Original Paper

Language Specificity in Speech Perception: Perception of Mandarin Tones by Native and Nonnative Listeners

Tsan Huang[a]   Keith Johnson[b]

[a]University at Buffalo, SUNY Buffalo, Bufallo, N.Y., and [b]Department of Linguistics, University of California, Berkeley, Calif., USA

Abstract

The results reported in this paper indicate that native speakers of Mandarin Chinese rate the perceptual similarities among the lexical tones of Mandarin differently than do native speakers of American English. Mandarin listeners were sensitive to tone contour while English listeners attended to pitch levels. Chinese listeners also rated tones that are neutralized by phonological tone sandhi rules in Mandarin as more similar to each other than did English speakers – indicating a role of phonology in determining perceptual salience. In two further experiments, we found that some of these differences were eliminated when the listening task focused listeners' attention on the auditory properties of the stimuli, but, interestingly, a degree of language specificity remained even in the most purely psychophysical listening tasks with speech stimuli.

Copyright © 2011 S. Karger AG, Basel

1. Introduction

Research done in the past two decades or so [e.g. Kohler, 1990; Hura et al., 1992; Steriade, 2001] has found that phonological processes including segmental reduction, deletion, and assimilation can be viewed as perceptually tolerated articulatory simplification, where the direction of such processes is determined by perception[1]. If a contrast is perceptually weak in a certain position, there is a tendency toward either perceptual enhancement of the contrast by dissimilation, epenthesis, or metathesis, or toward loss of the contrast by assimilation or deletion [Hume and Johnson, 2003]. For example, vowel epenthesis between sibilants in some English plural noun forms

---

[1] The influence of perception on phonology was discussed by various midcentury researchers [e.g., Trubetzkoy, 1969/1939; Jakobson et al., 1952; Jakobson and Halle, 1956]. Phonological theory since Chomsky and Halle [1968] has been based on articulatorily defined distinctive features, but recent years have seen a revival of interest in the relationship between perception and phonology [e.g. the chapters in Hume and Johnson, 2001; see also recent work by Hansson, 2001; Frisch et al., 2004, and Rose and Walker, 2004, on the role of phonetic/phonological similarity in phonology].

Keith Johnson
Department of Linguistics, UC Berkeley
1203 Dwinelle Hall
Berkeley, CA 94720-2650 (USA)
Tel. +1 510 643 7617,
E-Mail keithjohnson@berkeley.edu

[e.g., *buses*, *bushes*, *judges*; cf. *cats*, *cans*), metathesis of /skt/ to [kst] in Faroese and Lithuanian [Hume and Seo, 2004], and manner dissimilation of two consecutive obstruents in Greek [e.g., /kt/ → [xt] or /xθ/ → [xt]; Tserdanelis, 2001] have all been analysed as serving to strengthen the syntagmatic contrast between neighboring segments, at the cost (in some cases) of paradigmatic contrast neutralization. On the other hand, n-lateralization (/nl/ → [ll] and /ln/ → [ll]) in Korean [Seo, 2001] and optional /h/ deletion in Turkish [Mielke, 2003] sacrifice the perceptually weak contrasts, leading to syntagmatic contrast neutralization in both cases.

These phonological patterns emerge (at least partly) from the action of perceptual processes in sound change [Ohala, 1981, 1993; Blevins, 2006], due to the listener's misperception and reinterpretation of sounds or sound sequences. Instead of correcting distorted phonetic forms based on knowledge of possible variants of the common underlying forms, the listener-turned-speaker may exaggerate the distortion, resulting in historical sound change [Ohala, 1981, p. 183]. Janson [1983, p. 24] further hypothesizes a five-stage sound change process involving interaction between perception and production, using as an example the change of /r/ to /ʀ/ in Norwegian. Guion [1998] proposed that the cross-linguistically common sound change of velar palatalization is perceptually conditioned. Examples of perceptual effects on sound change can also be found in historical tonogenesis and tone developments in the tone languages, where previously redundant pitch differences became contrastive when the conditioning segmental contrast was lost [e.g. Maspero, 1912; Haudricourt, 1954a, b; Maran, 1973; Matisoff, 1973; Hombert, 1978; Hombert et al., 1979; Svantesson, 2001].

Thus, the influence of perception on synchronic phonology and sound change is attested cross-linguistically and several of the patterns attributed to perceptual mechanisms appear to be universal. This leads to the hypothesis of a universal perceptual basis for sound change, and thus many of the synchronic sound patterns found in phonology. This hypothesis was made explicit by Steriade [2001], who proposed that constraints in Optimality Theory phonology [McCarthy and Prince, 1993] can refer to a universal scale of perceptual salience [which Steriade, 2001, called the p-map].

However, the universal p-map hypothesis runs counter to evidence suggesting that listeners' perception of speech sounds depends on their linguistic experience, so that the listener's native phonology has an impact on speech perception as well. The relationship between phonology and perception is therefore a bidirectional interplay [Hume and Johnson, 2001], rather than one in which perception shapes phonology in a unidirectional fashion, as the universal perceptual map hypothesis implies. In fact, the inventories of contrastive sounds, the phonotactics of sound combination, and the phonological rules operating in the listeners' native languages may all have an impact on speech perception. There are many examples of language-specific speech perception patterns in the research literature. For instance, Japanese listeners, whose language has only one liquid sound, perceive the English /ɹ-l/ distinction differently from American English (AE) speakers [Goto, 1971; Miyawaki et al., 1975; MacKain et al., 1981; Strange and Ditman, 1984; Logan et al., 1991; Yamada et al., 1992; Lively et al., 1993, 1994; Flege et al., 1996. Similarly, Werker and Tees [1984] found that Hindi speakers are more sensitive than English speakers to a contrast between voiceless unaspirated retroflex stops and dental stops ([ʈa] vs. [t̪a]), and Nthlakampx[2] speakers are more

---

[2] Also referred to as Thompson, Nthlakampx is one of the Inner Salish languages spoken by the Canadian First Nations people of British Columbia.

sensitive than English speakers to a contrast between velar and uvular ejectives ([k'i] vs. [q'i]). Another example of how phonemic inventory can influence perception comes from McGuire's [2007] study. He found that the Polish fricatives [ʂ] and [ɕ] both sound like [ʃ] to speakers of American English and that Americans had reduced perceptual sensitivity to the contrast even in very low-level discrimination tasks. McGuire [2007] also found that Polish and American listeners seemed to be giving different weights to the fricative place cues (in the noise itself and in the vowel formant transitions adjacent to the fricatives). In the same vein, Hume et al. [1999] obtained results suggesting that perceptual cue utilization is language-specific. They found that for both AE and Korean listeners excised consonant-vowel transitions provided more place information for consonant place identification than did excised stop bursts. However, the difference between the two kinds of stimuli was greater for Korean listeners. Hume et al. [1999] suggested that this is related to differences in phonological contrasts in these two languages, so that Korean listeners with a three-way (tense, lax and aspirated) stop consonant contrast, which is cued in part by the duration of aspiration, allocate more attention to the CV transition between the burst and the vowel onset than do AE listeners, who have only a two-way (unaspirated versus aspirated) stop contrast.

So, differences in languages' inventory of contrastive sounds are related to language-specific patterns of speech perception. Beyond this it has recently been demonstrated that patterns of allophony may also produce language-specific speech perception sensitivities [Gandour, 1983; Dupoux et al., 1997; Harnsberger, 2001; Hume and Johnson, 2003; Boomershine et al., 2008; Babel and Johnson, 2010]. Commenting on studies of perceptual consequences of allophony, Hume and Johnson [2003] suggested that allophonic neutralizing rules tend to reduce the perceptual distinction between two categories that are otherwise contrastive in a language. Experimental data provide supporting evidence for such a statement. Fox [1992] found that English listeners fared poorly in identifying or discriminating vowels in the context of /hVr(d)/. Fox [1992] suggests that knowledge of the phonological rule that neutralizes vowel contrast in this context may have affected the ability of listeners to make perceptual decisions about vowel quality.

In this paper, we present the results of a study on the language-specificity of lexical tone perception. Past studies have also suggested that lexical tone perception may be language-specific. Gandour [1983, 1984] and Lee et al. [1996] showed that for speakers of tone languages, differences in lexical tone inventory may play a role in tonal perception. In Gandour's [1983, 1984] study using 19 synthesized fundamental frequency (f0) stimuli (five level, four rising, four falling, three falling-rising, and three rising-falling contours), about 40 speakers from each of five languages – Taiwan Mandarin (four tones), Hong Kong Cantonese (six tones), Taiwan Southern Min (six tones), Thai (five tones), and English – made dissimilarity judgments on tonal pairs. Results show that the tones were rated significantly differently by tone versus non-tone language speakers, by Thai versus Chinese (Mandarin and Southern Min) speakers, and by Cantonese versus Mandarin and Southern Min speakers. Gandour [1983, 1984] attributed the perceptual differences, in part, to differences in tonal inventories. In particular, tone height was more important for English listeners, while Thai listeners attached most importance to the direction of f0 contour (i.e. rising versus falling). When the tone language groups were compared, Cantonese listeners utilized mainly the dimension of tone levels/heights, which is not surprising, given that four of the six Cantonese tones have basically level contours. Lee et al. [1996] used naturally recorded stimuli of Cantonese and Mandarin tones on word and nonword syllables and

had Cantonese, Mandarin (Taiwan and Mainland) and English (US) listeners partici-pate in two 'same'/'different' discrimination experiments. They found that Cantonese and Mandarin listeners were better at discriminating tones in their respective native dialect and that the tone language-speaking listeners were better able to discriminate the tones than the English group.

Just as allophony influences the perception of consonants and vowels, there is some evidence in the literature that tonal allophony influences tone perception. Gandour [1981, 1983, 1984] suggests that tone sandhi rules may also influence tonal perception. Using Individual Difference Multidimensional Scaling [Carroll and Chang, 1970], Gandour [1981] analyzed confusion data from native listener identification of naturally produced Cantonese tones. He found that the high falling tone was placed midway between the level and the contour tones in the perceptual tone space. He argued that this was due to the fact that this tone has a high level allotone in Cantonese. Although the allotone was not present in the stimuli, allophony still interfered with listeners' per-ception. The effect of the same allophonic alternation showed up in Gandour's [1983, 1984] experimental data, where Cantonese listeners perceived a /44/ (high level) con-tour to be similar to a /53/ (high falling) contour. In the same data, Mandarin listen-ers perceived the /44/ contour to be similar to /35/ (rising), which, as Gandour [1983, 1984] argued, was due to the existence of the allophonic rule that turns a rising tone to a high level in Mandarin [e.g. Chao, 1968; Cheng, 1973; see also §2].

