Journal of Phonetics 38 (2010) 616–624

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/phonetics

The influence of language experience on categorical perception of pitch contours

Gang Peng [a,b,*], Hong-Ying Zheng [a], Tao Gong [c], Ruo-Xiao Yang [a], Jiang-Ping Kong [d], William S.-Y. Wang [a]

[a] *Language Engineering Laboratory, Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong SAR*
[b] *Department of Linguistics and State Key Laboratory of Brain and Cognitive Sciences, The University of Hong Kong, Hong Kong SAR*
[c] *Department of Linguistics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany*
[d] *Department of Chinese Language and Literature, The Peking University, Beijing, PR China*

ARTICLE INFO

Article history:
Received 16 December 2009
Received in revised form
31 July 2010
Accepted 13 September 2010

ABSTRACT

Previous research on categorical perception of pitch contours has mainly considered the contrast between tone language and non-tone language listeners. This study investigates not only the influence of tone language vs. non-tone language experience (German vs. Chinese), but also the influence of different tone inventories (Mandarin tones vs. Cantonese tones), on the categorical perception of pitch contours. The results show that the positions of the identification boundaries do not differ significantly across the 3 groups of listeners, i.e., Mandarin, Cantonese, and German, but that the boundary widths do differ significantly between tone language (Mandarin and Cantonese) listeners and non-tone language (German) listeners, with broader boundary widths for non-tone language listeners. In the discrimination tasks, the German listeners exhibit only psychophysical boundaries, whereas Chinese listeners exhibit linguistic boundaries, and these linguistic boundaries are further shaped by the different tone inventories.

1. Introduction

We perceive speech sounds categorically—that is to say, we are more likely to notice the differences between categories than within categories. Categorical perception (CP) has the following characteristics (Repp, 1984): (1) In the labeling function, there is a sharp boundary between two categories; (2) in the discrimination function, accuracy peaks at the category boundary, but is at or near chance level within category; (3) the discrimination function can be predicted from the identification function.

Therefore, the major signature of CP is better discrimination across category boundaries than for equivalently separated stimuli within the same category. This ability has been evidenced in several modalities, such as auditory perception of speech (Liberman, Harris, Hoffman, & Griffith, 1957; Liberman, 1996), visual perception of colors (Bornstein, Kessen, & Weiskopf, 1976), and facial expressions (Etcoff & Magee, 1992). Besides human subjects, animals also exhibit CP. Morse and Snowden (1975) found that rhesus monkeys had CP on formant transitions, and Kuhl and Miller (1975) found that chinchilla had CP on voice onset time (VOT). A recent study reported that zebra finches could not only discriminate and categorize monosyllabic words that differ in their vowel, but also exhibited the ability to transfer this categorization to the same words spoken by novel speakers independent of the speaker's gender (Ohms, Gill, Van Heijningen, Beckers, & Ten Cate, 2010). This study showed that birds, like humans, use intrinsic and extrinsic speaker normalization to perform categorization. This finding may imply that there is no need to invoke special mechanisms for speaker normalization in human speech perception (Johnson, 2005; Moore; & Jongman, 1997; Wong; & Diehl, 2003). All in all, CP might be a very basic ability with which an organism organizes the world in which it lives, and this CP might be dynamic in nature, relating to dynamic normalization, such as speaker normalization. CP and normalization might both be involved in extracting invariant features from signals whose physical forms change continuously.

Earlier studies on speech CP mainly focused on segmental features. Liberman et al. (1957) reported that when people listened to sounds that varied along a formant transition continuum, they heard only *ba*'s, *da*'s, or *ga*'s, but nothing in between. Instead of changing gradually, the perceived quality jumped abruptly from one category to another at a certain point along a continuum. CP was also found for a voicing continuum in 1- and 4-month-old infants (Eimas, Siqueland, Jusczyk, & Vigorito, 1971). The infants reacted to a 20 ms VOT difference when accompanied by a phonemic difference (a change from /ba/ to /pa/), but hardly reacted to the same VOT difference when it was not accompanied by a phonemic difference.

Phonemic category in speech perception and production typically develops in human infants during their first year of life: infants have the ability to discriminate phonetic contrasts of all languages during the first 3 months, but this ability starts to

[*] Corresponding author at: Language Engineering Laboratory, Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong SAR. Tel.: +852 3163 4346; fax: +852 2603 5558.
E-mail address: gpeng@ee.cuhk.edu.hk (G. Peng).

0095-4470/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.wocn.2010.09.003

decline in foreign-language consonant perception and to increase in native-language consonant perception by around 11 months (Kuhl, 2004). The progressive development of CP from infancy to adolescence is presumably influenced by spoken communication. The increase in CP causes within-category differences to become less discriminated, thereby preventing non-relevant information, e.g. phonetic variations of the same phoneme, from reaching the mental lexicon. This should facilitate word recognition, especially under difficult listening conditions, although other factors, such as context cuing, the listener's guess and frequencies of occurrence, are also helpful. Therefore, CP, which develops as the infant ages, enhances communication. In addition to the behavioral evidence, Rivera-Gaxiola, Silva-Pereyra, and Kuhl (2005) discovered neural correlates of this dynamic language development via electrophysiological measurements, and pointed out that individual developmental differences might have an impact on language development.

The linguistic environment must have a crucial influence on the development of CP. There are thousands of languages in the world which make use of pitch patterns (Yip, 2002) to build words much as vowels and consonants are used. Chinese is perhaps the best known example of these (Wang, 1973), where 'ma1' (媽, mother), 'ma2' (麻, hemp), 'ma3' (馬, horse), and 'ma4' (罵, to scold) share the same segments, but differ in their pitch patterns, with the numbers '1', '2', '3', and '4' indicating different lexical tones (Wang, 1967, 1972, 1973; Peng, 2006).

