

The Influence of Linguistic Experience on the Cognitive Processing of Pitch in Speech and Nonspeech Sounds

Tessa Bent, Ann R. Bradlow, and Beverly A. Wright
Northwestern University

In the present experiment, the authors tested Mandarin and English listeners on a range of auditory tasks to investigate whether long-term linguistic experience influences the cognitive processing of nonspeech sounds. As expected, Mandarin listeners identified Mandarin tones significantly more accurately than English listeners; however, performance did not differ across the listener groups on a pitch discrimination task requiring fine-grained discrimination of simple nonspeech sounds. The crucial finding was that cross-language differences emerged on a nonspeech pitch contour identification task: The Mandarin listeners more often misidentified flat and falling pitch contours than the English listeners in a manner that could be related to specific features of the sound structure of Mandarin, which suggests that the effect of linguistic experience extends to nonspeech processing under certain stimulus and task conditions.

Keywords: Mandarin tone, pitch processing, nonspeech sounds, cross-language speech perception

Human infants come into the world with the ability to develop the speech sound categories of any language, but within the 1st year of life, they begin to tune in specifically to the sounds of the ambient language (e.g., Kuhl, 2000; Werker & Tees, 1984). Whether the effect of this perceptual tuning extends beyond the speech domain and influences the processing of nonspeech sounds is the question of interest here. Although exposure to language in general may affect auditory processing, this study focuses on whether the observed variation in sound structure across languages is reflected in variation across language groups in the processing of nonspeech sounds.

It is possible that exposure to the sound structure of a particular language has no influence on nonspeech sound processing. There are two interpretations for this potential finding. The first possibility is that cross-linguistic differences in sound structure may not be extensive enough to induce differences in the processing of nonspeech sounds. Alternatively, long-term cognitive representations and auditory processing of speech and nonspeech sounds may be separated early on, with the consequence that they do not

influence each other. This view is in accord with models of speech perception that postulate separate modules for speech and nonspeech representation and processing (e.g., Liberman & Mattingly, 1989). In contrast, the processing and cognitive representation of speech and nonspeech sounds may not be fully separated, in which case long-term exposure to the speech sounds of the ambient language may exert a measurable influence on nonspeech sound processing. This outcome would be in line with proposals for shared speech and nonspeech processing and representation mechanisms (e.g., Kewley-Port, Watson, & Foyle, 1988; Kluender, Diehl, & Killeen, 1987; Kuhl, 1981). Previous research on the effect of long-term linguistic experience on nonspeech perception has yielded apparently discrepant results. Several studies have demonstrated a clear effect of linguistic experience on the perception of speech sounds but no effect on the perception of relatively simple nonspeech sounds when presented in a discrimination task. For example, native Japanese listeners, unlike native English listeners, were poor at discriminating English /r/ and /l/ when presented as speech (three synthetic formants, in which only the third formant varied) but performed similarly to native English listeners when presented with an isolated third formant (Miyawaki et al., 1975). Similarly, native English listeners had difficulty discriminating Thai tonal contours in natural speech but performed as well as native tone-language listeners (Cantonese and Thai) when these stimuli were either low-pass filtered or presented as music (Burnham et al., 1996). Finally, pure-tone pitch discrimination thresholds did not differ between listeners from tone and nontone languages, which vary dramatically in their use of pitch contrasts for linguistic purposes (Burns & Sampat, 1980). It should be noted that most of these studies confounded the stimulus variables of speech status and complexity because most speech stimuli are more complex than the nonspeech stimuli.

In contrast to studies that have found no effect of linguistic experience on the processing of nonspeech sounds, one study showed an effect of language background on a test of nonspeech perception. Specifically, nontone-language listeners (English)

Tessa Bent and Ann R. Bradlow, Department of Linguistics, Northwestern University; Beverly A. Wright, Department of Communication Sciences and Disorders and Institute for Neuroscience, Northwestern University.

Tessa Bent is now at the Department of Psychology, Indiana University Bloomington.

An earlier version of this study was presented at the 145th meeting of the Acoustical Society of America, Nashville, Tennessee, April 2003. This work was supported by National Institutes of Health—National Institute on Deafness and Other Communication Disorders Grants R03 DC 03762 (Ann R. Bradlow), R01 DC 04453 (Beverly A. Wright), and F31 DC 006515 (Tessa Bent); by a Cognitive Science Graduate Fellowship for Interdisciplinary Research from Northwestern University; and by a grant-in-aid of research from Sigma Xi.

Correspondence concerning this article should be addressed to Ann R. Bradlow, Department of Linguistics, Northwestern University, 2016 Sheridan Road, Evanston, IL 60208. E-mail: abradlow@northwestern.edu

were better than tone-language listeners (Cantonese) at discriminating frequency differences in nonspeech tone complexes that varied minimally from synthetic syllables, which suggests an influence of language background on this task with these nonspeech stimuli (Francis & Ciocca, 2003).

