Poznań Studies in Contemporary Linguistics 47(1), 2011, pp. 133–145

doi: 10.2478/psicl-2011-0011

CATEGORIZING MANDARIN TONES INTO LISTENERS' NATIVE PROSODIC CATEGORIES: THE ROLE OF PHONETIC PROPERTIES

CONNIE K. SO[a] AND CATHERINE T. BEST[a,b]
[a]*MARCS Auditory Laboratories, University of Western Sydney*
[b]*Haskins Laboratories, New Haven*
[a]*c.so@uws.edu.au* [b]*c.best@uws.edu.au*

ABSTRACT

This study examined whether native speakers of non-tone languages (Australian English, and French) were able to perceive foreign Mandarin tones in a sentence environment according to their native prosodic categories. Results found that both English and French speakers were able to perceptually categorize foreign tones into their intonational categories (i-Categories), and that categorizations were based on the contextual phonetic similarities of the pitch contours they perceived between Mandarin tones and their native i-Categories. Results also showed that French speakers, but not English speakers, were able to detect the fine-detailed phonetic feature differences between Tone 3 and Tone 4 (low/falling tone vs. high-falling tone). The findings support a new extension of the Perceptual Assimilation Model (PAM; Best 1995) to suprasegmental phonology (So and Best 2008): that non-native prosodic categories (e.g. lexical tones) will be assimilated to the categories of listeners' native prosodic system (e.g. intonation). In addition, rhythmic differences among languages may also contribute to perception of non-native tones.

KEYWORDS: Lexical tone perception; cross language perception; Perceptual Assimilation Model (PAM); phonetic influence; rhythmic properties.

1. Introduction

Our linguistic experience not only guides our perception of non-native segments (Best et al. 1998, 2001; Flege et al. 1997; Polka 1995) but also non-native suprasegmentals, such as lexical tones (Lee et al. 1996; So 2006; So and Best 2010a; Wayland and Guion 2004). Recent studies have demonstrated that adult listeners perceive non-native tones in citation form according to the prosodic categories of their native language system, such as tone, pitch-accent, and intonation (So and Best 2008, 2010b, in preparation). This study examined further how native speakers of two non-tone languages (English, a stress-timed language, and French, a syllable-timed language) perceived Mandarin tones *in a sentence environment* in terms of their intonation categories. The results will widen our knowledge of how listeners categorize non-native tones in a contextual environment and how their performance may be affected by the rhythmic, as well as intonational, properties of their native language.

Prior studies showed that native speakers from non-tone languages are able to categorize and/or discriminate non-native (Mandarin) tones to some extent, although they tend to be less sensitive to some properties of the target non-native tones (Gandour 1983, 1984; Hallé et al. 2004; Leather 1983). For example, English listeners tended to focus on pitch height, while listeners from Chinese languages (e.g. Cantonese and Mandarin) focused on both pitch height and pitch direction when discriminating tones (Gandour 1983, 1984). In addition, native speakers from a tone language might also outperform those from a non-tone language (Lee et al. 1996; Wayland and Guion 2004), because their linguistic experience of using pitch variations might help them discriminate non-native tone better. However, in some cases, the linguistic experience of using lexical tones might actually *increase* listeners' difficulty in identifying non-native tones, and seemingly paradoxically, might lead them to perform more *poorly* than those from non-tone languages (So 2006; So and Best 2010a).

It has been suggested that the perception of non-native tones, similar to that for non-native segments, is greatly affected by the phonetic and phonological properties of listeners' native language (So 2006; So and Best 2010a). Further, listeners can assimilate non-native tones to the categories of their native prosodic systems, such as tone, pitch-accent, and intonation (So 2006; So and Best 2008, 2010a), in ways that appear consistent with the assumptions of the Perceptual Assimilation Model (PAM; Best 1995), specifically that listeners perceptually assimilate non-native categories to their native categories, and that they perceive non-native contrasts in terms of several assimilation types, such as Single Category (SC) assimilation, Category Goodness (CG) assimilation, and Two Category (TC) Assimilation. This raises an important question as to how adults perceive non-native lexical tones, especially if they lack lexical tone distinctions in their native language. Do they perceive non-native tones according to some properties of their native prosodic categories, such as pitch patterns of their intonational categories? For example, they may categorize a rising tone according to the rising pitch patterns of their intonational category, *Question*.