CP of pitch contours for subjects with different language backgrounds was first shown behaviorally for Mandarin lexical tones by Wang (1976), who demonstrated the existence of a linguistic boundary for native Chinese (tone language) subjects but a psychophysical boundary for native American English (non-tone language) subjects. Since then, several studies of CP of pitch contours have investigated the impact of language experience on pitch perception, mainly focusing on the contrast between native tone language and non-tone language subjects (Francis, Ciocca, & Ng, 2003; Hallé, Chang, & Best, 2004; Xu, Gandour, & Francis, 2006). As for tones, Abramson (1979) claimed that tone perception in Thai is not categorical. Nonetheless, there is no strict dichotomy between CP and continuous perception. We agree with Hallé, Chang, and Best (2004) that a more stringent test of how categorical is the perception of tones by native listeners of a tone language requires a comparison with the perception of tones by listeners of a non-tone language.

Different inventories of tones may further influence pitch perception. Gandour (1983) investigated the perceptual dimensions of tones and the effect of linguistic experience on perception of tones by listeners from five language groups (Cantonese, Mandarin, Taiwanese, Thai, and English). He found that listeners from these five groups could be classified into tone language vs. non-tone language listeners, Thai vs. Chinese listeners, and Cantonese vs. Mandarin and Taiwanese listeners on the basis of their patterns of dimension weights, indicating that the experience of different tone language speakers further affected pitch perception. Lee, Vakoch, and Wurm (1996) found that Cantonese listeners were better than both Mandarin and English listeners at discriminating Cantonese tones, and Mandarin listeners did better than both Cantonese and English listeners at discriminating Mandarin tones. In both cases, the tone language listeners did better than the English listeners. However, as far as we know, there is still a lack of studies on whether and how different tone inventories affect pitch perception in terms of CP.

Mandarin and Cantonese are two of the seven major Chinese dialects. Mandarin is spoken throughout China; Cantonese is spoken mainly in south China (Wang, 1973). While Mandarin and Cantonese each have many varieties of speech, in this paper Mandarin refers to the speech of Beijing while Cantonese refers to the speech of Hong Kong. Using the five tone letters proposed by Chao (1930), the four Mandarin lexical tones in citation form are 55 (high level tone, Tone 1), 35 (high rising tone, Tone 2), 214 (low falling rising tone, Tone 3), and 51 (high falling tone, Tone 4). Cantonese has 6 lexical tones (ignoring duration difference), 55 (high level tone, Tone 1), 35 (high rising tone, Tone 2), 33 (middle level tone, Tone 3), 21 (low falling tone, Tone 4), 23 (low rising tone, Tone 5), and 22 (low level tone, Tone 6) (Bauer & Benedict, 1997). Since the tone inventories of Mandarin and Cantonese are very different, we examine whether this difference further influences categorical pitch perception.

It is generally known that other suprasegmental features, especially intensity profile, are highly correlated with tone perception (Abramson, 1972; Howie, 1976; Liu & Samuel, 2004; Whalen & Xu, 1992; Zee, 1978), but in this study, we focus on just the primary cue, fundamental frequency (the physical correlate of pitch), for lexical tone perception, and fixed other features as constant. We have reexamined the effect of tone language experience on categorical pitch perception by comparing the identification and discrimination performances of tone language listeners (Mandarin and Cantonese) vs. non-tone language (German) listeners. Moreover, we have further explored the influence of different tone inventories on pitch perception by comparing identification and discrimination performances of Mandarin vs. Cantonese listeners.

2. Methods

2.1. Materials

Two types of continua (rising and falling) were constructed for both speech context, Mandarin syllable /i/, and nonspeech context, pure tone. Fig. 1 shows a schematic diagram of the pitch contours of the 11 stimuli for the rising continuum (on the left, following Wang, 1976) and the 11 stimuli for the falling continuum (on the right). In Mandarin, the syllable /i/ means "衣 (clothes)" with the high level tone, represented by stimulus Number 11 in both continua, means "姨 (aunt)" with the high rising tone, represented as stimulus Number 1 in the rising continuum, and means "意 (meaning)" with the high falling tone, represented as stimulus Number 1 in the falling continuum. /i/ represents different lexical meanings in Cantonese when associated with different tones, and also occurs in the German phonological system. Fig. 2 shows the wideband spectrogram for 3 speech stimuli. All the speech stimuli in the rising and falling continua were resynthesized by applying the pitch-synchronous overlap and add (PSOLA) method (Moulines & Laroche, 1995) implemented in Praat (Boersma, Paul, Weenink, & David, 2009) to the same Mandarin syllable, /i/, with a high level tone produced by a male native Mandarin speaker.

The major procedures of resynthesizing the stimuli are: (1) Adjust the duration of the target syllable to 500 ms, and fix the pitch contour to the level frequency 135 Hz. (2) Reduce the number of pitch points to 3, with one at the starting position, one at the 100 ms position, and one at the end position. (3) By dragging the above three pitch points accordingly, various stimuli can be resynthesized. As for the nonspeech stimuli, we first constructed a pure sine wave with duration 500 ms and frequency 135 Hz. Then various nonspeech stimuli were resynthesized according to the same procedures as above. The different starting frequencies for the rising stimuli and the end frequencies for the falling stimuli were determined by the formula $105\,\text{Hz}+3\,\text{Hz}\times(\text{stimulus Number}-1)$.

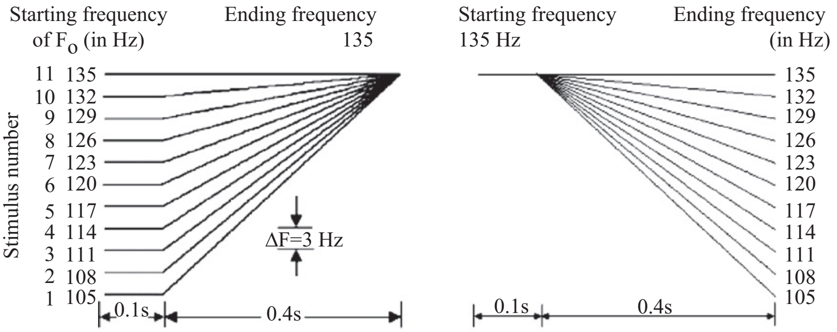

Fig. 1. Tone contours for the stimuli in the two types of continuum: left for the rising continuum, and right for the falling continuum.

