Just noticeable differences for pitch direction, height, and slope for Mandarin and English listeners

Allard Jongman, Zhen Qin, Jie Zhang, and Joan A. Sereno

Citation: The Journal of the Acoustical Society of America **142**, EL163 (2017);

View online: https://doi.org/10.1121/1.4995526

View Table of Contents: http://asa.scitation.org/toc/jas/142/2

Published by the Acoustical Society of America

Jongman *et al.*: JASA Express Letters [http://dx.doi.org/10.1121/1.4995526] Published Online 1 August 2017

Just noticeable differences for pitch direction, height, and slope for Mandarin and English listeners

Allard Jongman,[a] Zhen Qin, Jie Zhang, and Joan A. Sereno
Department of Linguistics, University of Kansas, 1541 Lilac Lane, Lawrence, Kansas 66045-3129, USA
jongman@ku.edu; qinzhenquentin2@ku.edu; zhang@ku.edu; sereno@ku.edu

Abstract: Previous studies on tones suggest that Mandarin listeners are more sensitive to pitch direction and slope while English listeners primarily attend to pitch height. In this study, just noticeable differences were established for pitch discrimination using a three-interval, forced-choice procedure with a two-down, one-up staircase design. A high rising and a high falling Mandarin tone were manipulated in terms of pitch direction, height, and slope. Results indicate that, overall, Mandarin listeners are more sensitive to pitch slope and English listeners to pitch height. However, these effects are modulated by both the direction (falling/rising) and slope of the pitch contours.

[RS]
Date Received: May 16, 2017 **Date Accepted:** July 11, 2017

1. Introduction

Mandarin Chinese (henceforth, Mandarin) distinguishes four lexical tones as in /pā/ "eight" (a high level pitch, Tone 1), /pá/ "to pull out" (a high rising pitch, Tone 2), /pǎ/ "to hold" (a low dipping pitch, Tone 3), and /pà/ "father" (a high falling pitch, Tone 4). These tonal contrasts are cued by both pitch height and the direction of pitch change within a syllable. Listeners with different language backgrounds use different cues to perceive lexical tones. Previous studies (e.g., Burnham and Mattock, 2007; Francis *et al.*, 2008; Gandour, 1983) found that while Mandarin listeners are more sensitive to pitch direction, English listeners attend more to pitch height as a cue. English listeners were also found to be less sensitive to pitch slope (Gandour and Harshman, 1978), although the relative contribution of slope seems to vary as a function of the number of dimensions manipulated (Gandour, 1983). Studies on the categorical perception of lexical tone suggest that speakers of non-tone languages may have a psychoacoustic advantage over speakers of tone languages. While Mandarin listeners are not sensitive to minor pitch changes in their tonal categorization, English listeners are able to detect subtle pitch variations that correspond to within-category differences for native Mandarin listeners (Leather, 1987; Stagray and Downs, 1993).

Few psychophysical studies have examined tone perception in listeners of tone and non-tone languages. Most notably, Liu (2013) measured just noticeable differences (JNDs) of tone discrimination for Mandarin and English listeners. Level, rising, and falling tones within or across tone boundaries were presented as standard stimuli, and tones at lower and higher $F0$ values at onset or offset served as test stimuli in a three-interval design. On each trial, a standard tone was presented followed by a standard and a test tone in random order.

Consistent with prior studies, Liu (2013) found that the discrimination pattern of tones within and across boundaries indicated that Mandarin listeners perceived lexical tones categorically, whereas the English listeners did not. The JND results also showed the effect of listeners' language background: while the Mandarin and English listeners did not differ in their performance for level tones, they did differ in their perception of falling and rising tones. Specifically, the Mandarin listeners had a significantly higher JND than the English listeners in discriminating pitch change at the offset of falling tones, and the English listeners showed a significantly higher JND than the Mandarin listeners in discriminating change at the onset of rising tones. Since position (i.e., onset vs offset) and pitch direction (i.e., falling vs rising) co-vary in Liu's

[a]Author to whom correspondence should be addressed.

design, it is unclear whether the two groups of listeners have different sensitivities to the onset versus offset of tone contours or to the rising versus falling tone contours.