A view of sound processing as varying along a continuum from “speech” to “nonspeech,” rather than being divided into a strict dichotomy between speech and nonspeech processing, provides an adequate account of these previous, seemingly discrepant results. It is possible that when nonspeech stimuli whose spectral and/or temporal properties bear some resemblance to speech stimuli are presented to listeners in a task that encourages classification into broad categories (as occurs in speech perception), then some response differences across listeners from different language backgrounds may be evident. Conversely, listeners from different language backgrounds may show identical responses to stimuli that are highly dissimilar to speech stimuli in a task that involves minimal influence of higher level categorical structure. This response pattern would indicate that long-term experience with the sounds of a particular language shapes some processing mechanisms on which both speech and nonspeech processing rely, whereas other basic hearing abilities are impervious to cross-linguistic sound structure differences.

In the present study, we explored the influence of linguistic experience on speech and nonspeech processing by testing listeners from two different language backgrounds, Mandarin and English, on a range of auditory perceptual tasks. The study included, at one extreme, a task in which listeners identified naturally produced syllables as members of linguistic categories and, at the other extreme, a task in which listeners made fine-grained discriminations between simple nonspeech sounds. The key condition was between these two extremes, in which listeners identified nonspeech sounds that were modeled after natural syllables as members of abstract categories. This cross-language approach complements investigations of the influence of short-term exposure to speech sounds on the processing of nonspeech sounds and vice versa. Studies of this type have found that speech and nonspeech sounds that are similar in their temporal and spectral characteristics can influence one another in perception (Stephens & Holt, 2003) and memory (Semal, Demany, Ueda, & Halle, 1996; Tremblay, Nicholls, Alford, & Jones, 2000). In the present study, we attempt to extend this general finding by investigating the effect of long-term linguistic experience on pitch processing in both speech and nonspeech sounds.

Method

Listeners

Thirteen native Mandarin speakers (7 men and 6 women; average age = 25 years, $SD = 2.21$) and 12 native English speakers (6 men and 6 women;

average age = 20 years, $SD = 1.44$) participated. Mandarin and English use pitch variation for different linguistic purposes and place these patterns on different-sized units. Mandarin (a tone language) has four different lexical tones: high level, rising, low dipping, and falling (Chao, 1968). A difference in tone signals a different lexical meaning. For example, *ma* with a high-level tone means “mother”, but *ma* with a falling tone means “curse.” In comparison, English (a nontone language) uses pitch systematically at the phrasal level. Rather than expressing lexical meaning, English pitch patterns express pragmatic meaning. For example, the sentence “Legumes are a good source of vitamins” is interpreted as a statement if there is a falling–rising pattern on *vitamins* but is interpreted as a question if there is a rising pattern on *vitamins* (Pierrehumbert, 1980). Within the system of English phrasal pitch patterns, the pitch contours are broadly comparable to those seen on Mandarin syllables (i.e., level, rising, dipping, falling).

All listeners had limited (< 3 years) or no musical experience and reported no speech or hearing impairments. The native English listeners were monolingual. The native Mandarin listeners all had come to the United States as adults and spoke English as a second language. All listeners were paid for their participation.

Stimuli and Procedure

Listeners performed three tests: Mandarin tone identification, nonspeech pitch discrimination, and nonspeech pitch contour identification. All testing was conducted in a sound-attenuated booth. In each test, audio files were played diotically over headphones (Sennheiser HD 580) through a computer sound card (SoundBlaster Live) using special-purpose experiment-running software (Superlab Pro 2.01; Cedrus Corp, 1996). No feedback was given during the experimental trials. However, a brief practice session with feedback was presented before the start of each task. The Mandarin tone identification task was always last so as not to encourage a speech-like processing strategy for the nonspeech tasks. The order of the nonspeech pitch discrimination test and the nonspeech pitch contour identification test was counterbalanced within listener groups.

Mandarin tone identification. The Mandarin tone stimuli were 64 consonant–vowel monosyllables (32 real Mandarin words and 32 nonsense syllables; see Table 1) that included consonants and vowels that occur in both Mandarin and English. One female native speaker of Mandarin was recorded reading all stimuli (using an Ariel Proport A/D sound board with a Shure SM81 microphone in a sound-attenuated booth at a 16-kHz sampling rate). The root-mean-squared amplitude was equated to 65 dB across all stimulus files.

