the two groups to become more consistent with one another (i.e. more Cantonese-like).

Subsequently, the group solution for each test was warped using the dimension weights calculated for each language and each test (Table 1). Following Arabie, Carroll, and DeSarbo (1987), warping was conducted by multiplying the coordinates of each point in the group space for a given test (pretest or posttest) by the square

Table 1
Dimension weights

	Cantonese		English		Mandarin	
	Dim. 1 (Height)	Dim. 2 (Direction)	Dim. 1 (Height)	Dim. 2 (Direction)	Dim. 1 (Height)	Dim. 2 (Direction)
Pretest	0.71	0.51	0.69	0.50	0.51	0.76
Posttest			0.84	0.33	0.62	0.65

root of each the corresponding subject (language) weights (English or Mandarin), thereby deriving separate representations of each language group's perceptual space for lexical tones at the time of the pretest and the posttest (Fig. 7).

From the results shown in Table 1 and Fig. 7, two observations can be made. First, although English and Mandarin listeners' tone identification performance was comparable on the pretest, the two groups did exhibit somewhat different patterns of weighting of the two major perceptual dimensions of tone, in a manner comparable to that observed by Gandour (1983) and Gandour and Harshman (1978). English listeners tended to give more weight to Height than they did to Direction (comparable to the pattern shown by Cantonese listeners), while Mandarin listeners showed the opposite distribution. Note that, because the two tests were scaled separately, it is not legitimate to compare dimension weights directly across tests. However, it is possible to compare the pattern of relative weighting of the two dimensions with respect to one another (i.e. their ratio) across tests.

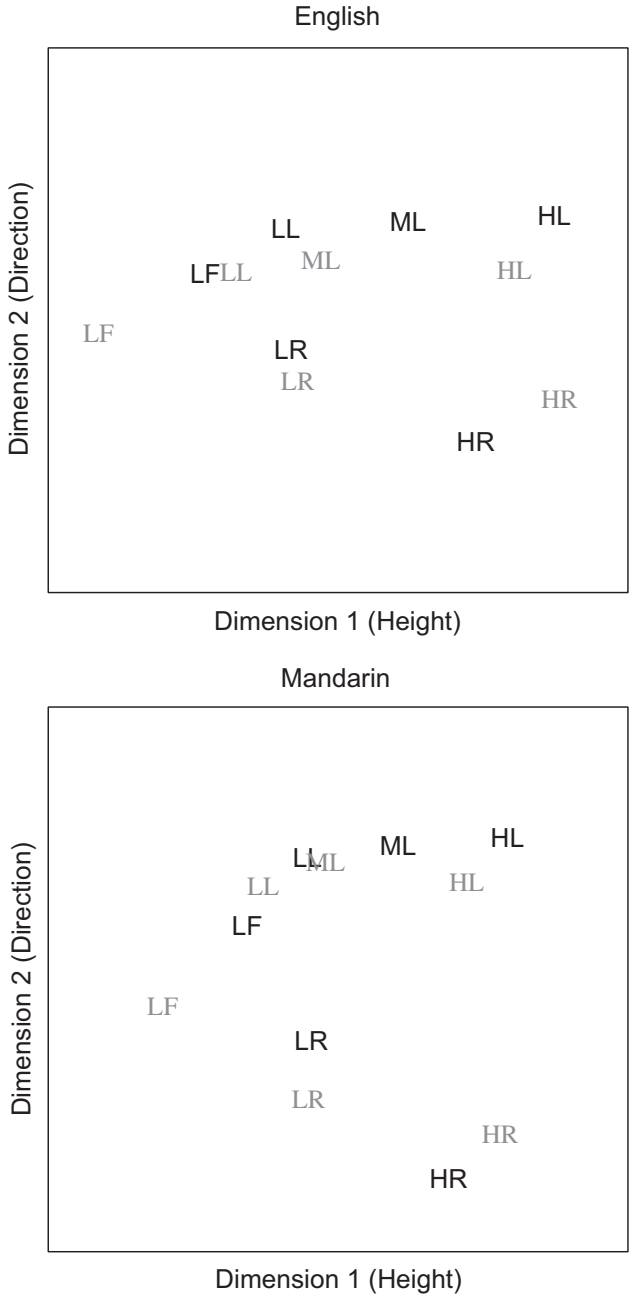

Fig. 7. Individual groups' tone spaces showing pretest (dark) and posttest (light) configurations.

Second, training to recognize Cantonese tones appears to have had a similar effect on both English and Mandarin listeners, encouraging an increase in the relative weight given to Height and a concomitant decrease in that given to Direction. However, the end result is somewhat different for each group: Mandarin listeners showed almost equal weight for Height and Direction following training, while English listeners appeared to give considerably more weight to Height than to Direction after training (even more than did Cantonese listeners). Differences in initial and final distribution of weight to each dimension may result from different patterns of native language experience with f0 contours, and may be reflected in different patterns of learning of individual tones.

3.3. Confusions

One disadvantage of the MDS analysis used here is that it depends on metalinguistic judgments of the similarity of two syllables presented outside of any sentential context. While the relationship between similarity judgments and more naturalistic processes of word recognition is not yet well understood, it is quite clear that tone perception can differ significantly in context as opposed to isolation (Francis, Ciocca, & Ng, 2003). Another way to investigate properties of perceptual space is to examine listeners' confusions on an identification task (e.g. Gandour, 1981). In the present case, the identification task is arguably more natural than the difference rating task because stimuli are presented in sentential context and listeners are asked to identify words (or tones), which is somewhat closer to the actual task of spoken word recognition. In the ideal case, the resulting confusion matrices could themselves be submitted to MDS analysis providing an even more accurate picture of listeners' tone space, as was done by Gandour (1981). Given that, in the present case, acoustic differences between some pairs of stimuli were very large, listeners were rather good at this task even before training began, resulting in a large number of empty cells in each matrix. Despite this, examination of the raw confusion scores, shown in Tables 2–5, provides some insight into differences in the perception and tone learning patterns of Mandarin and English listeners. For the sake of brevity, only those tones on which one or both groups of listeners showed significant improvement will be discussed here (Tables 2–5).