A recent study (So and Best 2008) has demonstrated that native English (NE) listeners can categorize non-native tones in citation form – i.e., on individual single words – in terms of their intonational categories (i-Categories). In general, Mandarin Tone 1 (*High level*) is categorized as *Flat Pitch*, Tone 2 (*mid-rising*) as *Question*, Tone 3 (*falling-rising*) as *Uncertainty* (although some NE listeners categorized it as *Question*), and Tone 4 (*high falling*) as *Statement*. The findings supported the new extension of PAM predictions, specifically that non-native tonal categories (e.g. lexical tones) will be assimilated to the categories of listeners' native prosodic system. The study also suggested that NE listeners assimilated the phonetic properties of Mandarin tones (e.g. pitch patterns) to those of English i-Categories, when both substantially shared similar phonetic features.

However, how do listeners categorize foreign tones when they are embedded in a sentential environment, which is more naturalistic and more relevant to, e.g., second language learning of a tone language? Do they categorize the foreign tones according to the prosodic categories of their native languages? In addition, it is well documented that the effects of tonal coarticulation (anticipation and carryover) will be involved in connected speech (Xu 1994, 1997). Therefore, do the contextually-varying phonetic characteristics (e.g. rising and falling pitch patterns) of foreign tones in sentences affect how listeners from non-tonal languages categorize them? Further, rhythm is another prosodic aspect on which languages can be classified, e.g. as having stress-timed versus syllable-timed rhythmic structures. Rhythmicity is an intrinsic characteristic of a language's prosodic system but refers to temporal (timing) rather than spectral patterning (F0) of syllables in the language. Therefore, it would be important and of theoretical interest to know whether there is any difference in the perceptual assimilations between native speakers of non-tonal languages with different rhythmic properties.

To answer the above questions, the present study examined the perception of Mandarin tones by native speakers of two non-tonal languages: English, a stress-timed or stress accented language (Beckman 1986), and French, a syllable-timed language without an accent system (Fox 2000). The new prediction of PAM for suprasegmentals (So and Best 2008) was tested by investigating how native speakers of these non-tone language groups categorized Mandarin tones in a sentence frame according to their native intonational categories (*Flat pitch, Question, Statement, and Exclamation*). Since the tones were in a sentential environment, listeners' categorizations should be based on the *contextual phonetic similarities* of the pitch contours they perceived between Mandarin tones (in the sentential form) and their native i-Categories. Thus, it was predicted that they would perceive Tone 1 as *Statement* (this level tone might be perceived as a tone with a slight falling movement) rather than *Flat Pitch*, Tone 2 as *Question*, Tone

3 (a low falling/low tone) as *Statement*, and Tone 4 as *Exclamation* (a falling tone with a more steeply falling pitch movement) rather than *Statement*.

2. Method

2.1. Participants

Thirty Australian NE speakers (18–24 years of age) and thirty native French (NF) speakers (21–37 years of age) were recruited as participants. They were all either undergraduate students at the University of Western Sydney, who received course credits after they completed the experiment, or residents living in Sydney at the time of the experiment who received AUD $40 for their participation. They had neither learned Mandarin or other lexical tone languages, nor received formal musical training, as previous studies have shown that listeners with musical training outperformed those without such training in both production and perception tasks with non-native tones (Alexander et al. 2005; Burnham and Brooker 2002; Gottfried and Riester 2000). Before they performed the experiment, they all passed a pure-tone hearing screening (250–8000 Hz at 25 dB HL).

