The effect of language experience on perceptual normalization of Mandarin tones and non-speech pitch contours

Xin Luo^{a)} and Krista B. Ashmore

Department of Speech, Language, and Hearing Sciences, Purdue University, 500 Oval Drive, West Lafayette, Indiana 47907

(Received 24 September 2013; revised 10 April 2014; accepted 14 April 2014)

Context-dependent pitch perception helps listeners recognize tones produced by speakers with different fundamental frequencies (f_0 s). The role of language experience in tone normalization remains unclear. In this cross-language study of tone normalization, native Mandarin and English listeners were asked to recognize Mandarin Tone 1 (high-flat) and Tone 2 (mid-rising) with a preceding Mandarin sentence. To further test whether context-dependent pitch perception is speech-specific or domain-general, both language groups were asked to identify non-speech flat and rising pitch contours with a preceding non-speech flat pitch contour. Results showed that both Mandarin and English listeners made more rising responses with non-speech than with speech stimuli, due to differences in spectral complexity and listening task between the two stimulus types. English listeners made more rising responses than Mandarin listeners with both speech and non-speech stimuli. Contrastive context effects (more rising responses in the high- f_0 context than in the low- f_0 context) were found with both speech and non-speech stimuli for Mandarin listeners, but not for English listeners. English listeners' lack of tone experience may have caused more rising responses and limited use of context f_0 cues. These results suggest that context-dependent pitch perception in tone normalization is domain-general, but influenced by long-term language experience.

© 2014 Acoustical Society of America. [<http://dx.doi.org/10.1121/1.4874619>]

PACS number(s): 43.71.An, 43.66.Ba, 43.66.Hg [PBN]

Pages: 3585–3593

I. INTRODUCTION

Perception of vowels and consonants strongly depends on surrounding phonetic context. For example, in a selective adaptation paradigm (e.g., [Morse et al., 1976](#)), repeated presentation of an adapting stimulus before the identification of a stimulus set would shift the boundary of the identification function toward the adaptor, possibly due to the fatigue of acoustic feature detectors, changing response criteria (response bias), or auditory contrast. An anchoring paradigm (e.g., [Sawusch and Nusbaum, 1979](#); [Sawusch et al., 1980](#)) has been used to explore the various explanations for vowel adaptation. It was found that the vowel category boundaries shifted toward the anchor that occurred more often than the other vowels in the anchoring condition, relative to the boundaries obtained with each vowel occurring equally often. The anchoring results did not agree with the feature detector fatigue and response bias hypotheses, but may be accounted for by auditory contrast.

Phoneme recognition has also been tested with a preceding sentence or phoneme carrier. In general, there are contrastive context effects on phoneme recognition, with more low-frequency responses in a high-frequency context than in a low-frequency context. Context-dependent vowel recognition (e.g., [Ladefoged and Broadbent, 1957](#)) was considered as the result of speaker normalization, in which listeners may recover the speaker's vocal tract size and vowel formant space from the context, and then use such speaker information to

calibrate target vowel recognition (see also [Johnson, 1990](#)). The effect of preceding liquid on stop consonant recognition (e.g., [Mann, 1980](#)) was thought to perceptually compensate for the acoustic assimilation caused by speech co-articulation, possibly based on the knowledge or recovery of vocal tract dynamics and constraints. Although context effects make phoneme recognition robust against speaker variability or speech co-articulation, the perceptual process may not require context information about speaker identity, vocal tract size, articulatory gesture, or phonetic space. For example, the contrastive context effects on consonant and vowel recognition reported by [Mann \(1980\)](#) and [Ladefoged and Broadbent \(1957\)](#) have been replicated by [Lotto and Kluender \(1998\)](#) and [Watkins and Makin \(1994\)](#), respectively, using non-speech contexts to model the formant structures of speech contexts. These results suggest that phonetic normalization may arise from general auditory processing of spectral contrast between context and target, but not necessarily from speaker-, phoneme-, or speech-specific processing.

In tonal languages such as Mandarin, the four lexical tones used to contrast word meanings are mainly characterized by different pitch heights and pitch contours (Tone 1: high-flat; Tone 2: mid-rising; Tone 3: low-falling-rising; Tone 4: high-falling). Because different speakers have different ranges of fundamental frequency (f_0 ; the main acoustic correlate of pitch), listeners may have to use pitch cues in the context to resolve tone ambiguities and perform tone normalization. Contrastive context effects have been found for Mandarin contour tone recognition (e.g., [Moore and Jongman, 1997](#); [Huang and Holt, 2009](#)), although the effects seem smaller than those for Cantonese level tone recognition

^{a)}Author to whom correspondence should be addressed. Electronic mail: luo5@purdue.edu

(e.g., Wong and Diehl, 2003; Francis *et al.*, 2006). Compared to level tone recognition, contour tone recognition may rely more on the intrinsic pitch patterns of target tones.