3.3.1. Low Falling (21) tone

On the pretest Mandarin listeners tended to identify the 21 (Low Falling) tone as 21 (Low Falling, 42.2%), 23 (Low Rising, 35.6%), 22 (Low Level, 12.2%), and rarely as 25 (High Rising 10%). Following training Mandarin listeners were much better at identifying this tone (80.0%), and only confused it occasionally with the 23 (Low Rising) and 25 (High Rising) tones (5.6% each) and the 22 tone (Low Level, 8.9%).

On the pretest English listeners identified the 21 (Low Falling) tone as either 21 (Low Falling, 61.0%) or 22 (Low Level, 32.0%), with only a few responses of 23 (Low Rising, 4.0%) or 33 (Mid-Level, 3.0%), suggesting that they were already more focused on the average f0 height of the syllable than were the Mandarin listeners. Following training, English listeners' only mis-identifications of this tone were as a 22 (Low Level) tone (26.0%), suggesting again that f0 height rather than direction was the primary determinant of their choice.

Table 2
Mandarin pretest confusions

Token	Identified as					
	21	22	23	33	25	55
21	0.422	0.122	0.356	0.000	0.100	0.000
22	0.267	0.256	0.256	0.156	0.056	0.011
23	0.189	0.067	0.522	0.011	0.211	0.000
33	0.044	0.067	0.011	0.744	0.000	0.133
25	0.000	0.000	0.044	0.000	0.956	0.000
55	0.000	0.000	0.000	0.078	0.000	0.922

Table 3
English pretest confusions

Token	Identified as					
	21	22	23	33	25	55
21	0.610	0.320	0.040	0.030	0.000	0.000
22	0.210	0.180	0.020	0.590	0.000	0.000
23	0.070	0.120	0.450	0.300	0.050	0.010
33	0.000	0.030	0.050	0.890	0.010	0.020
25	0.000	0.000	0.030	0.000	0.970	0.000
55	0.000	0.000	0.010	0.130	0.000	0.860

Table 4
Mandarin posttest confusions

Token	Identified as					
	21	22	23	33	25	55
21	0.800	0.089	0.056	0.000	0.056	0.000
22	0.289	0.444	0.133	0.133	0.000	0.000
23	0.278	0.100	0.544	0.011	0.067	0.000
33	0.000	0.211	0.000	0.689	0.000	0.100
25	0.011	0.000	0.011	0.000	0.978	0.000
55	0.000	0.011	0.000	0.067	0.000	0.922

Table 5
English posttest confusions

Token	Identified as					
	21	22	23	33	25	55
21	0.740	0.260	0.000	0.000	0.000	0.000
22	0.080	0.510	0.050	0.360	0.000	0.000
23	0.010	0.130	0.770	0.070	0.020	0.000
33	0.000	0.010	0.000	0.990	0.000	0.000
25	0.000	0.000	0.000	0.000	1.000	0.000
55	0.000	0.000	0.000	0.050	0.000	0.950

3.3.2. Low Level (22) tone

Mandarin listeners appeared to have great difficulty identifying the 22 tone consistently on the pretest, calling it 21 (Low Falling, 26.7%), 23 (Low Rising, 25.6%), 33 (Mid-Level, 15.6%), 25 (High Rising, 5.6%) and even 55 (High Level, 1.1%). Following training this tone was no longer mis-identified as either of the two high tones (25 or 55) but there was also a slight increase in the proportion of (incorrect) 21 (Low Falling) identifications (28.9%).

English listeners were somewhat more consistent in identifying this tone, though on the pretest they were actually less accurate than Mandarin listeners. In this case, the primary mis-identifications of this tone were as 33 (Mid-Level, 59.0%) and 21 (Low Falling, 21.0%). Following training English listeners were much more adept at recognizing this tone (51.0%), and mis-identified it less often as a 33 (Mid-Level) tone (36.0%).

3.3.3. Low Rising (23) tone

Most of the confusions that Mandarin listeners exhibited when identifying the 23 (Low Rising) tone were with either the 25 (High Rising, 21.1%) or 21 (Low Falling, 18.9%) tones. This supports the hypothesis that these listeners based their initial judgments more on direction of f0 change (for the 25 mis-identifications) or at

least on the perception of some change in f0 in any direction (for the 21 mis-identifications) than on f0 height per se. The observation that training served primarily to reduce the number of 25 (High Rising) mis-identifications (to 6.7%) but actually *increased* the number of 21 (Low Falling) confusions (to 27.8%) further supports the hypothesis that training induced Mandarin listeners to increase the weight they gave to the f0 height dimension as the expense of direction of f0 change. Moreover, the increasing frequency of confusions with the 21 tone suggest that Mandarin listeners may have been trying to force the Cantonese 23 tone into their native 214 category perhaps because the Cantonese 23 tone often begins with a slight fall (see discussion by Francis et al., 2003) and both the Cantonese and Mandarin tones exhibit a moderate rise in the middle of the talker's f0 range.

English listeners initially identified this tone as a 33 (Mid-Level) tone (30.0% of the time), and otherwise spread their mis-identifications more or less evenly through the lower end of the tone inventory. The proportion of 22 (Low Level) responses (12.0%) was greater than that of 21 (Low Falling) responses (7.0%), which was slightly larger than 25 (High Rising, 5.0%) or 55 (High Level, 1%) responses. This again supports the hypothesis that English listeners were not attending strongly to differences in direction of f0 change (therefore confusing a changing with a level tone) but were influenced by average f0 (along which dimension the 23 and 33 tones are most similar).

4. Discussion

4.1. Overall performance

English and Mandarin listeners' performance on identifying individual Cantonese tones did not differ significantly on the pretest. The observation that native speakers of a tone language and a non-tone language show qualitatively similar patterns of perception for non-native tone categories supports the conclusions of previous researchers that the mere presence or absence of lexical tone contrasts in the native language is not in itself sufficient to determine cross-language perception of lexical tones. Instead, it seems that it is necessary to take into account the f0 patterns that listeners have been exposed to in their native language irrespective of their function as cues to tone as opposed to intonational categories.[6] Although the two groups' overall performances, in terms of proportion of correct responses, were comparable, their differences, in terms of the kinds of errors they made before training and the ways in which they improved as a result of training, provide insight into the applicability of existing models of cross-language perception and perceptual learning to the problem of acquisition of non-native tones.