2.2. Stimuli

The stimuli for this study were produced by three native female Mandarin speakers (mean age: 24 years). They were asked to produce the four Mandarin tones on the syllable /fu/ in a statement frame. (In Chinese PinYin: *xia4 yi1 ge4 shi4 X zi4*, where the number indicates the tone on the word, and X indicates the target word; the English gloss is 'the next one is the X word'.) The syllable /fu/ was selected because its pronunciation is close to the one for the English word *fool,* and similar to that of the French word *fou,* which means 'crazy'. Five tokens of each target word (/fu/ with each of the four tones) were produced by each speaker. Among them, 3 samples per tone word per speaker were verified perceptually by another three native Mandarin speakers (mean age: 27.7 years) to ensure the selected stimuli were intelligible to native Mandarin speakers. All of the perceptual stimuli were correctly identified by the native speakers (see Figure 1). Note that Tone 3 (rising-falling) is produced as a low level or a low falling tone in connected speech, rather than the "dipping" pattern often found in citation-form productions.

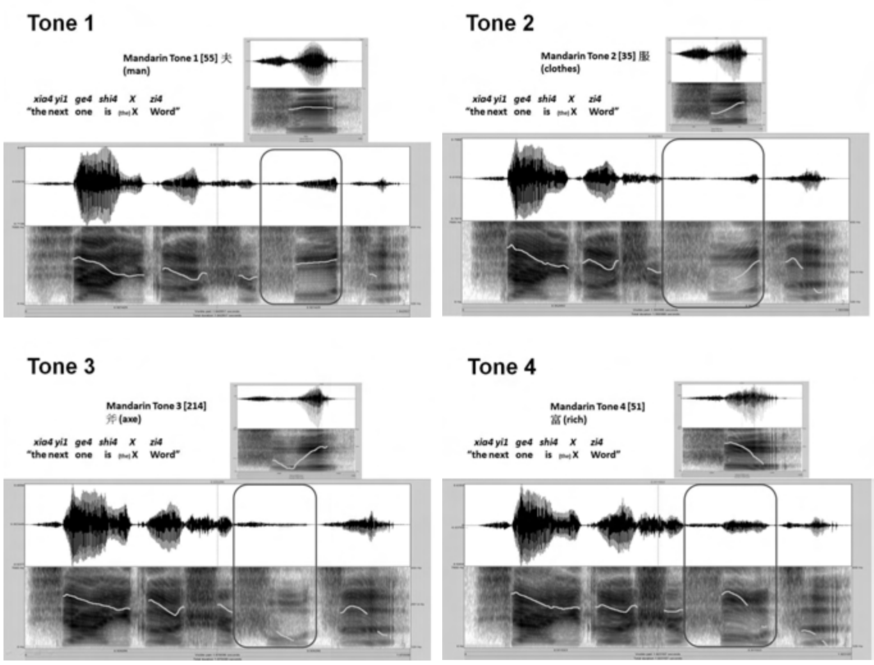

Figure 1. Samples of the four Mandarin stimuli: Tone 1 (upper left), Tone 2 (upper right), Tone 3 (lower left), and Tone 4 (lower right). In each cell, the upper row shows the target tone in citation form (for comparison), the meaning of the target word and the sentence frame used for this study. The bottom row shows the pitch contour of the target tone in a sentential environment. The rectangle boxes highlight the target tones.

2.3. Procedure

Participants were asked to categorize randomized individual presentations of 72 trials of the target word (/fu/-tones) in these stimulus sentences (3 speakers × 4 tones × 3 tokens per tone × 2 repetitions) into four English/French i-Categories – *Flat pitch*, *Question*, *Statement*, and *Exclamation*. In this study, the same i-Categories were provided to both the NE and NF groups, as these four i-Categories are common to both languages, and share similar pitch contours (Hardison 2004; Ladd 1996; Post 2002). During the experiment, the tokens of the stimulus sentences were presented individually from a PC, on the screen of which five buttons were provided, corresponding to the four i-Categories and a

5th button labelled as *Unknown*. Listeners were instructed to select *Unknown* when they could not identify a target-word's tone into any i-Category. Note that although the labels were in English (i.e., no French translations), all French listeners were also fluent in English. Careful instructions and a trial block of 8 samples had been given to both NE and NF listeners before the experiment to ensure they all understood that they were asked to categorize the perceived target tones into the i-Categories of their own native language. None of the French participants had any difficulties with doing this.