The present study examined the impact of phonological inventory and allophonic rules on speech perception comparing tone perception performance of listeners who were native speakers of Beijing Mandarin and American English. We investigated whether differences in tonal inventory and neutralization processes may affect tone perception.

Our study extended previous research by investigating the time course of language specificity in speech perception. Babel and Johnson [2010] suggested that responses formed early in processing are less influenced by the linguistic system than are responses formed later. They suggested that the universal p-map hypothesis is supported by the finding that language experience seems to have a limited effect early in speech pro-cessing. We tested this hypothesis for tone perception. In one experimental paradigm we permitted listeners time to reflect on their answers in a degree-of-difference rating task, while in another paradigm listeners were required to respond very quickly and our measure of phonetic similarity was response time. We also tested perception of both natural speech stimuli and nonspeech sine wave analogs of speech.

Our results reveal language specificity in tone perception late in processing (in a similarity rating task), as has been found many times before. We also found a cer-tain amount of language specificity early in processing (in reaction time tasks). The language-specific perceptual patterns found in the present study are a reflection of pho-nological category representation. This suggests that the p-map hypothesis should be tempered because perceptual salience is a function not only of raw acoustic/auditory contrast but also of the listener's phonological systems.

2. Tones and Tone Sandhis in Mandarin Chinese

Standard (Beijing) Mandarin has four lexical tones. Chao [1968, p. 26] describes them as high level /55/, mid-rising /35/, low falling-rising /214/, and high falling /51/. (The numbers indicate the idealized pitch values of the tones on a five-level scale

[Chao, 1930].)[3] For notational purposes in this paper, we shall refer to them as T55, T35, T214 and T51, respectively[4].

In tone sandhi processes, underlying full tones may be modified under the influence of their tonal environment in Mandarin [see, e.g. Chao, 1968; Kratochvil, 1968; Cheng, 1973; Chen, 2000; Duanmu, 2007]. As described by Chao [1968, p. 27], T214 of Mandarin becomes T35 when immediately followed by another T214 (because Mandarin T214 is traditionally numbered as tone 3 in Chinese linguistics, this rule is known as the Tone 3 Sandhi Rule):

(1) Tone 3 Sandhi Rule: /T214/ → [T35]/_T214

Since an underlying /T35.T214/ sequence is also realized as [T35.T214], the paradigmatic contrast between T35 and T214 is lost before a following T214, creating many homophonous surface pairs. Thus, /hao^{214}.mi^{214}/ 'good rice' is indistinguishable from /hao^{35}.mi^{214}/ 'millimeter', because both surface as [hao^{35}.mi^{214}]. It is believed that this process was already in place in the 16th century [Mei, 1977].

Chao [1968] also discusses a second tone sandhi rule, where T35 becomes T55 when following a T55 or T35 and preceding a full-toned syllable (because T35 is numbered tone 2 in Chinese linguistics, this rule is known as the Tone 2 Sandhi Rule). Chao [1968, pp. 27, 28] considers this rule to be 'of minor importance'. Unlike the T214 sandhi, the T35 rule is not taught to second-language learners.

(2) Tone 2 Sandhi Rule: /T35/ → [T55] / {T55,T35} _ Tx, where Tx = any of the four tones

According to Chao [1968], the middle position of a trisyllabic phrase is relatively weak prosodically. As a result, the low pitch onset of the sandhi-affected T35 gets deleted and the pitch contour simplified to [T55]. Note that the affected T35 does not have to be an underlying /T35/. For instance, a sequence of /T214.T214.T214/ may first undergo the T214 sandhi (applied twice iteratively from left to right), resulting in a hypothetical sequence of (T35.T35.T214), which is then affected by the T35 sandhi, yielding the actually observed sequence [T35.T55.T214]. Two familiar examples given by Chao [1968, p. 28] are: /cong^{55}.you^{35}.bing^{214}/ → [cong^{55}.you^{55}.bing^{214}] '(Chinese) onion pancakes', and /hao^{214}.ji^{214}.zhong^{214}/ → [hao^{35}.ji^{55}.zhong^{214}] 'quite a few kinds'. This sandhi also leads to paradigmatic neutralization: the contrast between T55 and T35 is lost here.

The reason why the Tone 2 Sandhi Rule is treated differently by Chao [1968] is that it is a lower-level phonetic rule [see also Xu, 2001; Kochanski et al., 2003]: in the prosodically weak position, the f0 contour of the medial T35 coarticulates with the preceding (strong) T55 or T35 by losing its low f0 onset. Similarly in the model of Kochanski et al. [2003, p. 627], 'all tonal variations of a lexical tone are generated from the lexically determined tonal templates' and the f0 at each time point calculated 'as a function of the nearby templates and their prosodic strengths'. Nevertheless, the T35

---

[3] Based on our small recorded database of 10 Beijing speakers, the tones are more likely /44, 24, 212, 51/ acoustically. The contours are nevertheless similar to Chao's [1930] description. Note that when the contour is fully realized, T214 may or may not have creaky voice in the middle. In Howie's [1976] study, T55 has a mid-vowel average around 150 Hz; T35 rises from about 115 Hz (with the reflection point near the 40% point of the total duration) to 160 Hz; T214 starts at around 110 Hz, dips down to about 90 Hz (around mid vowel) and rises again to 115 Hz, and T51 starts at around 170 Hz and falls steadily to about 110 Hz.

[4] The Mandarin tones are most commonly numbered from 1 through 4. We felt that it would be helpful to readers of this paper to refer to the tones using symbols from the International Phonetic Alphabet (IPA), rather than the traditional Chinese linguistics tone numbers. The mapping between tone number and IPA transcription is: tone 1 = T55, tone 2 = T35, tone 3 = T214, and tone 4 = T51.

---

sandhi has much in common with the T214 sandhi: both are conditioned by the phonetic environment; both result in simplification of the original tonal contour, and both result in category-changing alternations and lead to paradigmatic neutralization. As a recent study by Kochetov and Pouplier [2008] finds, even processes such as Korean stop place assimilation, traditionally viewed as obligatory with category-changing outputs, show gradient effects. This calls into question the view of a rigid, clear-cut division between phonological processes and low-level phonetic realization rules. At any rate, since in the surface forms the T55 vs. T35 contrast is neutralized, a phonetic [T55] will have at least two possible interpretations for the underlying form (i.e. /T55/ and /T35/), which is predicted to cause perceptual confusion of the relevant tones for native speakers.[5] Our studies will test whether the T35 sandhi has as strong an effect as that of the T214 sandhi.

Past studies [e.g. Howie, 1976; Shen and Lin, 1991] have found that T35 may be confusable with T214 for native Chinese listeners (as might be expected if there were a perceptual effect of the Tone 3/T214 Sandhi Rule, which neutralizes the difference between T35 and T214). However, these studies did not test directly for the T214 sandhi effect and their results are inconclusive. In Howie's [1976, p. 222] perceptual experiment using natural speech stimuli, the syllable /bao35/ was misidentified as /bao214/ due to 'an unusually low pitch range' at around 85 Hz (the average being 115 Hz) for the first 50% of the vowel duration. In the identification test using synthetic speech, 43.3% (or 26 out of 60) of Howie's [1976] /bao214/ tokens were misheard as /bao35/. But that was because the carrier phrase had the wrong neutral-toned syllable (originally occurring after /bao55/) following the test syllable of /bao214/. Shen and Lin [1991] tried to identify the perceptual cues that distinguish T35 and T214 with two tonal continua resynthesized from a female voice, a two-way forced choice labeling task, and native Mandarin listeners. They found that '[t]he phonetic contrast between Tones and is cued by the degree of the initial fall and by the timing of the turning point' [Shen and Lin, 1991, p. 149], and that a steeper initial fall in pitch (from 190 down to 160 Hz vs. from 190 to 175 Hz) and a later reflection point attracted more T214 responses.

3. Experiment 1: Rated Perceptual Differences among Mandarin Tones

The three experiments in the present study were designed to test the impact of native language experience on tone perception – exploring both the impact of tonal inventory and phonological sandhi processes. The experiments contrast performance by native speakers of Mandarin Chinese and AE speakers. Experiment 1 was designed to explore language specificity in tone perception, and in particular, the role of tone categories and tone sandhi rules in perception. Experiments 2 and 3 explored these effects further by using tasks and stimuli that tap different levels of perceptual processing.

Previous studies have shown that it is feasible to test tone perception with nontone language speakers. For example, Kiriloff [1969] found that, when asked to ignore the segmental element of the syllable and focus on the tones, nonnative speakers' performance was quite good with an average of 87.5% correct identification. Although

---

[5] It is hard to know how this surface [T55] may be linked to the true underlying form of /T214/ (as in phrases like /hao214.ji214.zhong214/ 'quite a few kinds'), which is removed one more step away from the surface tone.

English is a nontonal language, its stress-based prosodic system does utilize pitch as one way to distinguish stress accents, which may be realized as high, low, rising or falling contours and can thus be very similar to the Mandarin tones [see e.g. Beckman, 1984]. Although AE speakers may not have specific labels for the stimulus tones, a paired comparison task does not require that listeners have names for the items being compared [cf. Lee et al., 1996, where a small native speaker advantage was found].

On the other hand, without lexical tone categories in their lexicon, AE listeners may actually enjoy a perceptual advantage; that is, they may be able to detect subtle pitch differences, which may be missed by Chinese listeners' categorical perception of tone. Wang [1976] found that Mandarin-speaking listeners perceived synthesized stimuli along a level to rising contour continuum categorically, dismissing rises smaller than 9 Hz as negligible within-category variations. Similarly, Stagray and Downs [1993] reported that Mandarin-speaking listeners had significantly larger difference limens for frequency than English-speaking listeners around 125 Hz, which approximates pitch utilized in a male voice. Their Mandarin listeners also had poorly shaped psychometric functions close to chance response level, as opposed to regular ogive-shaped functions in AE listeners' data. Stagray and Downs [1993, p. 156] concluded that Mandarin listeners had poorer differential sensitivity for frequency because some variations in the stimulus tones were perceived 'as being within the same pitch range of a learned, level tone-phoneme category'. However, these indications of relative insensitivity to fine-grained pitch patterns contrast with the finding of Krishnan et al. [2005] that Mandarin tones give rise to a more accurate brain-stem frequency-following response, which suggests that early auditory processing of tone is facilitated by linguistic experience with tones.