4.1.1. Category-based explanations

It must be stated at the outset that all conclusions made here about mappings between Cantonese and Mandarin tone categories can only be speculative in nature. In order to determine the actual pattern of assimilation, it would be necessary to carry out a true study of cross-language mapping involving overt decisions by listeners on the similarity or goodness of fit between L2 tokens and L1 categories (e.g., as is done for consonant categories in Guion, Flege, Akahane-Yamada, & Pruitt, 2000). However, the observation that performance was best on three tones that have identifiable phonological counterparts in Mandarin is consistent with the hypothesis that cross-language perception, prior to training, is determined in large part by the degree to which non-native sounds can be assimilated to native category representations. For example, Mandarin listeners were quite good at identifying the Cantonese 55 (High Level) and 25 (High Rising) tones, presumably because these tones map unequivocally onto the Mandarin 55 (Tone 1) and 25 (Tone 2) tones, respectively. Similarly, Mandarin listeners' ability to correctly identify the 33 (Mid-Level) tone even on the pretest suggests that perhaps this tone might in fact map sufficiently well onto the Mandarin 214 (Tone 3) or neutral tone category.

[6]It is also worth considering the relative importance (functional load) of pitch-based suprasegmental patterns in the listeners native language. It is possible that speakers of a language with a more consistent stress pattern than English (e.g. French, cf. Hallé et al., 2004), or a weaker reliance on intonation (e.g. Dutch) might exhibit a weaker effect of native language categories on the perception or acquisition of non-native lexical tones.

However, Mandarin listeners' initial performance on the 23 and 21 tones are not quite as compatible with a pure category assimilation model. Their performance on the 23 (Low Rising) tone was worse than would have been expected if this tone were assimilated to the Mandarin 214 (Tone 3). While it is possible that the 23 tone was simply not sufficiently similar to any Mandarin category to be assimilated (an *uncategorizable* contrast in the terminology of PAM), it seems more likely that the moderate but not complete similarity between the 23 and the 214 tone led to interference of the native Mandarin category (as might be predicted by the SLM), an account supported by the observation that the Mandarin listeners failed to show any improvement on this tone over the course of training. Similarly, the fact that the 21 (Low Falling) tone was identified quite poorly to begin with, despite the presence of an (allophonic) 21 category in Mandarin, could be explained in terms of interference due to perceiving a familiar allophone in an unexpected context, but might also suggest that a model of cross-language perception that relies entirely on mappings between L2 and L1 *categories* is insufficient to account for the present data.

Similar results obtain for the English listeners. These participants were also quite good at identifying the 25 (High Rising) tone, possibly due to its assimilation to English question intonation (although this contour would be unexpected in the middle of an English sentence). However, performance on the 23 (Low Rising) tone was not as good as might have been expected. It is possible that the difference between a standard yes–no intonation pattern (L*HH% in ToBI notation) and one aimed more at soliciting agreement (H*HH%) may be less than categorical (cf. Hallé et al.'s argument that intonational categories are not perceived as categorically as tone ones), but such a claim would, of course, have to be investigated in English to be answered definitively. A more parsimonious solution would be to conclude that the 23 tone is simply too different from any native English category to be successfully assimilated, or, alternatively, that it is sufficiently similar to be interfered with by the presence of the native category. Overall, the variety of different conclusions that might be drawn depending on the assumed degree of similarity between the native and non-native categories in turn raises the question of how such similarity is to be determined within a multi-dimensional tone space.

English listeners' poor performance on the 21 (Low Falling) tone might likewise suggest that this contour is not sufficiently similar to any native intonational categories to be assimilated. However, low falling tones are commonly attested in English (e.g. in a tag response question, especially ones expressing disbelief/hostility, as in the first word of the response in the exchange "I worked all night on this assignment." "Did you?") so it is not clear why the Mandarin 21 tone does not assimilate. On the other hand, the distinction between a low falling and a high falling tone in English seems to be quite gradient (cf. Cruttenden, 1997, pp. 91–92) suggesting, that this intonational pattern, while present in the listeners' phonological inventory, may not be sufficiently categorical to support assimilation of a non-native tone category.

Ultimately, while some patterns of performance of both Mandarin and English listeners' can be explained in terms of assimilation to native categories, and much more might be explainable following a definitive study of the cross-language mappings (both between Cantonese and Mandarin tones, and between Cantonese tones and English intonational categories), not all of the results of either group are entirely amenable to this approach.

4.1.2. Confusions and perceptual dimensions

The patterns of confusion exhibited by listeners both before and after training support the idea that prior experience with linguistically meaningful f0 patterns, whether tone or intonational, influences tone perception and learning. Mandarin and English listeners showed comparable initial identification of the 23 (Low Rising) tone, but their patterns of mis-identification were quite different, indicating a clear influence of native language experience. Mandarin listeners were more likely to confuse the 23 (Low Rising) tone with the 25 (High Rising) tone than were English listeners, who in turn were more likely to mis-identify the 23 (Low Rising) tone as a 33 (Mid-Level) tone. This pattern supports the hypothesis that Mandarin listeners were focusing on Direction of f0 change (similar for 23 and 25, but different for 23 and 33), while English listeners were focusing on overall f0 Height (similar for 23 and 33, but different for 23 and 25), consistent with the findings of Gandour (1983) and Gandour and Harshman (1978).

In addition, Mandarin listeners frequently mis-identified a falling tone (21) as a rising one (either 23 or 25), which suggests that Mandarin listeners may have been focused on the changing nature of the f0 contour

without (initially) paying attention to the actual direction of change. Such a strategy could be quite effective in Mandarin. Although there is an apparent Direction-based contrast between the 25 (Tone 2, Rising) and 51 (Tone 4, Falling) Mandarin tone contours, these tones also differ strongly in terms of the syllable duration (25 is quite long, 51 is quite short). Thus, even listeners who treated these two tones as similar along a changing/not-changing axis of contrast could still distinguish them easily according to a long/short duration dimension. Finally, Mandarin listeners may have had a comparatively easy time learning the 21 (Low Falling) tone because of their allophonic experience with this contour that results from the tone sandhi rule that converts that any Tone 3 (214, Dipping) tone followed by another non-Tone 3 syllable into a 21 f0 pattern instead (Li & Thompson, 1989). Alternatively, Mandarin listeners' improvement on the 21 (Low Falling) tone could, instead, have resulted from a learned increase in weighting of the height dimension, rather than some change in attention to direction of f0 change, because the 21 (Low Falling) tone is, on average, lower in overall f0 than any other Cantonese tone.