Identification accuracy scores for Mandarin tones were obtained using a one-interval identification task. Five repetitions of the 64 stimuli were presented in random order. Listeners were instructed to listen to each monosyllable and then press one of four buttons on a response box to indicate whether the syllable had a level, rising, low-dipping, or falling tone. The four buttons were marked both with verbal descriptions (i.e., *level*, *rising*, *low-dipping*, or *falling*) and with symbols (i.e., $\bar{\quad}$, \uparrow , \downarrow , \searrow). Before the start of the experimental trials, participants were given 16 practice trials with real Mandarin words that were not included in the experimental trials (i.e., each of the four tones for the following four

Table 1
Stimuli Used in the Mandarin Tone Identification Task in Standard Pinyin Transcription

Stimulus type	Tone			
	High level	Rising	Low dipping	Falling
Word	di, duo, ti, tuo, li, la, ni, na	di, duo, ti, tuo, li, la, ni, na	di, duo, ti, tuo, li, la, ni, na	di, duo, ti, tuo, li, la, ni, na
Nonword	lia, lai, te, ne, nou, nu, nei, nai	dei, diu, dui, le, lia, te, tei, nu, nei, nai	die, diu, dui, le, lai, te, tei, ne, nou	dei, die, diu, lia, tei

syllables: /ma/, /fie/, /da/, /lao/). Participants were given feedback during the practice trials but not during the experimental trials.

Nonspeech pitch discrimination. The pitch discrimination stimuli included a standard stimulus of 150 Hz and 16 target stimuli that varied from 150.5 Hz to 158 Hz in 0.5-Hz steps. The stimulus frequency range was between the average fundamental frequency ranges for adult male and female speakers (Peterson & Barney, 1952) and therefore within the fundamental frequency range covered by adult speakers. All stimuli were 400 ms in duration, including 15-ms linear amplitude ramps at stimulus onset and offset. Pitch discrimination thresholds were measured with both pure tones and saw-tooth wave pulse trains. The saw-tooth wave-shaped pulses were 100 μ s in duration, and the time between the pulses was varied to elicit the changes in pitch (Zeng, 2002). We included these two types of stimuli to test pitch discrimination with stimuli that ranged in their similarity to speech. Pure tones bear very little resemblance to the glottal waveform of natural speech, whereas saw-tooth wave pulse trains are a closer approximation of the natural glottal waveform for speech.

We measured pitch discrimination thresholds using the method of constant stimuli and a two-interval forced-choice task. Listeners indicated whether the higher frequency stimulus was presented in the first or the second interval. The interstimulus interval was 600 ms. The pure-tone and saw-tooth wave pulse-train stimuli were presented in separate blocks comprising 320 randomized trials each. Each target was presented 20 times. On 50% of the trials, the target frequency was first, and on the other 50%, the standard was first. The discrimination threshold was defined as the smallest frequency difference at which the listener achieved at least 75% correct discrimination. Participants were given 18 practice trials with feedback before both the pure-tone and the pulse-train tasks. In these practice sessions, the type of signal was matched with the signal in the following experimental trials (pure tone or pulse train), and listeners were given practice with target stimuli with frequencies ranging from 157 Hz to 158 Hz. No feedback was provided during the experimental trials.

Nonspeech pitch contour identification. The pitch contour stimuli were sine waves that were either flat, rising, or falling in frequency (modeled after Gottfried & Reister, 2000). The 14 flat tones varied between 200 Hz and 350 Hz (200, 210, 220, 230, 240, 250, 260, 270, 280, 300, 320, 330, 340, and 350 Hz). There were 14 total rising stimuli. Seven of these rising stimuli started at 250 Hz and then rose linearly 5, 10, 15, 20, 30, 40, or 50 Hz. The other 7 started at 300 Hz and also rose linearly 5, 10, 15, 20, 30, 40, or 50 Hz. Similarly, there were 14 total falling stimuli. Seven of the falling pitch contours started at 250 Hz and fell linearly by 5, 10, 15, 20, 30, 40, or 50 Hz. The other 7 started at 300 Hz and also fell linearly by 5, 10, 15, 20, 30, 40, or 50 Hz. All stimuli were 400 ms in duration, including 15-ms linear amplitude ramps at onset and offset.

Percent correct identification scores were obtained using a single-interval identification task. Listeners were instructed to listen to each stimulus and then press a button to indicate whether they perceived the stimulus as flat, rising, or falling in pitch. The response buttons were marked with both verbal descriptions (i.e., *flat*, *rising*, or *falling*) and symbols (i.e., —, /, or \). Stimuli were presented in two blocks of 210 trials, with five repetitions of each stimulus in each block (for a total of 10 presentations of each stimulus and 420 total trials). Participants were given 12 practice trials with feedback. No feedback was provided during the experimental trials.

Results

The general pattern of results indicated that long-term linguistic experience influences the processing of nonspeech sounds when the stimuli and task bear some resemblance to speech and speech processing, respectively. There were large listener group differences for the Mandarin tone identification test, none for the nonspeech pitch discrimination test, and some for the nonspeech pitch contour identification test.

