Effects of language experience and expectations on attention to consonants and tones in English and Mandarin Chinese

Mengxi Lin
Linguistics Program, Purdue University, West Lafayette, Indiana 47907-2038

Alexander L. Francis[a)]
Department of Speech, Language, and Hearing Sciences, Purdue University, West Lafayette, Indiana 47907-2038

(Received 28 June 2013; revised 11 September 2014; accepted 30 September 2014)

Both long-term native language experience and immediate linguistic expectations can affect listeners' use of acoustic information when making a phonetic decision. In this study, a Garner selective attention task was used to investigate differences in attention to consonants and tones by American English-speaking listeners (N = 20) and Mandarin Chinese-speaking listeners hearing speech in either American English (N = 17) or Mandarin Chinese (N = 20). To minimize the effects of lexical differences and differences in the linguistic status of pitch across the two languages, stimuli and response conditions were selected such that all tokens constitute legitimate words in both languages and all responses required listeners to make decisions that were linguistically meaningful in their native language. Results showed that regardless of ambient language, Chinese listeners processed consonant and tone in a combined manner, consistent with previous research. In contrast, English listeners treated tones and consonants as perceptually separable. Results are discussed in terms of the role of sub-phonemic differences in acoustic cues across language, and the linguistic status of consonants and pitch contours in the two languages. © *2014 Acoustical Society of America.*
[http://dx.doi.org/10.1121/1.4898047]

PACS number(s): 43.71.Hw [BRM] Pages: 2827–2838

I. INTRODUCTION

Both intonation and lexical tones are suprasegmental, but they differ in key ways that make tones seem more similar to segments than to intonation: Intonational contours span phrases of varying lengths, while lexical tones extend over much smaller linguistic units (typically over single syllables or words). Moreover, while tone and intonation are both instantiated largely in terms of fundamental frequency (*f*0) contours, tone, unlike intonation, plays a linguistic role in lexical decision similar to that of segmental phonemes. In addition, long-term experience with making lexical decisions on the basis of tones may result in categorical perception of *f*0 continua similar to the categorical perception of segments (Francis *et al.*, 2003), while the categorical perception of intonation is less clear-cut (Ladd and Morton, 1997). Neurological evidence also supports a greater similarity between tones and segments than between tones and intonation: Tonal language speakers exhibit greater left hemisphere activation for processing tones, but predominantly right hemisphere activation for processing longer stretches of *f*0 contour related to intonation, in a manner similar to non-tonal language listeners' left-hemisphere processing of segments and right-hemispheric processing of intonation (Gandour *et al.*, 2003). Based on the segment-like linguistic role of lexical tones, it is possible that speakers of tonal languages may treat *f*0 patterns more like segmental properties than do speakers of non-tonal languages, especially when listening to single-syllable utterances. If this is the case, a logical extension of the question is whether tone language listeners maintain the close association between *f*0 and segments when listening to a non-tonal language that has been learned as a second language (L2). That is, do tone language speakers change their *f*0-directed listening strategy when listening to a non-tonal language? Will they give up the perceptual integration between segments and *f*0 in the context of a non-tonal language in which *f*0 is less critical for lexical decisions?

In English, the *f*0 contour associated with intonation unfolds over the course of a phrase or utterance, and may serve a wide range of prosodic and affective purposes. *F*0 is rarely used contrastively at the lexical level except in the case of lexical stress, and even this property necessarily extends over multiple syllables, Moreover, *f*0 is not the sole cue to lexical stress, and its role in cuing perception of lexical stress may be inextricably related to the role of stressed syllables as attractors of intonational accent. In contrast, tone languages such as Mandarin Chinese use *f*0 patterns phonemically[1] For instance, in Mandarin Chinese, lexical differences can be denoted solely by differences in the *f*0 contours of a given syllable: The syllable [ma] means "mother" with a high *f*0 contour (Tone 1), "hemp" with a rising *f*0 contour (Tone 2), "horse" with a dipping *f*0 contour (Tone 3), and "scold" with a falling *f*0 contour (Tone 4). It has been argued that segmental and tonal information are more closely integrated for Mandarin Chinese listeners than they are for English listeners (Repp and Lin, 1990), and in particular, that lexical tones play a role in lexical access that is equally as important as segmental information under most

a)Author to whom correspondence should be addressed. Electronic mail: francisa@purdue.edu

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 128.210.206.145 On: Tue, 18 Nov 2014 17:52:48

typical conditions of language use (Liu and Samuel, 2007). Although there is a growing literature on differences in production and perception of *f*0 between speakers of tonal and non-tonal languages, little is known about how Mandarin/English bilinguals may change their perception of *f*0 movement depending on the ambient language.