Given that position and tone contour are not separately examined in Liu's (2013) study, additional research is necessary to establish and compare the JNDs of Mandarin and English listeners. The primary goal of the current study is to examine whether language background influences listeners' use of pitch direction, height, and slope in tone perception. Extending Liu (2013), JNDs of pitch contour change were measured for Mandarin and English listeners. Given previous findings on the use of pitch cues (e.g., Burnham and Mattock, 2007; Francis *et al.*, 2008; Gandour and Harshman, 1978; Gandour, 1983), we predict that Mandarin and English listeners will show different sensitivities to pitch direction, height, and slope at the psychophysical level. While Mandarin listeners are predicted to show a lower JND than English listeners for pitch direction and slope, English listeners are predicted to show a lower JND than Mandarin listeners for pitch height.

2. Method

2.1 Listeners

Twenty-one native speakers of Mandarin and 21 native speakers of English participated in our experiment. All were students at the University of Kansas. The Mandarin listeners were Mandarin-English bilinguals who started learning English after age 12 in China and had lived in the U.S. for less than 4 years. Thirteen Mandarin listeners reported their English proficiency as advanced, with a self-reported estimate of frequency of daily English usage ranging from 20%–80% (mean: 57%). The remaining 8 Mandarin listeners reported their English proficiency as intermediate, with estimates of their frequency of daily usage ranging from 50%–95% (mean: 74%). In addition, 11 of the 21 Mandarin listeners reported having played a musical instrument (including voice), with a range from 0.5–11 yrs (mean: 3.45 yrs).

The English listeners were monolingual speakers with no previous exposure to Mandarin or any other tone language. Nineteen reported having played a musical instrument, with a range from 1 to 15 yrs (mean: 6.1 yrs). None of our participants reported any speech or hearing impairment. All participants were paid for their participation.

2.2 Stimuli

First, a young female native speaker of Mandarin produced the vowel /a/ with a high falling tone (Tone 4) and with a high rising tone (Tone 2). Since we adopted a three-interval, forced-choice procedure in our experiment, standard and test tones were used as stimuli. Then, a standard contour falling from 200 Hz at tone onset to 170 Hz at tone offset and another standard contour tone rising from 200 Hz at tone onset to 230 Hz at tone offset were resynthesized from the natural falling tone (Tone 4) and the rising tone (Tone 2), respectively, using the time-domain "pitch synchronous overlap add" method in Praat. The standard tones had a duration normalized to 295 ms (original durations of the falling and rising tones were 213 and 377 ms, respectively) and intensity to 70 dB sound pressure level (SPL) (original falling tone: 69 dB SPL; rising tone: 67 dB SPL). Finally, test tones were generated by systematically manipulating the pitch contour of the falling and rising standard tones in terms of pitch height and pitch slope.

For pitch height, tones that were parallel to the standard falling and rising tones in pitch contour but started (and ended) at lower and higher $F0$ values (i.e., ranging from 1 to 15 Hz in 1 Hz steps) were resynthesized as test stimuli. The use of 1 Hz steps ensured that we would not be underestimating our listeners' sensitivity (Flanagan and Saslow, 1958). The range of 1–15 Hz was partially based on Liu (2013) who had used a range of 1–20 Hz. Based on his results, it was clear that 1–15 Hz would be sufficient to obtain reliable JND estimates.

The test stimuli for the falling tone (200–170 Hz) had onsets ranging either from 215 to 200 Hz (higher condition) or from 200 to 185 Hz (lower condition), as illustrated in the top-left panel in Fig. 1. The test stimuli for the rising tone (200–230 Hz) had onsets ranging either from 200 to 215 Hz (higher condition) or from 185 to 200 Hz (lower condition), as shown in the top-right panel.

