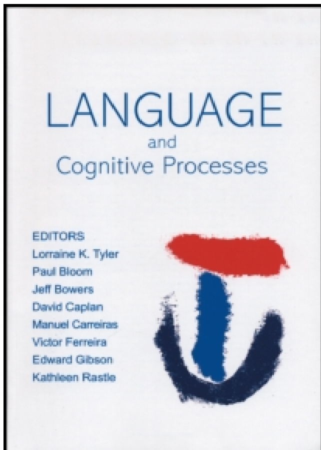

Language and Cognitive Processes
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713683153

Processing dependencies between segmental and suprasegmental features in Mandarin Chinese

Yunxia Tong [ab]; Alexander L. Francis [a]; Jackson T. Gandour [a]
[a] Department of Speech Language Hearing Sciences, Purdue University, West Lafayette, IN, USA
[b] National Institute of Mental Health, National Institutes of Health, Bethesda, MD, USA

First Published on: 01 December 2007

To cite this Article: Tong, Yunxia, Francis, Alexander L. and Gandour, Jackson T. (2007) 'Processing dependencies between segmental and suprasegmental features in Mandarin Chinese', Language and Cognitive Processes, 23:5, 689 — 708

To link to this article: DOI: 10.1080/01690960701728261
URL: http://dx.doi.org/10.1080/01690960701728261

LANGUAGE AND COGNITIVE PROCESSES
2008, 23 (5), 689–708

Psychology Press
Taylor & Francis Group

Processing dependencies between segmental and suprasegmental features in Mandarin Chinese

Yunxia Tong
Department of Speech Language Hearing Sciences, Purdue University, West Lafayette, IN, and National Institute of Mental Health, National Institutes of Health, Bethesda, MD, USA

Alexander L. Francis and Jackson T. Gandour
Department of Speech Language Hearing Sciences, Purdue University, West Lafayette, IN, USA

The aim of this study was to examine processing interactions between segmental (consonant, vowel) and suprasegmental (tone) dimensions of Mandarin Chinese. Using a speeded classification paradigm, processing interactions were examined between each pair of dimensions. Listeners were asked to attend to one dimension while ignoring the variation along another. Asymmetric interference effects were observed between segmental and suprasegmental dimensions, with segmental dimensions interfering more with tone classification than the reverse. Among the three dimensions, vowels exerted greater interference on consonants and tones than vice versa. Comparisons between each pair of dimensions revealed greater integrality between tone and vowel than between tone and consonant. Findings suggest that the direction and degree of interference between segmental and suprasegmental dimensions in spoken word recognition reflect differences in acoustic properties as well as other factors of an informational nature.

Correspondence should be addressed to Jackson T. Gandour, Purdue University, Department of Speech Language Hearing Sciences, 1353 Heavilon Hall, 500 Oval Drive, West Lafayette, IN 47907-2038, USA. E-mail: gandour@purdue.edu

Contract grant sponsor: National Institutes of Health; Contract grant number: R01 DC04584-05 (JG). Thanks to four anonymous reviewers and the Associate Editor for their insightful comments and suggestions on earlier versions of the manuscript and Yisheng Xu for his assistance with statistical analysis.

http://www.psypress.com/lcp DOI: 10.1080/01690960701728261

INTRODUCTION

The speech signal simultaneously carries information about phonetic segments (i.e., consonants and vowels) and suprasegmental features (e.g., pitch). Of importance for theories of speech perception is the manner in which the two types of information are extracted from the acoustic waveform during processing. Dependencies between segmental and suprasegmental dimensions have been estimated by using the Garner speeded classification paradigm (Carrell, Smith, & Pisoni, 1981; Lee & Nusbaum, 1993; Miller, 1978; Repp & Lin, 1990; Tomiak, Mullennix, & Sawusch, 1987; Wood, 1974, 1975). The paradigm includes two prototypical tasks (Garner, 1974, 1976; Garner & Felfoldy, 1970). In the baseline task, subjects are asked to make speeded classification judgements on the basis of the target dimension (e.g., /b/ and /g/) while the nontarget dimension is held constant (e.g., 100 Hz voice fundamental frequency (F_0)). In the orthogonal or filtering task, the task is the same except that the values along the irrelevant dimension vary (e.g., 100 and 200 Hz) on a trial-to-trial basis. Any decrement in performance in the orthogonal task as compared to the baseline task must result from the variability in the irrelevant dimension. This difference in performance between filtering and baseline tasks is called Garner interference.

Integrality between dimensions

Converging experimental results suggest that there are two major types of dimensional pairings: integral and separable. Dimensions are considered separable if the target feature can be identified with the same speed and accuracy in both baseline and orthogonal tasks. If the dimensions are processed integrally, irrelevant trial-to-trial variability in the nontarget dimension leads to poorer target performance in the orthogonal than in the baseline condition. Garner interference can occur at the level of perceptual feature processing, semantic processing, or response selection. It arises when the processing of a feature on one dimension is influenced by the context (i.e., trial-to-trial variability) of a feature in the other dimension (Melara & Marks, 1990; Melara, Marks, & Potts, 1993). Similar to the interference effect in the Stroop paradigm (Stroop, 1935), the additional processing demand or conflict in the Garner paradigm comes from intrinsic properties of the stimulus instead of an externally imposed (secondary) task, as in a dual-task paradigm, or from a conflict in spatial locations between a stimulus and response, as in the Simon paradigm (Simon, Craft, & Webster, 1973). However, in the Garner paradigm, unlike the Stroop paradigm, the conflict comes from between-trial variability in the irrelevant dimension (Melara & Marks, 1990) rather than from within-trial conflict (or congruency) between the two dimensions that leads to competing responses in the Stroop paradigm (MacLeod, 1991).

The term integrality refers to the degree to which two separate dimensions are dependent upon one another. Of relevance to this study is the observation that dimensional dependencies may be asymmetric. With certain pairs of dimensions, the interference effect is observed only with respect to one of the dimensions. Garner defined this pattern as reflecting a distinct type of dimensional interaction: asymmetric separable or, alternatively, asymmetric integral (Garner, 1974, 1976). For example, Wood (1974) tested two speech dimensions in speeded classification: F_0 of the vowel phoneme (high or low) and place of articulation of the consonant phoneme (/b/ or /g/) of a consonant-vowel (CV) syllable (e.g., /bæ/). When pitch was the target dimension, there was no interference from orthogonal variation of consonant. However, when the consonant was the target dimension, substantial interference came from the irrelevant variation in the nontarget pitch dimension. This experiment demonstrated that separability is not necessarily symmetrical. One dimension (e.g., place of articulation of consonants) may be separable from another (e.g., pitch) even when the converse is not true. On the other hand, some dimensions are symmetrically separable. For example, a mutual dependency was found between vowel quality and pitch (Carrell et al., 1981; Miller, 1978). Thus, how dimensions interact with one another appears to depend on the type of information being analysed. In the aforementioned studies, pitch information was never linguistically relevant. Yet integrality between dimensions can depend crucially on a listener's knowledge of their linguistic function (Tomiak et al., 1987). The question then arises as to whether interactions between segmental and suprasegmental information may vary depending on the linguistic function of pitch.