Studies have shown that bilingual listeners change the way they perceive the acoustic properties that define phonetic categories depending on the language they think they are listening to (Elman *et al*., 1977; Flege and Eefting, 1987; Garcia-Sierra *et al*., 2009; Hazan and Boulakia, 1993), but these studies have been conducted only with segmental properties. It is possible that Mandarin/English bilingual listeners may change how they attend to segments and *f*0 depending on the language they expect (English or Mandarin). Specifically, they may treat *f*0 as less integrated with the segmental properties of English syllables as compared with Mandarin ones. Alternatively, Mandarin listeners may not change their treatment of *f*0 contours even when listening in a non-tonal language, especially when the English intonational *f*0 contours are easily assimilated to native representations of lexical tones (cf. the discussion of equivalence classification by Flege, 1995; also Francis *et al*., 2008; Halle *et al*., 2004). In this case, Mandarin/English bilingual listeners may show similar integration of segmental and tonal information in both English and Mandarin.

A. Garner speeded classification paradigm

The Garner speeded classification paradigm (Garner, 1974, 1976) has been widely used to examine the relative integrality or separability of the intrinsic dimensions of stimuli during perceptual processing of speech (Carrell *et al*., 1981; Lee and Nusbausm, 1993; Miller, 1978; Repp and Lin, 1990; Tomiak *et al*., 1987; Tong *et al*., 2008; Wood, 1974) since it effectively reveals the interaction between different sources of information within a single signal (Melara and Marks, 1990). Participants in the Garner paradigm identify the value of the target dimension in two conditions.

(1) In the baseline condition, the target dimension (e.g., vowel) varies while the non-target dimension (e.g., consonant) is held constant: Participants are presented with repeated tokens of [pi] and [pai] in random order, and their task is to identify which vowel they hear.
(2) In the orthogonal condition, the target dimension varies as before, but now the non-target dimension also varies from trial to trial: Participants hear tokens of [pi], [pai], [ti], and [tai] in random order but the task is still to identify the vowels.

When response time (RT) is the same in the orthogonal as in the baseline condition, processing of the two dimensions is considered to be separable, since selective attention to the target dimension is unaffected by the varying non-target dimension. In contrast, a longer RT[2] in the orthogonal condition suggests that the target dimension cannot be attended to separately from the co-varying non-target dimension. Listeners cannot ignore the variability present in the non-target dimension, which results in an increase in processing time. In this case, it is said that listeners process the target dimension integrally with the non-target dimension.

B. Perceptual integrality between segments and tones/*f*0

Dimensional integrality is not always symmetric: The processing of two dimensions may be either symmetrically or asymmetrically integral (Garner, 1974, 1976). That is, when Dimension A is integral with Dimension B, Dimension B could still be separable from Dimension A. Asymmetric integrality also includes cases in which the integrality between two dimensions is bi-directional, but there is greater interference from one dimension on the processing of another than vice versa. The presence of asymmetric integrality suggests that the dimension that causes the greatest interference may dominate the combined percept.

Results of previous studies on processing interaction between segments and lexical tones are mixed. In Mandarin Chinese, some studies have shown symmetric integrality between consonantal and tonal information (Lee and Nusbaum, 1993): Irrelevant variation of tone slowed consonant classification and vice versa. Others have reported asymmetric integrality between vowels and tone (Repp and Lin, 1990), with variability in the vowel interfering more with tone processing than vice versa. Tong *et al*. (2008) found that both consonant and vowel were asymmetrically processed with tone: Vowel variability interfered with tone and consonant more than vice versa, and consonant interfered more with tone than vice versa, supporting the findings of Repp and Lin (1990) but contradicting those of Lee and Nusbaum (1993).