For pitch slope, tones that were identical to the standard falling and rising tones in average pitch height but had either a shallower or steeper slope were resynthesized as test stimuli. These stimuli were derived by adjusting the onset and offset of the standard tone in opposite directions by 1 to 15 Hz in 1 Hz steps. The test

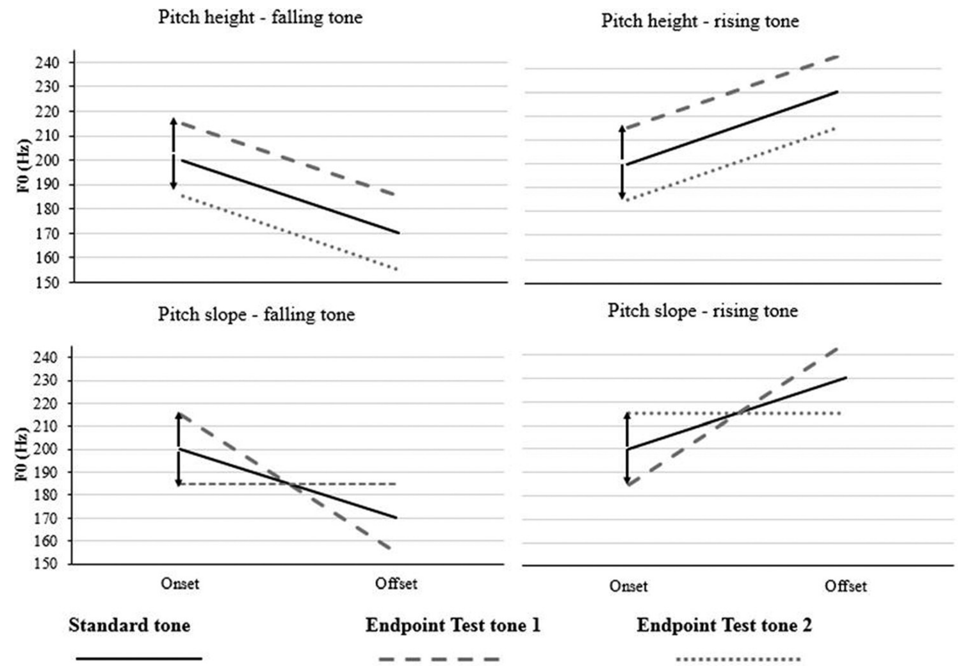

Fig. 1. Schematic representation of *F*0 contours varying in pitch height (top panels) and pitch slope (bottom panels) for the falling tone (left panels) and rising tone (right panels).

stimuli for the falling tone (200–170 Hz) had onsets ranging from 200 to 185 Hz and offsets ranging from 170 to 185 Hz in the shallower condition and had onsets ranging from 200 to 215 Hz and offsets ranging from 170 to 155 Hz in the steeper condition, as illustrated in the bottom-left panel in Fig. 1. The test stimuli for the rising tone (200–230 Hz) had onsets ranging from 200 to 215 Hz and offsets ranging from 230 to 215 Hz in the shallower condition and had onsets ranging from 200 to 185 Hz and offsets ranging from 230 to 245 Hz in the steeper condition, as shown in the bottom-right panel.

Thus, there were eight conditions in total: Four for the pitch height stimuli (i.e., falling-higher, falling-lower, rising-higher, and rising-lower), and four for the pitch slope stimuli (i.e., falling-shallower, falling-steeper, rising-shallower, and rising-steeper).