Mandarin Tone Identification

On the Mandarin tone identification test, the native Mandarin listeners identified the Mandarin tones more accurately than the native English listeners (see Figure 1; $M = 99\%$ correct, $SD = 1\%$, vs. $M = 59\%$ correct, $SD = 21\%$), $t(23) = 40.25$, $p < .01$. The Mandarin listeners exhibited a small range of performance (96% to 100% correct) because of a ceiling effect, whereas the English listeners' scores varied widely and did not overlap with any of the Mandarin listeners' scores (26% to 88% correct).

Nonspeech Pitch Discrimination

On the nonspeech pitch discrimination test, there were no significant group differences for either signal type (see Figure 2). Although the two groups of listeners performed very differently on the Mandarin tone identification test, their performance on the nonspeech pitch discrimination test was not significantly different. It is noteworthy that the within-group range of performance for the English listeners was greater than for the Mandarin listeners; however, the overall means did not differ across the groups. These results demonstrate that the influence of long-term linguistic experience did not extend to the processing of these simple nonspeech signals in a two-interval two-alternative forced-choice task requiring fine-grained discrimination.

Most previously reported thresholds of frequency discrimination with standards of 125 to 250 Hz (frequencies close to our measured target of 150 Hz) are clustered around 1 Hz (e.g., Harris, 1952; Moore, 1973; Wier, Jesteadt, & Green, 1977). Our average threshold values for the Mandarin and English listeners were 2.6 Hz and 3.5 Hz, respectively, for the pure-tone stimuli and 2.5 Hz and 3.6 Hz, respectively, for the pulse-train stimuli. A variety of factors could account for the difference between the previously reported values and the current values, including differences in listeners' experience with the task and stimuli, stimulus amplitude or duration, and/or specific task conditions.

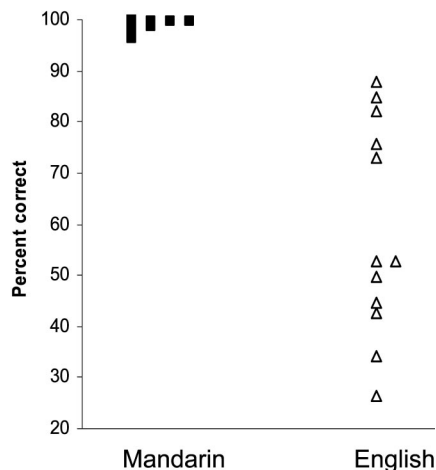


Figure 1. Individual listeners' percentage correct scores on the Mandarin tone identification task. Mandarin listeners (filled squares, left; $n = 13$) more accurately identified the Mandarin tones than English listeners (open triangles, right; $n = 12$).

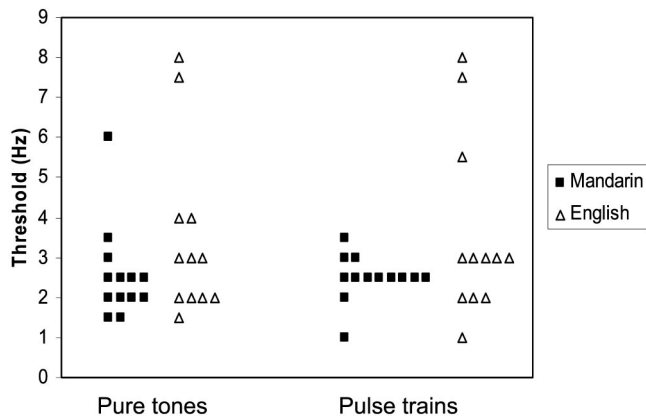


Figure 2. Mandarin (filled squares; $n = 13$) and English (open triangles; $n = 12$) listeners' discrimination thresholds did not differ for either 150-Hz pure tones (left) or pulse trains (right).

Nonspeech Pitch Contour Identification

On the nonspeech pitch contour identification test, the listener groups differed in how accurately they responded to some of the falling and flat stimuli but not to the rising stimuli. However, on the basis of a signal detection analysis on a subset of the data, where determinable, the differences between listener groups were due to differences in response bias rather than to changes in sensitivity (see the *Nonspeech Pitch Contour: Signal Detection Theory Analysis* section). Therefore, we report only a brief descriptive account of the full set of identification data, followed by a signal detection theory analysis. Although we do not include the details of the statistical analysis in this description, we fully analyzed the data, and the description is based on those analyses. All differences mentioned here were significant at the $p < .05$ level, with Bonferroni corrections for multiple tests where required.