Studies using the Garner paradigm to investigate processing interaction between segments and *f*0 in the context of English suggest the presence of different processing patterns than in Mandarin Chinese, but results are also varied. Wood (1974) found asymmetric integrality between syllable *f*0 (high vs low *f*0) and consonant ([b] vs [g]): Consonant processing was affected by the irrelevant variability of *f*0, but *f*0 processing was independent from irrelevant variability of consonant. Wood (1974) interpreted these results as suggesting that the *f*0 dimension taps basic auditory-sensory capability and, thus, is processed ahead of dimensions at the phonetic level. In contrast, Miller (1978) tested vowels against *f*0 and found that vowel and *f*0 were symmetrically integrated. To reconcile these findings with those of Wood (1974), Miller (1978) suggested that processing interactions between segments and *f*0 depend on the type of segment, such that *f*0 is symmetrically integral with vowels, but asymmetrically integral with consonants. However, this conclusion was subsequently contradicted by later studies, in which English listeners were found to exhibit symmetric interference between consonant and *f*0 (Repp and Lin, 1990; Lee and Nusbaum, 1993), and asymmetric interference between vowel and *f*0, with vowel interfering more with *f*0 than vice versa (Repp and Lin, 1990). In short, prior studies on the integrality of *f*0 with segments in English have found both symmetric and asymmetric processing interaction between

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 128.210.206.145 On: Tue, 18 Nov 2014 17:52:48

consonant and *f*0, and between vowel and *f*0, and it is unclear why this is so.

One factor of note in these studies is the assumption that *f*0 contours are not linguistically meaningful to English listeners. A growing body of research highlights the role of *f*0 as a cue to vowel height (Hoeme and Diehl, 1994; Whalen and Levitt, 1995), and of onset *f*0 as a cue to obstruent consonant voicing (Kingston and Diehl, 1994; Llanos *et al*., 2013) in many non-tonal languages including English. Therefore, it may be premature to dismiss the possibility that *f*0 may be integral with segments even in non-tonal languages, at least under certain circumstances.

It is possible that some differences observed across studies may result from (unintentional) differences in participants' expectations about the stimuli they were listening to. In a study by Tomiak *et al*. (1987), two groups of participants were instructed to listen to noise-tone analogs of fricative-vowel syllables, but one group was told to treat the stimuli as speech sounds and the other group was told the sounds were not speech. Results showed that participants processed the two dimensions integrally when treating them as speech sounds, but separately when they thought the noise-tone analogs were non-speech. Thus, listeners' expectation regarding the linguistic function of stimuli may affect the integration of segments and *f*0, raising the question of whether some of the differences between Mandarin and English listeners observed in previous studies might result from differences in the way listeners were instructed to treat *f*0 (as non-linguistic syllable pitch vs lexical tone) rather than from native language properties per se. On the other hand, if listeners are aware that the relative integrality of vowels, consonants, and tones varies across languages, then Mandarin speakers who are also familiar with English may show a different pattern of processing between segments and *f*0 when listening to English than they would when listening to Mandarin.

The goal of the present study is to extend the investigation of processing interactions between segmental and tonal dimensions by using stimuli that are linguistically meaningful along all dimensions for all listeners in a Garner speeded classification paradigm, and to determine whether bilingual Mandarin/English listeners change the way they treat tone depending on what language they are listening to.

II. METHOD

A. Subjects

A total of 57 listeners aged from 18 to 30 yr participated in the experiment. Among them, 20 Chinese listeners (9 female, 11 male) completed in the experiment in Mandarin Chinese, and another 17 Chinese listeners (12 female, 5 male) took the experiment in an English environment, while 20 American listeners (13 female, 7 male) completed the experiment in English. The three groups will be referred to as Chinese in Chinese, Chinese in English, and English in English hereafter. All participants were undergraduate or graduate students recruited from the Purdue University community under a protocol approved by the Human Research Subjects Protection Program and were compensated for their participation.