2.3 Procedure

Using a three-interval, forced-choice procedure with a two-down, one-up tracking algorithm (Levitt, 1971), JNDs of tone pitch discrimination were measured for each condition following Liu (2013). It should be noted that while Liu manipulated only the onset or the offset of the tones, we manipulated both simultaneously. For example, a JND of 7 Hz for the slope condition would mean a 7 Hz difference at the onset and a 7 Hz difference in the opposite direction at the offset. The experiment used a staircase script adopted from Hairston and Maldjian (2009) in E-Prime. On each trial, a standard tone was presented in the first interval, followed by another standard tone and a test tone randomly ordered in the second and third intervals. Similar to Liu (2013), a 400 ms inter-stimulus interval was selected, which ensured that perception was based on language-specific phonological (rather than auditory) memory (e.g., Werker and Logan, 1985). Listeners were instructed to focus on the pitch contour of the first tone and to judge which of the last two tones was different from the first one. The listeners were required to respond within 10 s after each trial.

The stimuli were blocked within each condition with four blocks for pitch height and four blocks for pitch slope. For each block, the *F*0 change (at the onset and offset; positive or negative) started at 15 Hz and was adjusted in 5 Hz steps for the first three reversals and in 1 Hz steps after that. The threshold for each block was based on the average onset *F*0 change corresponding to the last two reversals in the adaptive track. There were three repetitions for each condition and it took the listeners 50–60 min to complete the entire experiment. The blocks in each repetition were randomly ordered for each listener.

3. Results

3.1 Language, cue, and direction

We first examined the effects of Language (Mandarin, English), Cue (pitch height, pitch slope), and Direction (falling, rising). We are primarily interested in potential differences between the Mandarin and English listeners as revealed by significant interactions between Language and Cue and Language and Direction. Pitch height and pitch slope are not directly compared since they do not have the same units of measurement.

A three-way repeated-measures analysis of variance (ANOVA) with Language as between-subjects factor and Cue and Direction as within-subjects factors was conducted on JNDs of the Mandarin and English listeners in the pitch height and pitch slope conditions. The results did not show a main effect of Language [F (1, 40) = 0.07, $p = 0.80$], indicating that the overall JND for Mandarin listeners (6.84 Hz) was not significantly different from that for the English listeners (7.00 Hz). There was also no main effect of Direction [F (1, 40) = 0.01, $p = 0.99$], indicating that the overall JND for falling tones (6.91 Hz) was not different from that for rising tones (6.91 Hz).

However, the results revealed a significant interaction between Language and Cue [F (1, 40) = 5.0, $p = 0.03$], suggesting that the Mandarin and English listeners had different JNDs in the pitch height and pitch slope conditions (see top panel of Fig. 2). Paired t-tests revealed no significant difference in the JND of either the pitch height condition [t (40) = 0.56, $p = 0.58$] between the Mandarin listeners (7.05 Hz) and the English listeners (6.68 Hz) or the pitch slope condition [t (40) = −1.05, $p = 0.29$], with JNDs of 6.63 and 7.30 Hz for the Mandarin and English listeners, respectively. While the individual JND comparisons across participant groups did not reach significance, the overall significant interaction between Language and Cue indicated that Mandarin and English listeners do not behave in the same way when listening to differences in pitch height or pitch slope. As shown in Fig. 2, in terms of Cue, the Mandarin listeners