Following training, Mandarin listeners' increased weighting of f0 height may also have helped with their identification of the 22 (Low Level) tone because it was no longer mis-identified as either of the two high tones (25 or 55), and there was a slight increase in the proportion of (incorrect) 21 (Low Falling) identifications (28.9%) of this tone. It is also possible to account for this development in terms of a change in weighting of direction of f0 change (improving the distinctiveness of rising vs. level contours), since the major improvement in identification of the 22 (Low Level) results from a reduction in the number of 23 (Low Rising) responses. Although the proportion of weight given to this dimension is, overall, smaller following training, it is possible that localized changes in weight along the dimension may result in improved performance on specific contrasts even in the absence of an overall increase in dimensional weighting (see Francis & Nusbaum (2002) for discussion of localized warping models of phonetic learning).

English listeners' confusion of a 22 with a 33 tone also appears at first to suggest a lack of attention to f0 height, but the observation that the 33 (Mid-Level) tone was almost never identified as a 22 (Low Level) tone on the pretest (3.0%) suggests an alternative interpretation. Recall that English intonational contours have been argued to be composed of combinations of low and high tones (only) (Pierrehumbert, 1980; see also discussion of general autosegmental-metrical theories by Ladd, 1996, Chap. 3). Thus, English listeners' attention to f0 height may support the distinction of two, but not three, tones. In the present case, listeners may have perceived level tones as either high (e.g. the 55, High Level tone) or low (encompassing both the 33, Mid-Level, and 22, Low Level tones) and responded 33 (Mid-Level) to all low tones. Posttest performance, suggests that English listeners' increased attention to f0 height may reflect an increased ability to distinguish fine-grained differences beyond just their native "high- vs. low-" contrast.

4.1.3. The importance of categories vs. perceptual dimension weights

The results presented here lend qualified support to the idea that both tone and intonational categories influence the cross-language perception of lexical categories, in similar ways, consistent with a category assimilation model and in contradiction to a strict levels of representation argument. However, while category-based assimilation can account for many of the behaviors observed in both non-native listener groups before training, we would argue that, to fully understand the effects of training requires the adoption of a feature-oriented perspective. Hallé et al. (2004) illustrate the problem quite clearly. For tone language speakers, non-native tones should follow the principles established by cross-language studies of segmental phonology. Mandarin listeners have tone categories which cause them to direct selective attention to critical features of the f0 contour, and therefore should map Cantonese tones onto their native (Mandarin) tone system. Their initial (cross-language) perception should be predictable from this mapping, and the manner in which they learn should involve selective reweighting of acoustic features. In contrast, L1 speakers of a non-tone language should not be able to categorize non-native tones (in a strict interpretation of PAM), because they have no native tone categories to assimilate them to. Despite this, Hallé et al. (2004) reject the idea that tone categories are completely non-assimilable: Non-tone language speakers must still be able to map non-native tones onto some kind of native perceptual space, because tone contours are still used in non-tone languages, e.g. for intonation. Ultimately, one must conclude that non-tone language speakers possess some sort of mental representation of intonational contours, but, according to Hallé et al. (2004) these representations are simply weaker than those of true tone categories, in the sense of exerting less influence on lexical perception (inducing a lesser degree of categorical perception).

In contrast, many current theories of intonation (e.g. Beckman et al., 2005; Ladd, 1996) treat into national contours as linguistic categories on par with other phonological categories such as vowels, consonants, and lexical tones, suggesting that native intonational categories should have the same kind *and degree* of influence on the perception and acquisition of non-native tones as do native tone categories. Hallé et al. (2004) reject this notion on the grounds that their results show a different degree of categorical perception of lexical tones for French (non-tone language) as compared with Mandarin (tone language) speakers, suggesting that intonational categories are not as strongly defined as tone ones, and therefore have less influence on the perception of tone contrasts.

Indeed, there are clear linguistic differences between intonation and tone. Intonational contours typically extend over multiple syllables and are primarily determined by pragmatic and discourse considerations, while lexical tones are confined to single-syllable domains and are governed by lexical and morphological constraints and more or less immediate context. Moreover, recent neuroimaging evidence suggests that the two kinds of systems (intonation and tone) are processed in very different ways by the brain. For example, Gandour, Xu et al. (2003) showed that Chinese listeners processing intonational prosody (extending over three syllables) exhibited heightened activation in the frontoparietal regions of both left and right hemispheres, while processing lexical tones induced heightened activation only in the left hemisphere. Even more strikingly, Gandour et al. (2004) showed a difference in hemispheric processing for identical stimuli depending on whether listeners were processing the same f0 contour as a cue to lexical tone or as a cue to intonation.

It is possible that such differences correspond to a (language independent) phenomenon that causes long-term representations of intonational categories to simply be weaker (and more right-hemispheric) than those of tone categories. On the other hand, it is also possible that the differences in categorical perception observed by Hallé et al. (2004) result from the influence of specific native categories (whether tone or intonational) on the perception of specific acoustic dimensions of contrast. According to this hypothesis, the French listeners in Hallé et al.'s (2004) study represent their native intonational categories just as categorically as Mandarin listeners do their native tone categories. However, while the French listeners differentiate their categories according to one dimension (e.g. Height), the Mandarin listeners differentiate theirs along a different one (e.g. Direction). Since the experimenters chose to distinguish their experimental stimuli only along the Direction dimension, they built in a bias favoring the Mandarin over the French strategy for categorization. Thus, the reason that French listeners in the Halle et al. study showed only weak categorical perception of the Mandarin tone contrast is simply that the specific *dimension* used in the study is highly salient for distinguishing the Mandarin tone, but not any French intonational, categories. The difference between French and Mandarin listeners need not be that Mandarin listeners have strong representations of f0-based categories while French listeners do not. Rather, the difference may simply be that Mandarin listeners give a great deal of weight to the Direction dimension when categorizing f0-based categories, while French listeners do not.