Experiment 1 investigated the confusability of Mandarin Chinese tones with a subjective degree-of-difference rating task, which we assume provides listeners with time to engage in linguistic processing.

3.1. Methods

Participants

Twenty-one (13 female, 8 male, average age 27) Chinese and 30 (15 female, 15 male, average age 20) AE listeners participated in experiment 1. All Chinese listeners were from Beijing. Only 3 of these 21 Chinese listeners in experiment 1 also participated in experiments 2 and 3. The AE listeners spoke a Midland variety of American English. None of the listeners reported any history of speech or hearing problems. The Chinese listeners were paid a small amount of money for their participation, whereas AE listeners earned course credits.

Stimuli

The stimuli were recordings of the four monosyllabic words /ba55, ba35, ba214, ba51/ (fig. 1) produced by a male speaker of Beijing Mandarin in his early thirties. He produced several instances of each of the four monosyllabic words in isolated word reading, and we selected one typical instance of each word for use as the exemplar of that tone category in this experiment.

Procedure

All participants were tested in front of a computer in a quiet room, using E-Prime (Psychology Software Tools, Inc., Pittsburgh, Pa., USA). The stimuli were played through headphones. In each trial the listener was presented with a pair of syllables and asked to listen carefully for tonal differences and rate the degree of difference subjectively by pressing the five keys on the response box labeled
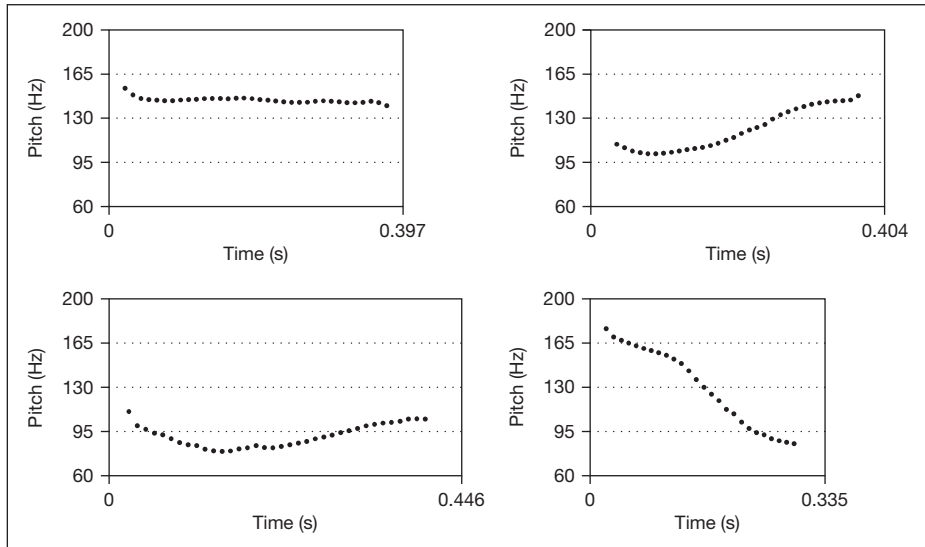
Fig. 1. f0 traces of tones T55 (upper left panel), T35 (upper right), T214 (lower left), and T51 (lower right) as produced by a male Beijing speaker. The segmental makeup is /ba/. Lengths of the x axes correspond to the relative lengths of the tones. These speech tokens were used in experiments 1 and 3. They also served as templates for the synthetic sine wave tones used in experiment 2. (Note that being produced in a prepausal position, the T214 tonal contour is fully realized, although the final rise is still not to the level of '4' as the traditional analysis indicates, rendering T214 a rather low tone.)

Table 1. Rating scale described for the listeners on the instruction sheet

Very similar	Moderately similar	Somewhat different	Moderately different	Very different
1	2	3	4	5

'1' through '5'. The scale was described for the listeners in the format shown in table 1. They were especially encouraged to use the full scale.

There are 16 possible combinations of the 4 stimulus syllables (ba⁵⁵ vs. ba⁵⁵, ba⁵⁵ vs. ba³⁵, ba⁵⁵ vs. ba²¹⁴, ba⁵⁵ vs. ba⁵¹, and so on). Each of these 16 stimulus pairs (12 nonidentical pairs, and 4 that involve presentation of the same sound file twice) was presented 12 times to listeners in 6 blocks of 32 trials. The order of the trials was randomized separately for each listener.

3.2. Results

Data from 17 AE and 17 Chinese listeners' data were analyzed. Data from 4 of the Mandarin listeners were discarded because they did not used the response scale appropriately – responding only '1' or '5' and not giving any intermediate rating values. To achieve an equal number of participants in each language group we also randomly
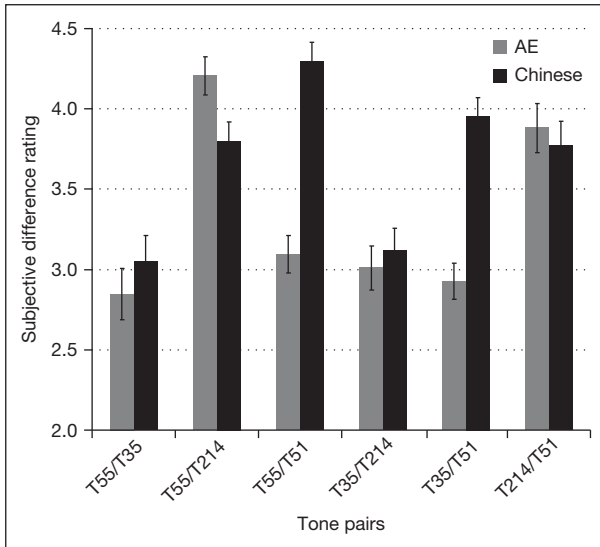
Fig. 2. Subjective degree-of-difference ratings by Chinese and AE listeners. '1' = 'Very similar', '5' = 'very different'. Error bars show one standard error.

Table 2. Significant t test comparisons of AE and Chinese listeners' ratings

Tone pair	t	p
T55/T214	2.43	0.02
T55/T51	−7.2	<0.001
T35/T51	−6.56	<0.001

selected 17 of the original 30 AE participants. The average difference rating for each tone pair for the two groups of listeners is shown in figure 2.[6]

A repeated-measures analysis of variance with between-subjects factor 'language' and a within-subjects factor 'tone pair type' found a marginally significant 'language' main effect [$F (1, 32) = 1.03$, $p = 0.023$, partial $\eta^2 = 0.15$]. There was also a significant effect of 'tone pair type' [$F (2.82, 90.3) = 37.8$, $p < 0.0001$, $\eta^2 = 0.541$]. (Since Mauchly's Test of Sphericity was significant, $p < 0.001$ and epsilon <0.75, the Greenhouse-Geisser correction is used here.) The interaction of 'listener language' by 'tone pair type' was significant as well [$F (2.82, 90.3) = 21.6$, $p < 0.0001$, $\eta^2 = 0.403$].

The differences seen between the two language groups were tested in planned comparisons of the rating data using independent samples t tests. As reported in table 2, the largest rating disparity lies with pairs T55/T51 and T35/T51, which were more distinctive for the Chinese listeners.

[6] In the analysis of this experiment we collapsed the data across the order of presentation in the tone pair presentations – that is, data for trials /ba55/ vs. /ba51/ are collapsed with data for trials /ba51/ vs. /ba55/. Stimulus order was not significant in any main effect or interaction in a larger analysis that included stimulus order contrasts.

Table 3. Confusability ranking of tone pairs in the AE and Chinese listeners' degree-of-difference rating data

	AE listeners	Chinese listeners
Most similar	all other pairs	T35/T214, T55/T35
Most different	T55/T214, T214/T51	all other pairs

Within-subject pairwise comparison (analysis of variance, ANOVA) revealed significant differences among tone pairs for both listener groups. For Chinese listeners, the pairs T55/T35 and T35/T214 were rated as significantly ($p < 0.01$) more 'similar' than were the other tone pairs, and no significant difference was found among any other tone pairs. For AE listeners, the most distinctive pairs T55/T214 and T214/T51 were rated as significantly ($p < 0.01$) more 'different' than any of the other pairs. There was also a marginally significant difference between T55/T214 and T214/T51 ($p = 0.017$) for the nonnative listeners (table 3).

3.3. Discussion

As we have seen, the degree-of-difference ratings were very different for the Chinese and AE listeners. The common characteristic of the pairs rated by AE listeners as most different (i.e. T55/T214 and T214/T51) is a high tone versus low tone contrast. The marginally significant difference between T55/T214 and T214/T51 ($p = 0.017$) further reveals that the f0 onset and/or offset values matter for AE listeners: since the pitch offset in T214 is actually close to the pitch level of '2' acoustically, pairs involving T55 and T214 have very distinctive pitch onsets and offsets. But this is not the case with T51 and T214, as T51 has a low offset that matches the onset and offset of T214. For AE listeners, apparently a tone that has a high pitch level in it somewhere is similar to another tone that has a high pitch level, even though the two may have quite different pitch contours (e.g. T35 and T51 were rated as relatively similar to each other). More generally, matching f0 onset and/or offset values may also have played a role in perceived tone similarity for AE listeners (i.e. T55/T35, T55/T51, T35/T214 and T35/T51).

On the other hand, pairs T55/T51 and T35/T51 were among the most dissimilar of tone pairs for Chinese listeners, having slightly higher ratings even than T55/T214 and T214/T51. A probable explanation is that these listeners were sensitive to the pitch contour shapes, which are very different for the tones involved in pairs T55/T51 (high level vs. falling) and T35/T51 (rising vs. falling). As the within-group pairwise comparisons show, the only two pairs rated by the Chinese listeners as significantly more similar than the other tone pairs are T35/T214 and T55/T35. This is a very interesting result because these are the two tone pairs that interact with each other in tone sandhi rules in Mandarin (see §2 above).