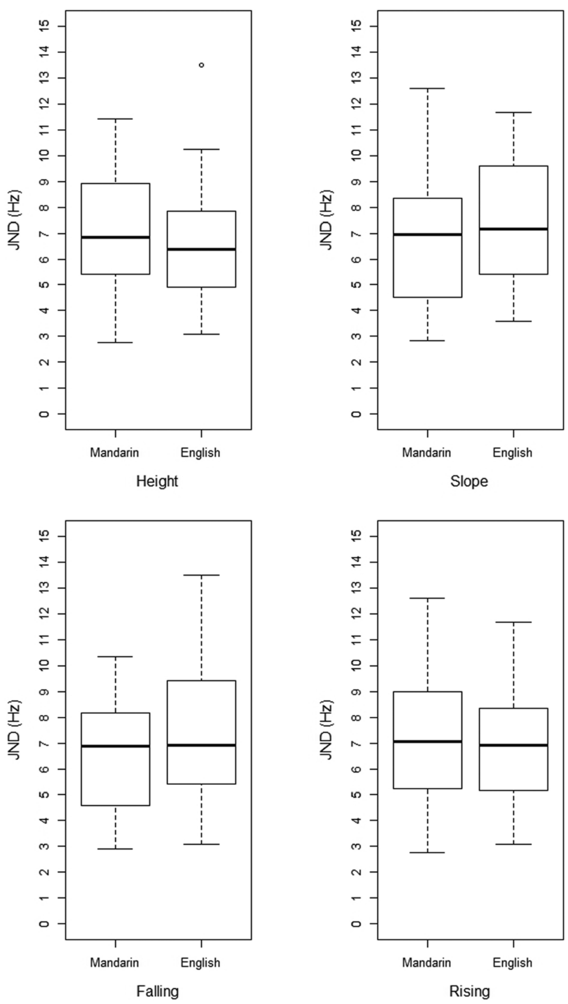

Fig. 2. JNDs of the Mandarin and English listeners in the pitch slope versus pitch height conditions (top panel) as well as in the falling versus rising tones condition (bottom panel).

had higher JNDs (less sensitivity) than the English listeners for pitch height whereas the Mandarin listeners showed lower JNDs (greater sensitivity) than the English listeners for pitch slope.

Moreover, the results showed a significant interaction between Language and Direction [F (1, 40) = 8.1, $p = 0.007$], indicating that the Mandarin and English listeners had different JNDs for the falling and rising tones (see bottom panel of Fig. 2). Follow-up tests revealed a marginally significant difference in the JND both for the Mandarin listeners [t (20) = 2.0, $p = 0.059$] between falling (6.54 Hz) and rising (7.13 Hz) tones and for the English listeners [t (20) = 2.0, $p = 0.056$] between falling (7.29 Hz) and rising (6.69 Hz) tones. In terms of Direction, while the Mandarin listeners had lower JNDs (greater sensitivity) than the English listeners for falling tones, the Mandarin listeners showed higher JNDs (less sensitivity) than the English listeners for rising tones.

3.2 Comparison within pitch height and pitch slope conditions

Next, we investigated the effect of test stimulus manipulation within the pitch height and the pitch slope conditions. The pitch height condition consisted of four blocks varying in pitch direction (falling, rising) as well as extent (higher vs lower). The pitch slope condition consisted of four blocks varying in pitch direction (falling, rising) as well as steepness (shallower vs steeper). Figure 3 shows the JNDs of the Mandarin and the English listeners in each condition for pitch height and pitch slope stimuli.

For pitch height, a three-way repeated-measures ANOVA with Language (Mandarin vs English) as between-subjects variable, and Direction (falling vs rising) and Extent (higher vs lower) as within-subjects variables was conducted on JNDs of the Mandarin and English listeners. The results did not show significant main effects of Language [F (1, 40) = 0.07, $p = 0.80$] or Direction [F (1, 40) = 0.5, $p = 0.48$], indicating that the overall JND for falling tones (6.95 Hz) was not significantly different from that for rising tones (6.78 Hz) in the pitch height condition. But the results did show a significant effect of Extent [F (1, 40) = 33.4, $p = < 0.001$], indicating that the overall JND for the higher condition (5.75 Hz) was significantly smaller than that for the lower condition (7.97 Hz).

Moreover, there was a significant interaction between Language and Direction [F (1, 40) = 4.3, $p = 0.045$], suggesting that Mandarin and English listeners were differentially sensitive to falling and rising tones. While the overall interaction was significant, none of the individual comparisons reached significance (Mandarin group [t(20) = −1.04, $p = 0.31$]; English group [t(20) = 1.85, $p = 0.08$]); (Falling tones [t(40) = −0.20, $p = 0.84$]; Rising tones [t(40) = 1.30, $p = 0.20$]).