Moore and Jongman (1997) tested the recognition of Tone 2–Tone 3 (mid-rising and low-falling-rising) series that varied acoustically in either the timing of pitch contour turning point, the f_0 difference between the onset and turning points (Δf_0), or both. Natural phrases produced by two speakers with partially overlapping f_0 ranges were used as the preceding context. It was found that for target tones varying only in Δf_0 , the high- f_0 context led to more Tone-3 (low-falling-rising) responses, while the low- f_0 context led to more Tone-2 (mid-rising) responses. The authors thought that the contrastive context effects may be due to a speaker-contingent process, in which the context f_0 range was used as a cue for speaker identity, and target tone recognition varied with the perceived speaker identity. However, Huang and Holt (2009) proposed that tone normalization may arise from general auditory processing of spectral (pitch) contrast as occurs in phonetic normalization. In Huang and Holt (2009), target stimuli were Tone 1–Tone 2 (high-flat and mid-rising) series synthesized by varying the onset f_0 of a male speaker's Tone 1. A preceding sentence from the same speaker was modified to have either a high or a low mean f_0 . Even with no change in speaker identity, speech contexts contrastively affected tone recognition, leading to more Tone-2 (mid-rising) responses in the high- f_0 context than in the low- f_0 context. Besides, non-speech contexts consisting of harmonic complex tones or pure tones at the mean f_0 s of speech contexts also contrastively affected tone recognition, which supported the explanation of tone normalization based on general auditory processing of pitch contrast instead of speaker, phonetic, or speech processing.

The role of language experience in tone normalization is unclear in the literature, partially due to methodological issues. Fox and Qi (1990) found that native Mandarin and native English listeners exhibited similar borderline effects of a Mandarin Tone-1 (high-flat) or Tone-2 (mid-rising) precursor on recognition of Tones 1 and 2. The Tone-2 context they used may have provided listeners with the f_0 range information to calibrate target tone recognition throughout the test and, thus, may not be ideal for testing tone normalization even in native Mandarin listeners. Wong (1998) found that for Cantonese–English bilinguals, English precursors also produced contrastive context effects on Cantonese level tone recognition, although the effects were smaller than those elicited by Cantonese precursors. In contrast, Jongman and Moore (2000) found different patterns of normalization for recognition of Mandarin Tone 2 (mid-rising) and Tone 3 (low-falling-rising) in Mandarin and English listeners. For Mandarin listeners, the contrastive context effects were significant only when context and target stimuli varied in the same single acoustic dimension (either Δf_0 or f_0 turning point), and their tonal language experience may have helped them disambiguate target tone contrasts. However, for English listeners, significant contrastive context effects were only observed when target stimuli varied in both Δf_0 and f_0 turning point (i.e., with more salient intrinsic acoustic changes). Such effects may thus be the results of acoustic

discrimination rather than tone normalization. Different from previous studies of tone normalization, Jongman and Moore (2000) tested lexical tones that differed in both the spectral (i.e., Δf_0) and temporal (i.e., f_0 turning point) dimensions. English listeners without tonal language experience may not be able to separate the two dimensions and use the spectral and temporal cues independently.

This study aimed to clarify the effect of language experience on Mandarin contour tone normalization. Both native Mandarin and native English listeners were tested using speech context and target stimuli that were similar to those in Huang and Holt (2009), but different from those in previous cross-language studies of Mandarin tone normalization (Fox and Qi, 1990; Jongman and Moore, 2000). The target Tone 1–Tone 2 series varied only in a single spectral dimension (onset f_0), and their recognition by native Mandarin listeners was known to be contrastively affected by the mean f_0 of both speech and non-speech contexts (Huang and Holt, 2009). If tone normalization is primarily due to general auditory processing of pitch contrast, similar contrastive context effects are expected for English listeners, who may rely on pitch discrimination instead of phonetic processing to recognize unfamiliar Mandarin tones. However, it is also possible that tonal language experience of Mandarin listeners may aid in their phonetic processing of target tones and result in stronger contrastive context effects (Jongman and Moore, 2000).

To further test whether context-dependent pitch perception is speech-specific or domain-general, both language groups were also asked to identify flat and rising pitch contours with a preceding high- or low- f_0 flat pitch contour. Both context and target stimuli in the pitch contour identification (PCI) test were non-speech harmonic complex tones with the same f_0 values as those in speech stimuli. Compared to tone recognition with non-speech contexts tested by Huang and Holt (2009), psychophysical PCI with non-speech contexts may be an arguably “purer” test of context effects arising from general auditory processing and may shed new light on the general pitch contrast basis of tone normalization. Using similar designs, contrastive context effects have been found for speech contexts and non-speech targets (Stephens and Holt, 2003), and for non-speech contexts and targets (Aravamudhan *et al.*, 2008) that were spectrally similar to vowels or consonants. Accordingly, our hypothesis is that for both Mandarin and English listeners, context effects on pitch perception would be similar with speech or non-speech stimuli.