For the rising pitch contours (Figure 3, top row), the listener groups performed similarly. In contrast, the Mandarin listeners misidentified the falling pitch contours more often than did the English listeners (Figure 3, middle row). However, this effect was observed only for pitch falls with a moderate amount of change. The stimuli that fell by 15 Hz were identified correctly more often by the English listeners than by the Mandarin listeners. For the contours with small amounts of change, both groups of listeners misidentified the vast majority of stimuli, and for the contours with relatively large amounts of change, both listener groups identified the stimuli very accurately. Finally, for the flat contours, the Mandarin listeners' responses were less accurate than those of the English listeners for the stimuli at the extremes of the frequency range tested (Figure 3, bottom row). The Mandarin listeners' overall rate of correct responses was higher for stimuli with intermediate frequencies than for stimuli with relatively low or high frequencies. In contrast, the English listeners' overall correct response rate was constant across stimulus frequencies. In summary, for the nonspeech pitch contours, there was no group difference for the rising stimuli, but the English listeners were more accurate at identifying certain falling and flat contours than the Mandarin listeners.

Nonspeech Pitch Contour: Signal Detection Theory Analysis

To determine whether the differences between the two groups of listeners for the falling stimuli were due to differences in sensitivity (a sensory factor) or differences in response bias (a decision factor), we carried out a signal detection theory analysis on a subset of the nonspeech pitch contour stimuli. Our choice was to analyze the rising and falling stimuli with reference to the two flat stimuli that were at the frequencies of the starting points for the rising and falling stimuli—that is, at 250 Hz and 300 Hz. The total stimulus set only included two flat stimuli that could be matched in a parallel stimulus–response matrix with the falling or rising stimuli; therefore, only two of the flat stimuli were entered into the analysis, and the other 12 flat stimuli are not analyzed here. In this matrix, the falling responses to the falling stimuli were considered hits, and the falling responses to the flat stimuli were considered false alarms. Similarly, the rising responses to the rising stimuli were considered hits, and the rising responses to the flat stimuli were considered false alarms.

The results from this analysis indicate that the observed differences in response accuracy for the falling stimuli were due to differences in response bias rather than sensitivity. All sensitivity and response bias scores are shown in Table 2. Separate analyses were conducted for the stimuli starting at 250 Hz and 300 Hz. Across participant groups, the cumulative sensitivity (cumulative d' ; Macmillan & Creelman, 1991, p. 211) between the rising and falling stimuli that changed by the same amount were compared. The cumulative d' value was obtained by summing the d' value for the sensitivity between the rising and flat stimuli and the d' value for the sensitivity between the flat and falling stimuli. For example, the cumulative d' for the stimuli changing by 5 Hz with a starting frequency of 250 Hz was obtained by summing the d' for the flat 250-Hz pitch contour and the contour that fell by 5 Hz with the d' for the flat 250-Hz pitch contour and the contour that rose by 5 Hz. Although d' increased with increasing frequency change for both groups, no differences in cumulative d' were observed between the groups for any amount of change at either of the starting frequencies, which indicates that the language experience of the listener groups did not differentially influence their sensitivity to these stimuli in an identification task.

Two separate three-factor analyses of variance (ANOVAs) were performed for the rising and falling response bias scores (β), with listener group (English or Mandarin) as the between-groups factor and starting frequency (250 Hz or 300 Hz) and amount of change (5, 10, 15, 20, 30, 40, or 50 Hz) as the within-group factors. For the ANOVA on the rising stimuli, there were no significant main effects or interactions except for a significant effect for amount of change, $F(6, 138) = 136.45$, $p < .01$, which was due to the response bias scores decreasing as the amount of change increased. For the ANOVA on the falling stimuli, there was no three-way interaction, but all main effects and two-way interactions were significant except for the Starting Frequency \times Listener Group interaction: listener group, $F(1, 138) = 11.41$, $p = .01$; starting frequency, $F(1, 138) = 11.97$, $p < .01$; amount of change, $F(6, 138) = 180.09$, $p < .01$; Amount of Change \times Listener Group, $F(6, 138) = 3.10$, $p < .01$; Starting Frequency \times Amount of Change, $F(6, 138) = 13.40$, $p < .01$. Separate two-factor ANOVAs were then performed for the two starting frequencies.