All Chinese participants except one were born and raised in Mainland China and all self-reported as being native speakers of standard Mandarin Chinese, but some also spoke an additional local language or nonstandard regional dialect of Mandarin (languages: Cantonese, Gan, Xiang, Wu; nonstandard Mandarin dialects: Hebei, Henan, Liaoning, Sichuan). They started learning English between age 8 to 14 and had been learning and using it for between 11 and 21 yr (mean: 14.5). One participant from the Chinese in English group was an early bilingual speaker of English and Mandarin (born in the United States), and her data were discarded, leaving 16 listeners in the Chinese in English group. Participants of the third group were all native speakers of American English, with no prior experience speaking or learning Mandarin Chinese or any other tonal language. None of the participants had a history of speech or hearing disorder by self report.

B. Stimuli

Previous studies of integrality between *f*0 and segments used only a limited set of sounds: Voiceless unaspirated stops (typically [p] and [t], which are phonologically voiced and represented as /b/ and /d/) and monophthongal vowels (typically [a], [ɑ], [æ], [u]). Some studies used only level tones, some only contour tones, and some both. For the present study, selection of segments and tones was primarily constrained by the need to develop syllables that could serve as real words in both languages, and this required using different consonants and vowels than have previously been used. However, given that recent research suggests that listeners may change the relative importance assigned to acoustic cues in different phonetic contexts (McMurray and Jongman, 2011), the present stimuli also serve to broaden the range of tokens represented in the literature.

Eight syllables, [p^hi], [p^hai], [t^hi], and [t^hai] with Mandarin lexical tone 2 and 4 were used. All syllables constitute real English words, namely, *pea*, *pie*, *tea*, and *Thai*.[3] They are also lexical items in Mandarin Chinese: [p^hi2] *skin*, [p^hi4] *secluded*, [p^hai2] *to put in order*, [p^hai4] *to send*, [t^hi2] *to lift*, [t^hi4] *to replace*, [t^hai2] *to carry*, and [t^hai4] *very* (see lexical frequency in Table I). The tone pairs differ in their initial and final *f*0, and in contour, with Tone 2 featuring a mid-to-high rising *f*0 contour and Tone 4 a high-to-low falling one. These tones were used because their *f*0 direction generally resembles rising and falling English intonation patterns, respectively: In the Tones and Breaks Indices (ToBI) transcription system of the English intonation inventory (Beckman *et al*., 2005), the *f*0 contour of an English yes-no question is denoted as L*H%, and a neutral declarative intonation pattern as H*L%, where "H" and "L" represents high and low pitch, respectively, and % indicates boundary tone. This means that the English intonation for yes-no question generally features a low-to-high rising *f*0 pattern and the declarative intonation a high-to-low falling one, quite similar to Mandarin Tone 2 and Tone 4. Note that it is unnecessary for these Mandarin tones and English intonation patterns to

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 128.210.206.145 On: Tue, 18 Nov 2014 17:52:48

TABLE I. Lexical frequency and goodness ratings of the stimuli.[a]

Mandarin Syllable	Mandarin Frequency	English Syllable	English Frequency	Goodness (English word)	Goodness (English intonation)
[p^hi2]	67978	[p^hi]	1737	Pea: 5.4	Question: 5.6
[p^hi4]	17136			Pea: 6.0	Statement: 5.6
[p^hai2]	21447	[p^hai]	8917	Pie: 5.6	Question: 5.6
[p^hai4]	76750			Pie: 5.4	Statement: 6
[t^hi2]	400263	[t^hi]	20739	Tea: 5.8	Question: 5.8
[t^hi4]	34214			Tea: 5.8	Statement: 5.8
[t^hai2]	127337	[t^hai]	18795	Thai: 5.4	Question: 5.8
[t^hai4]	277746			Thai: 5.6	Statement: 5.6

[a]The frequency counts of the Chinese syllables were obtained from the frequency statistics of Chinese syllable wish tones calculated by Da (2010), which was based on the total frequency of the most commonly used 3500 Chinese characters. The frequency counts of the English syllables were obtained from the Corpus of Contemporary American (Davis, 2008), a corpus that contains 450 million words. The goodness rating of the English words and intonations were obtained through the rating of five native speakers of American English on a seven-point scale with 1 representing "extremely bad," and 7 representing "extremely good."

be phonetically identical, because non-tonal language listeners are able to map unfamiliar tones to native intonational patterns when both share similar *f*0 contours (Braun and Johnson, 2011). By using *f*0 contours that may represent tones in Chinese as well as intonational patterns in English, the *f*0 dimension is linguistically meaningful to all participants.