In sum then, Mandarin listeners were giving a phonological dissimilarity judgment – tones that are always distinct from each other in Mandarin phonology were given a difference rating of about 4, while tones that are sometimes neutralized by a phonological tone sandhi process were given a difference rating of 3. AE listeners, on

the other hand, were giving a phonetic dissimilarity judgment based on the initial pitch of the tone. If the initial pitch differed maximally (T55 vs. T214, or T214 vs. T51) the pair was given a difference rating of 4 and all other nonidentical tone pairs were given a difference rating of 3.

The question of interest then is: what will happen when we shift Mandarin listeners away from their phonological listening mode? We have seen that phonological processes have an impact on their off-line judgments of tone similarity, and what we explored in the two remaining experiments reported here is how Mandarin and AE listeners differ from each other in more low-level auditory perceptual tasks. Experiment 2 used a speeded, low-uncertainty discrimination task with the same stimuli used in experiment 1. Experiment 3 repeated experiment 2 but with sine wave analogs of the Mandarin tones.

4. Experiment 2: AX Discrimination of Natural Speech Tones

In this experiment AE and Mandarin listeners were again presented with pairs of Mandarin tone words, and we measured the dissimilarity of the tones. However, this experiment was designed to tap a lower level of perceptual processing that might not be influenced by the phonological patterning of tones in Chinese. Listeners were asked to make a simple 'same' or 'different' decision, responding as quickly as possible, and we used reaction time (RT) as a measure of perceptual similarity [Shepard, 1978]. To encourage listeners to make quick decisions based as much as possible on low-level sensory properties of the stimuli, we also used a low-uncertainty design, so that within each block of trials listeners were faced with only a single pair of stimuli. In this way we sought to be able to compare Mandarin and AE listeners when both groups are processing the stimuli in a phonetic way.

4.1. Methods

Participants and Stimuli

Eleven (6 female, 5 male, average age 30) Chinese and 13 (8 female, 5 male, average age 21) AE listeners participated in experiment 2. (Two of the AE listeners' data were later randomly thrown out so that each language group had 11 participants.) All Chinese listeners were from Beijing. The AE listeners spoke a Midland variety of American English. None of the listeners reported any history of speech or hearing problems. The Chinese listeners were paid a small amount of money for their participation, whereas AE listeners earned course credits.

The same four recorded syllables that were used in experiment 1 were also used in this experiment.

Procedure

All participants were tested in front of a computer in a quiet room, stimuli were played by the computer through headphones, and listeners entered their responses using a custom response box (stimulus presentation and response times and responses were controlled by a script using the E-prime experiment control software). Stimuli were presented in pairs as in experiment 1, but in this experiment the listener was given an AX 'same'/'different' discrimination task, responding 'same' to identical trials, and 'different' for nonidentical pairs. The interstimulus interval was 100 ms. Such a short interstimulus interval tends to block high-level linguistic perception [see e.g. Pisoni, 1973].

Table 4. Sensitivity d' values for Chinese and AE listeners' AX discrimination response accuracy data in experiment 2, using naturally recorded speech tones

	Chinese listeners		AE listeners		
	d'	95% range (±)	d'	95% range (±)	
T55/T35	3.95	0.42	3.70	0.42	
T55/T214	4.16	0.46	3.22	0.36	*
T55/T51	4.06	0.45	3.22	0.36	*
T35/T214	4.37	0.52	3.23	0.36	*
T35/T51	4.32	0.56	3.44	0.39	
T214/T51	4.32	0.50	3.41	0.38	*

Significant difference between the two listener groups is indicated with an asterisk.

The experiment uses a low-uncertainty design [Watson et al., 1976]. In this approach, a limited set of stimuli is presented in each trial block so the listener's perceptual attention can be narrowly focused on small distinctions. Each block tested the discrimination of only two tones (e.g. T55 and T35 might be tested in block 1, T35 and T51 in block 2, and so on). In each block, each of the four possible combinations of the two tones was repeated twice per cycle and each block contained five cycles (e.g. block 1 might have four ordered pairs, T55-T55, T55-T35, T35-T55, and T35-T35, all of which were repeated 10 times, yielding 4 pairs × 2 repetitions/cycle × 5 cycles = 40 pairs). There were 40 × 6 blocks = 240 trials in total. The order of the trials within cycles and the order of the blocks were randomized separately for each participant. There was a brief practice session, of four example tone pairs, at the beginning of the experiment.

Written and oral instructions were given to the two groups of listeners in their respective native language. After a listener responded correctly, a feedback message, detailing his/her RT and percentage correct, was displayed on the screen for 1.5 s. After a 2-second delay, the next trial was played. Both RT and response accuracy were recorded. Experiment 2 was run within the same 1-hour session right after experiment 3, with a short break between the experiments.[7]

4.2. Results

As expected with the low-uncertainty design, error rates were quite low, so our main data analysis will focus on RT. Nonetheless, we can note that an analysis of perceptual sensitivity found that d' [a measure computed for response accuracy data that factors out false alarms and misses; MacMillan and Creelman, 2005] for the Chinese listeners was nearly perfect for all pairs (d' ≥ 4, except for T55/T35 where d' = 3.95) and there were no significant differences among the tone pairs for the Chinese listeners. Interestingly, the Chinese listeners performed better than the AE listeners, and significant differences ($p < 0.05$) in d' values were found with tone pairs T55/T214, T55/T51, T35/T214 and T214/T51. Table 4 lists all d' values and 95% confidence limits for both groups of listeners.

---

[7] For the sake of a clear exposition we are presenting experiment 2 before experiment 3. However, all of the subjects who participated in experiment 2 also participated in experiment 3. The order of the experiments was the same for all subjects – they did the nonspeech experiment (exp. 3) first, and then did the speech experiment (exp. 2) because we wanted to insure that subjects did not think of the nonspeech stimuli as speech.
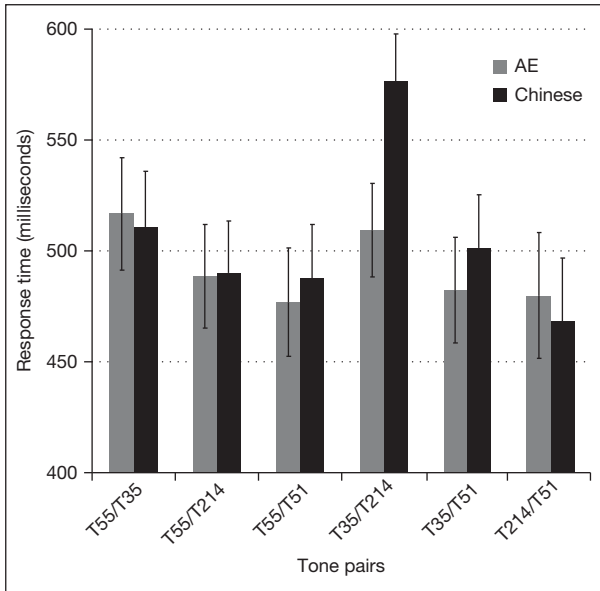
Fig. 3. RTs (in milliseconds) for the correct 'different' responses from the AX limited set discrimination task using natural speech stimuli. No significant language effect was found in the repeated-measures ANOVA. (Error bars show one standard error.)

RTs for correct 'different' responses were analyzed in a repeated-measures ANOVA (the general linear model implemented in SPSS). The design is as follows: the between-subjects variable 'listener language' has two levels (American English versus Chinese), while the within-subject variable 'tone pair type' has six levels (T55/T35, T55/T214, T55/T51, T35/T214, T35/T51 and T214/T51). For each tone pair, median RT values were determined for each individual participant.

Median RT values for each collapsed 'different' tone pair (i.e. T55/T35, T35/T214, etc.) were determined for each participant and these were submitted to a repeated-measures analysis of variance. The RT data are shown in figure 3. The language main effect was not significant [$F (1, 20) = 0.21$, $p = 0.65$, partial $\eta^2 = 0.01$], indicating that overall AE and Chinese listeners responded to the stimuli at about the same speed. The within-subject tone pair type effect was significant [$F (4.6, 91.9) = 6.6$, $p < 0.0001$, partial $\eta^2 = 0.247$]. (Since Mauchly's Test of Sphericity was marginal, $p = 0.074$ and epsilon > 0.75, the Huynh-Feldt correction is reported here.) The 'tone pair type' by 'listener language' interaction effect was marginally significant, [$F (4.6, 91.9) = 2.1$, $p = 0.078$, partial $\eta^2 = 0.095$]. While the RT functions for the two groups are largely parallel, the RT for T35/T214 in the Chinese listeners' data is noticeably longer. When a planned between-group comparison using the independent samples T test was performed on the RT data of all tone pairs, the RTs for T35/T214 in Chinese listeners' data were found to be significantly longer than those in AE listeners' data ($p = 0.036$).

Separate post-hoc tests revealed that for the Chinese listeners T35/T214 (with the longest RT) was the most confusable pair and significantly different (at the 99% confidence level) from all other pairs except T55/T35. No other significant difference was found. This is summarized in table 5. For AE listeners, no significant RT difference was found between any two pairs. This means that the significant within-subject main effect of 'tone pair type' came from the Chinese listeners' data.

Table 5. Confusability ranking of tone pairs in Chinese listeners' data

Most confusable	T35/T214
'Middle'	T55/T35
Least confusable	T55/T51, T55/T214, T35/T51, T214/T51

4.3. Discussion

In experiment 1 we found that Mandarin listeners based their phonological dissimilarity judgments on phonological patterning, so that tones that are sometimes neutralized by a tone sandhi process were rated as being more similar to each other than were other tone pairs. We hypothesized that this pattern of results was due to the 'linguistic' level being tapped by the similarity rating task, and that using a speeded discrimination task might put AE listeners and Mandarin listeners on a more equal footing – dealing with the syllables as auditory objects rather than as items in a linguistics system. In experiment 2, however, the differences between language groups were only partly eliminated. For Chinese listeners, our RT data showed that the tones related to each other by a sandhi rule (T35 and T214 are neutralized by the Tone 3 Sandhi Rule, and T55 and T35 are neutralized by the Tone 2 Sandhi Rule) had longer discrimination RTs than did the other tone pairs. The AE listeners in experiment 1 were influenced by initial pitch of the tone. If the initial pitch differed maximally (T55 vs. T214, or T214 vs. T51) the pair was rated as more different than otherwise. This pattern was not repeated in experiment 2, as our RT data showed that the AE listeners responded about equally quickly to all of the tone pairs. It is possible that the lack of an effect of tone shape on the AE response time data may have reflected a floor effect, the listeners having practiced the paradigm in experiment 3 before their participation in experiment 2. If this was the case, it is interesting that the Chinese listeners still showed slower RTs for phonologically neutralized tones.