II. METHODS

A. Subjects

Subjects were 10 native Mandarin listeners (4 females, 6 males; 26–31 years old) and 13 native English listeners (11 females, 2 males; 20–25 years old) recruited from students at Purdue University. None of the English listeners had any prior exposure to Mandarin or other tonal languages. No subject, Mandarin or English, had more than five years of formal musical training. Hearing thresholds in both ears of all subjects were below 25 dB hearing level (HL) at octave

frequencies between 0.25 and 8 kHz. This study was reviewed and approved by the Institutional Review Board of Purdue University. All subjects provided informed consent and were compensated for their participation.

B. Stimuli

Mandarin target and context stimuli were resynthesized from speech recordings of a male native Mandarin speaker at a 44 100-Hz sampling rate with 16-bit resolution. The target was a Mandarin syllable /yi/, which means “cloth” in Tone 1 and “aunt” in Tone 2. An utterance of the syllable in Tone 1 was selected based on clarity from the recordings. The f_0 contour of the 506-ms syllable was manipulated to be one of nine linear functions using Praat 5.3.17 (Boersma and Weenink, 2012). Target onset f_0 ranged from 160 to 200 Hz in 5-Hz steps, while target offset f_0 was fixed at 200 Hz, resulting in a nine-step series of target stimuli varying perceptually from Tone 1 to Tone 2. These f_0 values were modeled after natural Mandarin Tones 1 and 2 produced by the same male speaker. Preceding context was a semantically neutral Mandarin sentence: “请听下个词”/qing3 ting1 xia4 ge4 ci2/meaning “please listen to the next word,” which had all four Mandarin tones. The originally recorded sentence was 1319 ms and had a mean f_0 of 165 Hz with a range from 123 to 229 Hz. The entire f_0 contour of this sentence was shifted to create a low- f_0 context with a mean f_0 of 160 Hz or a high- f_0 context with a mean f_0 of 200 Hz using Praat 5.3.17 (Boersma and Weenink, 2012). The two mean f_0 values in context corresponded to the lowest and highest target onset f_0 s. Context sentences and target syllables were matched in root mean square (RMS) level and were separated by 50 ms. Figures 1(a) and 1(b) show the waveform, spectrogram, and pitch contour of an example speech stimulus, which is the context sentence with the high mean f_0

(200 Hz) followed by the target syllable with the lowest onset f_0 (160 Hz).

Non-speech target and context stimuli were both harmonic complex tones with the first four equal-amplitude harmonics (i.e., f_0 , $2f_0$, $3f_0$, and $4f_0$). Nine 500-ms target stimuli created a perceptual series from a flat to a rising pitch contour (i.e., non-speech analogs of Mandarin Tones 1 and 2, respectively). Target onset f_0 ranged from 160 to 200 Hz in 5-Hz steps, while target offset f_0 was fixed at 200 Hz. Target f_0 transitioned linearly from its onset to its offset frequency. Preceding context was also 500 ms and had a flat pitch contour with either a low (160 Hz) or high (200 Hz) f_0 . Compared to the speech context stimuli, the non-speech context stimuli had a shorter duration with no f_0 variations. Non-speech contexts with fixed f_0 s as long as sentences have been shown to contrastively affect Mandarin listeners’ tone recognition (Huang and Holt, 2009). It is, however, unlikely that listeners need such long context stimuli to perceive/use the fixed context f_0 cues. Non-speech contexts with fixed f_0 s as short as words, such as those used in this study, are also expected to have a contrastive effect on Mandarin listeners’ PCI. Non-speech context and target stimuli both had 10-ms onset and offset Hanning-window ramps and were separated by 50 ms. They also had the same RMS level as the speech stimuli. Figures 1(c) and 1(d) show the waveform, spectrogram, and pitch contour of an example non-speech stimulus, which is the high- f_0 (200 Hz) context followed by the target pitch contour with the lowest onset f_0 (160 Hz).

C. Procedures

For Mandarin listeners, Mandarin tone recognition with speech stimuli and PCI with non-speech stimuli were tested in a counterbalanced order on two different days within a week. For English listeners, PCI with non-speech stimuli was first tested as part of a previous study. Only 7 out of the

FIG. 1. (Color online) Waveform (top panels), spectrogram (bottom panels with the frequency axis labeled on the left), and pitch contour (solid curves in the bottom panels with the frequency axis labeled on the right) of an example speech (left panels) and non-speech stimulus (right panels). Both example stimuli have the context with the high mean f_0 (200 Hz) followed by the target with the lowest onset f_0 (160 Hz).

13 English listeners were available for the Mandarin tone recognition test with speech stimuli more than 8 months later. The PCI data from all 13 English listeners and the tone recognition data from 7 of them were reported and analyzed in Sec. III.

All stimuli were presented via the basic loudspeaker of a GSI-61 audiometer (Grason-Stadler, Eden Prairie, MN) at 70 dBA to each listener in a double-walled, sound-treated booth. Mandarin tone recognition and PCI were both tested with a two-alternative, forced-choice task, using two buttons on a computer screen to show the two response choices for each trial. For Mandarin tone recognition, one button was labeled with “Tone 1,” while the other was labeled with “Tone 2.” For PCI, one button was labeled with a flat line denoting a flat pitch contour, while the other was labeled with a rising diagonal line denoting a rising pitch contour. Subjects chose the target tone or pitch contour by clicking on one of the two response buttons. The percentage of Tone-2 responses was recorded for each target stimulus in each context condition of the tone recognition test, while the percentage of rising responses was recorded for the PCI test.