The influence of native experience (whether tone or intonational) can be understood in terms of the relative strength of long-term memory representations within the context of the multi-store memory model of speech perception proposed by Xu et al. (2006) *and* the effects these representations have on perceptual weighting of the features of tone (see also earlier work by Lee, Vakoch, & Wurm, 1996). Both tone and non-tone language speakers must store long-term memory representations of f0-based phonological categories (tone and/or intonational), and these representations affect the distribution of selective attention to the features of f0. In cases of clear correspondence between native and non-native categories, the non-native category assimilates to the native category representation. But in cases in which there is no clear correspondence between categories, cross-language perception is determined by the relative weight given to specific perceptual features—weights that are set by the demands of native tone categorization. Acquisition involves changing the relative weighting of features from the initial native pattern to one more closely approximating that of native speakers of the L2, as well (eventually) developing specific category representations for L2 phonemes.[7]

[7]In this regard, it is worth noting that there may well be other salient features of tonal and/or intonational categories beyond those of Height and Direction. For example, recent research on both tonal and intonational categories has begun to suggest an important role for the *alignment* of f0 events with respect to the segmental information in the syllable (For intonation, cf. Arvaniti & Garding, in press; Atterer & Ladd, 2004; Grabe, Post, Nolan, & Farrar, 2000; Mennen, 2004. For tone, see Xu, 1998, 1999; Xu & Liu, 2006). However, the application of such findings to studies of cross-language perception of lexical tone (or intonation) will require considerably more work simply to identify and codify the relevant dimensions of contrast in each case before clear predictions can be formulated, let alone tested.

4.2. Other factors of note

4.2.1. Differences in amount of training

It is important to note that the two groups of listeners received different amounts of training, although the total time spent on training tasks was comparable (10 h), as was posttest performance. English listeners were able to complete roughly twice as many training trials in the same amount of time as Mandarin listeners. It is possible that, by doubling the amount of training that Mandarin listeners received, they might have improved more than the English listeners. However, the patterns of learning reflected in the MDS and confusion analyses are more complex than can be explained simply in terms of different amounts of training; to explain these it appears to be necessary to consider each groups' native language experience with f0-based linguistic distinctions.

Indeed, the difference in the number of exposures to training stimuli across groups may initially seem more significant than it actually turns out to be in the present case. Studies of L2 acquisition quite clearly demonstrate that differences degrees of experience can lead to qualitatively different outcomes in performance (Bongaerts, 1999). Moreover, different *kinds* of training clearly result in differential performance, as has been shown by a variety of studies in which different groups of subjects were exposed to these stimuli but received different feedback and consequently developed different response patterns (Francis et al., 2000; Francis & Driscoll, 2006; Goldstone, 1994; Polley, Steinberg, & Merzenich, 2006). Might it then be possible that providing the American listeners with twice as many training trials also somehow resulted in their receiving a different *kind* of training?[8] Might providing the Mandarin listeners with the same number of trials have resulted in more similar performance across groups?

Although this possibility is intriguing, we believe that it does not hold in the present case, for the following reasons. First, both groups reached a similar level of overall performance with the amount of training given. Had the Mandarin listeners received *more* training than this, they presumably would have improved to a greater degree than the English speakers, and this improvement would presumably have involved increased accuracy on those tones (21–23, and 33) on which they had not yet reached ceiling. On these tones, the present amount of training already resulted in Mandarin listeners performing better than American listeners on the 21 tone, and worse on the other three tones. Thus, while additional training trials would presumably have served to make the Mandarin response patterns more similar to those of the English speakers on the 22, 23, and 33 tones, the two groups would actually have become *more different* on the 21 tone. Since it does not seem plausible that increased training would significantly *reduce* performance on this tone (alone), which is what would be required to cause both groups to be more similar on all tones after training, we must conclude that, at the very least, this difference between the two groups cannot be ascribed to differences in number of training trials.

Second, studies that examine the influence of *type* of training on perceptual learning typically shape subjects' perceptions by giving different feedback to different groups in response to the same stimulus (Francis et al., 2000; Francis & Driscoll, 2006; Goldstone, 1994; Polley et al., 2006) and/or by providing subjects with a different distributions of training stimuli (Holt & Lotto, 2006; Maye, Werker, & Gerken, 2002). However, in the present case, both groups received exactly the same distribution of stimuli, and the category label assigned to each stimulus was identical in both groups (e.g. if a subject called a 21 tone a 22, that was considered wrong regardless of native language). Moreover, unlike most previous laboratory training studies focusing on differential weighting of individual dimensions, training in the present experiment was not necessarily restricted to a single stimulus dimension. The natural stimuli used here differed from one another along a variety of dimensions, and listeners were free to choose among them to identify those that, alone or in combination, sufficiently indicated differences between Cantonese tone categories. The observation that the two groups of learners differed in terms of the dimensions that they learned best actually supports the most important insight to be derived from the present results: Listeners' prior linguistic experience influenced the choice of which dimensions were learned (better), almost as if Mandarin speakers had assigned themselves to one task (e.g. learning height) while English speakers self-assigned to the other (direction), despite the fact that both groups were exposed to exactly the same distribution of stimulus properties and exactly the same

[8]We are grateful to an anonymous reviewer for pointing out this possibility.

corresponding labels for each stimulus. Thus, the proposition that the two groups received different *kinds* of training is only true insofar as one accepts our basic conclusion—that native language experience differentially affected the stimulus dimensions that participants learned to use for categorization. Although listeners in both groups received identical feedback on every stimulus, each group of listeners' individual native language experience caused them to use that feedback to reinforce learning of different stimulus properties in each group.