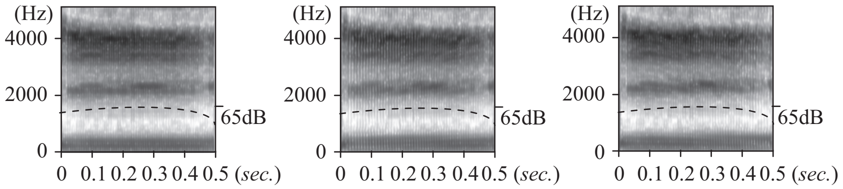

Fig. 2. Resynthesized speech (Mandarin /i/ syllables). The panels show the wideband spectrogram of the Number 11 speech stimulus (left), the Number 1 stimulus of the rising speech continuum (middle), and the Number 1 stimulus of the falling speech stimuli (right). The thin dashed lines indicate the intensity profiles.

To make the loudness of nonspeech stimuli comparable to that of speech stimuli, the speech stimuli were delivered at 65 dB and the nonspeech stimuli at 85 dB, according to subjective judgment by two independent researchers. Moreover, the overall shape of the intensity profile of the nonspeech stimuli was closely matched to that of the speech counterparts.

2.2. Participants

Twenty native listeners of Mandarin at Peking University in Beijing (10 female and 10 male, mean age = 23.2 yr, SD = 2.2), 19 native listeners of Cantonese at the Chinese University of Hong Kong (10 female and 9 male, mean age = 21.6 yr, SD = 1.6), and 20 native German listeners at the Max Plank Institute in Leipzig, Germany (11 female and 9 male, mean age = 28.5, SD = 4.6) were recruited for this study. Except for one Cantonese listener, who failed to complete the discrimination test due to fatigue, all other listeners completed both the identification and discrimination tests. No participant reported any speech, language, or hearing difficulty. All participants were paid for their participation, and gave informed consent in compliance with a protocol approved by the Survey and Behavioral Research Ethics Committee of The Chinese University of Hong Kong.

2.3. Procedure

2.3.1. Identification test

Two sounds (the two endpoints of a continuum) were played multiple times (3–8 times according to participants' request) to participants at the beginning of the test for the specific continuum (i.e. in front of the practice block), and the participants were asked to remember these two representative sounds (the level-end one was defined as 'Sound 1', while the contour-end one was defined as 'Sound 2'). The stimuli in each continuum were presented to the participants in random order, and the participants were asked to press key '1' when they heard 'Sound 1' and to press key '2' when they heard 'Sound 2' (two-alternative forced choice, 2AFC). By doing so, the three groups could share the same set of instructions, minimizing the bias due to different sets of instructions. Once a response was collected or 3500 ms had elapsed from the onset of the stimulus, whichever came first, the next stimulus was presented automatically following a 300 ms pause. The 11 stimuli were repeated twice in a block. There were 5 such testing blocks for each tone continuum. There was an additional practice block before the testing blocks for each tone continuum.

The same procedure was repeated for each of the four continua: speech rising continuum, speech falling continuum, nonspeech rising continuum, and nonspeech falling continuum. The order of continuum presentation was counterbalanced across participants.

2.3.2. Discrimination test

In order to fit the whole experiment into one experimental session, participants carried out AX discrimination tests only for the two speech continua, i.e., discrimination tasks were not performed for the nonspeech continua. Stimuli were presented in pairs with a 500 ms inter-stimulus interval (ISI). For each speech continuum, a total of 29 pairs were presented in random order. Of these pairs, 18 consisted of two different stimuli separated by two steps (in our pilot experiments, we have found 1-step differences, i.e. 3 Hz, to be too difficult for subjects to discriminate) on the speech continuum (*different pairs*), in either forward (1-3, 2-4, 3-5, 4-6, 5-7, 6-8, 7-9, 8-10, 9-11) or reverse order (3-1, 4-2, 5-3, 6-4, 7-5, 8-6, 9-7, 10-8, 11-9), and 11 consisted of the 11 stimuli on the speech continuum each paired with itself (*same pairs*). The above 29 pairs were repeated seven times, resulting in 203 pairs in total being presented to each participant for each speech continuum. After hearing each pair, participants were instructed to judge whether the two stimuli were the same or different, and to respond by pressing a key ('*v*' for "same", and '*n*' for "different": these two keys were chosen because they are more or less in the lower middle of the keyboard.). Once a response was collected or 2500 ms had elapsed from the onset of the second stimulus in a pair, whichever came first, the next stimulus was presented automatically following a 300 ms pause. Accuracy scores were shown after each block to encourage participants to try their best. The order of continuum presentation was also counterbalanced across participants.

2.4. Data analysis

To investigate the effects of language experience and domain specificity (speech or nonspeech) on identification and discrimination performance, we obtained individual measures for each participant based on three essential characteristics of CP: position of category boundary, width of category boundary, and discrimination peak.

2.4.1. Identification scores

Given a particular stimulus, the identification score was defined as the percentage of responses with which participants identified that stimulus as being either 'Sound 1' or 'Sound 2'. Boundary position and boundary width were assessed by probit analyses of individual identification curves (Finney, 1971): The boundary position was defined as the 50% crossover points, and the boundary width was defined as the linear distance between the 25th and 75th percentiles as determined by the mean and standard deviation obtained from probit analysis (Hallé, Chang, & Best, 2004). Please note: We have replaced 0% with 0.1%, and 100% with 99.9% for individual identification curves at both ends, in order to fit the asymptotic property of probit function.

2.4.2. Discrimination scores

To obtain the discrimination scores of each pair, calculation was made using the formula described by Xu et al. (2006). Each cohort comprised all the trials involving the four types of pairwise comparison (*AB*, *BA*, *AA*, and *BB*, for stimuli *A* and *B* separated by two steps), where *AB* and *BA* were "different" pairs, and *AA* and *BB* are "same" pairs. Adjacent comparison cohorts contained overlapping *AA* or *BB* trials. The score (or accuracy) *P* for each comparison cohort was defined by

$$P = P(''S''|S) \cdot P(S) + P(''D''|D) \cdot P(D)$$

where $P(''S''|S)$ denotes the percentage of "same" responses to all "same" pairs, i.e., the correct responses. Likewise, $P(''D''|D)$ denotes the percentage of "different" responses to all "different" pairs. $P(S)$ and $P(D)$ are the percentages of "same" and "different" pairs in each cohort, respectively.