Mandarin tone recognition was tested with speech stimuli in four sessions with a short break in between. In the first and third sessions, tone recognition without context was tested to check if the manipulation of target onset f_0 was sufficient for subjects to identify both Mandarin Tones 1 and 2. The second and fourth sessions tested tone recognition with context. In each tone recognition session without context, the 9 target stimuli were presented 10 times in random order, resulting in a total of 90 tokens per session. In each tone recognition session with context, the 9 target stimuli preceded by either the high- or low- f_0 context were presented 10 times

in random order, resulting in a total of 180 tokens per session. The choice of the high- or low- f_0 context randomly varied from trial to trial. Tone recognition results with or without context were averaged across the two sessions. No feedback was provided during any test sessions. Before the tone recognition test, English listeners were briefly instructed that Mandarin Tones 1 and 2 have a high-flat and mid-rising pitch contour, respectively. PCI was tested with non-speech stimuli in the same procedure, except that subjects were asked to choose between flat and rising pitch contours instead of Mandarin Tones 1 and 2.

III. RESULTS

A. Tone recognition or PCI without context

Figure 2 shows average tone recognition (top panels) and PCI responses (bottom panels) without context as a function of target onset f_0 for Mandarin (left panels) and English listeners (right panels). The responses shifted from Tone 2 to Tone 1 or from rising to flat as the target onset f_0 increased from 160 to 200 Hz. Psychometric functions in all four panels have the typical S-shape, suggesting that both Mandarin and English listeners were able to identify isolated target tones and pitch contours. One-way repeated-measures (RM) analyses of variance (ANOVAs) found significant effects of target onset f_0 on the rationalized arcsine transformed percentages (Studebaker, 1985) of Tone-2 and rising PCI responses without context for both Mandarin and English listeners [Fig. 2(a): $F_{8,72} = 123.12$, $p < 0.001$, partial $\eta^2 = 0.93$; Fig. 2(b): $F_{8,72} = 90.14$, $p < 0.001$, partial $\eta^2 = 0.91$; Fig. 2(c): $F_{8,48} = 65.11$, $p < 0.001$, partial $\eta^2 = 0.92$; Fig. 2(d): $F_{8,96} = 130.02$, $p < 0.001$, partial $\eta^2 = 0.92$]

FIG. 2. Percentage of Tone-2 responses with speech stimuli (top panels) and rising responses with non-speech stimuli (bottom panels) without context as a function of target onset f_0 for Mandarin (left panels) and English listeners (right panels). Symbols represent the mean, while error bars represent the standard deviation across subjects.

FIG. 3. Perceptual boundary between Tones 1 and 2 with speech stimuli and between flat and rising contours with non-speech stimuli without context for Mandarin and English listeners. Symbols represent the mean, while error bars represent the standard deviation across subjects.

$$y = \frac{100}{1 + e^{-(x-x_0)/b}} \quad (1)$$

A two-parameter sigmoid function as shown in Eq. (1) was used to fit the tone recognition or PCI function of each subject without context. The parameter b is in inverse proportion

to the slope of function and indicates the sharpness of perceptual boundary between Tones 1 and 2 or between flat and rising contours, and x_0 is the perceptual boundary that corresponds to the target onset f_0 with 50% Tone-2 or rising responses. Each parameter of the best-fit sigmoid function was analyzed using a two-way mixed-design ANOVA with language group and stimulus type as the factors. The function slope b was not significantly different between language groups ($F_{1,21}=0.74$, $p=0.40$, partial $\eta^2=0.033$) or stimulus types ($F_{1,15}=1.88$, $p=0.19$, partial $\eta^2=0.11$). The interaction between the two factors was not significant for b ($F_{1,15}=1.65$, $p=0.22$, partial $\eta^2=0.099$). For the perceptual boundary x_0 , the effects of language group ($F_{1,21}=4.34$, $p=0.05$, partial $\eta^2=0.17$) and stimulus type ($F_{1,15}=6.39$, $p=0.02$, partial $\eta^2=0.30$) were both significant, but the interaction between the two factors was not significant ($F_{1,15}=0.28$, $p=0.61$, partial $\eta^2=0.018$). As shown in Fig. 3, English listeners had higher perceptual boundaries or more rising responses than Mandarin listeners with both speech and non-speech stimuli. Both language groups had higher perceptual boundaries or more rising responses with non-speech than with speech stimuli.

B. Tone recognition or PCI with context

Figure 4 shows average tone recognition (top panels) and PCI responses (bottom panels) with the low- f_0

FIG. 4. Percentage of Tone-2 responses with speech stimuli (top panels) and rising responses with non-speech stimuli (bottom panels) with the low- f_0 (downward triangles) and high- f_0 contexts (upward triangles) as a function of target onset f_0 for Mandarin (left panels) and English listeners (right panels). Symbols represent the mean, while error bars represent the standard deviation across subjects. For clarity of illustration, error bars are shown in only one direction.