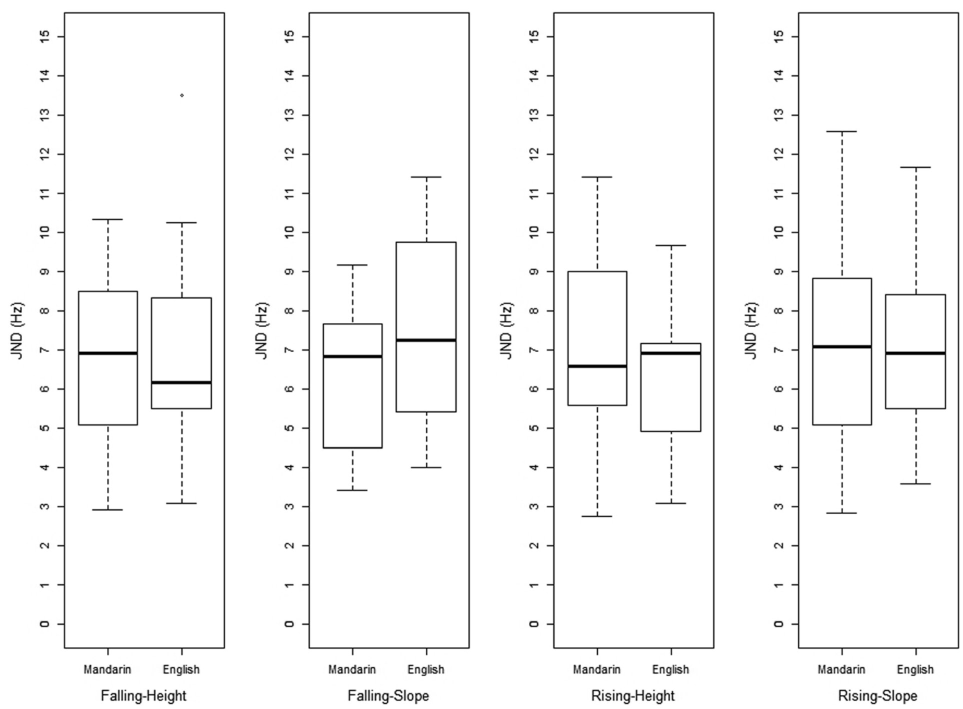

Fig. 3. JNDs of the Mandarin and English listeners for pitch height and pitch slope as a function of direction (falling, rising).

For pitch slope, a three-way repeated-measures ANOVA with Language (Mandarin vs English) as between-subjects variable and with Direction (falling vs rising) and Steepness (shallower vs steeper) as within-subjects variables was conducted on JNDs of the Mandarin and English listeners. There was no significant main effect of Language [F (1, 40) = 0.07, p = 0.80] or Direction [F (1, 40) = 0.31, p = 0.58], indicating that the overall JND for falling tones (6.88 Hz) was not significantly different from that for rising tones (7.05 Hz) in the pitch slope condition. But there was a significant effect of Steepness [F (1, 40) = 30, $p < 0.001$], indicating that the overall JND for the steeper condition (5.84 Hz) was significantly lower than that for the shallower condition (8.09 Hz).

Moreover, there was a significant interaction between Language and Direction [F (1, 40) = 4.9, p = 0.03], as well as between Language and Steepness [F (1, 40) = 15.3, $p < 0.001$]. A significant three-way interaction between Language, Direction, and Steepness [F (1, 40) = 6.6, p = 0.01] was also found.