All stimuli were modeled on natural productions by a male native speaker of Mandarin Chinese reading each of the eight syllables five times. Stimuli were synthesized with a Klatt speech synthesizer (Klatt, 1980) implemented in Praat 5.3.11 (Boersma and Weenink, 2011). All stimuli were 430 ms in duration. Synthesis parameters for the consonantal, vocalic, and tonal properties of the stimuli were held mostly constant for all syllables sharing a consonant, vowel, or tone, respectively, except that some differences were allowed in order to accommodate differences required by other properties (e.g., aspiration/vowel portion durations differed following [p^h] and [t^h] because duration of aspiration is a secondary cue to stop consonant place of articulation, but overall syllable duration was held constant). For the tone dimension, the *f*0 contour started at the end of aspiration, which was 120 ms for syllables with the consonant [p^h] and 105 ms for those with [t^h]. Tone 2 (rising) began at 130 Hz, dropped to 120 Hz in the next 70 ms (a slight dip after onset *f*0 is a typical features of Mandarin Tone 2), and then gradually rose to 170 Hz. Tone 4 (falling) started at 187 Hz and maintained the same *f*0 level for 95 ms, and then gradually decreased to 130 Hz. Synthesis parameters of vowels are listed in Table II.

Since the synthesized stimuli were modeled on Mandarin productions, they were evaluated by five native speakers of American English to determine whether they were sufficiently English-like. Raters listened to each stimulus as often as they wished and then rated its goodness as an example of either "pea," "pie," "tea," or "Thai" and also rated its goodness as a example of a "yes-no question" or a "statement," All ratings were made on a seven-point scale with one representing "extremely bad" and seven representing "extremely good." Each stimulus was rated once for word and once for intonation by each listener. Means (across listeners) for each rating for each syllable are shown in Table I. A mean rating of at least 5.5 for all words and 5.8 for all intonations suggest that the synthetic stimuli were at least acceptable examples of the English words and intonation patterns that they were intended to represent.

For each of the three target dimensions (consonant, vowel, and tone), the eight stimuli were grouped into four sets for the baseline condition and two sets for each of the orthogonal conditions. In the baseline condition, the values of the target dimension across stimuli varied while those of the other two remained constant (e.g., [p^hi2] and [p^hai2] in one vowel judgment task). In the orthogonal condition, the values of the non-target dimension varied independently of the target dimension, while the third dimension remained constant (e.g., [p^hi2], [p^hi4], [p^hai2], and [p^hai4] in the vowel judgment task with orthogonal variation of tone). A complete list of the stimuli sets for all conditions is given in Table III.

C. Procedure

The experiment consisted of three speeded classification tasks, judging vowel, consonant, and tone, respectively. The procedure was the same across the three subject groups, except for the ambient language. For the Chinese in Chinese group, all participants were recruited by the first author who is a native speaker of Mandarin Chinese, and all oral and written correspondence was in Mandarin including the original posters advertising the experiment. In addition, the orthographical representations of all information within the

TABLE II. Synthesis parameters of the stimulus vowels.