(downward triangles) and high- f_0 contexts (upward triangles) as a function of target onset f_0 for Mandarin (left panels) and English listeners (right panels). The rationalized arcsine transformed percentages of Tone-2 or rising PCI responses in each panel were analyzed using a two-way RM ANOVA with target onset f_0 and context f_0 as the two factors. For Mandarin listeners, tone recognition with speech stimuli [Fig. 4(a)] was significantly affected by both target onset f_0 ($F_{8,72} = 92.12$, $p < 0.001$, partial $\eta^2 = 0.91$) and context f_0 ($F_{1,9} = 21.20$, $p = 0.001$, partial $\eta^2 = 0.70$). The interaction between the two factors was also significant ($F_{8,72} = 4.10$, $p < 0.001$, partial $\eta^2 = 0.31$). Mandarin listeners gradually reduced Tone-2 responses with increasing target onset f_0 . The high- f_0 preceding sentence led to more Tone-2 responses for Mandarin listeners than the low- f_0 preceding sentence, as was expected for contrastive context effects on tone recognition. However, *post hoc* Bonferroni t -tests showed that the context effects were significant ($p < 0.01$) only for perceptually ambiguous target tones with onset f_0 ranging from 175 to 190 Hz.

Mandarin listeners' PCI with non-speech stimuli [Fig. 4(b)] had similar patterns of context effects as their tone recognition with speech stimuli. The effects of target onset f_0 ($F_{8,72} = 110.09$, $p < 0.001$, partial $\eta^2 = 0.92$) and context f_0 ($F_{1,9} = 5.86$, $p = 0.039$, partial $\eta^2 = 0.40$), and the interaction between the two factors ($F_{8,72} = 8.21$, $p < 0.001$, partial $\eta^2 = 0.48$) were all significant. Again, Mandarin listeners' PCI functions had the typical S-shape. For perceptually ambiguous target pitch contours with onset f_0 ranging from 180 to 190 Hz, Mandarin listeners' PCI responses were also significantly affected by context f_0 in a contrastive manner (*post hoc* Bonferroni t -tests: $p < 0.001$), with more rising responses in the high- f_0 context than in the low- f_0 context.

For English listeners, Mandarin tone recognition with speech stimuli [Fig. 4(c)] was significantly affected by target onset f_0 ($F_{8,48} = 75.24$, $p < 0.001$, partial $\eta^2 = 0.93$), but not by context f_0 ($F_{1,6} = 5.84$, $p = 0.052$, partial $\eta^2 = 0.49$). The two factors did not significantly interact with each other ($F_{8,48} = 2.03$, $p = 0.063$, partial $\eta^2 = 0.25$). English listeners' tone recognition functions also had the typical S-shape. However, the f_0 of the preceding sentence had no clear effects on English listeners' tone recognition. This was in contrast with the strong context effects on tone recognition for Mandarin listeners.

English listeners' PCI with non-speech stimuli [Fig. 4(d)] had similar patterns of results as their tone recognition with speech stimuli. The effect of target onset f_0 was significant ($F_{8,96} = 95.72$, $p < 0.001$, partial $\eta^2 = 0.89$). However, the effect of context f_0 on PCI was not significant ($F_{1,12} = 3.92$, $p = 0.071$, partial $\eta^2 = 0.25$) and there was no significant interaction between the two factors ($F_{8,96} = 1.06$, $p = 0.40$, partial $\eta^2 = 0.081$).

Each subject's tone recognition or PCI function in the high- or low- f_0 context was then fit with the sigmoid function in Eq. (1) to estimate the perceptual boundary and its shifts with context f_0 , language group, and stimulus type. This analysis was used to compare the context effects between language groups and stimulus types. Figure 5 shows perceptual boundaries with speech (top panels) and non-speech stimuli (bottom panels) as a function of context f_0 for Mandarin (left panels) and English listeners (right panels). A three-way mixed-design ANOVA with context f_0 , language group, and stimulus type as the factors was used to analyze the perceptual boundaries. The effect of context f_0 was significantly contrastive ($F_{1,23} = 39.20$, $p < 0.001$, partial $\eta^2 = 0.63$), whereby the high- f_0 context led to higher

FIG. 5. Perceptual boundary between Tones 1 and 2 with speech stimuli (top panels) and between flat and rising contours with non-speech stimuli (bottom panels) as a function of context f_0 for Mandarin (left panels) and English listeners (right panels). Symbols represent the mean, while error bars represent the standard deviation across subjects.