Paired t-tests to explore the Language by Direction interaction showed a marginally significant effect of Direction for the Mandarin group [t (20) = −2.0, p = 0.058], suggesting that the Mandarin listeners' JND for falling tones (6.20 Hz) was lower than that for rising tones (7.06 Hz) while the comparison did not reach significance for the English group (falling: 7.56 Hz; rising: 7.05 Hz) [t (20) = 1.15, p = 0.26]. Independent t-tests showed an effect of Language for the falling tones: Mandarin listeners (6.20 Hz) had significantly lower JNDs than English listeners (7.56 Hz) [t (40) = −2.1, p = 0.04] while this comparison did not reach significance for the rising tones [t (40) = 0.01, p = 0.99] (Mandarin: 7.06 Hz; English: 7.05 Hz).

Exploring the Language by Steepness interaction, paired t-tests showed a significant effect of Steepness for the Mandarin group [t (40) = 5.6, $p < 0.001$] with lower JNDs in the steeper conditions (4.70 Hz) than the shallower conditions (8.55 Hz) but this difference did not reach significance for the English group [t (40) = 1.45, p = 0.16] (steeper: 6.98 Hz; shallower: 7.62 Hz). An independent t-test showed a significant effect of Language for the steeper conditions, with lower JNDs for the Mandarin (4.70 Hz) than the English listeners (6.98 Hz) [t (40) = −3.4, p = 0.002] but not for the shallower conditions [t (40) = 1.1, p = 0.28] (Mandarin: 8.55 Hz; English: 7.62 Hz).

For the three-way interaction between Language, Direction, and Steepness, independent t-tests showed that the Mandarin listeners (3.83 Hz) had a significantly lower JND than the English listeners (6.09 Hz) in the Falling-Steeper condition [t (30.3) = −3.7, $p < 0.001$], and that Mandarin listeners (5.57 Hz) had a significantly lower JND than the English listeners (7.87 Hz) in the Rising-Steeper condition [t (40) = −2.6, p = 0.01]. Moreover, English listeners (6.22 Hz) had a significantly lower JND than Mandarin listeners (8.54 Hz) in the Rising-Shallower condition [t(40) = 2.4, p = 0.02] while this effect did not reach significance in the Falling-Shallower condition [t (40) = −0.50, p = 0.62] (Mandarin: 8.56 Hz; English: 9.03 Hz).

To sum up, for pitch height, while listeners were more sensitive to higher pitch stimuli, the height of the contours (higher vs lower) was not found to modulate the Mandarin and English listeners' different sensitivities to tones. In contrast, for pitch slope, the results indicate that the Mandarin and English listeners' different sensitivities to tones are modulated by the pitch direction (falling vs rising) and steepness of the contours (steeper vs shallower). For pitch slope, the Mandarin listeners were more sensitive to the falling tones and the tones in steeper conditions compared with the English listeners.

4. Discussion and conclusion

The current study examined Mandarin and English listeners' sensitivity to pitch direction, height, and slope at the psychoacoustic level by measuring JNDs. Our findings extend those of Liu (2013) who reported JNDs of tone discrimination in terms of onset and offset position for English and Mandarin listeners, but did not separately test the two groups of listeners' sensitivity to pitch height and pitch slope. Consistent with previous studies on tone perception, we found that English listeners showed greater sensitivity to pitch height than Mandarin listeners. We also found that Mandarin listeners were more sensitive to pitch slope than English listeners [see Gandour and Harshman (1978) for similar results for Thai versus English listeners]. This pattern of results supports previous explanations based on the linguistic relevance of tonal characteristics. While slope is a dimension that is linguistically relevant to speakers of Mandarin and many other tone languages, speakers of non-tonal languages such as English focus on more general non-linguistic characteristics of tone such as pitch height (average pitch) (e.g., Gandour and Harshman, 1978; Gandour, 1983).