3. Results

3.1. Identification and discrimination curves

Identification and discrimination curves are shown in Figs. 3 and 4. The estimated boundary position and width, obtained by probit analysis, are shown in Table 1.

3.2. Position of category boundary

A three-way mixed design repeated measures analysis of variance (ANOVA) was conducted to determine the impact of language *group* (Mandarin, Cantonese, and German), tone *continuum* (rising and falling continua), and stimulus *context* (speech and nonspeech) on the boundary *position*, with *group* as the between-subject factor, and *continuum* and *context* as the within-subject factors. Corrections for violations of sphericity were made, where appropriate, using the Greenhouse–Geisser method, and when necessary, Tukey's HSD post hoc test was applied to make pairwise comparison. All effects were reported as significant at $p<0.05$.

A significant main effect of *context* ($F(1, 56)=5.16$; $p=0.027$) indicated that the perceptual boundary positions were significantly different across speech and nonspeech contexts, with the nonspeech boundary positions occurring consistently at a smaller stimulus number, as shown in Table 1. As depicted in Fig. 1, the stimuli with smaller numbers have steeper slope, with stimulus Number 1 having the steepest rising pitch contour in the rising continua, and the steepest falling pitch contour in the falling continua. The smaller number of the boundary positions means that participants required a steeper pitch slope to perceive the stimulus as a contour pitch. There was no significant interaction between *group* × *context*, further indicating that participants from the three language groups generally treated the contrast between speech and nonspeech in the same way in terms of perceptual boundary positions.

A strong significant main effect of *continuum* ($F(1, 56)=92.83$; $p<0.001$) indicated that the perceptual boundary positions were significantly different for the rising continua and the falling continua, with the boundary positions in the rising continua at a consistently smaller stimulus number than the falling continua counterparts, as shown in Table 1.

There was a significant interaction effect between *group* and *continuum* ($F(2, 56)=4.05$; $p=0.023$). This indicated that the variation of boundary positions for the rising and falling continua differed in the three language groups. The estimated marginal means (Table 2) were used to determine the nature of this interaction. Fig. 5 shows that the difference in boundary position between the rising and falling continua was larger for the Cantonese group than for the other two groups. This indicates that the Cantonese listeners differed in perceiving the two directions (rising vs. falling) of stimuli to a greater extent than the other two groups of listeners.

There was no significant main effect of *group* ($F(2, 56)=0.111$, $p=0.895$), indicating that the mean boundary positions across the three language groups were in general the same (about 5.9), shown in the last column of Table 2.

3.3. Width of category boundary

A similar statistical analysis was applied to boundary *width*. A significant main effect was found only for *group* ($F(2, 56)=6.35$; $p=0.003$). Tukey's HSD post hoc pairwise comparisons of the three groups indicated that both the Mandarin group (mean=1.65, $p=0.012$) and the Cantonese group (mean=1.60, $p=0.007$) had significantly narrower boundary widths than the German group (mean=2.25), but that the boundary widths for the Mandarin and Cantonese groups were not significantly different.

The only significant interaction effect was found for *group* × *context* ($F(2, 56)=4.52$; $p=0.015$). The estimated marginal means (Table 3) were used to determine the nature of this interaction. The interaction figure (Fig. 6) shows that the difference in boundary width between the speech and nonspeech continua was larger for the Mandarin group than for the other two groups. This indicates that Mandarin participants differed most in perceiving speech from nonspeech in terms of boundary width. A within-group paired *t*-test on boundary width, contrasting speech and nonspeech, was found to be significant only for the Mandarin group ($t=2.46$; $p<0.024$; $df=19$).

3.4. Discrimination peaks

As for the rising continuum, as shown in the left panel of Fig. 7, the accuracy for the two tone language groups reached their maxima at pair 6–8. One-way ANOVA revealed a significant difference in accuracy at this position across the three language groups ($F(2, 55)=3.92$; $p=0.026$). Tukey's HSD post hoc comparison revealed that the accuracy for the Mandarin group, 69.6%, was significantly greater than that for the German group, 60.2%