3. Results

3.1. Tonal categorization of Native English (NE) speakers

NE listeners' tonal categorizations for each tone (in %) are shown in Figure 2. Individual t-tests were carried out to test each i-Category mean for each target tone, against chance level (20%). The results confirmed that the means of the following i-Categories for their target tones were all significantly above the chance level (20%): *Question* [$t(26) = 1.880, p < .05$] and *Statement* for Tone 1 [$t(29) = 7.880, p < .01$], *Question* [$t(29) = 5.799, p < .01$] and *Statement* [$t(28) = 4.950, p < .01$] for Tone 2, *Question* and *Statement* for Tone 3 [$t(28) = 2.738$ and 6.538, $ps < .01$], and *Statement* and *Exclamation* for Tone 4 [$t(28) = 9.029$ and 4.194, $ps < .01$]. A Chi-square test revealed a significant association between Tones (4) and i-Categories (5), χ^2 (12) = 165.794, $p < 0.001$. A further mixed design 2-way ANOVA[1] (Tone × i-Category) found no significant effect of Tone (*n.s.*), but a significant effect of i-Category [$F(3,389) = 51.952, p < .001$] on listeners' mean assimilation percentage (%). The Tone × i-Category interaction was also significant [$F(9,389) = 7.722, p < .001$].

Individual 1-way ANOVAs for the four tones were carried out to investigate the i-Category effect for each tone target. It was found that the i-Category effect was significant for each tone: Tone 1 [$F(3,105) = 21.615, p < .0001$], Tone 2 [$F(3,92) = 13.077, p < .001$], Tone 3 [$F(3,95) = 13.65, p < .001$], and Tone 4 [$F(3, 97) = 30.005, p < .001$]. Post-hoc HSD Tukey tests further indicated the following results for each tone. For Tone 1, the mean percentage (%) of *Statement* assimilations (41%) was significantly greater than each of the other counterparts: *Flat Pitch* (13%), *Question* (23%), and *Exclamation* (20%) assimilations (*ps*

---

[1] Analysis was performed without "Unknown" responses, which contributed to 3.33% (72 counts) of total responses (2160 counts).

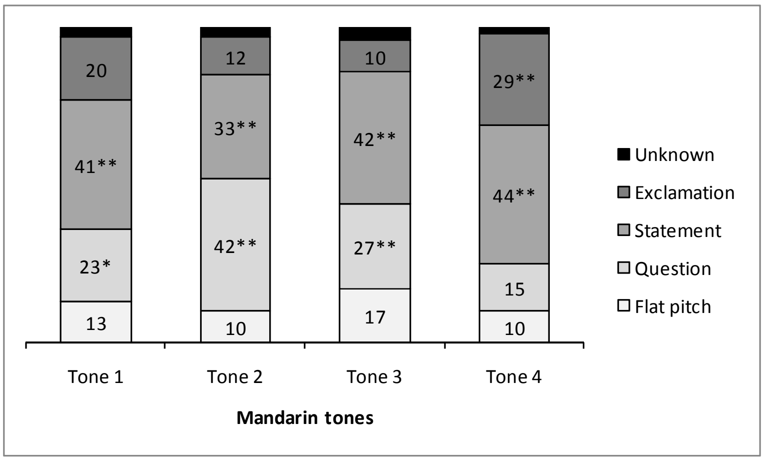

Figure 2. Listeners' tonal categorizations for each tone (in %).
The total number of responses for each tone category was 540.
Categories that were used 5% or less are not labelled.
The symbols * (p<.05) and ** (p<.01) show that the mean
of the chosen i-Category is above the chance level (20%).