Chinese tones

In Chinese (Mandarin), suprasegmental (tone) as well as segmental (consonant, vowel) information is lexically relevant. Chinese has four lexical tones: e.g., ma^1 'mother', ma^2 'hemp', ma^3 'horse', ma^4 'scold' (Chao, 1968). Tones 1 to 4 can be described phonetically as high level [55], high rising [35], low falling rising [214], and high falling [51], respectively (Howie, 1976). Perceptual data on Chinese tones indicate that variation in F_0 yields high levels of recognition for isolated tones (Howie, 1976; Tseng, 1990). Other acoustic cues include amplitude (Whalen & Xu, 1992) and duration (Fu, Zeng, Shannon, & Soli, 1998; Xu, Tsai, & Pfingst, 2002).

Previous Garner interference studies of Chinese tones

Earlier crosslanguage, speeded classification studies of Chinese suggest that both Chinese and English listeners show mutual integrality between consonantal and tonal information (Lee & Nusbaum, 1993; Repp & Lin, 1990). However, only Chinese listeners display an asymmetric pattern of interference between vowels and tones (Repp & Lin, 1990). Vowels have a greater interference effect on tones than vice versa. Such data suggest that the nature of dimensional interactions involving suprasegmental information does not depend simply on the acoustic characteristics of the stimuli but also on its phonological status. Lee and Nusbaum (1993) claimed a symmetric integration between consonants and tones. However, they did not perform a direct comparison between the interference effects.

These studies have examined interactions between two pairs of dimensions only: tone vs. consonant and tone vs. vowel. But a Mandarin syllable consists of *three* components that may interact with one another: onset (consonant), rime (vowel), and tone. By investigating all three dimension pairs concurrently, we are able to compare processing interactions not only between segmental and suprasegmental dimensions (tone vs. consonant; tone vs. vowel), but also between two segmental dimensions (consonant vs. vowel).

It is important to compare all three dimension pairs within a single experiment in order to clarify the roles of acoustic cues and linguistic properties in determining dimensional asymmetry. Consonants and vowels differ in terms of their relative duration and the order in which their information unfolds over the course of a syllable. Yet both are segmental properties. Vowels and tones, on the other hand, are similar in duration and position in the syllable. But vowels are segmental; tones are suprasegmental. Acoustically, consonants and tones must both vary over time, whereas vowels (e.g., monophthongs) may remain comparatively static. Thus, a finding of asymmetric integrality suggesting that vowels are processed differently from consonants and tones could possibly be attributed to the relatively static nature of their core acoustic properties. Faster events are more vulnerable to interference, demanding greater commitment of processing resources; conversely, slower/longer events are more resistant to interference. On the other hand, a finding of asymmetric integrality between consonants and vowels vs. tones would seem to support a role for linguistic status. Perhaps more *informative* segmental features demand greater processing commitment than less informative suprasegmental properties.

A potential confound in speeded classification tasks is that dimensional asymmetry can simply be a result of the relative discriminability between the two dimensions (Carrell et al., 1981; Lee & Nusbaum, 1993). If the values along one dimension are more discriminable than the other, small changes in the irrelevant dimension will not be detected. As a result, classification along the less discriminable dimension will be greatly affected by variation along the more discriminable dimension. To investigate dimensional interactions, it is therefore crucially important to select values from the dimensions that are approximately equal in discriminability. In previous speeded classification tasks involving lexical tone (Lee & Nusbaum, 1993; Repp & Lin, 1990), crosslanguage comparisons of the dimensional interactions were not veridical because of unequal discriminability of the tonal dimension for English listeners.

Present study

First, we conducted a preliminary experiment (not reported herein) to evaluate the relative discriminability of the consonant, vowel, and tone dimensions in Chinese. The stimulus pairs we chose for the three dimensions were found to be equal in discriminability for native Chinese listeners.

Using a two-choice speeded classification task (Garner, 1974), our aim in this paper was to map out interactions between segmental and suprasegmental dimensions of speech in a tone language (Mandarin Chinese). We assessed the degree of integrality or separability between two kinds of segmental dimensions (consonant: place of articulation; vowel: height + rounding) and a suprasegmental dimension (tone: direction of pitch contour). Subjects were required to classify stimuli on the target or relevant dimension, while trying to ignore the nontarget or irrelevant dimension. Each task was comprised of two basic conditions: *baseline* or *control*, in which the irrelevant dimension was held constant; and *orthogonal*, in which the relevant and irrelevant dimensions are varied orthogonally. If the processing of the target dimension entails processing of the nontarget dimension, listeners are expected to have difficulty selectively attending to only one dimension, and their performance is expected to worsen in the orthogonal condition. Such interference effects allow us to assess the degree of integrality between segmental and suprasegmental dimensions and to examine whether the level of integrality is the same (symmetric) or different (asymmetric) between dimensions.

By examining processing interactions among consonants, vowels, and tones, we were able to draw inferences about the nature and extent of their interactions with attention and memory processes. Moreover, we were able to reconcile contradictory results from earlier studies, and provide fresh insights into processing interactions among Mandarin consonants, vowels, and tones that carry implications relevant to speech perception in general.

METHODS

Subjects

Participants were 20 native speakers of Mandarin Chinese from mainland China (9 male; 11 female). Their mean age and level of education were 28 and 19 years, respectively. All subjects were enrolled at Purdue University at the time of being tested. All subjects exhibited normal hearing sensitivity (pure-tone air conduction thresholds of 20 dB HL or better in both ears at frequencies of 0.5, 1, 2, and 4 kHz). All subjects gave informed consent in compliance with a protocol approved by the Institutional Review Board of Purdue University. None of them had participated in the preliminary discrimination experiment.