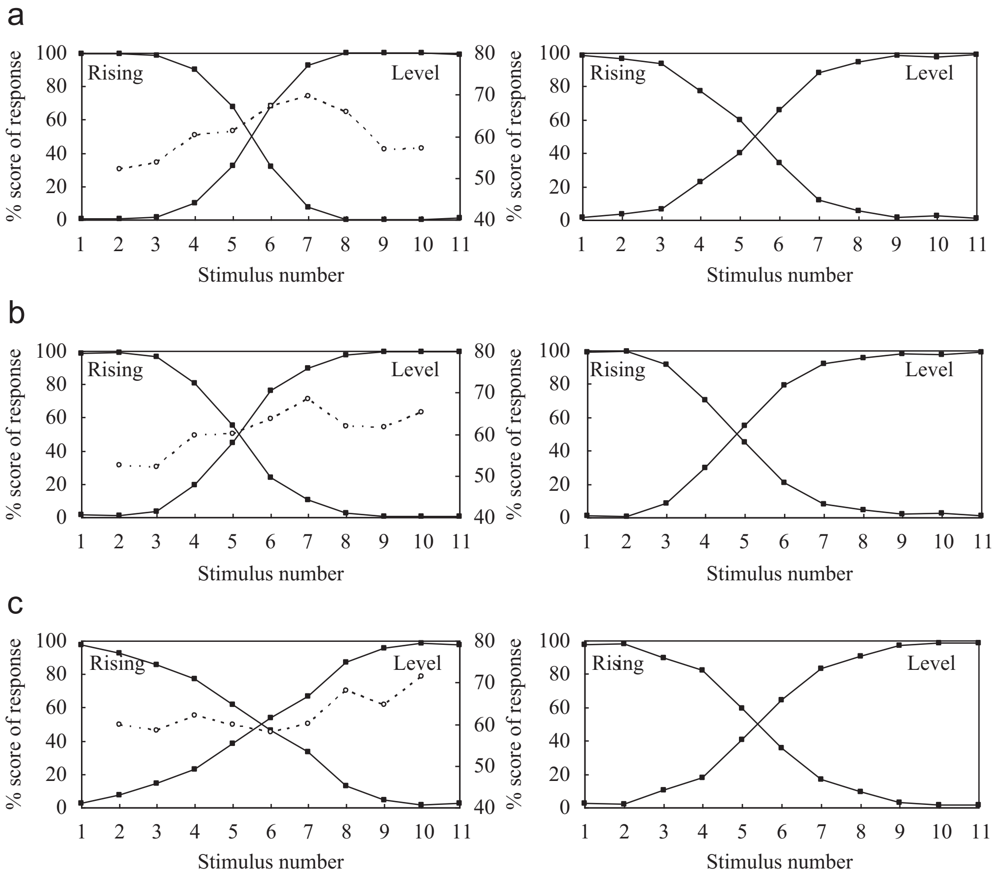

Fig. 3. Identification (solid lines) and discrimination (dashed lines) curves pooled across participants for the rising continua, with the left column of panels for the speech continuum, the right column of panels for the nonspeech continuum. The left *y*-axis indicates the percentage score of responses, while the right *y*-axis, where present, indicates the discrimination accuracy. (a) Mandarin participants for the rising continua, (b) Cantonese participants for the rising continua and (c) German participants for the rising continua.

($p = 0.033$), and that the accuracy for the Cantonese group, 68.4%, was marginally significantly greater than that for the German group ($p = 0.08$), but no significant difference between the two tone language groups ($p = 0.95$) was observed.

The German group reached an accuracy maximum toward the level end, i.e., pair 9–11. However, it was interesting to observe that the Cantonese group also reached a local maximum at this position. One-way ANOVA revealed a strongly significant difference in accuracy of this pair across the three language groups ($F(2,\ 55) = 9.85$; $p < 0.001$). Tukey's HSD post hoc comparison revealed that the accuracy for the German group, 71.4%, was significantly higher than that for the Mandarin group, 57.1% ($p < 0.001$), and that the accuracy for the Cantonese group, 65.2%, was also significantly higher than that for the Mandarin group ($p = 0.047$), but there was no significant difference between the German and Cantonese groups ($p = 0.16$).

As for the falling continuum, as shown in the right panel of Fig. 7, the Mandarin group reached an accuracy maximum at pair 6–8, but there was no significant difference in accuracy at this position across the three language groups ($F(2,\ 55) = 1.52$; $p = 0.23$). Similarly, the Cantonese group reached an accuracy maximum at pair 7–9, but there was no significant difference in accuracy at this position across the three language groups ($F(2,\ 55) = 1.13$; $p = 0.33$).

The German group still reached an accuracy maximum toward the level end, i.e., pair 9–11, for the falling continuum. It was also interesting to observe that the Cantonese group also had a much higher accuracy at this position than the Mandarin group. One-way ANOVA revealed a strongly significant difference in accuracy at this position across the three language groups ($F(2,\ 55) = 16.12$; $p < 0.001$). Tukey's HSD post hoc comparison revealed that the accuracy for the German group, 74.8%, was significantly higher than that of the Mandarin group, 58.6% ($p < 0.001$), and that the accuracy for the Cantonese group, 71.4%, was significantly higher than that of the Mandarin group ($p < 0.001$). However, there was no significant difference between the German and Cantonese groups ($p = 0.52$).

4. Discussion

This study has examined pitch contour perception by three groups of listeners: one German group, and two Chinese groups (Cantonese and Mandarin). Our results confirm some of the findings reported previously that have contrasted native tone language listeners and non-tone language listeners. Given the different tone inventories of Mandarin and Cantonese, the influence of these two tone systems on pitch contour perception was indeed reflected in the identification and discrimination curves for these two tone language groups.

4.1. Influence of language experience reflected by the identification curves

In the present study, different language groups did not show significantly different boundary positions for either the rising continua (around 5.32) or the falling continua (around 6.47), as shown in Table 2. This result is consistent with the result reported by Xu, Gandour, and Francis (2006), where native English listeners