Vowels	Time Point	F1 (Hz)	F2 (Hz)	F3 (Hz)
[p^hi]	Beginning	300	1400	2200
	45 ms	350	2200	3100
[t^hi]	Beginning	300	2000	3100
	45 ms	350	2330	3100
[p^hai]	Beginning	720	1300	2400
	100 ms	820	1560	2300
	100 ms	620	1700	2300
	End	500	1800	2300
[t^hai]	Beginning	720	1300	2400
	100 ms	820	1560	2300
	100 ms	620	1700	2300
	End	500	1800	2300

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 128.210.206.145 On: Tue, 18 Nov 2014 17:52:48

TABLE III. Arrangement of stimuli sets along the three dimensions in all conditions, and response choices in the two language contexts.[a]

Target Dimension	Non-target Dimension	Condition	Stimuli Set	Response Choices in Chinese	Response Choices in English
Tone	- (T/B)	Baseline	p^hi2, p^hi4	麻 ([ma2] "hemp")	? (question)
				骂 ([ma4] "scold")	! (statement)
			p^hai2, p^hai4	麻 ([ma2] "hemp")	? (question)
				骂 ([ma4] "scold")	! (statement)
			t^hi2, t^hi4	麻 ([ma2] "hemp")	? (question)
				骂 ([ma4] "scold")	! (statement)
			t^hai2, t^hai4	麻 ([ma2] "hemp")	? (question)
				骂 ([ma4] "scold")	! (statement)
	Vowel (T/V)	Orthogonal	p^hi2, p^hi4, phai2, phai4	麻 ([ma2] "hemp")	? (question)
				骂 ([ma4] "scold")	! (statement)
			t^hi2, t^hi4, t^hai2, t^hai4	麻 ([ma2] "hemp")	? (question)
				骂 ([ma4] "scold")	! (statement)
	Consonant (T/C)	Orthogonal	p^hi2, p^hi4, t^hi2, t^hi4	麻 ([ma2] "hemp")	? (question)
				骂 ([ma4] "scold")	! (statement)
			p^hai2, p^hai4, t^hai2, t^hai4	麻 ([ma2] "hemp")	? (question)
				骂 ([ma4] "scold")	! (statement)
Vowel	- (V/B)	Baseline	p^hi2, p^hai2	衣 [i1] "clothes")	ee (as in "tea")
				哀 ([ai1] "sorrow")	ai (as in "Thai")
			p^hi4, p^hai4	衣 [i1] "clothes")	ee (as in "tea")
				哀 ([ai1] "sorrow")	ai (as in "Thai")
			t^hi2, t^hai2	衣 [i1] "clothes")	ee (as in "tea")
				哀 ([ai1] "sorrow")	ai (as in "Thai")
			t^hi4, t^hai4	衣 [i1] "clothes")	ee (as in "tea")
				哀 ([ai1] "sorrow")	ai (as in "Thai")
	Tone (V/T)	Orthogonal	p^hi2, p^hai2, p^hi4, p^hai4	衣 [i1] "clothes")	ee (as in "tea")
				哀 ([ai1] "sorrow")	ai (as in "Thai")
			t^hi2, t^hai2, t^hi4, t^hai4	衣 [i1] "clothes")	ee (as in "tea")
				哀 ([ai1] "sorrow")	ai (as in "Thai")
	Consonant (V/C)	Orthogonal	p^hi2, p^hai2, t^hi2, t^hai2	衣 [i1] "clothes")	ee (as in "tea")
				哀 ([ai1] "sorrow")	ai (as in "Thai")
			p^hi4, p^hai4, t^hi4, t^hai4	衣 [i1] "clothes")	ee (as in "tea")
				哀 ([ai1] "sorrow")	ai (as in "Thai")
Consonant	- (C/B)	Baseline	p^hi2, t^hi2	趴 ([pha1] "lean over")	p (as in "pie")
				他 ([tha1] "he")	t (as in "tie")
			p^hi4, t^hi4	趴 ([pha1] "lean over")	p (as in "pie")
				他 ([tha1] "he")	t (as in "tie")
			p^hai2, t^hai2	趴 ([pha1] "lean over")	p (as in "pie")
				他 ([tha1] "he")	t (as in "tie")
			p^hai4, t^hai4	趴 ([pha1] "lean over")	p (as in "pie")
				他 ([tha1] "he")	t (as in "tie")
	Tone (C/T)	Orthogonal	p^hi2, t^hi2, p^hi4, t^hi4	趴 ([pha1] "lean over")	p (as in "pie")
				他 ([tha1] "he")	t (as in "tie")
			p^hai2, t^hai2, p^hai4, t^hai4	趴 ([pha1] "lean over")	p (as in "pie")
				他 ([tha1] "he")	t (as in "tie")
	Vowel (C/V)	Orthogonal	p^hi2, t^hi2, p^hai2, t^hai2	趴 ([pha1] "lean over")	p (as in "pie")
				他 ([tha1] "he")	t (as in "tie")
			p^hi4, t^hi4, p^hai4, t^hai4	趴 ([pha1] "lean over")	p (as in "pie")
				他 ([tha1] "he")	t (as in "tie")