Stimuli

Eight syllables were generated using a Klatt cascade/parallel formant synthesiser (Klatt, 1980; Klatt & Klatt, 1990). Each syllable constituted a real Chinese word consisting of a consonant (/b/ or /d/), vowel (/a/ or /u/), and tone (2 or 4): /ba^2/'pull out', /ba^4/'father', /da^2/'reach', /da^4/'big', /bu^2/ 'mold', /bu^4/'no', /du^2/'read', and /du^4/'stomach'. The stimulus set was balanced for consonant (2), vowel (2), and tone (2). The consonant pair /b/ vs. /d/ differed in consonant place of articulation, manifested primarily by differences in the second formant (F_2) transition. The same pair was used by Lee and Nusbaum (1993) and Repp and Lin (1990). The vowel pair /a/ vs. /u/ differed in vowel height and rounding. The steady-state portion of this vowel quality contrast was signalled principally by differences in F_1 and F_2. The same vowel pair was used by Repp and Lin (1990). The tone pair 2 vs. 4 provided a contrast between rising and falling F_0 contours (Figure 1). This tonal pair was optimal for tonal comparisons because of the relatively high dissimilarity between T2 (rising) and T4 (falling) as reflected by their positions in a multidimensional perceptual space (Gandour, 1983; Huang,

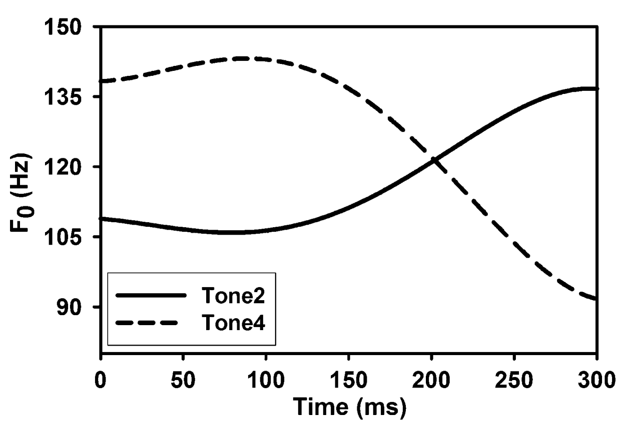

Figure 1. Voice fundamental frequency contours (F_0) superimposed on synthetic Chinese syllables with Tone 2 (solid line) and Tone 4 (dashed line). All syllables were 300 ms in duration.

2004; Hume & Johnson, 2001), coupled with the fact that these two tones are not related by tone sandhi (Chao, 1968).

The synthesis parameters for the consonant and vowel of /ba^2/ vs. /ba^4/ syllables were identical except for their tonal contours. Similarly, /da^2/ vs. /da^4/, /bu^2/ vs. /bu^4/, /du^2/ vs. /du^4/ differed only in their F_0 contours. All stimuli were 300 ms in duration (Lee & Nusbaum, 1993; Repp & Lin, 1990). Tones 2 and 4 were modelled after the mean F_0 contours from eight male speakers producing /ma/ syllables in isolation (Xu, 1997), and normalised to 300 ms (Figure 1). Tone 2 began at 109 Hz and rose curvilinearly to 137 Hz, while Tone 4 began at 138 Hz and fell curvilinearly to 92 Hz (see Figure 1). Synthesis parameters for the vowels /a/ and /u/ were based on formant values from Howie (1976). The amplitude of each syllable was ramped up from 40 to 60 dB in the first 10 ms of the stimulus, and gradually dropped to 55 dB in the next 120 ms, then fell to 45 dB at the end of the syllable. For /ba/ syllables, the starting and steady-state frequencies for the first three formants (F_1, F_2, and F_3) were 500 and 800 Hz, 900 and 1150 Hz, and 2200 and 2500 Hz, respectively. The duration of the formant transition was 30 ms. For /da/ syllables, the starting and steady-state frequencies, respectively, were 650 and 800 Hz for F_1, 1400 and 1150 Hz for F_2, and 2800 and 2500 Hz for F_3. The duration of the formant transition was 80 ms. For /bu/ syllables, F_1 was held constant at 350 Hz. The starting and steady-state frequencies were 475 and 650 Hz for F_2, and 2200 and 2500 Hz for F_3. The duration of the formant transition was 40 ms. For /du/ syllables, F_1 was held constant at 350 Hz. The starting and steady-state frequencies were 1150 and 650 Hz for F_2, and 2800 and 2500 Hz for F_3. The duration of the formant transition was 80 ms.

To ensure the intelligibility of the synthetic speech stimuli, an 8-alternative forced choice labelling paradigm was conducted with five Mandarin Chinese listeners. Results showed 100% correct identification for all eight stimuli, and no listener reported that the synthetic syllables sounded unnatural.

All speech stimuli were digitally normalised to the same peak intensity. They were presented binaurally at 75 dB SPL by means of a computer playback system (E-Prime 1.1) through a pair of headphones (SONY MDR-7506).

For each target dimension, stimuli were grouped into four sets for the control condition and two sets for each of the orthogonal conditions (Table 1). In each control condition, subjects heard two stimuli in which the values of the target dimension varied and the values of the other two dimensions were fixed. In each orthogonal condition, subjects heard four stimuli in which the values on the nontarget dimension as well as the target dimension were varied. The same stimuli were included in both a control condition and its corresponding orthogonal condition. Thus, each stimulus served as its own control across conditions, making it possible to evaluate interference effects of the orthogonal condition relative to its control.

TABLE 1
Dimensions, conditions, and stimulus sets

Dimension			*Stimulus set*			
Target	*Nontarget*	*Condition*	*1*	*2*	*3*	*4*
Tone	Consonant	Orthogonal (T/C)	/ba^2/ /ba^4/ /da^2/ /da^4/	/bu^2/ /bu^4/ /du^2/ /du^4/		
	Vowel	Orthogonal (T/V)	/ba^2/ /ba^4/ /bu^2/ /bu^4/	/da^2/ /da^4/ /du^2/ /du^4/		
	—	Control (T/B)	/ba^2/ /ba^4/	/bu^2/ /bu^4/	/da^2/ /da^4/	/du^2/ /du^4/
Consonant	Tone	Orthogonal (C/T)	/ba^2/ /da^2/ /ba^4/ /da^4/	/bu^2/ /du^2/ /bu^4/ /du^4/		
	Vowel	Orthogonal (C/V)	/ba^2/ /da^2/ /bu^2/ /du^2/	/ba^4/ /da^4/ /bu^4/ /du^4/		
	—	Control (C/B)	/ba^2/ /da^2/	/bu^2/ /du^2/	/ba^4/ /da^4/	/bu^4/ /du^4/
Vowel	Tone	Orthogonal (V/T)	/ba^2/ /bu^2/ /ba^4/ /bu^4/	/da^2/ /du^2/ /da^4/ /du^4/		
	Consonant	Orthogonal (V/C)	/ba^2/ /bu^2/ /da^2/ /du^2/	/ba^4/ /bu^4/ /da^4/ /du^4/		
	—	Control (V/B)	/ba^2/ /bu^2/	/da^2/ /du^2/	/ba^4/ /bu^4/	/da^4/ /du^4/

Procedure

The experimental paradigm consisted of three speeded classification tasks. In each task, subjects were instructed to classify the syllables according to the target dimension (tone, consonant, vowel) and to press the appropriate response button as quickly as possible. The assignment of responses to buttons was counterbalanced across subjects. There were three conditions per task (B = baseline or control): one *control* and two orthogonal conditions. The control conditions for the tone, consonant, and vowel tasks, in order, were T/B, C/B, and V/B. The *orthogonal* conditions (target/nontarget dimensions) were T/C & T/V for the tone task, C/T & C/V for the consonant task, and V/C & V/T for the vowel task (Table 1).