$< .05$). For Tone 2, the mean percentage of *Question* assimilations (42%) was significantly greater than those of *Flat Pitch* (10%), and *Exclamation* (12%) assimilations (*ps* $< .01$), but did not differ significantly from that of *Statement* assimilations (33%; *n.s.*), which was selected significantly more often than *Flat Pitch* and *Exclamation* assimilations (*ps* $< .01$). For Tone 3, the mean percentage of *Statement* assimilations (42%) was significantly greater than those of *Flat Pitch* (17%), *Question* (27%), and *Exclamation* (10%) assimilations (*ps* $< .01$). In addition, the mean percentage of *Question* assimilations was significantly greater than that of *Exclamation* assimilations ($p < .05$). For Tone 4, the mean percentage of *Statement* assimilations (44%) was significantly greater than those of *Flat Pitch* (10%), *Question* (15%), and *Exclamation* (29%) assimilations (*ps* $< .01$). The mean percentage of *Exclamation* assimilations was significantly greater than those of *Flat Pitch* and *Question* assimilations (*ps* $< .01$).

3.2. Tonal categorization of Native French (NF) speakers

NF listeners' tonal categorizations for each tone (in %) are shown in Figure 3. Individual t-tests were carried out to test each i-Category mean for each target

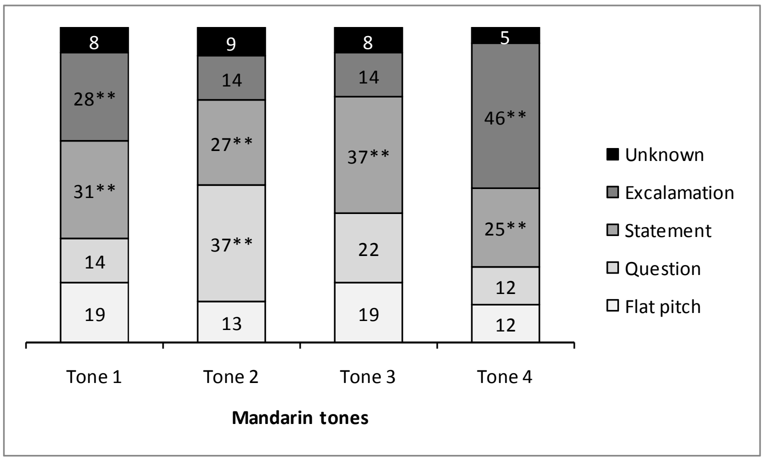

Figure 3. Listeners' tonal categorizations for each tone (in %).
The total number of responses for each tone category was 540.
Categories that have 5% or less are not labelled. The symbol ** (p<.01)
shows that the mean of the i-Category is above the chance level (20%).

tone, against chance of 20%. The results confirmed the means of the following i-Categories for their target tones were all significantly above the chance level (20%): *Exclamation* [$t(27) = 3.691$, $p < .001$] and *Statement* for Tone 1 [$t(29) = 4.198$, $p < .001$], *Question* [$t(29) = 4.286$, $p < .001$] and *Statement* [$t(28) = 3.270$, $p < .001$] for Tone 2, *Statement* for Tone 3 [$t(29) = 4.957$, $p < .001$], and *Statement* [$t(26) = 3.463$, $p < .001$] and *Exclamation* [$t(29) = 6.461$, $p < .001$] for Tone 4. A Chi-square test revealed a significant association between Tones (4) and i-Categories (5), χ^2 (12) = 262.22, $p < 0.001$. A further mixed design 2-way ANOVA[2] (Tone × i-Category) found no significant effect of Tone (*n.s.*), but a significant effect of i-Category [$F(3, 403) = 13.045$, $p < .001$] on listeners' mean assimilation percentage (%). The Tone × i-Category interaction was also significant [$F(3,403) = 10.176$, $p < .001$].