So, in experiment 2 we found a significant difference between Chinese and AE listeners in how discriminable T35 is from T214. These tones were more confusable to the native Chinese speakers who use them in their daily speech. This finding suggests that the phonological alternation between T35 and T214 in Mandarin may have had an influence on these listeners' low-level speech perception. Thus, even in a 'low-level' speech perception task, the Chinese listeners' experience with the phonology of Mandarin had an impact on the perceptual similarity of tones. Experiment 3 was designed to further test the linguistic basis of auditory perception by presenting nonspeech tokens that contain tonal information. We used sine wave tone analogs for this purpose.

5. Experiment 3: AX Discrimination Task with Sine Wave Tones

5.1. Methods

Participants, Stimuli and Procedure
The same Mandarin and AE listeners who participated in experiment 2 also participated in this experiment.

The stimuli were simple sine wave analogs of the four natural speech stimuli that were used in experiments 1 and 2. The stimuli were generated with a synthesizer adapted from the C-code generously shared by Alex Francis and Howard Nusbaum at the University of Chicago. Specifically, the frequency of a single time-varying sinusoidal wave followed the trajectory of the f0 of each of the four recorded natural speech monosyllabic words /ba55, ba35, ba214, ba51/ that we used as stimuli in the previous experiments (fig. 1). The amplitude of the sinusoid was modeled on the amplitude contour of each of the speech syllables and the overall impression of the synthetic sinusoidal stimuli was that they were like low-pass filtered speech, but with the pure-tone quality of a sine wave.

After running the experiment, we noticed that the falling portion of the T51 contour was somewhat delayed when the pitch traces of the sine wave tone and natural tone were aligned at vowel onset. The sine wave tone was also about 1/7 (or 40 ms) shorter with a flat intensity contour, while the intensity decreased sharply in the last 1/3 of the natural speech monosyllable (total duration = 340 ms). These properties of the falling pitch contour and intensity envelopes of the sine wave T51 may have contributed to an increase in the perceived similarity between T55 and T51 in this experiment.

The same low-uncertainty speeded AX discrimination task, with feedback, that we used in experiment 2 was also used in this experiment.

5.2. Results

Error rates were low for both listener groups in experiment 3, though not as low as they were in experiment 2. The Chinese listeners' overall error rate for 'different' pairs was 5%, with pairs T51-T55 (9%) and T35-T214 (13%) drawing the most errors. But these larger numbers were due to a very high error rate of 1 listener in each case (50 and 80%, respectively). When these outliers were disregarded, the error rates for these pairs conformed to the overall error rate. For AE listeners, the overall error rate was 7%. The most errors were made with pairs T35-T55 (12%), T55-T51 (11%) and T51-T55 (9%). But again, these were attributable to 1 or 2 listeners' high error rate in each case.

The sensitivity measure d' was again computed for the response accuracy data. For Chinese listeners, T35/T214 was found to be significantly different ($p < 0.05$) from T55/T214 and T35/T51. Thus, T35/T214 was significantly more difficult to discriminate than T55/T214 and T35/T51 for the native listeners. In the AE listeners' accuracy data, no significant differences were found among d' values. Overall, AE listeners' performance was better than the Chinese listeners. They did significantly better ($p < 0.05$) than the Chinese listeners with pairs T55/T35 and T214/T51, with a marginal difference for T35/T214. Table 6 lists d' values and the 95% confidence limits for both listener groups.

RTs for correct 'different' responses were analyzed in a repeated-measures ANOVA (the general linear model implemented in SPSS). The design was the same as in experiment 2: the between-subjects variable 'listener language' has two levels (American English versus Chinese), while the within-subject variable 'tone pair type' has six levels (T55/T35, T55/T214, T55/T51, T35/T214, T35/T51 and T214/T51). For each tone pair, median RT values were determined for each individual participant. As can be seen in figure 4, the overall pattern of RT was similar for the two groups. There was a significant effect of the within-subject factor of 'tone pair type' [$F_{(4.35, 87)}$ = 16, $p < 0.0001$, partial $\eta^2 = 0.44$]. (Since Mauchly's Test of Sphericity was significant, $p = 0.02$ and epsilon >0.75, the Huynh-Feldt correction is used here.) No significant between-subject main effect of 'listener language' was detected by ANOVA in the RT data [$F_{(1, 20)} = 0.003$, $p = 0.957$, partial $\eta^2 < 0.001$]. But the 'tone pair type' by
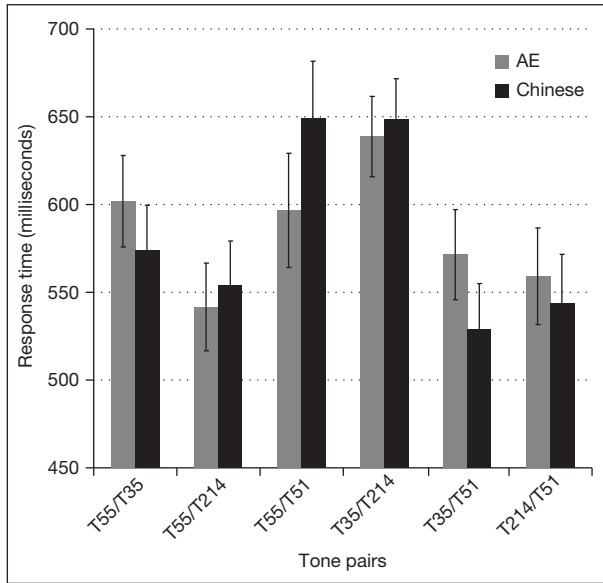
Fig. 4. Response time plot of response time data for the experiment of AX limited stimulus set discrimination task with sine wave tones for Chinese and AE listeners. (Error bars show one standard error.)

Table 6. Sensitivity d' values for Chinese and AE listeners' response accuracy data from experiment 3 using sine wave tones

	Chinese listeners		AE listeners		
	d'	95% range (±)	d'	95% range (±)	
T55/T35	3.18	0.38	4.15	0.57	*
T55/T214	3.45	0.43	3.36	0.38	
T55/T51	2.98	0.36	3.65	0.42	
T35/T214	2.59	0.32	3.25	0.36	(*)
T35/T51	3.49	0.42	3.61	0.41	
T214/T51	3.05	0.38	3.99	0.47	*

Significant difference between the two listener groups is indicated with an asterisk.

'listener language' interaction had a marginally significant effect [F (4.35, 87) = 2.7, p = 0.032, partial η^2 = 0.12], indicating that, at least for some tone pairs, AE listeners' RTs were different from Chinese listeners'.

Pairwise comparison (ANOVA) of the within-subject factor of tone pairs for AE listeners' RT data showed that pairs T55/T214 (with the shortest RTs, thus most distinctive) and T35/T214 (with the longest RTs, thus most confusable) differed significantly (p = 0.002 at the 99% confidence level). Pairs T55/T35, T55/T51, T35/T51 and T214/T51 fell in the middle, with T55/T35 being marginally different from T55/T214 (p = 0.038) and T35/T51 from T35/T214 (p = 0.032). In the Chinese listeners' data, RTs to

Table 7. Confusability ranking of tone pairs

	AE	Chinese
Most confusable	T35/T214	T35/T214, T55/T51
'Middle'	T55/T35, T55/T51	
	T35/T51, T214/T51	
Most distinctive	T55/T214	T55/T35, T55/T214, T35/T51, T214/T51

T35/T214, T55/T51 were significantly longer than those to the other pairs (T55-T35, T55-T214, T35-T51, T214-T51), which were not reliably different from each other. These patterns are summarized in table 7.

5.3. Discussion

Although there were some differences between AE and native Chinese listeners in the d' analysis of response sensitivity, these differences should be taken with a grain of salt because, overall, performance was very high on this task. We found more coherent and important results in our analysis of the response time data. In the response time analysis of variance we found significant differences in how long it took listeners to respond 'different' to tone pairs, depending on which specific tones were being contrasted with each other. There was substantial similarity between Chinese and AE listeners on this measure (the Pearson's product-moment correlation of the Chinese and AE RTs was 0.78). However, there was a marginally significant interaction between 'tone pair type' by 'listener language', indicating a small language effect even in this study of nonspeech stimuli using a low-uncertainty listening task. When the RT data were further probed with pairwise comparisons of the within-subject factor of tone pair types, different confusability rankings were derived for the Chinese and AE listener groups (table 7). While tones T35 and T214 were confusable for both groups, the Chinese listeners also found T55 and T51 quite confusable. On the other hand, AE listeners perceived T55/T214 to be the most distinctive (i.e. with the shortest RTs), while this tone pair did not stand out in the Chinese listeners' data.

There are noticeable differences between the results from experiments 2 and 3. First of all, the overall RT was shorter in experiment 2 for both Chinese and AE listeners, which may be seen as a practice effect [Werker and Logan, 1985], as experiment 2 was run after experiment 3 in a single listening session. This effect was more pronounced in the AE listeners' RT data. RTs for pairs T55/T51, which were short in experiment 2, were relatively long in experiment 3. Thus, the construction of the synthetic stimulus tone T51, which made it more acoustically similar to T55, may have also contributed to longer RTs for the T55/T51 comparison for both groups of listeners.

Another interesting point to note here is that the AE listeners performed better than the Chinese listeners in experiment 3 (using synthetic stimuli), while the pattern was reversed in experiment 2 (using natural speech stimuli), as revealed by the sensitivity measure d'. This may be because the natural speech stimuli made the discrimination task more similar to normal linguistic processing for the Chinese listeners.

6. General Discussion

The results from the three experiments in our study revealed cross-linguistic differences as well as similarities between the Chinese and the AE listeners. The basic patterns in the rating data from the subjective degree-of-difference rating task (experiment 1) are: Chinese listeners had rated as more similar only two pairs of tones (i.e. T35/T214 and T55/T35); on the other hand, AE listeners had rated as more distinctive only two pairs of tones (i.e. T55/T214 and T214/T51). It seems clear in experiment 1 that the Chinese listeners' tone perception was influenced by the tone sandhi rules of their native language: tone pairs T35/T214 and T55/T35 were rated as most similar because these tones are involved in the T214 or T35 sandhi (see §2), which leads to contextual neutralization of the tonal contrast between T35 and T214 or that between T55 and T35.