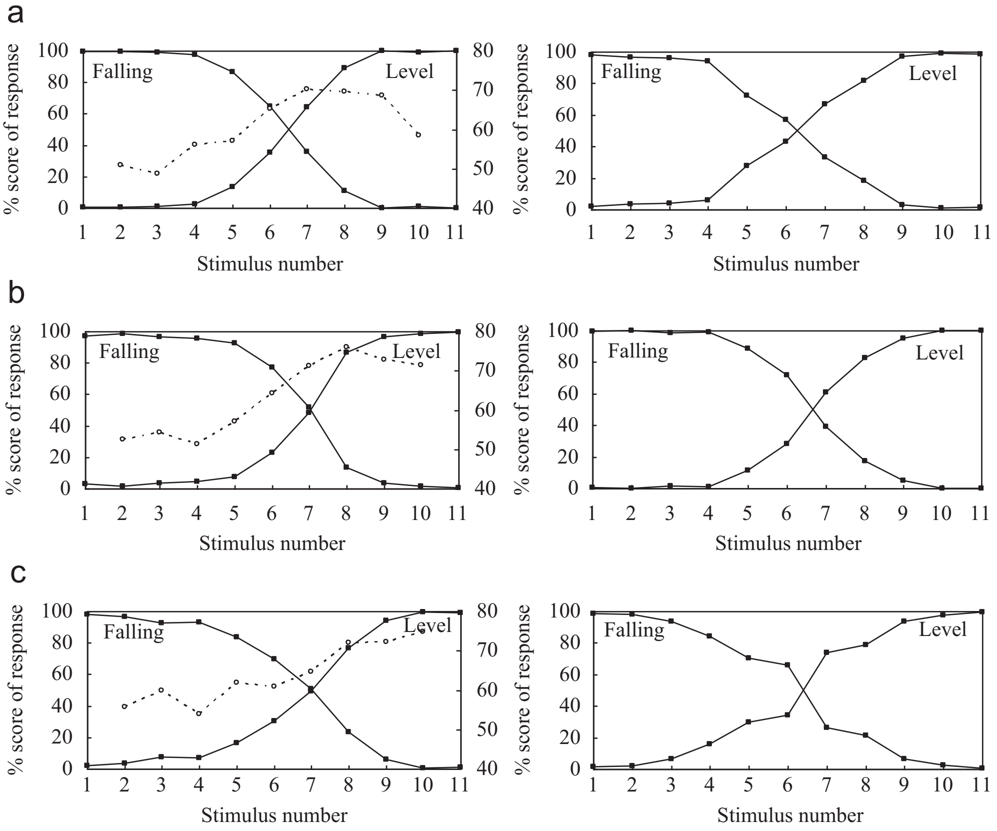

Fig. 4. Identification (solid lines) and discrimination (dashed lines) curves pooled across participants for the falling continua, with the left column of panels for the speech continuum, the right column of panels for the nonspeech continuum. The left *y*-axis indicates the percentage score of responses, while the right *y*-axis, where present, indicates the discrimination accuracy. (a) Mandarin participants for the falling continua, (b) Cantonese participants for the falling continua and (c) German participants for the rising continua.

Table 1
Derived categorical boundary position and width for each subgroup (SP=speech; NS=nonspeech).

Continuum	Mandarin		Cantonese		German	
	Position	Width	Position	Width	Position	Width
SP_Rising	5.46	1.21	5.17	1.49	5.61	2.54
SP_Falling	6.43	1.50	6.74	1.66	6.61	2.10
NS_Rising	5.31	1.85	4.96	1.78	5.42	2.17
NS_Falling	6.21	2.04	6.65	1.49	6.17	2.21

Table 2
Marginal means of the boundary positions of the rising and falling continua across language groups.

Language groups	Rising	Falling	Mean
Mandarin	5.38	6.32	5.85
Cantonese	5.07	6.70	5.89
German	5.52	6.39	5.96
Average	5.32	6.47	5.90

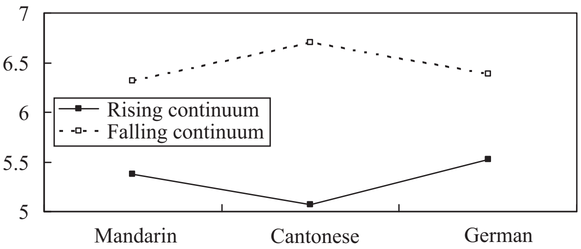

Fig. 5. The interaction of the marginal means of the boundary positions for the rising and falling continua across language groups.

Table 3
Marginal means of the boundary width of the speech and nonspeech continua across language groups.

Language groups	Speech	Nonspeech	Mean
Mandarin	1.35	1.95	1.65
Cantonese	1.58	1.63	1.60
German	2.31	2.19	2.25

and Mandarin listeners showed the same boundary position. They also reported that the boundary positions of nonspeech (harmonic tones) were found more toward the level end (boundary position <4 for nonspeech, and >4 for speech). (Note that Stimulus number 4 was the middle stimulus in their continuum, and Number 1 was perfectly level in their study, but was steepest in our study.) In order to explain this result, they assumed that the level end of the continuum was represented by a flat contour (without F0 movement), and that this stimulus was therefore likely to serve as an anchor point (we will refer to this later as the "anchor hypothesis"). They argued that speech is more complex than their nonspeech stimuli, and assumed that this greater complexity would reduce the overall pitch sensitivity (we will refer to this later as the "complexity hypothesis"). Therefore, listeners required more steps (i.e., steeper slope in the pitch contour) to make a "rising" response to the speech stimuli as compared to nonspeech.

However, in our experiment, the boundary positions of nonspeech (here, pure tone) were found more toward the contour end (the stimuli with steepest slope in the pitch contour), and the pure tones were even less complex than the harmonic tones. This finding is not consistent with the finding of Xu, Gandour, and Francis (2006). We argue that the "richness of harmonics" facilitates pitch perception. Both speech and harmonic tones have richer harmonic structure than pure tones; indeed, pure tones have no harmonic structure at all. Considering the "anchor hypothesis", the "complexity hypothesis", and the "richness of harmonics hypothesis" together, and assuming that the "richness of harmonics" can override "complexity", we infer that listeners required more steps for nonspeech stimuli (total absence of harmonic structure) to be identified as contour sounds in our study. By doing so, we can also explain why listeners required fewer steps for harmonic tones (with harmonic structure like speech, but less complex) in Xu, Gandour, and Francis (2006).

We also note that two harmonics of the harmonic tones in Xu, Gandour, and Francis (2006), were above 1000 Hz. Although we did compensate our pure tone stimuli by 20 dB in loudness (speech stimuli were delivered at 65 dB, while nonspeech stimuli were delivered at 85 dB in this study), humans are much more sensitive in the frequency range of 1000–5000 Hz than 100–150 Hz (Fletcher & Munson, 1933). Therefore, the inconsistency in our respective results might also be due to the different frequency ranges for the nonspeech stimuli in these two studies. This issue needs further examination.

For all the listeners, the boundary positions for the falling continua were more toward the level end compared to those for the rising continua. This might be largely due to the physical properties of the stimuli. As shown in Fig. 1, in the rising continua, the pitch contours of the stimuli were brought closer and closer toward the end, i.e., they were acoustically more similar toward the end of the stimulus, but in the falling continuum, the pitch contours of the stimuli were further apart toward the end, i.e., acoustically less similar toward the end of the stimulus. The lifetime of auditory sensory memory has been estimated to be about 300 ms (Cowan, 1984, 1987), so the participants' perception might rely more on the later part of each stimulus (the duration of our stimuli is 500 ms). Therefore, listeners might require more steps from the perfectly level stimulus for a rising stimulus to be perceived as rising.