Individual 1-way ANOVAs for the four tones were carried out to investigate the i-Category effect for each tone target. It was found that the i-Category effect was significant for each tone: Tone 1 [$F(3,102) = 5.037$, $p < .001$], Tone 2 [$F(3,102) = 10.327$, $p < .001$], Tone 3 [$F(3, 102) = 7.132$, $p < .001$], and Tone 4 [$F(3, 95) = 20.742$, $p < .001$]. Post-hoc HSD Tukey tests further indicated the

[2] Analysis was performed without "Unknown" responses, which contributed to 7.5% (162 counts) of total responses (2160 counts).

following results for each tone. For Tone 1, the mean percentage (%) of *Statement* assimilations (31%) was significantly greater than *Question* (14%; $p < 0.01$) and *Flat Pitch* assimilations (19%; $p < 0.05$). In addition, the mean of *Exclamation* assimilations (28%) was also significantly greater than that of *Question* assimilations ($p < 0.05$). For Tone 2, the mean percentage of *Question* assimilations (37%) was significantly greater than those of *Flat Pitch* (13%), and *Exclamation* (14%) assimilations (*ps* < .001), but did not differ significantly from that of *Statement* assimilations (27%; *n.s.*). The mean % of *Statement* assimilations was significantly greater than that of *Flat Pitch* assimilations ($p < 0.01$). For Tone 3, the mean percentage of *Statement* assimilations (37%) was significantly greater than those of *Flat Pitch* (19%), *Question* (22%), and *Exclamation* (14%) assimilations (*ps* < .01). For Tone 4, the mean percentage of *Exclamation* assimilations (46%) was significantly greater than those of *Flat Pitch* (12%), *Question* (12%), and *Statement* (25%) assimilations (*ps* < .001). The mean percentage of *Statement* assimilations was significantly greater than that of *Flat Pitch* assimilations ($p < .01$).

4. Discussion

The results, as expected, clearly showed that both NE and NF speakers were able to categorize Mandarin tones (embedded in a sentence frame) into their native prosodic/intonational categories. Their selections depended on *contextual phonetic similarities* between the pitch contours of the prosodic categories of Mandarin and those of English and of French (see Table 1 for the summary of their tonal categorizations).

Table 1. Summary of the categorizations of the four Mandarin tones by the native English and French speakers. The symbols, ✓, × and **?** indicate the prediction was supported, not supported and partially supported.

Predictions	Aus. English	Support	French	Support
Tone 1 → Statement	Statement	✓	Statement,	✓
			Exclamation	?
Tone 2 → Question	Question,	✓	Question,	✓
	Statement	×	Statement	×
Tone 3 → Statement	Statement	✓	Statement	✓
Tone 4 → Exclamation	Statement	×	Exclamation	✓

For Australian English speakers, Tone 1 was more perceived as a *Statement*, Tone 2 was more perceived as a *Question*. However, NEs also assimilated Tone 2 to their *Statement* i-category, which might reflect their perception of the overall falling pitch movement of the target word within the sentence frame (this will be discussed later in this section). Tone 3 was perceived more as a *Statement,* but it was sometimes perceived as a *Question* (also to be discussed later in this section). Tone 4 was perceived mainly as *Statement*, although the pitch contour of Tone 4 involves a greater falling movement and steeper slope (a fall of -162.09 Hz from its maximum pitch, on average) than that of Tone 3 (a fall of -51.64 Hz, on average). NEs were clearly able to perceive the falling pitch movement, but failed to perceive the fine-gained phonetic difference in rate/extent of F0 decline. Similarly, native French speakers (NFs) were able to assimilate Mandarin tones into their i-Categories. NFs perceived Tone 1 as both *Statement* and *Exclamation*. This may due to the fact that both share the falling pitch feature (either with the *Statement* i-category or with the sentence environment), and the involvement of the high pitch of Tone 1 together with the descending (falling) pitch direction (due to the sentence frame) might obscure their perception of the target word's tone clearly. NFs perceived Tone 2 primarily as a *Question*, then as a *Statement* (this will be discussed later in this section). They assimilated Tone 3 to *Statement*, and Tone 4 to *Exclamation;* both these patterns are different from those of the NEs.