Experiment 2 found that the effect of tone sandhi on Chinese listeners' tone perception is quite remarkable in strength, because the T35/T214 similarity was even present in a simple AX low-uncertainty discrimination task using natural speech stimuli: it took the Chinese listeners significantly longer (than the AE listeners) to make the 'same'/'different' discrimination decision for T35/T214. Again, had tone perception been affected by phonetic similarity only, Chinese listeners' RTs for T35/T214 should have conformed to the overall RT difference between the two groups of listeners, that is, slightly longer than AE listeners' just as in the other tone pairs, but not significantly longer. After all, Chinese listeners are more experienced with Mandarin tone distinctions than AE listeners, as is evidenced by their higher d' values obtained from the accuracy data in experiment 2. So, there is no reason for them to perform significantly worse in this particular pair of tones than AE listeners. In addition, within-subject pairwise comparison revealed that T35/T214 had significantly longer RTs than some other pairs of tones for the Chinese listeners. This also points to the effect of the T214 sandhi, as T35 and T214 were not significantly more similar than any other pairs of tones for AE listeners. The relatively long – albeit nonsignificant – RT for T35/T55 in the Chinese listeners' data may be a hint of the weaker tone 2 sandhi effect [see (2) in §2 above].

The two groups of listeners behaved even more alike in the low-uncertainty AX discrimination test using sine wave tones synthesized from natural speech tone templates (experiment 3), where no obvious phonological effect was found. It is possible that, with the segmental makeup taken away and with just four stimulus tone tokens repeated over and over in experiment 3, it was easy to focus attention on the acoustic properties of the stimuli. As a result, the data reflect mainly auditory perception. Although in post-test questioning the Chinese listeners reported that they heard Chinese tones in this experiment, it is rather doubtful that there was any lexical activation involved in this task. Nevertheless, even in this experiment using synthetic stimulus tones, there were still some slight differences in how the two groups reacted to pairs T55-T214, T51-T55, and T51-T214 (shorter RTs for AE than for Chinese listeners) and T35-T51 (longer RTs for AE listeners), which resulted in a small effect of the 'tone pair type' by 'listener language' interaction. We also noted that the unnaturalness in the synthetic stimulus tone T51 affected Chinese listeners' perception more than that of AE listeners, which shows that perception in this task may not have been completely based on auditory discriminability; otherwise, one would not expect naturalness to have any impact here. As the RT data (fig. 4) show, even in this simple task, the RTs

for T35/ T214 were still longer relative to most of the other tone pairs in the Chinese listeners' data.

As we noted above, when time was allowed for reflection in experiment 1, AE listeners adopted a strategy, perhaps based on English intonation, to listen for pitch levels and to disregard contours. This enabled them to distinguish more easily the two pairs of tones involving a high vs. low pitch contrast (i.e. T55/T214 and T214/T51). Interestingly, it is less clear to what extent this strategic approach to the stimuli was used in the low-uncertainty psychophysical tasks (experiments 2 and 3), as AE and Chinese listeners acted more alike there, except for the sandhi pair T35/T214 in experiment 2. This suggests that in a task similar to normal language use situations (experiment 1) it may be impossible for a listener to not be influenced by his/her native language, even in cases where the stimuli are nonnative.

7. Conclusion

As is evident from the experimental results reported above, linguistic experience can lead to language-specific patterns in speech perception. Regarding the suggestion [Steriade, 2001] that a universal map of perceptual salience influences phonological patterning in language, there is no doubt that general auditory capacities do not differ much among language-learning children with normal hearing ability, no matter how diverse their linguistic backgrounds are. It is thus reasonable to assume that universal patterns in perception exist at a very early point in language acquisition.

However, our data, together with other data on language specificity in speech perception, suggest that the universal auditory map is modified by linguistic experience. One way to account for the role of experience has been demonstrated in the models proposed by Guenther and colleagues [Guenther and Gjaja, 1996; Guenther et al., 1999; Guenther and Bohland, 2002; Guenther et al., 2004] and Bauer et al. [1996]. A central component in these models is an auditory cortical map, whose formation is influenced by stimulus input and type of training. In particular, Guenther et al. [1999] found that categorization training in psychophysical experiments using nonspeech-like band-pass-filtered acoustic noise in different frequency ranges led to smaller cortical representation of (hence, decreased sensitivity to) stimuli in the training range, while discrimination training led to larger cortical representation (hence, increased sensitivity) in the training range. Functional magnet resonance imaging studies by Guenther and Bohland [2002] and Guenther et al. [2004] provided supporting evidence for a model of auditory cortical map formation. Greater temporal lobe activation was recorded when subjects heard nonprototypical examples of American English /i/ than when they heard prototypical examples of /i/ [Guenther and Bohland, 2002; Guenther et al., 2004]. If we may further interpret Guenther and colleagues' results, their prototypical examples of /i/ can be seen as stimuli from the 'categorization training range', except that the training was not done under laboratory conditions but during a listener's lifetime experience with his native language. Such an auditory warping certainly serves the linguistic purpose well, as it directs neural activities to distinguishing between-category (phonological) differences and to ignoring irrelevant within-category (low-level phonetic) differences. As pointed out by Guenther and Gjaja [1996], this would also enable a unified neural model for speech modalities and other sensory and motor modalities. These processes are probably at work very early in language acquisition

leading to cross-linguistic differences in perception such as those noted for the vowel [i] for Swedish-learning and English-learning infants [Kuhl et al., 1992].

On the other hand, Johnson's [2004] lexical distance model, although not explicitly denying the existence of auditory warping, tries to account for language-specific effects by referring to activation in the lexicon. In the lexical distance model, incoming signals are compared with phonetically detailed forms in the lexicon directly. Consequently, language-specific perceptual effects may simply emerge from the lexicon. The model computes overall perceptual distance (d) from two sources: (1) inherent auditory similarities between two stimuli (d_a) and (2) aggregated average difference in lexical activations by the two stimuli (d_l, computed as the difference in the amounts of activation of the lexicon caused by these stimuli, with a constant k gating the influence of this lexical distance on perception under different experimental conditions); i.e. $d = d_a + k \times d_l$. Because of the way the overall perceptual distance is computed, it is claimed that the model has the ability to distinguish discrimination performance from categorization performance. Discrimination performance can be found in a minimal-uncertainty task of limited stimulus set or speeded AX discrimination with a short inter-stimulus interval such as experiments 2 and 3 in our study (no lexical access assumed, perceptual distance computed exclusively from auditory distance). And categorization performance is found in tasks involving higher memory load such as AXB identification, the degree-of-difference rating as in experiment 1 in our study (where it is hypothesized that lexical forms may be consulted for similarity judgments). Johnson's [2004] fricative perception data from a rating task as well as a speeded AX discrimination task by Dutch and AE listeners support this hypothesis. The fact that the language effect is the most significant in the degree-of-difference rating task for both Chinese and AE listeners in our study provides further supporting evidence for the different degrees of lexical activation posited in Johnson's [2004] model.

Neither the neural model [Guenther and Gjaja, 1996; Guenther and Bohland, 2002] nor the lexical distance model [Johnson, 2004] explicitly discusses the issue of how neutralization rules may affect discrimination of two contrastive sounds (or tones) that are neutralized in certain phonetic environments. Within the neural model of Guenther and Gjaja [1996] and Guenther and Bohland [2002], we may imagine a 'noisy' training condition under which stimuli categorized into an abstract representation of A may sometimes have to be categorized as A or a second category B (e.g. [T35] to either /T35/ or /T214/ because of the /T214.T214/ → [T35.T214] neutralization rule). As a result of this double-identity status of certain speech sounds, the contrast between the relevant tone (or phoneme) categories may be weakened and category boundaries less well defined. Within Johnson's [2004] lexical distance model, because of the cross-representation of two tones or sounds (e.g. /T35/ and /T214/ in our case), a [T35] or a [T214] input may activate lexical items containing either /T35/ or /T214/. Consequently, the difference in lexical activation, i.e. the lexical distance, between /T35/ and /T214/ is predicted to be smaller than if there is no such neutralization rule. In both accounts, the sandhi pattern affects phonological category representation, making the surface ordering of the tones irrelevant. Thus, neither account should predict an order effect between T35-T214 and T214-T35, which is consistent with the findings reported here.

In summary, the findings of the three experiments reported here support the hypothesis that speech perception is language-specific, specifically that tone sandhi processes influence native listener tone perception at various levels. These findings

are consistent with the results reported in Gandour [1981, 1983, 1984], Deutsch et al. [2006] and Krishnan et al. [2005]. The fact that language specificity showed up in even the low-uncertainty, speeded AX discrimination supports the hypothesis that linguistic experience may lead to auditory warping [Guenther and Gjaja, 1996; Guenther and Bohland, 2002]: a neural map, once formed and warped by one's native phonological category representation, should be reflected in perceptual patterns even before higher-level linguistic processing is involved.

Findings by some recent developmental studies by Kuhl [1991] and Kuhl et al. [1992, 2006] further support this view. A longitudinal study by Rivera-Gaxiola et al. [2005] followed infants (of American English-speaking families) from 7 to 11 months of age. It was found that although infants at 11 months of age were still able to discriminate nonnative contrasts, their performance was better at 7 months. On the other hand, their ability to process native contrasts improved over time. Our interpretation of such findings is that the neural map favoring native contrasts may have already started taking shape during this early period in life. Similarly, Tsao et al. [2006; see also Kuhl et al., 2006] found that for infants between 6 and 12 months of age, discriminability of native contrasts increased while that for nonnative contrasts decreased with age. They conclude that 'infants develop language-specific processing around their first birthday'.

In light of these findings, we would suggest that the auditory cortical map replaces the universal perceptual map (the latter of which may be assumed for prelinguistic infants under 6 months of age [cf. Kuhl, 1991; Kuhl et al., 2006; Tsao et al., 2006]) as one is continuously exposed to a particular language and develops phonological representations and rules, and that it is this neural map that determines perceptual similarity between two stimulus sounds.[8] It is thus not surprising for language specificity to show up in such a simple task as AX discrimination.