Each subject participated in six runs, two each for the C, V, and T tasks. The order of runs was randomised for each subject. Each run lasted about 7.5 minutes, containing six randomly presented blocks (59.4 s) with intervening 10.6 s rest intervals (Figure 2). The six blocks within each run included all three conditions of one task, with two blocks for each condition. Given two runs per task, there was a total of four blocks per condition. Each block contained one set of the stimuli listed in Table 1. For the control condition, each set of stimuli appeared once. Because there were only two sets of stimuli for each orthogonal condition, each set was repeated twice in the experiment.

There were 33 trials in each block. Trial duration was 1.8 s including the stimulus (300 ms) and response interval (1500 ms). Instructions were delivered to subjects via headphones immediately preceding each run for their respective classification task.

Data analysis

In a pilot study, it was observed that subjects' RTs were always longer during the first few trials of a block. Therefore, the first three trials in each block were treated as practice, and excluded from analysis. In this way, variance across subjects, tasks, and conditions was stabilised.

Response accuracy and mean RT in correct response trials were calculated for each subject, task, and condition. The interference effect was evaluated as the difference in RT between each orthogonal condition and the corresponding control condition (e.g., T/V – T/B). To confirm the existence of the

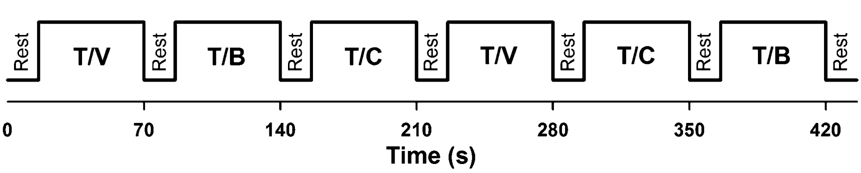

Figure 2. A schematic illustration of the sequence and timing of a run used for the tone task. The six blocks include all three conditions of the tone task, two blocks per condition presented in random order. T/B = control condition: tone target/baseline (i.e., C and V fixed); T/C = orthogonal condition: tone target/consonant nontarget; T/V = orthogonal condition: tone target/vowel nontarget; Rest = silent interval between blocks. T = tone; C = consonant; V = vowel; B = baseline. Similar sequences were used for the consonant and vowel tasks.

interference effect, RT in each orthogonal condition was compared with the control condition by using a paired *t* test.

Integrality was measured as the strength of mutual interference effects between a pair of dimensions. Two separate analyses were performed to examine the integrality between dimensions and the asymmetry within each pair of dimensions. Each pair corresponds to two conditions; e.g., the degree of integrality between T and C is derived from the interference effects in T/C as compared with C/T. To compare the integrality *between* dimension pairs T-C, T-V, and C-V, RT was analysed using a repeated measures ANOVA that included a within-subjects factor Condition (T/C, C/T, T/V, V/T, C/V, V/C) nested within a fixed factor Dimension Pair (T-C, T-V, C-V). To compare asymmetry *within* dimension pairs, RT was examined by using paired *t* tests.

In addition, RT and accuracy in the baseline condition were analysed using a one-way ANOVA to confirm that all three dimensions were equal in classification.

RESULTS

Comparison of the baseline conditions

A one-way ANOVA showed that the three dimensions (C, V, T) did not differ significantly in classification for Chinese listeners, d': $F(2, 56) = 0.42$ n.s.; RT: $F(2, 56) = 0.72$ n.s. (Figure 3), indicating equal sensitivity along all three dimensions of contrast.

Existence of interference effects

For each orthogonal condition (T/C, C/T, T/V, V/T, C/V, V/C), RT was significantly longer than its corresponding control condition, T/C > T/B: $t(19) = 8.38$, $p < .001$; T/V > T/B: $t(19) = 14.59$, $p < .001$; C/T > C/B: $t(19) = 8.44$, $p < .001$; C/V > C/B: $t(19) = 13.05$, $p < .001$; V/T > V/B: $t(19) = 7.43$, $p < .001$; V/C > V/B: $t(19) = 10.57$, $p < .001$, indicating the presence of interference effects regardless of dimension.

Integrality between dimensions

ANOVA results revealed a significant main effect of Dimension Pair (T-C, T-V, C-V) on RT, $F(2, 113) = 4.53$, $p < .05$ (Figure 4). Post hoc Tukey–Kramer adjusted comparisons revealed that the integrality between C-V and T-V were almost equal, $t(113) = 0.27$ *ns*, whereas the integrality between T-C was not as strong as the other two, T-C vs. C-V: $t(113) = -2.73$, $p < .05$; T-C vs. T-V: $t(113) = -2.46$, $p < .05$.