4.2.2. Cognitive load

One possible explanation for the difference between the two groups' training demands may be found in the identification task itself. Although both groups heard and saw the same frame sentences written in Chinese characters and phonetic transcription on each trial (in both training and testing), this visual display provided considerably more information to Mandarin listeners (who could read Chinese) than to English listeners. This greater amount of information may in fact have interfered with performance, in part because the Mandarin pronunciation of many Chinese characters differs from the Cantonese. For example, both the characters for *chair* and *ear*, which in Cantonese have a 25 (High Rising) and 23 (Low Rising) tone, respectively, have a 214 (Tone 3, Dipping) tone in Mandarin. Thus, Mandarin listeners not only had to learn to hear the difference between the two kinds of rising tones, but they had to learn to ignore their native assignment of the same tone contour to each of the two characters. While the perceptual task was comparable for Mandarin and English listeners, the response task was not, as English listeners began the study with no preconceived notions of what tone contours might properly be associated with any given character (and, in fact, could simply ignore the Chinese characters altogether).

In contrast, the fact that nearly all of the Chinese characters used in the present study had the same *meaning* (though not necessarily the same pronunciation) in Mandarin as in Cantonese could in principle have provided something of a facilitory effect, since lexical knowledge can improve speech perception (Samuel, 1997, 2001). One factor that might be expected to mitigate against such influence is that there are significant differences between many of the modern "simplified" characters used in the Mainland of the People's Republic of China and the traditional characters still in use in Hong Kong, Macau, and Taiwan, and in the present study. On the other hand, the original characters are still used for calligraphic purposes in the PRC and it is considered a point of pride for many educated Chinese speakers to be able to read them (and all of our Beijing participants were highly educated medical school students). In short, it is possible that differences between traditional and simplified characters may have made it more difficult for Mandarin listeners to easily read the Chinese text on the screen, but they almost certainly had some degree of familiarity with the characters displayed. Thus, the ultimate consequence of the use of Chinese characters on the response screen, while intended to facilitate learning and make the task more like actual word recognition and learning, is unknown. It may have made the task more difficult, though only for those able to read Chinese, and it may have made the task easier, though only for those able to read traditional characters without difficulty. We assume that some combination of these factors contributed to the nearly double time-per-trial for the Mandarin listeners, but how, and to what degree, is not determinable.

5. Conclusions

Both the specific structure of a listener's native tone (or intonational) category inventory as well as experience-dependant patterns of general perceptual weighting contribute to a better understanding of cross-language perception and acquisition of lexical tones. The assimilatory influence of well-defined native categories (e.g. the English question intonation (L*HH%) and the Mandarin 55 (Tone 1) and 25 (Tone 2) lexical tone categories) clearly explains situations in which non-native listeners are able to identify L2 tone categories (i.e. the Cantonese 55, High Level, and 25, High Rising, tones) at close to native levels of accuracy. However, there is also a clear advantage to considering the influence of native language categories from the perspective of relative feature weighting. Not only does this allow us to draw on the full power of existing theoretical perceptual mechanisms, it also supports a simple resolution to the question of how poorly assimilated tones might be perceived as well as providing a framework for understanding the process of perceptual learning of non-native sounds.

However, the overall patterns of weights given to different perceptual dimensions may only begin to capture cross-language differences in perceptual predispositions. In addition to such general measures, it may also be necessary to investigate localized "stretching" and "shrinking" along particular dimensions. For example, we have proposed that the difference between English listeners' pretest and posttest performance on level tones might result from a tendency to distinguish between two levels of f0 height on the pretest, while distinguishing three levels on the posttest. Such a change could be described as a local increase in distance within one of the two original categories (effectively "splitting" one category into two, cf. Francis & Nusbaum, 2002; Pisoni, Aslin, Perey, & Hennessy, 1982) without affecting the important observation that both pretest and posttest patterns derive from an overall strong weighting of the f0 height dimension. Finally, the apparent contribution of Mandarin listeners' prior allophonic experience with a 21 (Low Falling) f0 contour to improved learning of this tone, but not to initial recognition, suggests the need for further research on the importance of the phonological status of perceptual categories in determining their influence on cross-language perception and learning non-native speech sounds.

Acknowledgments

This work was conducted while the first author was a postdoctoral fellow and the third author was a visiting research assistant professor in the Department of Speech and Hearing Sciences at the University of Hong Kong. We would like to thank See Lok Kan, Ka Man Wong, Choi Hung Law, Ho Yin Leung and Elaine Eramela for help with data collection. Some of these results were presented at the 147th Meeting of the Acoustical Society of America, New York, NY, May 24–28, 2004. We thank Ken de Jong and Rajka Smiljanic for insightful comments and assistance with understanding the ToBI literature, and Ken de Jong, Susan Guion, Amanda Seidl, and an anonymous reviewer for helpful comments and suggestions on earlier versions of this paper. This project was funded through a Hong Kong University Seed Fund grant awarded to the first two authors.

Appendix A. Stimuli

Sentence 1

下	–	個	字	係	—
ha22	jat5	ko33	tsi22	hai22	—
next	one	classifier	character	is	—

The next word is ____

Sentence 2

請	指	下	—	俾	我	睇
ts^{h}ing35	tsi35	ha23	—	pei35	ngo23	t^{h}ai35
please	point to		—	for	me	look

Please point to ___ for me to see

Sentence 3

我	會	讀	—	俾	你	聽
ngo23	wui23	tuk22	—	pei35	nei23	t^{h}en55
I	will	read	—	for	you	hear

I will read ___ for you to hear

Sentence 4

邊	個	係	—	字?
pin55	ko33	hai22	—	tsi22
which	classifier	is	—	character

Which one is the character ___?

Sentence 5

個		字	係	邊	度?
ko33	___	tsi22	hai35	pin55	tou22
Classifier	___	character	is	where	

Where is the character ___?