perceptual boundaries or more rising responses than the low- f_0 context. The effect of language group was significant ($F_{1,21} = 5.22$, $p = 0.033$, partial $\eta^2 = 0.20$), with English listeners having higher perceptual boundaries or more rising responses than Mandarin listeners. The effect of stimulus type was significant as well ($F_{1,15} = 14.68$, $p = 0.002$, partial $\eta^2 = 0.50$), and there were higher perceptual boundaries or more rising responses with non-speech than with speech stimuli. Among the two- and three-way interactions, only the interaction between context f_0 and language group was significant ($F_{1,23} = 16.60$, $p < 0.001$, partial $\eta^2 = 0.42$), suggesting that the context effects on tone recognition or PCI were greater for Mandarin listeners than for English listeners. Notably, the interaction between context f_0 and stimulus type was not significant ($F_{1,15} = 0.17$, $p = 0.68$, partial $\eta^2 = 0.011$), suggesting that the context effects on tone recognition were similar in magnitude to those on PCI for both Mandarin and English listeners.

The function slopes with speech and non-speech stimuli as a function of context f_0 for Mandarin and English listeners were also analyzed using a three-way mixed-design ANOVA. The effect of context f_0 was significant ($F_{1,22} = 5.38$, $p = 0.030$, partial $\eta^2 = 0.20$), whereby the high- f_0 context led to steeper functions than the low- f_0 context. However, the effects of language group ($F_{1,22} = 1.39$, $p = 0.25$, partial $\eta^2 = 0.060$) and stimulus type ($F_{1,15} = 0.51$, $p = 0.48$, partial $\eta^2 = 0.033$) were not significant. Also, none of the interactions among the factors was significant ($p > 0.88$, partial $\eta^2 < 0.002$).

IV. DISCUSSION

A. Tone recognition or PCI without context

Although isolated tone recognition with speech stimuli and PCI with non-speech stimuli both generated a sigmoidal response function for Mandarin and English listeners, the perceptual boundary between response categories was significantly higher with non-speech than with speech stimuli and for English than for Mandarin listeners. The shift of perceptual boundary with different stimulus types was consistent with the results of [Xu et al. \(2006\)](#). However, [Xu et al. \(2006\)](#) did not find different perceptual boundaries for the two language groups. Instead, they showed that English listeners had shallower response functions or weaker categorical perception than Mandarin listeners with either speech or non-speech stimuli. The methods of [Xu et al. \(2006\)](#) were similar to ours, except that they tested a lower f_0 range from 102 to 130 Hz with shorter stimuli (300 ms), and asked Mandarin or English listeners to choose between flat and rising pitch contours no matter what type of stimuli were tested. The different task instructions in the two studies may have altered the response behaviors of Mandarin and English listeners, as well as the perceptual differences between the two language groups.

In this study, English listeners had higher perceptual boundaries or more Tone-2 responses than Mandarin listeners in isolated tone recognition, possibly because English listeners lacked tonal language experience. Mandarin listeners identified the target tones with speech stimuli based on their

long-term representations of Mandarin tones, in which Tone 1 may not necessarily be completely flat in pitch and may include pitch contours with relatively small f_0 variations. For example, acoustic analyses of the multi-speaker Mandarin tones used in [Luo and Fu \(2004\)](#) actually showed an average f_0 increase of 15 Hz in Tone 1. In contrast, English listeners without tonal language experience may perform tone recognition simply based on pitch discrimination and choose Tone 2 whenever a rising pitch is perceived. The response strategy may have led to more Tone-2 responses or higher perceptual boundaries in English listeners than in Mandarin listeners. The same effect of language experience was found for non-speech PCI, suggesting that Mandarin listeners' non-speech PCI was, at least partially, influenced by their overlearned pitch representations of Mandarin tones, and thus showed lower perceptual boundaries or less rising responses than PCI of English listeners.

For both Mandarin and English listeners, PCI functions with non-speech stimuli had higher perceptual boundaries than tone recognition functions with speech stimuli. A possible explanation for the shifted perceptual boundaries in both language groups is related to differences in spectral complexity between speech and non-speech stimuli ([Xu et al., 2006](#)). In this study, speech stimuli contained both low-order resolved and high-order unresolved harmonics, while non-speech stimuli only had the first four resolved harmonics that have been shown to contribute more to pitch perception than unresolved harmonics (e.g., [Shackleton and Carlyon, 1994](#)). Because both stimuli were equalized in the overall RMS level, non-speech stimuli would have more spectral energy in low-order resolved harmonics and may thus produce a more salient pitch perception than speech stimuli. Also, the formant structure of speech stimuli may interfere with pitch perception of the f_0 contour ([Carrell et al., 1981](#); [Repp and Lin, 1990](#)). If pitch salience reduces with spectrally more complex speech stimuli, both Mandarin and English listeners may need larger f_0 increases to perceive similar pitch changes and, in turn, have lower perceptual boundaries with speech than with non-speech stimuli. For Mandarin listeners with tonal language experience, the different nature of listening tasks with speech and non-speech stimuli may have also contributed to the different perceptual boundaries. Compared to tone recognition based on tonal language experience, psychophysical PCI simply based on pitch discrimination may have more rising responses because the tested stimulus set contained only one truly flat pitch contour.