The different experimental tasks in the present study also brought out different degrees of language specificity (from the strongest tone sandhi effects found in our degree-of-difference rating experiment 1, to a weaker effect found in the speeded discrimination tasks using natural speech and synthetic stimuli in experiments 2 and 3), suggesting that speech perception may also be influenced by the degrees of lexical activation involved in different tasks [Johnson, 2004]. This led us to the speculation that there may exist two types of linguistic effects: the first, found in tasks involving mainly auditory processing (as in our experiments 2 and 3), comes from the auditory cortical map, where category boundaries between segments and lexical tones are defined by a particular segmental or tonal system; the second, found in tasks involving higher-level linguistic processing (as in our experiment 1), comes from the more conscious consultation of phonological or tonal rules, as well as the auditory cortical map. In essence, this is not very different from Johnson's [2004] lexical activation model, where the auditory and lexical effects interact to various degrees in different tasks. The only difference is that in Johnson's [2004] model, auditory distance is computed based on pure auditory similarity that is universal to speakers of any language, whereas in our

---

[8] This is not to say that one may completely lose the ability to learn foreign sounds. As many studies [e.g. Buonomano and Merzenich, 1998; Kilgard and Merzenich, 1998; Stiles, 2000; Zhang et al., 2006] have shown, our neural system exhibits remarkable plasticity, even late in life. Therefore, new neural maps may be formed when new sound categories are acquired. Or perhaps modifications can be made to existing maps to accommodate these new categories, by rearranging the map.

view this purely universal auditory similarity may not exist, due to differences in the auditory cortical maps warped by the listeners' respective phonologies.

Our data also show that tone sandhi rules may play a role in defining tone category representation, as evidenced by the fact that the Chinese listeners have different perceptual patterns from those of the AE listeners, especially in the case of T214 and T35, where the Chinese listeners actually showed a disadvantage in discriminating these tones. We argued that this was because the boundary between these tone categories is blurred due to the Tone 3 Sandhi Rule. A similar, although weaker, effect was seen between T55 and T35, as a result of the Tone 2 Sandhi Rule. Another point to note here is the difference in the tone 3 and tone 2 sandhi effects. Such a difference in the strength of the effect of these sandhi rules may be an indication that unlike category-changing phonological rules (the Tone 3/T214 Sandhi Rule in this case), the lower-level phonetic realization rules (e.g. the Tone 2/T35 Sandhi Rule) do not play as important a role in forming phonological category representations. Consequently, the effect is not seen in simple AX discrimination, where the perceptual patterns are largely determined by category representations on the auditory cortical map. Extending our discussion to phonological studies in general, if the goal of phonology is to capture and represent linguistic cognitive processes in a theory, one has to take the position that not all phonological contrasts are the same, and that allophonic alternations should be taken into account in a more realistic model of phonological representation [cf. Hume and Johnson, 2003 and Frisch et al., 2004].

Although the focus of much phonological research is on certain common phonological processes (e.g. palatalization) and sound changes (e.g. tonal genesis) are observed across-linguistically, it should be noted that there are also language-particular processes and sound changes. Take Rugao Chinese for example. In this southern Mandarin dialect with an inventory of high level, rising, falling-rising, and falling tones similar to the inventory found in standard Mandarin, the second of two consecutive falling-rising tones is phonetically downstepped, instead of being neutralized with the rising tone [Huang, in press]. Such language-particular processes will have different effects on perception, similar to what we have seen in the experimental results reported here. It is further predicted that these language-specific perceptual patterns will in turn contribute to the reshaping of the respective synchronic phonological systems. Thus, in Rugao Mandarin the category between the rising tone and the falling-rising tone is better defined than that in Beijing Mandarin. Furthermore, these language-specific perceptual patterns are predicted to condition different historical sound changes. For instance, most of the modern Chinese dialects are assumed to have descended from a common ancestral language, i.e. Middle Chinese [Downer, 1963; Norman, 1973; Baxter, 1992]. Yet the tonal systems of the modern dialects can be very different, with some dialects having only three tones [e.g. Yantai Mandarin; Qian, 1982], instead of the more common number of four. Apparently a merger of two tone categories happened in some dialects but not the others. Thus, while we observe that there exist language-universal phonological patterns and processes, it is also necessary to acknowledge the presence of language-specific processes and perceptual patterns.

Acknowledgments

discussions in the design phase of this research and comments on the paper. Many thanks also to Dr. Cao Wen of Beijing Language and Culture University, Virgina Novak and Kris Pokorny for helping us collect data.

References

Babel, M.; Johnson, K.: Accessing psycho-acoustic perception with speech sounds. Lab. Phonol. *1:* 179–205 (2010).

Bauer, H.-U.; Der, R.; Hermann, M.: Controlling the magnification factor of self-organizing feature maps. Neural Computation *8:* 757–771 (1996).

Baxter, W.H.: A handbook of old Chinese phonology (Mouton de Gruyter, Berlin 1992).

Beckman, M.E.: Toward phonetic criteria for a typology of lexical accent; PhD diss. Cornell University (1984).

Blevins, J.: A theoretical synopsis of evolutionary phonology. Theoret. Ling. *32:* 117–166 (2006).

Boomershine, A.; Hall, K.C.; Hume, E.; Johnson, K.: The impact of allophony vs. contrast on speech perception; in Avery, Dresher, Rice, Contrast in phonology: perception and acquisition (Mouton de Gruyter, New York 2008).

Buonomano, D.V.; Merzenich, M.M.: Cortical plasticity: from synapses to maps. Annu. Rev. Neurosci. *21:* 149–186 (1998).

Carroll, J.D.; Chang, J.-J.: Analysis of individual differences in multi-dimensional scaling via an n-way generalization of 'Eckart-Young' decomposition. Psychometrika *35:* 283–319 (1970).

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

Chao, Y.-R.: A grammar of spoken Chinese (University of California Press, Berkeley 1968).

Chen, M.Y.: Tone sandhi patterns across Chinese dialects (Cambridge University Press, Cambridge 2000).

Cheng, C.-C.: A synchronic phonology of Mandarin Chinese (Mouton, The Hague 1973).

Chomsky, N.; Halle, M.: The sound pattern of English (Harper & Row, New York 1968).

Deutsch, D.; Henthorn, T.; Marvin, E.; Xu, H.-S.: Absolute pitch among American and Chinese conservatory students: prevalence differences, and evidence for a speech-related critical period. J. acoust. Soc. Am. *119:* 719–722 (2006).

Downer, G.B.: Chinese, Thai, and Miao-Yao; in Shorto, Linguistics comparison in Southeast Asia and the Pacific (London 1963).

Duanmu, S.: The phonology of standard Chinese; 2nd ed. (Oxford University Press, Oxford 2007).

Dupoux, E.; Pallier, C.; Sabastian, N.; Mehler, J.: A destressing 'deafness' in French? J. Memory Lang. *36:* 406–421 (1997).

Flege, J.E.; Takagi, N.; Mann, V.: Lexical familiarity and English language experience affect Japanese adults' perception of /r/ and /l/. J. acoust. Soc. Am. *99:* 1161–1173 (1996).

Fox, R.A.: Perception of vowel quality in a phonologically neutralised context; in Tokura, Vatikiotis-Bateson, Sagisaka, Speech perception, production and linguistic structure, pp. 21–42 (Ohmsha/IOS Press, Tokyo 1992).

Frisch, S.; Pierrehumbert, J.B.; Broe, M.: Similarity avoidance and the OCP. Natural Lang. linguistic Theory *22:* 179–228 (2004).

Gandour, J.T.: Perceptual dimensions of tone: evidence from Cantonese; in J. Chinese Ling. *9:* 20–36 (1981).

Gandour, J.T.: Tone perception in Far Eastern languages. J. Phonetics *11:* 149–176 (1983).

Gandour, J.T.: Tone dissimilarity judgments by Chinese listeners. J. Chinese Ling. *12:* 235–261 (1984).

Goto, H.: Auditory perception by normal Japanese adults of the sounds 'l' and 'r'. Neuropsychologia *9:* 317–323 1971).

Guenther, F.H.; Bohland, J.W.: Learning sound categories: a neural model and supporting experiments. Acoust. Sci. Technol. *23:* 213–220 (2002).

Guenther, F.H.; Gjaja, M.: The perceptual magnet effect as an emergent property of neural map formation. J. acoust. Soc. Am. *100:* 1111–1121 (1996).

Guenther, F.H.; Husain, R.T.; Cohen, M.A.; Shinn-Cunningham, B.G.: Effects of categorical and discrimination training on auditory perceptual space. J. acoust. Soc. Am. *106:* 2900–2912 (1999).

Guenther, F.H.; Nieto-Castanon, A.; Ghosh, S.S.; Tourville, J.A.: Representation of sound categories in auditory cortical maps. J. Speech Lang. Hear. Res. *47:* 46–57 (2004).

Guion, S.G.: The role of perception in the sound change of velar palatalisation. Phonetica *55:* 18–52 (1998).

Hansson, G.O.: Theoretical and typological issues in consonant harmony; PhD diss. University of California, Berkeley (2001).

Harnsberger, J.D.: The perception of Malayalam nasal consonants by Marathi, Punjabi, Tamil, Oriya, Bengali, and American English listeners: a multidimensional scaling analysis. J. Phonet. *29:* 303–327 (2001).

Haudricourt, A.-G.: De l'origine des tons en Vietnamien. J. Asiatique *242:* 69–82 (1954a).

Haudricourt, A.-G.: Comment reconstruire le Chinois achaïque. Word *10:* 351–364 (1954b).

Hombert, J.-M.: Consonant types, vowel quality, and tone; in Fromkin, Tone: a linguistic survey, pp. 77–111 (Academic Press, New York 1978).

Hombert, J.-M.; Ohala, J.J.; Ewan, W.G.: Phonetic explanations for the development of tones. Language *55:* 37–58 (1979).

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

Huang, T.: Contextual and pitch range effects on tonal realization in Rugao Chinese. J. Chinese Ling. (in press).

Hume, E.; Johnson, K.: A model of the interplay of speech perception and phonology; in Hume, Johnson, The role of perception in phonology (Academic Press, New York 2001).

Hume, E.; Johnson, K.: The impact of partial phonological contrast on speech perception. Proc. 15th Int. Congr. Phonet. Sci., 2003.