Sentence 6

我	想	要	個		字
ngo23	soeng35	jiu33	ko33	___	tsi22
I	want		classifier	___	character

I want the character ___

Sentence 7

我	識		呢	個	字
ngo23	sik55	—	li55	ko33	tsi22
I	know	—	determiner	classifier	character

I know the character ___

A.1. Individual characters

/se/

55 (High Level)	25 (High Rise)	33 (Mid-Level)	21 (Low Fall)	23 (Low Rise)	22 (Low Level)
些	寫	瀉	蛇	社	射
some	write	spill	snake	society	sheet

/si/

55 (High Level)	25 (High Rise)	33 (Mid-Level)	21 (Low Fall)	23 (Low Rise)	22 (Low Level)
思	史	試	時	市	事
thought	history	try	time	market	event

/jau/

55 (High Level)	25 (High Rise)	33 (Mid-Level)	21 (Low Fall)	23 (Low Rise)	22 (Low Level)
休	柚	幼	由	有	又
rest	grapefruit	young	from	have	and

/fan/

55 (High Level)	25 (High Rise)	33 (Mid-Level)	21 (Low Fall)	23 (Low Rise)	22 (Low Level)
分	粉	訓	焚	奮	份
separate	noodle	command	burn	(work) hard	portion

/fu/

55 (High Level)	25 (High Rise)	33 (Mid-Level)	21 (Low Fall)	23 (Low Rise)	22 (Low Level)
夫	苦	富	符	婦	負
husband	bitter	rich	appropriate	woman	negative

/ji/

55 (High Level)	25 (High Rise)	33 (Mid-Level)	21 (Low Fall)	23 (Low Rise)	22 (Low Level)
醫	椅	意	兒	耳	二
cure	chair	meaning	son	ear	two

References

Arabie, P., Carroll, D., & DeSarbo, W. S. (1987). *Three way scaling*. London: Sage Publications.

Arvaniti, A., & Garding, G. (in press). Dialectal variation in the rising accents of American English. In J. Hualde, & J. Cole (Eds.) *Papers in laboratory phonology*, vol. 9. Mouton de Gruyter. Available from 〈http://ling.ucsd.edu/~arvaniti/publications.html〉; Last accessed July 20, 2007.

Atterer, M., & Ladd, D. R. (2004). On the phonetics and phonology of "segmental anchoring" of F0: Evidence from German. *Journal of Phonetics*, *32*, 177–197.

Bauer, R. S., & Benedict, P. K. (1997). *Modern Cantonese phonology*. Berlin: Mouton de Gruyter.

Beckman, M. E., & Hirschberg, J. (1994). *The ToBI annotation conventions*. Online MS. Available at 〈http://www.ling.ohio-state.edu/~tobi/ame_tobi/annotation_conventions.html〉; Last accessed June 6, 2007.

Beckman, M. E., Hirschberg, J., & Shattuck-Hufnagel, S. (2005). The original ToBI system and the evolution of the ToBI framework. In S.-A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 9–54). Oxford: Oxford University Press.

Best, C. T. (1995). A direct realistic view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience. Issues in cross-language research* (pp. 171–204). Baltimore: York Press.

Boersma, P., & Weenink, D. (2004). *Praat: Doing phonetics by computer version 4.2.04, downloaded May 10, 2004 from* 〈www.praat.org〉.

Bongaerts, T. (1999). Ultimate attainment in L2 pronunciation: The case of very advanced late L2 learners. In Birdsong, D. (Ed.) *Second language acquisition and the critical period hypothesis* (pp. 133–159). Mahwah, NJ: Lawrence Erlbaum Associates.

Chen, Y., & Xu, Y. (2006). Production of weak elements in speech—Evidence from F0 patterns of neutral tone in Standard Chinese. *Phonetica*, *63*, 47–75.

Ciocca, V., Francis, A. L., Aisha, R., & Wong, L. (2002). The perception of Cantonese lexical tones by prelingually deaf cochlear implantees. *Journal of the Acoustical Society of America*, *111*(5), 2250–2256.

Cruttenden, A. (1997). *Intonation* (2nd ed.). Cambridge: Cambridge University Press.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience. Issues in cross-language research* (pp. 233–277). Baltimore: York Press.

Fok Chan, Y. Y. (1974). *A perceptual study of tones in Cantonese*. Hong Kong: Centre of Asian Studies, University of Hong Kong.

Francis, A. L., Baldwin, K., & Nusbaum, H. C. (2000). Effects of training on attention to acoustic cues. *Perception & Psychophysics*, *62*, 1668–1680.

Francis, A. L., Ciocca, V. C., & Ng, B. K. C. (2003). On the (non)categorical perception of lexical tones. *Perception & Psychophysics*, *65*, 1029–1044.

Francis, A. L., & Driscoll, C. J. (2006). Evidence for a training-induced shift in hemispheric processing of voice onset time. *Brain and Language*, *98*, 310–318.

Francis, A. L., & Nusbaum, H. C. (2002). Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 349–366.

Fu, Q.-J., Zeng, F.-G., Shannon, R. V., & Soli, S. D. (1998). Importance of tonal envelope cues in Chinese speech recognition. *Journal of the Acoustical Society of America*, *104*, 505–510.