Listeners' categorizations of the Mandarin tones in a sentence environment were affected by both the *overall descending pitch tendency of the sentence frame*, and *the contextually-varying phonetic characteristics* (e.g., rising and falling pitch contours) *of the target word's tones within the sentences*. The former characteristic has clearly exerted some perceptual influence on NEs and NFs, and caused them sometimes to perceive Tone 2, the rising tone, as their *Statement* category (NE: 33%; NF: 27%). The latter characteristic, however, actually also reflects tonal coarticulation effects, both anticipation and carry-over effects (Xu 1994, 1997), which might also obscure listeners' categorizations of the target tones themselves (as opposed to the sentence contour) to some extent. For example, listeners perceived Tone 1 (high level) as their statement category (which generally has a falling pitch contour in both English and French). In addition, listeners sometimes perceived Tone 3 as they had perceived Tone 2. Tone 3 is generally produced with a low level/falling pitch contour in a sentence environment, and the production of the word following the target word, *zi*, involves a higher pitch at its onset, because it has a high falling tone (Tone 4). Thus, the anticipatory coarticulation effect might create an illusion of a rising pitch pattern on the target word to the listeners. This explanation may also apply to NEs' perception of Tone 1, which 23% of time was perceived in the same way as they had perceived Tone 2, i.e. as a rising tone.

The rhythmic properties of different language classes also appear to be influencing the perception of non-native tones by non-tone language speakers. Native speakers of French, a syllable timed language (Fox 2000) appear to perceive the fine-gained tonal phonetic features or feature changes across the target words in the sentence context better than the native speakers of English, a stress timed language (Fox 2000), possibly because Mandarin, like French but unlike English, is a syllable timed language (Chen et al. 2001; Smit 2004). That is, both French and Mandarin, but not English, syllables tend to maintain a regular timing interval (syllable duration) throughout an utterance. These similar rhythmic properties might help NFs locate the pitch contour of the target word better during the perception of the whole sentence. As a result, they were better able to perceive Tone 4 as *Exclamation*, a French i-Category that involves a greater falling movement and steeper slope relative to Statement, than English listeners were.

5. Conclusion

The results of the present study indicated that both NE and NF listeners assimilated non-native tones to their native intonational categories (Tone 1 and Tone 3 as *Statement*, Tone 2 as *Question*, and Tone 4 as *Statement* for NE speakers but as *Exclamation* for NF speakers) that share (contextual) phonetic similarities with those of Mandarin tones in a sentence environment. However, their perception appears to be affected to some extent by the effect of tonal coarticulation in connected speech. In addition, while NE listeners were unable to detect the fine phonetic difference between Tone 3 (involving a slight falling pitch pattern) and Tone 4 (involving a dramatic falling pitch pattern), NF listeners were better able to perceive the phonetic difference between Mandarin Tone 3 and Tone 4, which they perceived as similar to two native i-Categories, *Statement* and *Question*, whereas NE listeners perceived them as exemplars of a single native i-Category, *Statement*. The difference may be attributed to differences in the rhythmic (temporal) properties of their native languages. Thus, the overall results further affirm the assumption that non-tone listeners (NE and NF) assimilate non-native prosodic categories (e.g., tones) to their native prosodic categories based on the (contextual) phonetic similarities they perceive. Moreover, the present findings also suggest that the rhythmic (temporal) properties of the listener's native prosodic system play a role on the assimilation of non-native tones to native prosodic categories (i.e., interaction).

6. Acknowledgements

This work is supported by the Australian Research Council (PI: C. So). We thank Arman Abrahamyan and Asako Kafafy for their assistance with data collection.

REFERENCES

Alexander, J.A., P.C.M. Wong, and A.R. Bradlow. 2005. "Lexical tone perception by musicians and non-musicians". *Proceedings of Interspeech 2005*, Lisbon.

Beckman, M.E. 1986. *Stress and non-stress accent*. Dordrecht: Foris Publications.

Best, C.T. 1995. "A direct realist view of cross-language speech perception". In: Strange, W. (ed.), *Speech perception and linguistic experience: Issues in cross-language research*. Timonium, MD: York Press. 171–232.

Best, C.T., G.W. McRoberts and N.M. Sithole. 1988. "Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants". *Journal of Experimental Psychology: Human Perception and Performance* 14(3). 345–360.

Best, C.T., G.W. McRoberts and E. Goodell. 2001. "Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system". *Journal of the Acoustical Society of America* 109(2). 775–794.