[a]In the column of the non-target dimension, the letter before the slash represents the target dimension, and the letter after the slash the non-target dimension. For example, "V/T" denotes vowel as the target dimension with tone as the non-target dimension. "B" means baseline condition; for instance, V/B means vowel baseline task. In the response columns, only the characters, letters, or symbols before the parentheses were shown on the screen. The information in the parentheses was supplied in the instructions on screen and orally. Simplified forms of Chinese characters, instead of traditional forms, were used because all the Mandarin Chinese participants were originally from Mainland China where simplified Chinese characters represent the official writing system.

experiment, including instructions and stimuli, were written in the simplified Chinese characters that are the standard orthography of Mainland China. The other two groups participated in sessions that were conducted by native speakers of American English, and all orthographical representations of instructions and stimuli were also in English. When a participant arrived, the experimenter would chat with them for a few minutes in the designated language to make sure that they, especially the Chinese bilinguals, were operating in the target language mode (Grosjean, 1998).

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 128.210.206.145 On: Tue, 18 Nov 2014 17:52:48

The classification tasks made use of a two-alternative, forced choice paradigm in which recorded stimuli were presented once at a time in random order, and listeners indicated their categorization of the target dimension by selecting one of two alternative responses shown on the screen. Participants were instructed to make decisions by pressing the appropriate response key on the response box as quickly and accurately as possible. The assignment of response buttons was randomized across participants. All response choices for each language group are shown in Table III. In English, the response choices were indicated by a "p" and "t" for the consonant decisions [p^h] and [t^h], the digraphs "ee" and "ai" for [i] and [ai], and the symbols "!" and "?" for the falling and rising tones (statement and yes-no question intonations). Listeners were provided with specific instruction that the "!" and "?" symbols indicated whether the word was said as a statement or a yes-no question. Although the exclamation mark may be more likely to be associated with a narrow focus utterance (i.e., L + H* L-L%) rather than a simple neutral declarative utterance (H-L%), it adequately represents a general pattern of falling intonation in contrast with that of a yes-no question, using a symbol that is familiar to listeners and is visually more noticeable than a period (the other punctuation symbol commonly used in English that may be typically associated with a falling intonational pattern) when serving a response choice.

Because Mandarin orthography does not explicitly represent consonants, vowels, or tones, it was not possible to provide orthographic response choices representing each dimension as for participants tested in English. Instead, for participants tested in Mandarin, response choices were commonly used Chinese characters which shared the same feature along the target dimension as the stimuli but differed in the two non-target dimensions. Participants were asked to choose the character that shared the same consonant, vowel, or tone as in the stimulus they just heard. For instance, in one of the tone judgment tasks with vowel varying irrelevantly, the two response choices were *ma2* (麻 "hemp") and *ma4* (骂 "scold") with the first being the correct response to stimuli [p^hi2] and [p^hai2] and the second being the correct response to stimuli [p^hi4] and [p^hai4].

A potential problem with this kind of response choices was that listeners making responses by choosing Chinese characters had to ignore irrelevant dimensions of the response token when making a decision while this was not the case for those making responses by selecting English orthographic tokens as responses. For example, in a tone judgment task with irrelevantly varying vowels, listeners not only had to ignore the irrelevantly varying vowel in the auditory stimuli, they had to ignore the fact that both stimulus vowels were different from the vowel in the response items, and they also had to ignore differences between the consonants in the stimuli and the response tokens. It is possible that the additional processing demands elicited in this manner might affect accuracy and/or response time.