B. Tone recognition or PCI with context

Unlike previous studies of Mandarin tone normalization (e.g., [Fox and Qi, 1990](#); [Moore and Jongman, 1997](#); [Jongman and Moore, 2000](#); [Huang and Holt, 2009](#)), this study incorporated two possibly interacting factors of context-dependent pitch perception (i.e., language experience and stimulus type) into a single experimental design to better understand the nature of Mandarin tone normalization. The results showed evidence of strong contrastive context effects with both speech and non-speech stimuli for

Mandarin listeners, but not for English listeners, which suggest that perceptual normalization along the pitch dimension is domain-general (i.e., not speech-specific), but influenced by long-term language experience.

Context-dependent tone recognition of Mandarin listeners replicated the findings of [Huang and Holt \(2009\)](#) and confirmed that Mandarin listeners did use the mean f_0 of a preceding sentence to adjust their recognition of Mandarin contour tones, which differ from each other in both f_0 height and f_0 contour. [Wong and Diehl \(2003\)](#) found that listeners relied more on the f_0 s of the preceding syllables closer in time to the target syllable for tone normalization. Future studies should investigate how many preceding syllables are used to compute the mean f_0 for tone normalization, and how the tone types of the immediately preceding syllables may affect tone normalization. The context effects on tone recognition were contrastive, with more Tone-2 (mid-rising) responses in the high- f_0 context than in the low- f_0 context. As an instance of speaker normalization, this pattern of perceptual shifts in contour tone recognition had the same contrastive directionality as most of the previously reported context effects on vowel, consonant, or level tone recognition (e.g., [Ladefoged and Broadbent, 1957](#); [Mann, 1980](#); [Wong and Diehl, 2003](#)).

Context-dependent PCI of Mandarin listeners extended the findings of [Huang and Holt \(2009\)](#) by showing that non-speech contexts affected non-speech PCI in the same contrastive manner as non-speech or speech contexts affected Mandarin tone recognition. In either the high- or low- f_0 context, Mandarin listeners had significantly more rising responses or higher perceptual boundaries for non-speech PCI than for Mandarin tone recognition, which may be due to the differences in stimulus complexity and task nature, as mentioned earlier. Also, the non-speech contexts were always flat in pitch and may have led listeners to hear subsequent target pitch contours as more rising. In future studies, harmonic complex tones with the same f_0 variations as the speech contexts may be used to avoid this adaptation effect. Nevertheless, the perceptual shifts across the high- and low- f_0 contexts with non-speech stimuli were the same as with speech stimuli. Context-dependent pitch perception not only supported tone normalization for Mandarin listeners, but also generalized to their identification of non-speech pitch contours that resembled Mandarin tones in f_0 contours. The contrastive context effects on PCI in this study, together with the results of [Huang and Holt \(2009\)](#), suggest that phonetic, articulatory, or speaker information in either the context or target stimuli may not be necessary for tone normalization. The present PCI results of Mandarin listeners lend further support to the proposal that tone normalization may be due to general pitch contrast processing ([Huang and Holt, 2009](#)), which would, not surprisingly, elicit similar contrastive context effects on both non-speech PCI and Mandarin tone recognition for Mandarin listeners.

The major finding of this study was that English listeners did not have as robust tone normalization along the pitch dimension as Mandarin listeners. This was similar to the results of [Jongman and Moore \(2000\)](#) that Mandarin listeners used context f_0 range to adjust their identification of

Mandarin Tones 2 and 3 varying only in Δf_0 between the onset and turning points of pitch contour, whereas English listeners did not. [Jongman and Moore \(2000\)](#) speculated that the target tones they tested may not be salient enough for English listeners without tonal language experience. Thus, English listeners may have been left with limited perceptual resources to use context f_0 cues, leading to no context effects. Another possibility is that English listeners had no context effects because their target tone categories may not have been well-established. For example, [Xu et al. \(2006\)](#) found that tone recognition is more categorical for Mandarin listeners than for English listeners. As shown in [Aravamudhan et al. \(2008\)](#), categorization of non-speech targets that were spectrally similar to vowels was affected by non-speech contexts only after listeners have been trained to categorize the targets based on a fixed categorization boundary. Despite these possibilities, English listeners in this study did not seem to have weaker categorical perception of the Tone 1–Tone 2 series varying only in onset f_0 , and their tone boundaries were as sharp as those of Mandarin listeners. It is more likely that English listeners may have given less perceptual weighting to pitch cues in the context than Mandarin listeners, because such cues are not linguistically important in English. Also, the greater percentages of Tone-2 responses in English listeners may have limited the room for significant response shifts with context f_0 s. Note that English listeners' inability to perform tone normalization was not related to how well they understood or used the phonetic, speaker, or articulatory information in Mandarin speech context. This is because English listeners did not have context effects even for the identification of non-speech meaningless pitch contours, whereas Mandarin listeners did.