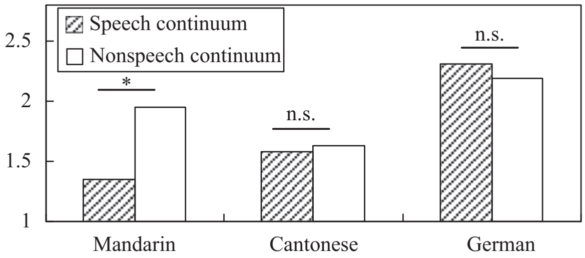

Fig. 6. The interaction figure of the marginal means of the boundary width of the speech and nonspeech continua across language groups, with '*' denoting significance at $p < 0.05$.

We can see from Figs. 3 and 4 that the boundary was sharper (i.e., narrower) for tone language (Mandarin and Cantonese) listeners than for non-tone language (German) listeners. This finding is highly consistent across several studies (Hallé, Chang, & Best, 2004; Wang, 1976; Xu, Gandour, & Francis, 2006). The differences in boundary width between speech and pure tones were more salient for Mandarin listeners. Considering only the Mandarin and German groups, our pattern of boundary slope was similar to that of Xu, Gandour, and Francis (2006): native tone language listeners exhibited a sharper boundary for the speech continuum than for the nonspeech continuum, but non-tone language listeners exhibited the reverse pattern. But why did the Cantonese group, as native tone language listeners, show almost the same boundary width for speech and nonspeech continua? We argue that, due to the richer tone inventory of Cantonese, Cantonese listeners make greater use of pitch information in producing and perceiving speech, further strengthening their ability in pitch perception, and that this ability is carried over to the nonspeech domain. Musicians with tone language experience have been reported to have significantly greater prevalence of absolute pitch (Deutsch, Henthorn, Marvin, & Xu, 2006; Deutsch, Dooley, Henthorn, & Head, 2009). We further hypothesize that the prevalence of absolute pitch of native Cantonese musicians would be even greater than that of native Mandarin musicians, because Cantonese has a much richer tone inventory. This is an interesting topic that merits further study.

4.2. Influence of language experience reflected by the discrimination curves

In the current study, the discrimination test was performed only for the two speech continua. The discrimination peaks of the German listeners were located at the level ends for both continua, reflecting the psychophysical boundaries, which is consistent with previous relevant findings (Hallé, Chang, & Best, 2004; Wang, 1976; Xu, Gandour, & Francis, 2006). The discrimination peak for the rising continuum of the tone language listeners was located at pair 6–8, with significantly higher accuracy than the German listeners attained at this position. This finding is consistent with the literature (Wang, 1976) in which similar stimuli were used. More interestingly, the Cantonese group showed a local maximum at the level end, which might reflect the influence of the Cantonese low rising tone, Tone 5, which rises in pitch only

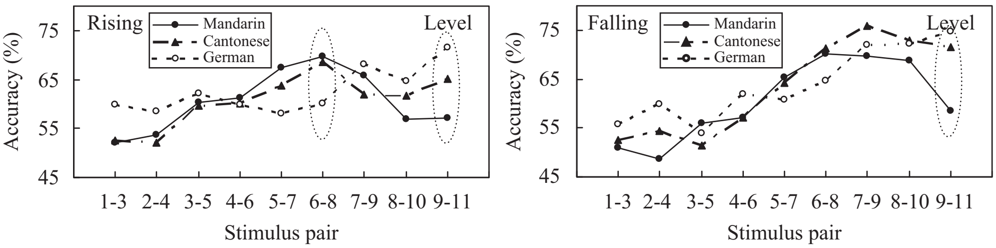

Fig. 7. Two-step discrimination scores pooled across participants in the two speech continua: rising (left) and falling (right).

slightly, shaping Cantonese listeners' sensitivity in distinguishing level vs. non-level along the rising direction.

The discrimination peak for the falling continuum of Mandarin listeners was also located at pair 6–8, but the discrimination peak of Cantonese listeners was located at pair 7–9. This might reflect the influence of the low falling tone in Cantonese, Tone 4 (only slightly falling in its pitch contour), shaping Cantonese listeners' sensitivity in distinguishing level vs. non-level along the falling direction.

Taking the two speech continua together, the Cantonese group performed similarly to the German group at the level end, i.e., the pair 9–11. We claimed above that the high accuracy at this pair for the German group reflects the psychophysical boundary. It was unclear whether the higher accuracy for the Cantonese group also reflects a psychophysical boundary. It was reported in Wang (1976) that the co-existence of separate linguistic and psychophysical boundaries only occurred for well-trained Chinese psychologists, who were familiar with psychophysical experiments. This point was confirmed with our Mandarin group, for which only linguistic boundaries were demonstrated. In our participant recruitment, we deliberately avoided participants from either psychological or linguistic fields. Therefore, it is unlikely that the higher accuracies for the Cantonese group at pair 9–11 merely reflect psychophysical boundaries, but rather that they do reflect the influence of linguistic tones.

It has been reported in the literature that the discrimination peaks were usually well aligned with identification boundaries, especially linguistic-related boundaries (Liberman et al., 1957; Wang, 1976; Xu, Gandour, & Francis, 2006), and that the discrimination could be predicted by the identification (Liberman et al., 1957; Xu, Gandour, & Francis, 2006). However, our discrimination peaks, except for the speech falling continuum for Mandarin listeners, were misaligned with the corresponding identification crossovers. We argue that this discrepancy might be due to the task requirements. In our task, the listeners were asked to identify whether the heard stimuli were more like 'Sound 1' or 'Sound 2', which might force the listeners to focus more on the physical properties of the stimuli. However, the discrimination peaks would inherently reflect the corresponding linguistic boundaries. Therefore the discrepancy here might be partially due to difference in linguistic vs. non-linguistic processing.

5. Conclusion

In this study, we have examined the influence of different language experience on the perception of pitch contours in the framework of CP. From the identification curves shown in Figs. 3 and 4, we see clearly that changes from one category to another are more abrupt for the two groups of tone language listeners. This is also reflected in the narrower boundary width shown in Table 1, and in Table 3 where the boundary widths for the rising and falling continua were pooled together, especially in the context of speech. Compared with the German group, the two Chinese groups perceived the contrast of level vs. rising, and level vs. falling pitch contours more categorically. From the discrimination curves shown in Fig. 7, the German group shows only the psychophysical boundary, while the two tone language groups show linguistic boundaries, and these linguistic boundaries are further shaped by the different tone inventories.