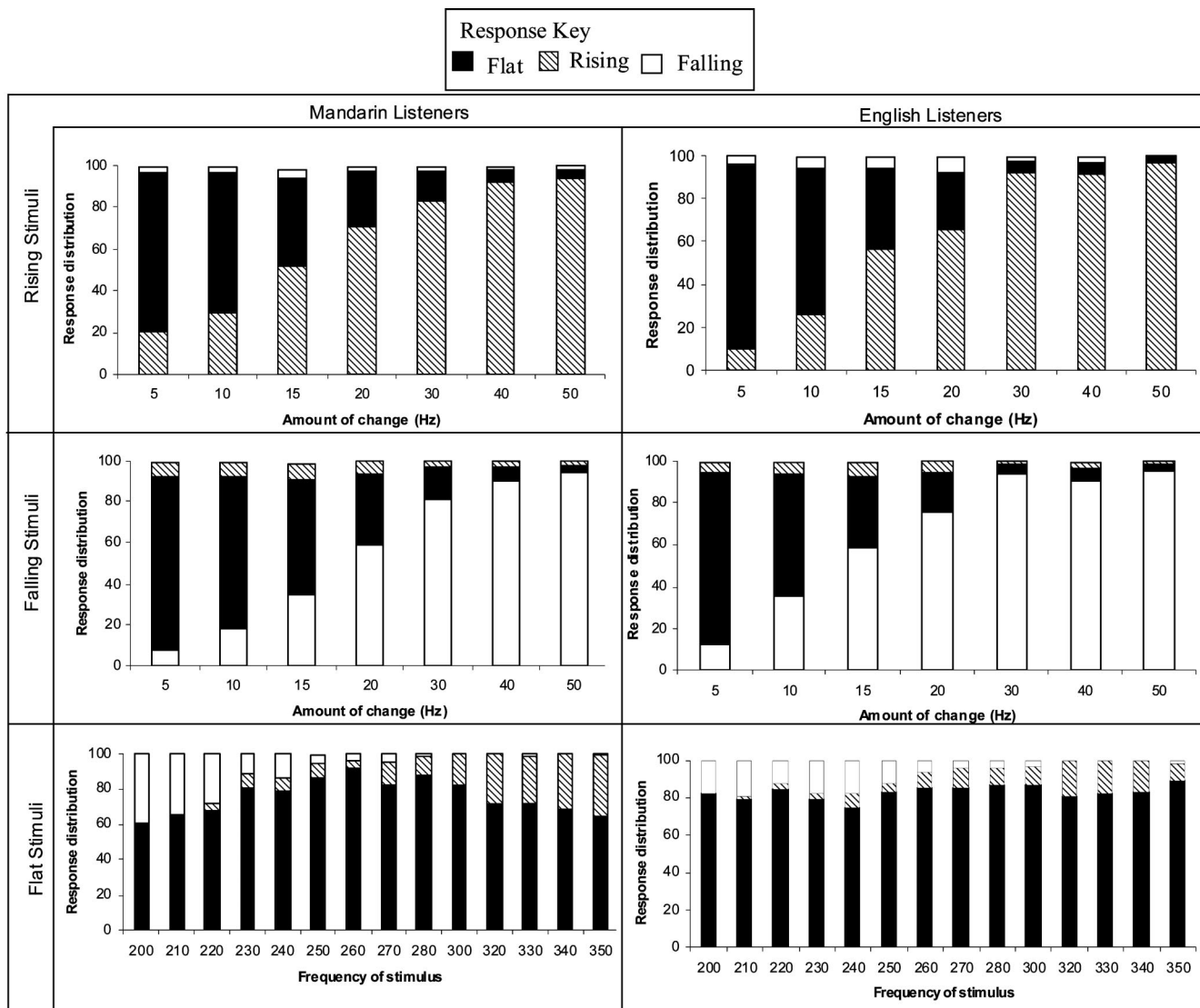


Figure 3. Mandarin listeners identified flat and falling pitch contours less accurately than English listeners but identified rising contours with the same accuracy. The responses for the Mandarin listeners (left side) and the English listeners (right side) are shown for the rising contours (top row), the falling contours (middle row), and the flat pitch contours (bottom row). Columns show the distribution of responses across the three response types. For the graphs with the flat stimuli, the *x*-axis represents the frequency of each stimulus. For the graphs with the rising and falling stimuli, the *x*-axis represents the amount in hertz by which the stimulus rose or fell. The values for the rising and falling stimuli represent the average across the two starting frequencies.

For the ANOVA with the falling stimuli starting at 300 Hz, there was a significant effect of listener group, $F(1, 138) = 10.60, p < .01$, with the Mandarin listeners overall having higher response bias scores than the English listeners. There was also a significant effect of amount of change, $F(6, 138) = 153.07, p < .01$, with response bias scores decreasing as amount of change increased. There was no Listener Group \times Amount of Change interaction. For all amounts of change, the Mandarin listeners had higher response bias scores than the English listeners. For the ANOVA with the falling stimuli starting at 250 Hz, there were significant main effects of listener group, $F(1, 138) = 6.71, p = .02$, and

amount of change, $F(6, 138) = 103.15, p < .01$, and an interaction between the two, $F(6, 138) = 2.83, p = .01$. English listeners had significantly lower response bias scores than Mandarin listeners for the pitch fall of 15 Hz, $t(23) = 2.93, p < .01$. Thus, in general, the observed listener group differences were due to different response biases rather than sensitivity differences, which suggests that linguistic experience shapes cognitive categorization responses rather than changing lower level perception. In other words, language categories, in this case Mandarin tones, can apparently influence the criteria people use when classifying non-speech sounds.