In summary, general auditory processing of pitch contrast interacts with long-term language experience in Mandarin tone normalization. The contrastive use of pitch cues in the context may be general, but operative depending on whether it is relevant to speech recognition in one's language experience. To better understand the effects of language learning on tone normalization, future studies may test tone recognition with context in English listeners who are Mandarin learners with various Mandarin experience. The percentages of Tone-2 or rising responses may gradually reduce and the context effects on tone recognition may enhance with more Mandarin experience.

ACKNOWLEDGMENTS

We are grateful to all subjects for their participation in this study. Research was supported in part by National Institutes of Health Grant No. R21-DC-011844.

Aravamudhan, R., Lotto, A. J., and Hawks, J. W. J. (2008). "Perceptual context effects of speech and nonspeech sounds: The role of auditory categories," *J. Acoust. Soc. Am.* **124**, 1695–1703.

Boersma, P. and Weenink, D. (2012). "Praat: Doing phonetics by computer (version 5.3.17) [computer program]," <http://www.fon.hum.uva.nl/praat/> (Last viewed August 10, 2013).

Carrell, T. D., Smith, L. B., and Pisoni, D. B. (1981). "Some perceptual dependencies in speeded classification of vowel color and pitch," *Percept. Psychophys.* **29**, 1–10.

- Fox, R. A., and Qi, Y. Y. (1990). "Context effects in the perception of lexical tone," *J. Chin. Linguist.* **18**, 261–283.
- Francis, A. L., Ciocca, V., Wong, N. K. Y., Leung, W. H. Y., and Chu, P. C. Y. (2006). "Extrinsic context affects perceptual normalization of lexical tone," *J. Acoust. Soc. Am.* **119**, 1712–1726.
- Huang, J., and Holt, L. L. (2009). "General perceptual contributions to lexical tone normalization," *J. Acoust. Soc. Am.* **125**, 3983–3994.
- Johnson, K. (1990). "The role of perceived speaker identity in *F0* normalization of vowels," *J. Acoust. Soc. Am.* **88**, 642–654.
- Jongman, A., and Moore, C. B. (2000). "The role of language experience in speaker and rate normalization processes," in *Proceedings of the Sixth International Conference on Spoken Language Processing*, Vol. I, pp. 62–65.
- Ladefoged, P., and Broadbent, D. E. (1957). "Information conveyed by vowels," *J. Acoust. Soc. Am.* **29**, 98–104.
- Lotto, A. J., and Kluender, K. R. (1998). "General contrast effects of speech perception: Effect of preceding liquid on stop consonant identification," *Percept. Psychophys.* **60**, 602–619.
- Luo, X., and Fu, Q.-J. (2004). "Enhancing Chinese tone recognition by manipulating amplitude envelope: implications for cochlear implants," *J. Acoust. Soc. Am.* **116**, 3659–3667.
- Mann, V. A. (1980). "Influence of preceding liquid on stop-consonant perception," *Percept. Psychophys.* **28**, 407–412.
- Moore, C. B., and Jongman, A. (1997). "Speaker normalization in the perception of Mandarin Chinese tones," *J. Acoust. Soc. Am.* **102**, 1864–1877.
- Morse, P. A., Kass, J. E., and Turkienicz, R. (1976). "Selective adaptation of vowels," *Percept. Psychophys.* **19**, 137–143.
- Repp, B. H., and Lin, H. B. (1990). "Integration of segmental and tonal information in speech perception: A cross-linguistic study," *J. Phonetics* **18**, 481–495.
- Sawusch, J. R., and Nusbaum, H. C. (1979). "Contextual effects in vowel perception I: Anchor-induced contrast effects," *Percept. Psychophys.* **25**, 292–302.
- Sawusch, J. R., Nusbaum, H. C., and Schwab, E. C. (1980). "Contextual effects in vowel perception II: Evidence for two processing mechanisms," *Percept. Psychophys.* **27**, 421–434.
- Shackleton, T. M., and Carlyon, R. P. (1994). "The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination," *J. Acoust. Soc. Am.* **95**, 3529–3540.
- Stephens, J. D. W., and Holt, L. L. (2003). "Preceding phonetic context affects perception of nonspeech," *J. Acoust. Soc. Am.* **114**, 3036–3039.
- Studebaker, G. A. (1985). "A 'rationalized' arcsine transform," *J. Speech Lang. Hear. Res.* **28**, 455–462.
- Watkins, A. J., and Makin, S. J. (1994). "Perceptual compensation for speaker differences and for spectral-envelope distortion," *J. Acoust. Soc. Am.* **96**, 1263–1282.
- Wong, P. C. M. (1998). "Speaker normalization in the perception of Cantonese level tones," M.S. thesis, University of Texas at Austin, Austin, TX.
- Wong, P. C. M., and Diehl, R. L. (2003). "Perceptual normalization for inter- and intra-talker variation in Cantonese level tones," *J. Speech Lang. Hear. Res.* **46**, 413–421.
- Xu, Y., Gandour, J., and Francis, A. (2006). "Effects of language experience and stimulus complexity on categorical perception of pitch direction," *J. Acoust. Soc. Am.* **120**, 1063–1074.