Acknowledgements

The work described in this paper was partially supported by grants from the Shun Hing Institute of Advanced Engineering of The Chinese University of Hong Kong (Project No. BME-8115020), from National Science Foundation of China (NSFC: 11074267), and from the Research Grant Council of Hong Kong. We thank the members of the Language Engineering Laboratory of The Chinese University of Hong Kong for many helpful discussions. We thank the editor and the two reviewers for their constructive help in improving the paper.

References

Abramson, A. S. (1972). Tonal experiments with whispered Thai. In A. Valdman (Ed.), *Papers in linguistics and phonetics to the memory of Pierre Delattre* (pp. 31–34). Mouton: The Hague.

Abramson, A. S. (1979). The noncategorical perception of tone categories in Thai. In B. Lindblom, & S. Ohman (Eds.), *Frontiers of speech communication* (pp. 127–134). London: Academic Press.

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

Boersma, P., & Weenink, D. (2009). Praat: Doing phonetics by computer. 〈http://www.fon.hum.uva.nl/praat/〉.

Bornstein, M. H., Kessen, W., & Weiskopf, S. (1976). Color vision and hue categorization in young infants. *Journal of Experimental Psychology: Human Perception & Performance*, *2*, 115–129.

Chao, Y.-R. (1930). A system of tone letters. *Le Maître Phonétique*, *45*, 24–27.

Cowan, N. (1984). On short and long auditory stores. *Psychological Bulletin*, *96*, 341–370.

Cowan, N. (1987). Auditory sensory storage in relation to the growth of sensation and acoustic information extraction. *Journal of Experimental Psychology: Human Perception and Performance*, *13*, 204–215.

Deutsch, D., Dooley, K., Henthorn, T., & Head, B. (2009). Absolute pitch among students in an American music conservatory: Association with tone language fluency. *Journal of Acoustical Society of America*, *125*(4), 2398–2403.

Deutsch, D., Henthorn, T., Marvin, E., & Xu, H. (2006). Absolute pitch among American and Chinese conservatory students: Prevalence differences, and evidence for a speech-related critical period (L). *Journal of Acoustical Society of America*, *119*(2), 719–722.

Eimas, P. D., Siqueland, E. R., Jusczyk, P. W., & Vigorito, J. (1971). Speech perception in infants. *Science*, *171*, 303–306.

Etcoff, N. L., & Magee, J. J. (1992). Categorical perception of facial expressions. *Cognition*, *44*, 227–240.

Finney, D. J. (1971). *Probit analysis* (3rd ed). Cambridge, UK: Cambridge University Press.

Fletcher, H., & Munson, W. A. (1933). Loudness, its definition, measurement and calculation. *Journal of Acoustical Society of America*, *5*, 82–108.

Francis, A. L., Ciocca, V., & Ng, B. K. (2003). On the (non)categorical perception of lexical tones. *Perception & Psychophysics*, *65*(7), 1029–1044.

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

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

Howie, J. M. (1976). *Acoustical studies of Mandarin vowels and tones*. Cambridge, UK: Cambridge University Press.

Johnson, K. (2005). Speaker normalization in speech perception. In D. B. Pisoni, & R. E. Remez (Eds.), *The handbook of speech perception* (pp. 363–389). Malden, MA, Oxford: Blackwell Pub.

Kuhl, P. K., & Miller, J. D. (1975). Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science*, *190*, 69–72.

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, *5*, 831–843.

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

Liberman, A. M. (1996). *Speech: A special code*. MIT Press.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*, 358–368.

Liu, S., & Samuel, A. G. (2004). Perception of Mandarin lexical tones when F0 information is neutralized. *Language and Speech*, *47*, 109–138.

Moore, C. B., & Jongman, A. (1997). Speaker normalization in the perception of Mandarin Chinese. *Journal of Acoustical Society of America*, *102*(3), 1864–1877.

Morse, P. A., & Snowden, C. T. (1975). An investigation of categorical speech discrimination by rhesus monkeys. *Perceptual Psychophysics*, *17*, 9–16.

Moulines, E., & Laroche, J. (1995). Non-parametric techniques for pitch-scale and time-scale modification of speech. *Speech Communication*, *16*, 175–205.

Ohms, V. R., Gill, A., Van Heijningen, C. A., Beckers, G. J., & Ten Cate, C. (2010). Zebra finches exhibit speaker-independent phonetic perception of human speech. *Proceedings of the Royal Society: B*, *277*, 1003–1009.

Peng, G. (2006). Temporal and tonal aspects of Chinese syllables: A corpus-based comparative study of Mandarin and Cantonese. *Journal of Chinese Linguistics*, *34*(1), 134–154.

Repp, B. H. (1984). Categorical perception: Issues, methods, findings. In: N.J. Lass (Ed.), Speech and language: Advances in basic research and practice (Vol. 10) (pp. 243–335). New York: Academic Press.

Rivera-Gaxiola, M., Silva-Pereyra, J., & Kuhl, P. K. (2005). Brain potentials to native- and non-native speech contrasts in seven and eleven-month-old American infants. *Developmental Science*, *8*(2), 162–172.

Wang, W. S.-Y. (1967). Phonological features of tone. *International Journal of American Linguistics*, *33*(2), 93–105.

Wang, W. S.-Y. (1972). The many uses of F0. In A. Valdman (Ed.), *Linguistics and phonetics to the memory of Pierre Delattre* (pp. 487–503). The Hague: Mouton.

Wang, W. S.-Y. (1973). The Chinese language. *Scientific American*, *228*, 50–63.

Wang, W. S.-Y. (1976). Language change. *Annals of the New York Academy of Sciences*, *208*, 61–72.

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

Wong, P. C. M., & Diehl, R. L. (2003). Perceptual normalization for inter- and intratalker variation in Cantonese level tones. *Journal of Speech, Language, and Hearing Research*, *46*, 413–421.

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

Yip, M. (2002). *Tone*. Cambridge, UK: Cambridge University Press.

Zee, E. (1978). Duration and intensity as correlates of F0. *Journal of Phonetics*, *6*, 213–220.