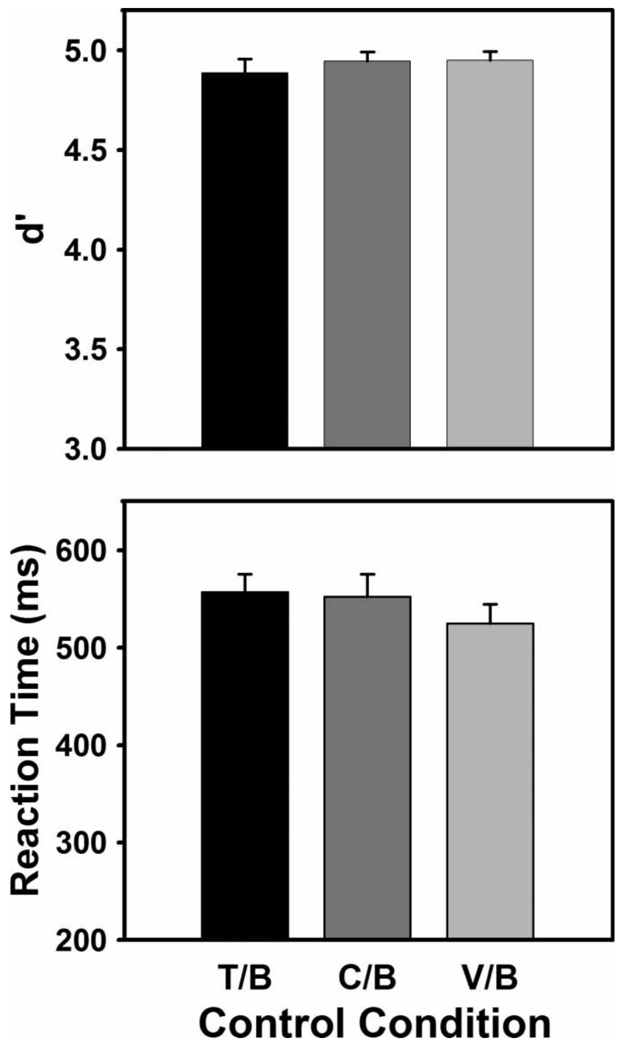

Figure 3. Average *d′* (top panel) and RT (bottom panel) from the three classification tasks in their respective control conditions. Both response accuracy and speed are comparable for Chinese listeners' classification judgements of tones, consonants, and vowels. T/B = control condition, tone target/baseline (C & V fixed); C/B = control condition, consonant target/baseline (T & V fixed); V/B = control condition, vowel target/baseline (T & C fixed). Error bars represent ± 1 *SE*.

Dimensional asymmetry

As indexed by RT, paired *t* tests of the interference effects showed significant asymmetry within each pair of dimensions (T/C vs. C/T; T/V vs. V/T; C/V vs.

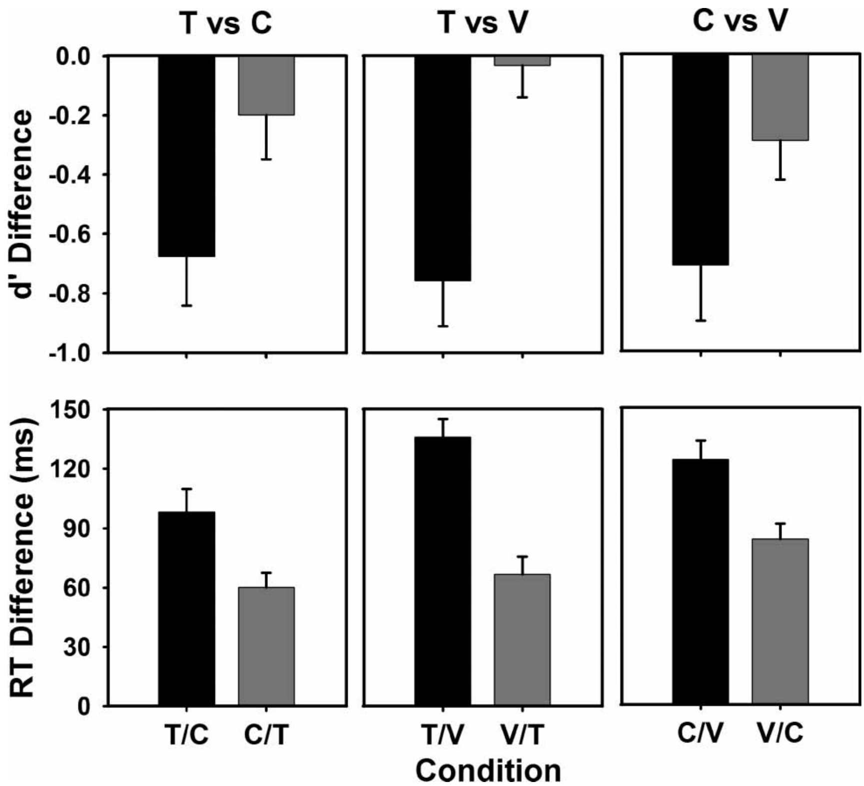

Figure 4. Interference effects from the classification tasks in the six orthogonal conditions, grouped into three dimension pairs: tone vs. consonant, tone vs. vowel, and consonant vs. vowel. These effects were measured by taking the difference in RT between each orthogonal and its corresponding control condition (e.g., the interference effect in the T/C condition was calculated as the difference in RT between T/C and T/B). T/C = tone target/consonant nontarget; C/T = consonant target/tone nontarget; T/V = tone target/vowel nontarget; V/T = vowel target/tone nontarget; C/V = consonant target/vowel nontarget; V/C = vowel target/consonant nontarget. Asymmetric integrality was evident in each pair of dimensions. In T vs. C and T vs. V, respectively, consonants and vowels interfere more with the processing of tones than vice versa. In C vs. V, vowels interfere more with the processing of consonants than vice versa. T = tone; C = consonant; V = vowel; B = baseline. Error bars represent ± 1 *SE*.

V/C) (Figure 4). In a comparison of segmental and tonal dimensions, C had a greater interference effect on T than vice versa, T/C > C/T; $t(19) = 3.33$, $p < .005$. Similarly, V had a greater interference effect on T than vice versa, T/V > V/T; $t(19) = 6.14$, $p < .001$. In a comparison of two segmental dimensions, V had a greater interference effect on C than vice versa, C/V > V/C; $t(19) = 4.08$, $p < .005$.

Similarly, as indexed by accuracy (d'), paired t tests of the interference effects showed significant asymmetry within each pair of dimensions, T/C >

C/T: $t(19) = 2.34$, $p < .05$; T/V > V/T: $t(19) = 4.38$, $p = .001$; C/V > V/C: $t(19) = 2.07$, $p = .052$.

DISCUSSION

The results of the Garner interference tasks demonstrate the presence of interference effects across the board (T/C, T/V, C/T, C/V, V/T, V/C) in Mandarin Chinese regardless of the type of dimension (segmental, suprasegmental) or the dimension that is selectively attended to (target, nontarget). The integrality between dimension pairs T-V and C-V are greater than C-T. Moreover, all interactions between the three paired dimensions are asymmetric. Vowels are most resistant to interference effects. Variations in the vowel interfere with classifications by tone (T/V > V/T) or consonant (C/V > V/C) more than vice versa. Tones are the most vulnerable to interference effects. Variations in either segmental dimension (T/V > V/T; T/C > C/T) exert a greater interference on tonal classifications than vice versa. Such asymmetric interactions are not likely due to differences in relative discriminability between the paired dimensions (Carrell et al., 1981) as revealed in both the discrimination (preliminary experiment) and identification (baseline condition) tasks.