In addition, listeners' language background modulated the sensitivity to pitch changes of falling versus rising tones. Specifically, our results showed that while the Mandarin listeners were more sensitive to falling tones than rising tones, the English listeners were more sensitive to rising tones than falling tones. Liu (2013) also reported an effect of language background dependent on pitch direction (falling vs rising). However, his English listeners showed greater sensitivity than the Mandarin listeners for the falling endpoint tone while the Mandarin listeners showed greater sensitivity for the rising endpoint tone and the rising boundary tone than the English listeners. While Liu's (2013) falling endpoint condition varies in slope, it also varies in height (average $F0$) because the onset is fixed. Since the stimuli in Liu's (2013) falling and rising conditions varied in both pitch slope and pitch height, it is difficult to directly compare the present results to his. However, the current findings suggest that direction interacts with steepness. In general, we find that Mandarin listeners are more sensitive than English listeners to steeper contours, whether they be rising or falling. In addition, English listeners are more sensitive to shallower contours but only in the rising condition.

Liu (2013) found that the Mandarin and English listeners showed different JNDs for the falling endpoint tone, the rising endpoint tone, as well as the rising boundary tone, but no significant difference was found between the two groups for the level tone, also demonstrating an asymmetry of contour tones versus level tones in the findings. Our results also show an asymmetry between pitch slope and pitch height. While the Mandarin and English listeners' sensitivity to tones is modulated by both the direction (falling vs rising) and steepness of the contours (steeper vs shallower) in the pitch slope condition, there was no difference between the Mandarin and English listeners' JNDs for either the rising tones or the falling tones in the pitch height condition. As a result, our findings indicate that listeners' sensitivity to falling tones and rising tones is modulated by both stimulus properties (e.g., changes in pitch slope and height) and language background. The stimuli in the pitch slope condition, encoding both changes in contour and height, are more pitch-informative than those in the pitch height condition. That might account for why listeners' different sensitivities to falling and rising tones are found in the pitch slope condition but not in the pitch height condition.

In sum, the present study indicates that, overall, Mandarin listeners are more sensitive to pitch slope and English listeners to pitch height at the psychoacoustic level. However, these effects are modulated by both the direction (falling/rising) and slope (steeper/shallower) of the pitch contours.

Acknowledgments

This research was partially supported by a grant from the University of Kansas General Research Fund. We thank Dr. Rajka Smiljanic and two anonymous reviewers for their helpful comments.

References and links

Burnham, D., and Mattock, K. (**2007**). "The perception of tones and phones," in *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, edited by O.–S. Bohn and M. J. Munro (John Benjamins, Amsterdam, The Netherlands).

Flanagan, J. L., and Saslow, M. G. (**1958**). "Pitch discrimination for synthetic vowels," J. Acoust. Soc. Am. **30**, 435–442.

Francis, A. L., Ciocca, V., Ma, L., and Fenn, K. (**2008**). "Perceptual learning of Cantonese lexical tones by tone and non-tone language listeners," J. Phonetics **36**, 268–294.

Gandour, J. T. (**1983**). "Tone perception in far Eastern languages," J. Phonetics **11**, 149–175.

Gandour, J. T., and Harshman, R. A. (**1978**). "Crosslanguage differences in tone perception: A multidimensional scaling investigation," Lang. Speech **21**, 1–33.

Hairston, W. D., and Maldjian, J. A. (**2009**). "An adaptive staircase procedure for the E-Prime programming environment," Comp. Meth. Progr. Biomed. **93**, 104–108.

Leather, J. (**1987**). "F0 pattern inference in the perceptual acquisition of second language tone," in *Sound Patterns in Second Language Acquisition*, edited by A. James and J. Leather (Foris Publications, Dordrecht, The Netherlands).

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

Liu, C. (**2013**). "Just noticeable differences of tone pitch contour change for English- and Chinese-native listeners," J. Acoust. Soc. Am. **134**, 3011–3020.

Stagray, J. R., and Downs, D. (**1993**). "Differential sensitivity for frequency among listeners of a tone and a nontone language," J. Chin. Ling. **21**, 143–163.

Werker, J. F., and Logan, J. S. (**1985**). "Cross-language evidence for three factors in speech perception," Percept. Psychophys. **37**, 35–44.