Table 2
Sensitivity (d') and Response Bias (β) Scores for Rising and Falling Pitch Contours

Starting frequency	Language	Hz						
		5	10	15	20	30	40	50
		Amount of fall						
250	English							
	d'	0.16	1.16	2.02	2.32	2.59	2.55	2.72
	β	1.12	0.62	0.19	0.04	-0.09	-0.07	-0.16
	Mandarin							
	d'	0.35	0.89	1.47	2.26	2.54	2.89	3.03
	β	1.30	1.03	0.74	0.34	0.20	0.03	-0.04
300	English							
	d'	0.16	0.66	1.26	1.98	2.97	2.80	2.94
	β	1.44	1.19	0.09	0.53	0.03	0.12	0.05
	Mandarin							
	d'	0.03	0.20	0.72	1.39	2.56	2.80	2.98
	β	1.63	1.55	1.29	0.95	0.36	0.25	0.16
		Amount of rise						
250	English							
	d'	0.06	0.70	1.67	1.94	2.82	2.74	3.00
	β	1.46	1.14	0.65	0.52	0.08	0.12	-0.02
	Mandarin							
	d'	0.25	0.44	1.27	1.92	2.21	2.58	2.63
	β	1.20	1.10	0.69	0.36	0.22	0.03	0.01
300	English							
	d'	0.12	0.61	1.50	1.86	2.70	2.74	2.85
	β	1.25	1.00	0.56	0.38	-0.05	-0.07	-0.12
	Mandarin							
	d'	0.21	0.63	1.18	1.73	2.28	2.53	2.56
	β	0.90	0.69	0.42	0.15	-0.13	-0.25	-0.27

Note. Cumulative d' is the sum of the rising and falling d' values for a given starting frequency. There were no observed differences between the Mandarin and English listeners on the measures of sensitivity or on the response bias scores for the rising pitch contours. However, Mandarin listeners had overall higher response bias scores than English listeners for the falling pitch contours starting at 300 Hz. Mandarin listeners also exhibited higher response bias scores on the falling pitch contours starting at 250 Hz for pitch falls between 10 and 30 Hz.

Summary of Results

Mandarin and English listeners participated in three tests: a Mandarin tone identification test, a nonspeech pitch discrimination test, and a nonspeech pitch contour identification test.

1. Mandarin listeners were more accurate at identifying Mandarin tones than English listeners.

2. No significant listener group differences were observed on the nonspeech pitch discrimination test in which the stimuli and task were highly distinct from speech and speech perception, respectively.

3. Mandarin listeners were less accurate than English listeners in the nonspeech pitch contour test in which the stimuli were modeled after speech stimuli (syllables) in their duration and frequency and in which the task more closely resembled speech perception by involving categorization rather than discrimination. Where determinable, these listener group differences were attributable to systematic differences in response bias across the Mandarin and English listener groups rather than to differences in sensitivity.

Discussion

A crucial issue raised by the results of this study is how exactly long-term experience with Mandarin versus English might have resulted in the observed response patterns for the nonspeech pitch

contour identification test. The listener group differences observed in this study are consistent with established facts about Mandarin tone production, which partly define the speech perception experience of the Mandarin listeners. The Mandarin listeners' differing responses to the rising and falling pitch contours might have been due to the typically larger pitch range for falling tones than for rising tones in Mandarin (e.g., Xu, 1994). On the basis of their experience with falling tones in which the pitch changes are relatively large and rising tones in which the pitch changes are relatively small, Mandarin listeners may set a different criterion for labeling stimuli as falling than for labeling stimuli as rising. They may be reluctant to categorize marginally falling pitch contours as falling but willing to categorize marginally rising pitch contours as rising. Although the pitch rises and falls are more extensive in naturally produced Mandarin than the rises and falls used in the present pitch contour identification test, the large systematic differences between Mandarin rising and falling tones might have influenced the criteria used for making responses to stimuli presented at near threshold levels. In contrast to the Mandarin listeners, the English listeners responded similarly (i.e., set generally consistent labeling criteria) for the rising and falling stimuli in this nonspeech pitch contour identification task. The Mandarin listeners' tendency to misidentify low-flat contours as falling and high-flat contours as rising may be explained by a listening strategy that

relies on attending to the end-point of the stimulus induced by patterns of Mandarin tone production in which underlying pitch targets are best approximated at the end of the syllable (Xu, 2001; Xu & Wang, 2001). This language-specific perceptual strategy might enhance the psychoacoustic (i.e., language-independent) tendency to label relatively high-frequency flat pitch contours as “rising” and relatively low-frequency flat pitch contours as “falling” (Schouten & Pols, 1989) and to match the pitch of a rising or falling pitch contour with a level pitch contour that corresponds to the ending frequency of the contour (Nabelek, Nabelek, & Hirsh, 1970).