Burnham, D. and R. Brooker. 2002. "Absolute pitch and lexical tones: Tone perception by non-musician, musician, and absolute pitch non-tonal language speakers". In: Hansen, J. and B. Pellom (eds), *The 7 th International Conference on Spoken Language Processing*, Denver. 257–260.

Chen, Y., M.P. Robb, H.R. Gilbert and J.W. Lerman. 2001. "A study of sentence stress production in Mandarin speakers of American English". *Journal of the Acoustical Society of America* 109(4). 1681–1690.

Flege, J., O.-S. Bohn and S. Jang. 1997. "The effect of experience on nonnative subjects' production and perception of English vowels". *Journal of Phonetics* 25. 437–470.

Fox, A. 2000. *Prosodic features and prosodic structure: The phonology of suprasegmentals*. New York: Oxford University Press.

Gandour, J.T. 1983. "Tone perception in Far Eastern languages". *Journal of Phonetics* 11. 149–175.

Gandour, J.T. 1984. "Tone dissimilarity judgments by Chinese listeners". *Journal of Chinese Linguistics* 12. 235–261.

Gottfried, T.L. and D.N. Riester. 2000. "Preliminary results from study of pitch perception and Mandarin tone identification". *Journal of the Acoustical Society of America* 108. 2604.

Hallé, P.A., Y.-C. Chang and C.T. Best. (2004). "Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners". *Journal of Phonetics* 32(3). 395–421.

Hardison, D.M. 2004. "Generalization of computer-assisted prosody training: Quantitative and qualitative findings". *Language Learning & Technology* 8. 34–52.

Ladd, D.R. 1996. *Intonational phonology*. Cambridge: Cambridge University Press.

Leather, J. 1983. "Speaker normalization in perception of lexical tones". *Jounral of Phonetics* 11. 373–382.

Lee, Y.-S., D.A. Vakoch and L.H. Wurm. 1996. "Tone perception in Cantonese and Mandarin: A cross-linguistic comparison". *Journal of Psycholinguistic Research* 25(5). 527–542.

Polka, L. 1995. "Linguistic influences in adult perception of non-native vowel contrasts". *Journal of Acoustical Society of America* 97(2). 1286–1296.

Post, B. 2002. "French tonal structures". In: Bel, B. and I. Marlien (eds.), *Proceedings of the Speech Prosody 2002 conference*, Aix-en-Provence. 583–586.

Smit, A.B. 2004. *Articulation and phonology resource guide for school-age children and adults*. New York: Thomson Delmar Learning

So, C.K. 2006. Effect of L1 prosodic background and AV training on learning Mandarin tones by speakers of Cantonese, Japanese, and English. (PhD dissertation, Simon Fraser University.)

So, C.K. and C.T. Best. 2008. "Do English speakers assimilate Mandarin tones to English prosodic categories?" *Proceedings of Interspeech 2008 incorporating SST 2008*, Brisbane.

So, C.K. and C.T. Best. 2010a. "Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences". *Language and Speech* 53(2). 273–293.

So, C.K. and C.T. Best. 2010b. "Discrimination and categorization of Mandarin tones by Cantonese speakers: The role of native phonological and phonetic properties". Paper presented at the 13th Australasian International Conference on Speech Science and Technology, Melbourne.

So, C.K. and C.T. Best. In preparation. "Discrimination and categorization of Mandarin tones by Cantonese, Japanese, and English speakers: The role of phonological and phonetic properties".

Wayland, R.P. and S.G. Guion. 2004. "Training English and Chinese listeners to perceive Thai tones: A preliminary report". *Language Learning* 54(4). 681–712.

Xu, Y. 1994. "Production and perception of coarticulated tones". *Journal of the Acoustical Society of America* 95(4). 2240–2253.

Xu, Y. 1997. "Contextual tonal variations in Mandarin". *Journal of Phonetics* 25. 61–83.

Corresponding author:
Connie K. So
MARCS Auditory Laboratories, University of Western Sydney
Locked Bag 1797
Penrith South DC, NSW 1797
Australia
c.so@uws.edu.au