Gandour, J. T. (1981). Perceptual dimensions of tone: Evidence from Cantonese. *Journal of Chinese Linguistics*, *9*, 20–36.
Gandour, J. T. (1983). Tone perception in far Eastern languages. *Journal of Phonetics*, *11*, 149–175.
Gandour, J., Dzemidzic, M., Wong, D., Lowe, M., Tong, Y., Hsieh, L., et al. (2003). Temporal integration of speech prosody is shaped by language experience: An fMRI study. *Brain & Language*, *84*, 318–336.
Gandour, J. T., & Harshman, R. A. (1978). Crosslanguage differences in tone perception: A multidimensional scaling investigation. *Language and Speech*, *21*, 1–33.
Gandour, J., Tong, Y., Wong, D., Talavage, T., Dzemidzic, M., Xu, Y., et al. (2004). Hemispheric roles in the perception of speech prosody. *Neuroimage*, *23*, 344–357.
Gandour, J., Xu, Y., Wong, D., Dzemidzic, M., Lowe, M., Li, X., et al. (2003). Neural correlates of segmental and tonal information in speech perception. *Human Brain Mapping*, *20*, 185–200.
Goldstone, R. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology*, *123*(2), 178–200.
Grabe, E., Post, B., Nolan, F., & Farrar, K. (2000). Pitch accent realization in four varieties of British English. *Journal of Phonetics*, *28*, 161–185.
Grabe, E., Rosner, B. S., García-Albea, J. E., & Zhou, X. (2003). Perception of English intonation by English, Spanish, and Chinese listeners. *Language & Speech*, *46*, 375–401.
Guion, S. G., Flege, J. E., Akahane-Yamada, R., & Pruitt, J. (2000). An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *Journal of the Acoustical Society of America*, *107*, 2711–2724.
Guion, S. G., & Pederson, E. (2007). Investigating the role of attention in phonetic learning. In O.-S. Bohn, & M. Munro (Eds.), *Second-language speech learning: The role of language experience in speech perception and production: A festschrift in honour of James E. Flege* (pp. 57–77). Amsterdam: John Benjamins.
Hallé, P. A., Chang, Y. C., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics*, *32*, 395–421.
Herman, R., & McGory, J. T. (2002). The conceptual similarity of intonational tones and its effects on intertranscriber reliability. *Language and Speech*, *45*, 1–36.
Holt, L. L., & Lotto, A. J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *Journal of the Acoustical Society of America*, *119*(5), 3059–3071.
Krishnan, A., Xu, Y., Gandour, J., & Cariani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Cognitive Brain Research*, *25*, 161–168.
Ladd, D. R. (1992). An introduction to intonational phonology. In G. J. Docherty, & D. R. Ladd (Eds.), *Papers in laboratory phonology II: Gesture, segment prosody* (pp. 321–334). Cambridge: Cambridge University Press.
Ladd, D. R. (1996). *Intonational phonology*. Cambridge: Cambridge University Press.
Lee, K. Y., van Hasselt, C. A., Chiu, S. N., & Cheung, D. M. (2002). Cantonese tone perception ability of cochlear implant children in comparison with normal-hearing children. *International Journal of Pediatric Otorhinolaryngology*, *63*(2), 137–147.
Lee, Y.-S., Vakoch, D. A., & Wurm, L. H. (1996). Tone perception in Cantonese and Mandarin: A cross-linguistic comparison. *Journal of Psycholinguistic Research*, *25*, 527–542.
Li, C. N., & Thompson, S. A. (1989). *Mandarin Chinese: A functional reference grammar*. Berkeley, CA: University of California Press.
Liu, S., & Samuel, A. G. (2004). Perception of Mandarin lexical tones when F0 information is neutralized. *Language and Speech*, *47*, 109–138.
Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*, B101–B111.
McCandliss, B. D., Fiez, J. A., Protopapas, A., Conway, M., & McClelland, J. L. (2002). Success and failure in teaching the [r]–[l] contrast to Japanese adults: Tests of a Hebbian model of plasticity and stabilization in spoken language perception. *Cognitive, Affective, & Behavioral Neuroscience*, *2*(2), 89–108.
Mennen, I. (2004). Bi-directional interference in the intonation of Dutch speakers of Greek. *Journal of Phonetics*, *32*, 543–563.
Norman, J. (1988). *Chinese*. Cambridge: Cambridge University Press.
Peng, S. C., Tomblin, J. B., Cheung, H., Lin, Y. S., & Wang, L. S. (2004). Perception and production of mandarin tones in prelingually deaf children with cochlear implants. *Ear & Hearing*, *25*(3), 251–264.
Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation*. Ph.D. thesis, MIT.
Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. R. Cohen, J. Morgan, & M. E. Pollack (Eds.), *Intentions in communication* (pp. 271–311). Cambridge, MA: The MIT Press.
Pierrehumbert, J., & Steele, S. A. (1989). Categories of tonal alignment in English. *Phonetica*, *46*, 181–196.
Pisoni, D. B., Aslin, R. N., Perey, A. J., & Hennessy, B. L. (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*(2), 297–314.
Pisoni, D. B., Lively, S. E., & Logan, J. S. (1994). Perceptual learning of nonnative speech contrasts: Implications for theories of speech perception. In H. C. Nusbaum, & J. C. Goodman (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 121–166). Cambridge, MA: The MIT Press.
Polley, D. B., Steinberg, E. E., & Merzenich, M. M. (2006). Perceptual learning directs auditory cortical map reorganization through top-down influences. *The Journal of Neuroscience*, *26*(18), 4970–4982.
Samuel, A. G. (1997). Lexical activation produces potent phonemic percepts. *Cognitive Psychology*, *32*, 97–127.
Samuel, A. G. (2001). Knowing a word affects the fundamental perception of the sounds within it. *Psychological Science*, *12*, 348–351.
Vance, T. J. (1976). An experimental investigation of tone and intonation in Cantonese. *Phonetica*, *33*, 368–392.

Wang, Y., Behne, D. M., Jongman, A., & Sereno, J. (2004). The role of linguistic experience in the hemispheric processing of lexical tone. *Applied Psycholinguistics*, *25*, 449–466.

Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America*, *106*(6), 3649–3658.

Ward, G., & Hirschberg, J. (1985). Implicating uncertainty: The pragmatics of fall–rise intonation. *Language*, *61*, 747–776.

Wayland, R. P., & Guion, S. G. (2004). Training English and Chinese listeners to perceive Thai tones: A preliminary report. *Language Learning*, *54*, 681–712.

Wei, C.-G., Cao, K., & Zeng, F.-G. (2004). Mandarin tone recognition in cochlear implant subjects. *Hearing Research*, *197*, 87–95.

Whalen, D. H., & Xu, Y. (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica*, *49*, 25–47.

Wong, P. C. M. (2002). Hemispheric specialization of linguistic pitch patterns. *Brain Research Bulletin*, *59*, 83–95.

Xu, Y. (1998). Consistency of tone–syllable alignment across different syllable structures and speaking rates. *Phonetica*, *55*, 179–203.

Xu, Y. (1999). Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics*, *27*, 55–105.

Xu, Y., Gandour, J., & Francis, A. L. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *Journal of the Acoustical Society of America*, *120*(2), 1063–1074.

Xu, Y., & Liu, F. (2006). Tonal alignment, syllable structure and coarticulation: Toward an integrated model. *Italian Journal of Linguistics*, *18*(1), Available from 〈http://www.phon.ucl.ac.uk/home/yi/publications.html〉. Last accessed July 20, 2007.

Yamada, R. A., & Tohkura, Y. (1992). Perception of American English /r/ and /l/ by native speakers of Japanese. In Y. Tohkura, E. Vatikiotos-Bateson, & Y. Sagisaka (Eds.), *Speech perception, production, and linguistic structure* (pp. 155–174). Tokyo: Ohmsha.