Relative discriminability across dimensions

Given the results from the preliminary experiment for the group of Chinese listeners, we can be reasonably confident that there are no significant differences in discriminability between the chosen values in the present stimuli either within or across dimensions (i.e., consonant: /b/ vs. /d/; vowel: /a/ vs. /u/; tone: /2/ vs. /4/). The results of the baseline condition in this Garner paradigm corroborate those of the preliminary discrimination experiment, showing similar RTs across all three baseline conditions (C/B, V/B, T/B). Thus, we may assume that in the absence of interference from other dimensions, Chinese listeners found tones, consonants, and vowels to be equally discriminable (Melara & Mounts, 1994), and moreover, that any dimensional asymmetry observed in the orthogonal conditions cannot be attributed to inherent differences in discriminability among the stimuli.

In this study, the asymmetric integrality observed between vowels and tones does not agree with the conclusion drawn by Repp and Lin (1990). Variation in vowels increased processing time for classifying tones as compared with the effects of tonal variation on processing time for classifying vowels. Although Repp and Lin reported a similar result, they observed longer RTs for tone than for vowel classification in the *baseline* condition. As noted by the authors themselves (p. 493), such differences in inherent (baseline) discriminability between tones and vowels may have confounded their determination of any processing interaction between the two dimensions. Given that patterns of perceptual integrality may change as the relative discriminability of dimensions is varied (Carrell et al., 1981; Garner, 1983), it is important to ensure that dimensions are matched for discriminability *first*. Any post hoc correction of RTs is statistically invalid.

Asymmetric integrality between dimensions

A major finding of the present study is that variation in segmental dimensions generates greater interference to a suprasegmental dimension than the reverse. One hypothesis is that tonal processing requires more time than segmental processing (Cutler & Chen, 1997). In a speeded response discrimination task, Cantonese listeners were faster and more accurate in their responses when the only difference between two syllables was segmental (consonant, vowel) as compared with suprasegmental (tone). Similarly, in a syllable-monitoring task, Mandarin listeners responded faster to mismatches in vowels as compared with tones when target syllables were presented in isolation (Ye & Connine, 1999). In this study, asymmetric integrality cannot be due to longer processing time for tones because in the baseline conditions, RTs and d' are comparable across the three dimensions. Nor is it likely that the differences in the temporal unfolding of acoustic cues for identifying tones relative to vowels cause the observed disparity in processing time. The high falling (T4) and low rising (T2) tones already differ by about 30 Hz at syllable onset. T2 and T4 are well identified (75% correct) from brief 100 ms segments extracted from tonal onset to offset (Whalen & Xu, 1992). Only 6–8 cycles of F_0 are required to identify pitch (Robinson & Patterson, 1995). At the onset of these stimuli, 6–8 cycles of F_0 occur within 56–74 ms (T2) and 43–58 ms (T4), roughly the same as the duration of the formant transitions that can cue both consonant place of articulation (Walley & Carrell, 1983) and vowel quality (Strange, 1989).

The orthogonal condition of the Garner paradigm is essentially a selective attention task. In selective attention, the degree of excitation of the target and the inhibition of distractor are determined by the *salience* of the dimensions (Melara & Algom, 2003). Salience refers to how much attention a stimulus or a dimension can capture in a certain task. Two dimensions compete for attention and greater salience of one dimension is achieved at the cost of weaker salience of the other. As a consequence, the greater the salience along the distractor dimension, the more difficult it is for observers to focus selectively on the target dimension. We argue that the imbalance in salience among dimensions is caused by at least two factors, namely, their *information value* and *acoustic properties.*

Information value. Higher-level linguistic factors have been shown to influence spoken word processing. Information-theoretic methods for quantifying human communication date back to the early 1950s (see Garner, 1988, for a review). According to such approaches, the informational value of a given signal can be quantified in terms of its probability of occurring in a communication system. The higher the probability is, the lower the information value.

Tsai (2000) is a database of 13,060 Mandarin characters which, in total, can be pronounced with 1,251 unique combinations of onset, rime, and tone, i.e., syllable + tone combinations. In order to determine how much information each subpart contributes, we can compute the size of the equivalence class that results from specifying only one subpart (Altman, 1990). For example, the equivalence class size (ECS) for the onset /p/ is defined as the number of syllable + tone combinations that begin with /p/. Of the 1251 syllable + tone combinations in the database, 56 (4.5%) begin with /p/, meaning that the ECS for /p/ onsets is 56. Simply by knowing that a syllable + tone combination begins with /p/, a listener has a 1/56 (4.5%) chance of identifying that particular combination correctly as compared to the base chance probability (1/1251). Thus, ECS provides an index of the relative informational value of a given subpart. The larger the ECS value, the less informative it is. By computing the ECS for each onset (rime, tone) separately, we were able to derive an average ECS across all onsets (rimes, tones). The average ECS (*SD* enclosed in parentheses) for onsets, rimes, and tones were, in order, 52.1 (12.7), 36.8 (23.0), and 310 (37.6). A one-way ANOVA of ECS with three levels (onset, rime, tone) revealed a significant main effect, $F(2,59) = 314.05$, $p < .001$. Post-hoc Tukey HSD pairwise comparisons (onset vs. tone; rime vs. onset; rime vs. tone) were all significant ($p < .03$). Since informational value is inversely proportional to ECS, we may infer that rimes are significantly more informative than consonants, and consonants, in turn, are significantly more informative than tones in Mandarin Chinese.

Such differences in informational value may account in part for the asymmetric integrality of segmental dimensions with tones based on the relative weight given to each element (onset, rime, tone) in making a lexical decision. Given the ECS values of onsets (52), rimes (37), and tones (310), it appears that an optimal strategy for recognising Chinese syllables *under conditions of limited resources* would be to focus attention on phonological units that are more informative, only attending to the less informative features if sufficient capacity remained (Lavie & De Fockert, 2005). In the orthogonal conditions herein, when two dimensions compete for attentional resources, priority is given to the more informative dimensions as shaped by listening strategies developed from long-term language experience. Top-down attentional processing can enhance the salience of a feature significantly (Bacon & Egeth, 1994; Buschman & Miller, 2007; Yantis & Egeth, 1999). We infer that the higher priority assigned to vowels relative to consonants and tones, and similarly, consonants relative to tones, leads to the imbalance in dimensional salience, and subsequently contribute to the observed patterns of asymmetric integrality.

Acoustics. The asymmetric interaction between vowel and consonant (C/V > V/C) and vowels and tones (T/V > V/T) can also be accounted for in terms of the relative *degree* of change over time (dynamic vs. static) in key acoustic properties of vowels (steady-state formant frequencies), tones (changing F_0 contours), and consonants (rapidly changing bursts and formant transitions).