From the present results, it appears that the influence of linguistic experience extends beyond speech processing and that the strength of this influence on nonspeech processing varies depending on stimulus and/or task factors. Thus, at least to some extent, the processing of speech and nonspeech signals relies on shared mechanisms that are subject to the influence of long-term experience in a particular language environment. It remains to be determined whether the effect of language background observed in the present study—and, hence, the overlap of speech and nonspeech processing—is due to task and/or stimulus similarities.

References

- Burnham, D., Francis, E., Webster, D., Luksaneeyanawin, S., Attapaiboon, C., Lacerda, F., & Keller, P. (1996). Perception of lexical tone across languages: Evidence for a linguistic mode of processing. In H. T. Bunnell & W. Idsardi (Eds.), *Proceedings of the Fourth International Conference on Spoken Language Processing* (Vol. 1, pp. 2514–2517). Wilmington, DE: Applied Science and Engineering Laboratories.
- Burns, E., & Sampat, K. (1980). A note on possible culture-bound effects in frequency discrimination. *Journal of the Acoustical Society of America*, *68*, 1886–1888.
- Cedrus Corp. (1996). Superlab Pro 2.01 [computer software]. San Pedro, CA: Author.
- Chao, Y. R. (1968). *A grammar of spoken Chinese*. Berkeley: University of California Press.
- Francis, A. L., & Ciocca, V. (2003). Stimulus presentation order and the perception of lexical tones in Cantonese. *Journal of the Acoustical Society of America*, *114*, 1611–1621.
- Gottfried, T. L., & Reister, A. M. (2000). The relation of pitch glide perception and Mandarin tone identification. *Journal of the Acoustical Society of America*, *108*, 2604.
- Harris, J. D. (1952). Pitch discrimination. *Journal of the Acoustical Society of America*, *24*, 750–755.
- Kewley-Port, D., Watson, C. S., & Foyle, D. C. (1988). Auditory temporal acuity in relation to category boundaries; speech and nonspeech stimuli. *Journal of the Acoustical Society of America*, *83*, 1133–1145.
- Kluender, K. R., Diehl, R. L., & Killeen, P. R. (1987, September 4). Japanese quail can learn phonetic categories. *Science*, *237*, 1195–1197.
- Kuhl, P. K. (1981). Discrimination of speech by nonhuman animals: Basic auditory sensitivities conducive to the perception of speech-sound categories. *Journal of the Acoustical Society of America*, *70*, 340–349.
- Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences of the USA*, *97*, 11850–11857.
- Lieberman, A. M., & Mattingly, I. G. (1989, January 27). A specialization for speech perception. *Science*, *243*, 489–494.
- Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. New York: Cambridge University Press.
- Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J., & Fujimura, A. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics*, *18*, 331–340.
- Moore, B. (1973). Frequency-difference limens for short duration tones. *Journal of the Acoustical Society of America*, *54*, 610–619.
- Nabelek, I. V., Nabelek, A. K., & Hirsh, I. J. (1970). Pitch of tone bursts of changing frequency. *Journal of the Acoustical Society of America*, *48*, 536–553.
- Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, *24*, 175–184.
- Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation*. Unpublished doctoral dissertation, Massachusetts Institute of Technology.
- Schouten, M. E. H., & Pols, L. C. W. (1989). Identification and discrimination of sweep formants. *Perception & Psychophysics*, *46*, 235–244.
- Semal, C., Demany, L., Ueda, K., & Halle, P.-A. (1996). Speech versus nonspeech in pitch memory. *Journal of the Acoustical Society of America*, *100*, 1132–1140.
- Stephens, J. D. W., & Holt, L. L. (2003). Preceding phonetic context affects perception of nonspeech. *Journal of the Acoustical Society of America*, *114*, 3036–3039.
- Tremblay, S., Nicholls, A. P., Alford, D., & Jones, D. M. (2000). The irrelevant sound effect: Does speech play a special role? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1750–1754.
- Werker, J., & Tees, R. (1984). Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America*, *75*, 1866–1878.
- Wier, C., Jesteadt, W., & Green, D. (1977). Frequency discrimination as a function of frequency and sensation level. *Journal of the Acoustical Society of America*, *61*, 178–184.
- Xu, Y. (1994). Production and perception of coarticulated tones. *Journal of the Acoustical Society of America*, *95*, 2240–2253.
- Xu, Y. (2001). Sources of tonal variations in connected speech. *Journal of Chinese Linguistics*, Monograph Series No. 17, pp. 1–31.
- Xu, Y., & Wang, Q. E. (2001). Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication*, *33*, 319–337.
- Zeng, F.-G. (2002). Temporal pitch in electric hearing. *Hearing Research*, *174*, 101–106.