Hume, E.; Johnson, K.; Seo, M.; Tserdanelis, G.; Winters, S.: A cross-linguistic study of stop place perception. Proc. 14th Int. Congr. Phonet. Sci., 1999, pp. 2069–2072.

Hume, E.; Seo, M.: From speech perception to optimality theory: metathesis in Faroese and Lithuanian. Nordic J. Ling. *27:* 35–60 (2004).

Hura, S.L.; Lindblom, B.; Diehl, R.L.: On the role of perception in shaping phonological assimilation rules. Lang. Speech *35:* 59–72 (1992).

Jakobson, R.; Fant, G.; Halle, M.: Preliminaries to speech analysis: the distinctive features and their correlates (Acoustics Laboratory, Massachusetts Institute of Technology, Cambridge 1952).

Jakobson, R.; Halle, M.: Fundamentals of language (Mouton, Gravenhage 1956).

Janson, T.: Sound change in perception and in production. Language *59:* 18–34 (1983).

Johnson, K.: Processes of speaker normalization in vowel perception; PhD diss. Department of Linguistics, Ohio State University (unpublished, 1988).

Johnson, K.: Cross-linguistic perceptual differences emerge from the lexicon; in Agwuele, Warren, Park, Proc. 2003 Texas Linguistics Soc. Conf.: Coarticulation in Speech Production and Perception, pp. 26–41 (Cascadilla Press, Sommerville 2004).

Kilgard, M.P.; Merzenich, M.M.: Cortical map reorganization enabled by nucleus basalis activity. Science *279:*1714–1718 (1998).

Kiriloff, C.: On the auditory perception of tones in Mandarin. Phonetica *20:* 63–67 (1969).

Kochanski, G.P.; Shih, C.; Jing, H.: Prosody modeling with soft templates. Speech Commun. *39:* 311–352 (2003).

Kochetov, A.; Pouplier, M.: Phonetic variability and grammatical knowledge: an articulatory study of Korean place assimilation. Phonology *25:* 1–33 (2008).

Kohler, K.: Segmental reduction in connected speech: phonological facts and phonetic explanations; in Hardcastle, Marchal, Speech production and speech modeling, pp. 69–92 (Kluwer Academic Publishers, Dordrecht 1990).

Kratochvil, P.: The Chinese language today: features of an emerging standard (Hutchinson, London 1968).

Krishnan, A.; Xu, Y.; Gandour, J.; Cariani, P.: Encoding of pitch in the human brainstem is sensitive to language experience. Brain Res cogn. Brain Res. *25:* 161–168 (2005).

Kuhl, P.K.: Human adults and human infants show a 'perceptual magnet effect' for the prototypes of speech categories, monkeys do not. Perception Psychophysics *50:* 93–107 (1991).

Kuhl, P.K.; Stevens, E.; Hayashi, A.; Deguchi, T.; Kiritani, S.; Iverson, P.: Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. Dev. Sci. *9:* F13–F21 (2006).

Kuhl, P.K.; Williams, K.A.; Lacerda, F.; Stevens, K.N.; Lindblom, B.: Linguistic experience alters phonetic perception in infants by 6 months of age. Science *255:* 606–608 (1992).

Lee, Y.-S.; Vakoch, D.A.; Wurm, L.H.: Tone perception in Cantonese and Mandarin: a cross-linguistic comparison. J. psycholing. Res. *25:* 527–542 (1996).

Lively, S.; Logan, J.; Pisoni, D.: Training Japanese listeners to identify English /r/ and /l/. II. The role of phonetic environment and talker variability in new perceptual categories. J. acoust. Soc. Am. *94:* 1242–1255 (1993).

Lively, S.; Pisoni, D.B.; Yamada, R.A.; Tohkura, Y.I.; Yamada, T.: Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories. J. acoust. Soc. Am. *96:* 2076–2087 (1994).

Logan, J.; Lively, S.; Pisoni, D.: Training Japanese listeners to identify English /r/ and /l/: a first report. J. acoust. Soc. Am. *89:* 874–886 (1991).

MacKain, K.S.; Best, C.T.; Strange, W.: Categorical perception of English /r/ and /l/ by Japanese bilinguals. Appl. Psycholing. *2:* 369–390 (1981).

Macmillan, N.A.; Creelman, C.D.: Detection theory: a user's guide; 2nd ed. (Erlbaum, Mahwah 2005).

Maran, L.: On becoming a tone language: a Tibeto-Burman model of tonogenesis; in Hyman, Consonant types and tone. Southern Calif. Occas. Papers Ling., No. 1, pp. 97–114 (1973).

Maspero, H.: Etudes sur la phonétique historique de la langue annamite: les initiales. Bull. Ecole fr. extrême Orient 12 (1912).

Matisoff, J.A.: Tonogenesis in Southeast Asia; in Hyman, Consonant types and tone. Southern Calif. Occas. Papers Ling., No. 1, pp. 71–95 (1973).

McCarthy, J.; Prince, A.: Prosodic morphology: constraint interaction and satisfaction. Rutgers Univ. Center Cogn. Sci. Techn. Rep. No. 3 (1993).

McGuire, G.L.: Phonetic category learning; PhD diss. Ohio State University (2007).

Mei, T.-L.: Tones and tone sandhi in 16th century Mandarin. J. Chinese Ling. *1977:* 237–260.

Mielke, J.: The interplay of speech perception and phonology: experimental evidence from Turkish. Phonetica *60:* 208–229 (2003).

Miyawaki, K.; Strange, W.; Verbrugge, R.; Lieberman, A.M.; Jenkens, J.J.; Fujimura, O.: An effect of linguistic experience: the discrimination of [r] and [l] by native speakers of Japanese and English. Perception Psychophysics *18:* 331–340 (1975).

Norman, J.: Tonal development in Min. J. Chinese Ling. *1:* 222–238 (1973).

Ohala, J.: The listener as a source of sound change; in Masek, Hendrik, Miller, Papers from the parasession on language and behavior: CLS, pp. 178–203 (Chicago Linguistics Society, Chicago 1981).

Ohala, J.J.: The phonetics of sound change; in Jones, Historical linguistics: problems and perspectives, pp. 237–278 (Longman, London 1993).

Pisoni, D.: Auditory and phonetic memory codes in the discrimination of consonants and vowels. Perception Psychophysics *13:* 253–260 (1973).

Qian, Z.: Yantai Fangyan Baogao (Report on the Yantai dialect, in Chinese). (Qilu Shushe, Jinan 1982).

Rivera-Gaxiola, M.; Silva-Pereyra, J.; Kuhl, P.K.: Brain potentials to native and non-native speech contrasts in 7- and 11-month-old American infants. Dev. Sci. *8:* 162–172 (2005).

Rose, S.; Walker, R.: A typology of consonant agreement as correspondence. Language *80:* 475–531 (2004).

Shepard, R.N.: The circumplex and related topological manifolds in the study of perception; in Shye, Theory construction and data analysis in the social sciences (Jossey-Bass, San Francisco 1978).

Seo, M.: A perception-based study of sonorant assimilation in Korean; in Hume, Johnson, Studies on the interplay of speech perception and phonology. Ohio State Univ. Working Papers Ling., No. 55, pp. 43–69 (2001).

Shen, X.S.; Lin, M.: A perceptual study of Mandarin Tones 2 and 3. Lang. Speech *34:* 145–156 (1991).

Stagray, J.R.; Downs, D.: Differential sensitivity for frequency among speakers of a tone and a nontone language. J. Chinese Ling. *21:* 143–163 (1993).

Steriade, D.: A perceptual account of directional asymmetries in assimilation and cluster reduction; in Hume, Johnson, The role of perception in phonology (Academic Press, New York 2001).

Stiles, J.: Neural plasticity and cognitive development. Dev. Neuropsychol. *18:* 237–272 (2000).

Strange, W.; Dittmann, S.: Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. Perception Psychophysics *36:* 131–145 (1984).

Svantesson, J.-O.: Tonogenesis in Southeast Asia – Mon-Khmer and beyond; in Shigeki Kaji, Proc. Symp.: Cross-Linguistic Studies of Tonal Phenomena, Tonogenesis, Japanese Accentology, and Other Topics, ILCAA, Tokyo University of Foreign Studies, 2001, pp. 45–58.

Trubetzkoy, N.S.: Principles of phonology (University of California Press, Berkeley 1969). Translated by C. Baltaxe from Trubetzkoy, N.S.: Grundzüge der Phonologie. Travaux du Cercle Linguistique de Prague 7 (Prague 1939).

Tsao, F.-M.; Liu, H.-M.; Kuhl, P.K.: Perception of native and non-native affricate-fricative contrasts: cross-language tests on adults and infants. J. acoust. Soc. Am. *120:* 2285–2294 (2006).

Tserdanelis, G.: A perceptual account of manner dissimilation in Greek; in Hume, Johnson, Studies on the interplay of speech perception and phonology. Ohio State Univ. Working Papers Ling., No. 55, pp. 172–199 (2001).

Wang, S.-Y.W.: Language change. Ann. N.Y. Acad. Sci. *280:* 61–72 (1976).

Watson, C.S.; Kelly, W.J.; Wroton, H.W.: Factors in discrimination of tonal patterns. II. Selective attention and learning under various levels of stimulus uncertainty. J. acoust. Soc. Am. *60:* 1176–1186 (1976).

Werker, J.F.; Tees, R.C.:Phonemic and phonetic factors in adult cross-language speech perception. J. acoust. Soc. Am. *75:* 1866–1878 (1984).

Werker, J.F.; Logan, J.S.: Cross-language evidence for three factors in speech perception. Perception Psychophysics *37:* 35–44 (1985).

Xu, Y.: Fundamental frequency peak delay in Mandarin. Phonetica *58:* 26–52 (2001).

Yamada, R.; Tohkura, Y.; Kobayashi, N.: Effect of word familiarity on nonnative phoneme perception: identification of English /r/, /l/, and /w/ by native speakers of Japanese; in James, Leather, Second language speech, pp. 103–117 (Mouton de Gruyter, The Hague 1992).

Zhang, Y.; Kuhl, P.K.; Imada, T.; Iverson, P.; Pruitt, J.; Stevens, E.B.; Kawakatsu, M.; Tohkura, Y.; Nemoto, I.: Neural signatures of phonetic learning in adulthood: a magnetoencephalography study. Neuroimage *46:* 226–240 (2009).