Static and dynamic cues have different decay rates in auditory working memory (Pisoni, 1973). Using nonspeech sounds, Mirman, Holt, and McClelland (2004) found that steady-state nonspeech stimuli were discriminated more accurately than those characterised by rapidly changing acoustic cues. They suggested that the perceptual trace of rapidly changing cues decays faster than that of steady-state sounds. Therefore, even though the three dimensions (C, V, T) are matched in perceptual discriminability, the working memory representation of a vowel persists more strongly than that of consonants or even tones. Acoustic feature variation automatically captures attention (Lee & Nusbaum, 1993; Molholm, Martinez, Ritter, Javitt, & Foxe, 2005; Tomiak et al., 1987; Watkins, Dalton, Lavie, & Rees, 2007; Wood, 1974), first signalling the change detectors in the auditory cortex, and then recruiting frontal and parietal regions for attention-switching. The greater the memory residual of an acoustic feature from the previous trial, the more salient the varying feature in the current trial. It is also known that vowels are louder, longer, and more robust to noise masking than consonants (Dorman, Kewley-Port, Bradley, & Turvey, 1977; Horii, 1970). In the orthogonal conditions, C/V and T/V, the trial-to-trial variation in the irrelevant vowel dimension is more noticeable, and consequently, leads to greater interference with the other two dimensions than vice versa. Thus, it appears that even relatively slow-varying acoustic cues (e.g., F_0 contours), as well as rapidly changing acoustic cues (e.g., formant transitions), can give rise to asymmetric patterns of interference when compared to more static spectral cues (e.g., steady-state vowel formants) that extend over a long time window. A logical extension of this research would be to determine the effects of dipthongs or tripthongs (slow + dynamic) relative to monophthongs (slow + steady-state).

The distinction between static and dynamic cues can also account for interactions between dimensions that come from different domains. In a comparison of a *phonetic* dimension (place of articulation: /b/ vs. /g/) and an *auditory* dimension (pitch: 104 Hz vs. 140 Hz), pitch had a greater interference effect on consonant than vice versa (Wood, 1974, 1975). Consonants were signalled by rapidly changing cues; the pitch levels were held constant throughout the duration of the sound. Regardless of whether the dimension is segmental or suprasegmental, these data support the view that the perceptual trace of dynamic cues decays faster than that of static cues.

Integration of top-down and bottom-up processes in the Garner paradigm. The trace decay approach, however, cannot fully account for the asymmetric interference of rapidly changing onsets on more slowly changing tones (T/C > C/T). Based on this account, we would predict precisely the opposite pattern of interference, such that the less rapidly changing (thus, more slowly decaying) tones should interfere more with the rapidly changing (thus, rapidly decaying) onsets than vice versa. Our findings indicate that higher-level linguistic strategies (top-down) derived from subjects' previous knowledge about informational value play a more important role in influencing dimensional salience than lower-level detection of changes in acoustic features (bottom-up).

The dependency of dimensional interactions on acoustic features or the listener's knowledge of linguistic function is consistent with previous results using the Garner paradigm (Lee & Nusbaum, 1993; Tomiak et al., 1987; Wood, 1974). More importantly, our findings highlight the dynamic interplay between top-down and bottom-up systems in contributing to the asymmetric dimensional interactions. It appears that variation in the irrelevant dimension initially entails a certain degree of salience through bottom-up acoustic feature analysis, but that top-down attentional strategies may capitalise on learned statistical knowledge of the target and distractor dimensions to override the weights of the bottom-up salience (Bacon & Egeth, 1994; Yantis & Egeth, 1999). Further research is necessary in order to specify more precisely the relative contribution of these two sources of influence on the weighting of onsets, rimes, and tones.

Broader implications for speech perception

One of the major findings in this study is that tone is the most vulnerable dimension to interference among the three dimensions. Cutler and Chen (1997) and Ye and Connine (1999) reported a similar tone disadvantage as compared to the processing of segmental information in Chinese. Their explanation is that "the perceptual acquisition of tone information lags behind that of vowel information" (Ye & Connine, 1999, p. 615). In this experiment, we still observe a significant asymmetry among the dimensions in spite of equal reaction times to all the stimuli.

Then what is the reason for tones being weaker than segments in constraining word recognition in Mandarin, as suggested by Cutler and Clifton (1999)? Our information theory account provides an explanation for the weakness of tone in constraining word recognition. Due to the small tonal inventory, each tone is associated with more words than consonants and vowels. Consequently, tonal information exerts fewer constraints on word recognition compared with segmental information. Converging evidence comes from an implicit priming task in Mandarin Chinese (Chen, Chen, & Dell, 2002). They showed that the syllable (onset + rime) by itself produced implicit priming effects, whereas tone-alone prime did not. This suggests that tone is less informative than segmental information in priming words. In sum, the perceptual disadvantage of tones is not merely due to their later availability in the auditory stream. Rather it is because of the lower priority of the tone dimension relative to segmental dimensions, which is largely due to the differences in information values.

Similarly, in the present study, the advantage of V over C and T is in part due to its higher information value within the context of the Mandarin lexicon, and its greater perceptual stability. Together with the fact that the integrality of dimension pairs T-V and C-V are greater than C-T, our results indicate that the vowel (rime) is pivotal in Chinese speech processing.

CONCLUSION

Using the Garner speeded classification task, asymmetrical dimensional dependencies are revealed among all three phonological units in Mandarin. This phenomenon likely depends on multiple factors including acoustic cues, information value, and context. Consonants and vowels are more informative in nature, and may capture more attention than tones when they are varying as irrelevant dimensions. Vowels, on the other hand, are relatively steady-state and occupy a longer time interval, and so are more resistant to interference than consonants. Even though tones are longer in duration than consonants, they are more easily disrupted by irrelevant consonant variability than vice versa, suggesting that the effects of informational value may outweigh acoustic cues under certain stimulus and task conditions.

Manuscript received November 2006
Revised manuscript accepted October 2007
First published online December 2007

REFERENCES

Altman, G. T. M. (1990). Lexical statistics and cognitive models of speech processing. In G. T. M. Altman (Ed.), *Cognitive models of speech processing* (pp. 211–235). Cambridge, MA: MIT Press.

Bacon, W. F., & Egeth, H. E. (1994). Overriding stimulus-driven attentional capture. *Perception and Psychophysics*, *55*(5), 485–496.

Buschman, T. J., & Miller, E. K. (2007). Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science*, *315*(5820), 1860–1862.

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

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

Chen, J.-Y., Chen, T.-M., & Dell, G. (2002). Word-form encoding in Mandarin Chinese as assessed by the implicit priming task. *Journal of Memory and Language*, *46*, 751–781.

Cutler, A., & Chen, H. C. (1997). Lexical tone in Cantonese spoken-word processing. *Perception and Psychophysics*, *59*(2), 165–179.

Cutler, A., & Clifton, C. (1999). Comprehending spoken language: A blueprint of the listener. In C. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 123–166). New York: Oxford University Press.

Dorman, M. F., Kewley-Port, D., Bradley, S., & Turvey, M. T. (1977). Vowel recognition: Inferences from studies of forward and backward masking. *Quarterly Journal of Experimental Psychology*, *29*, 483–497.

Fu, Q. J., Zeng, F. G., Shannon, R. V., & Soli, S. D. (1998). Importance of tonal envelope cues in Chinese speech recognition. *Journal of the Acoustical Society of America*, *104*(1), 505–510.

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

Garner, W. R. (1974). *The processing of information and structure*. Potomac, MD: Erlbaum.

Garner, W. R. (1976). Integration of stimulus dimensions in concept and choice processes. *Cognitive Psychology*, *8*, 98–123.

Garner, W. R. (1983). Asymmetric interactions of stimulus dimensions in perceptual information processing. In T. J. Tighe & B. E. Shepp (Eds.), *Perception, cognition, and development: Interactional analyses* (pp. 1–37). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Garner, W. R. (1988). The contribution of information theory to psychology. In W. Hirst (Ed.), *The making of cognitive science: Essays in honour of George A. Miller* (pp. 19–35). Cambridge, UK: Cambridge University Press.

Garner, W. R., & Felfoldy, G. L. (1970). Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology*, *1*, 225–241.

Horii, Y. (1970). A masking noise with speech-envelope characteristics for studying intelligibility. *Journal of the Acoustical Society of America*, *49*, 1849–1856.

Howie, J. (1976). *An acoustic study of Mandarin tones and vowels*. Cambridge, UK: Cambridge University Press.

Huang, T. (2004). *Language-specificity in auditory perception of Chinese tones*. Unpublished Dissertation, Ohio State University, Columbus, OH.

Hume, E., & Johnson, K. (2001). *The role of speech perception in phonology*. New York: Academic Press.

Klatt, D. H. (1980). Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America*, *67*, 971–995.

Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America*, *87*(2), 820–857.

Lavie, N., & De Fockert, J. (2005). The role of working memory in attentional capture. *Psychonomic Bulletin Review*, *12*(4), 669–674.

Lee, L., & Nusbaum, H. C. (1993). Processing interactions between segmental and suprasegmental information in native speakers of English and Mandarin Chinese. *Perception and Psychophysics*, *53*(2), 157–165.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, *109*(2), 163–203.

Melara, R. D., & Algom, D. (2003). Driven by information: A tectonic theory of Stroop effects. *Psychological Review*, *110*(3), 422–471.

Melara, R. D., & Marks, L. E. (1990). Perceptual primacy of dimensions: Support for a model of dimensional interaction. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(2), 398–414.

Melara, R. D., Marks, L. E., & Potts, B. C. (1993). Primacy of dimensions in color perception. *Journal of Experimental Psychology: Human Perception and Performance*, *19*(5), 1082–1104.

Melara, R. D., & Mounts, J. R. (1994). Contextual influences on interactive processing: Effects of discriminability, quantity, and uncertainty. *Perception and Psychophysics*, *56*(1), 73–90.

Miller, J. L. (1978). Interactions in processing segmental and supra-segmental features of speech. *Perception and Psychophysics*, *24*(2), 175–180.

Mirman, D., Holt, L. L., & McClelland, J. L. (2004). Categorization and discrimination of nonspeech sounds: differences between steady-state and rapidly-changing acoustic cues. *Journal of the Acoustical Society of America*, *116*(2), 1198–1207.

Molholm, S., Martinez, A., Ritter, W., Javitt, D. C., & Foxe, J. J. (2005). The neural circuitry of pre-attentive auditory change-detection: An fMRI study of pitch and duration mismatch negativity generators. *Cerebral Cortex*, *15*(5), 545–551.

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception and Psychophysics*, *13*(2), 253–260.

Repp, B. H., & Lin, H. B. (1990). Integration of segmental and tonal information in speech perception: A cross-linguistic study. *Journal of Phonetics*, *18*(4), 481–495.

Robinson, K., & Patterson, R. D. (1995). The stimulus duration required to identify vowels, their octave, and their pitch chroma. *Journal of the Acoustical Society of America*, *98*(4), 1858–1865.

Simon, J. R., Craft, J. L., & Webster, J. B. (1973). Reactions toward the stimulus source: Analysis of correct responses and errors over a five-day period. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *101*(1), 175–178.

Strange, W. (1989). Dynamic specification of coarticulated vowels spoken in sentence context. *Journal of the Acoustical Society of America*, *85*(5), 2135–2153.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 643–662.

Tomiak, G. R., Mullennix, J. W., & Sawusch, J. R. (1987). Integral processing of phonemes: evidence for a phonetic mode of perception. *Journal of the Acoustical Society of America*, *81*(3), 755–764.

Tsai, C. -H. (2000). *Mandarin syllable frequency counts for Chinese characters.* Retrieved July 17, 2006, from http://technology.chtsai.org/syllable/

Tseng, C.-Y. (1990). *An acoustic phonetic study on tones in Mandarin Chinese* (Vol. 94). Taipei: Institute of History & Philology, Academia Sinica.

Walley, A. C., & Carrell, T. D. (1983). Onset spectra and formant transitions in the adult's and child's perception of place of articulation in stop consonants. *Journal of the Acoustical Society of America*, *73*(3), 1011–1022.

Watkins, S., Dalton, P., Lavie, N., & Rees, G. (2007). Brain mechanisms mediating auditory attentional capture in humans. *Cerebral Cortex*, *17*(7), 1694–1700.

Whalen, D. H., & Xu, Y. (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica*, *49*(1), 25–47.

Wood, C. C. (1974). Parallel processing of auditory and phonetic information in speech discrimination. *Perception and Psychophysics*, *15*(3), 501–508.

Wood, C. C. (1975). Auditory and phonetic levels of processing in speech perception: Neurophysiological and information-processing analyses. *Journal of Experimental Psychology: Human Learning and Memory*, *104*(1), 3–20.

Xu, L., Tsai, Y., & Pfingst, B. E. (2002). Features of stimulation affecting tonal-speech perception: Implications for cochlear prostheses. *Journal of the Acoustical Society of America*, *112*, 247–258.

Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics*, *25*, 61–83.

Yantis, S., & Egeth, H. E. (1999). On the distinction between visual salience and stimulus driven attentional capture. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 661–676.

Ye, Y., & Connine, C. M. (1999). Processing spoken Chinese: The role of tone information. *Language and Cognitive Processes*, *14*(5–6), 609–630.