Because of this possibility, two other response strategies were initially considered but then rejected for specific reasons. One alternative might have been to present listeners with Chinese characters corresponding to the actual words used as stimuli (e.g., 皮 for [p^hi2] "skin" and 僻 for [p^hi4] "secluded"). However, this was undesirable because it would require the use of four response choices in the orthogonal condition, but only two in the baseline conditions, changing the response-processing demands in the two conditions and defeating the goals of the Garner paradigm. Another possibility might have been to use *Pinyin* (Romanization). However, because the Pinyin representations of [p^h] and [t^h] are "p" and "t" which are identical to the characters used in the English language context, this could have encouraged listeners to think about the English pronunciations of the Roman characters, potentially shifting listeners out of Chinese listening mode. Finally, iconic representations of *f*0 contours such as rising or falling arrows were not used in either language context to avoid encouraging listeners to ignore the linguistic significance of the phonetic features they were being asked to attend to.

Prior to the beginning of the experiment, to confirm that the Chinese subjects tested in Mandarin were familiar with all the Chinese characters used as response choices, they were given the six response characters to transcribe in *Pinyin*, including consonant, vowel and tone. All participants were able to accomplish this task on the first try without prompting.

For each judgment (consonant, vowel, tone), participants completed the five conditions that constituted the three task types in the following order: Baseline, Correlated[4] (non-target dimension 1), Correlated (non-target dimension 2), Orthogonal (non-target dimension 1), and Orthogonal (non-target dimension 2). Each baseline condition contained four blocks and each orthogonal condition two blocks (see Table III). Accordingly, every task was composed of 16 blocks. The order of judgments (vowel, consonant, tone) and the blocks within each condition was randomized for individual subjects. To familiarize participants with the experimental paradigm, a practice session was provided at the beginning of each task, using a subset of the stimuli used in the corresponding baseline condition.

There were 30 trials in each practice block, 30 in each baseline block, and 60 in each orthogonal block. The trials in each block consisted of equal number of repetitions of the corresponding stimuli (listed in Table III) in random order. Thus, within each block, each stimulus was repeated 15 times. In total, there were 430 trials for each type of judgment, and the entire experiment thus consisted of 1290 trials. The duration of each trial was 2 s, including 430 ms for the stimulus and 1570 ms as response interval. There was also an intervening 10 s rest interval between blocks. The length of each task was about 15 min, and the complete experiment lasted between 45–60 min.

III. RESULTS

A. Response accuracy

Response accuracy was calculated for each listener in each condition of each task. The overall average percentage of response accuracy was relatively high, ranging from 78%–90% for the Chinese in Chinese group, 70%–95% for the Chinese in English group, and 65%–90% for the English

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 128.210.206.145 On: Tue, 18 Nov 2014 17:52:48

in English group. Nonetheless, there existed considerable within-subject variability across tasks. For example, English Participant #17 scored above 80% in all tasks except in the consonant classification blocks where tone was the orthogonally varying dimension (C/T: 45%) and in the tone classification blocks with vowel as orthogonal variability (T/V: 53%). Scores this close to chance suggests that the participant found those blocks particularly difficult and therefore may have been simply guessing.

In order to make sure that participants were able to make correct decisions instead of merely guessing, a binomial test on response accuracy was conducted to determine the threshold percentage of correctness that entailed performance above chance. The threshold for baseline blocks was 60% ($N = 120$, $p = 1/2$, $\alpha = 0.05$), and that for orthogonal and correlated blocks was 57% ($N = 240$, $p = 1/2$, $\alpha = 0.05$). Therefore, any participant with less than 60% correct in any baseline block or less than 57% correct in any orthogonal block were removed from the data analysis. Altogether, four participants were discarded from the Chinese-in-Chinese group, five from the Chinese-in-English group, and five from the English-in-English group, leaving 16, 15 and 11 participants in each group, respectively. Mean RTs in the following analysis were calculated based on the correct responses of these remaining participants.

B. Comparison of baseline discrimination

One factor that may be confounded with dimensional interaction is the relative discriminability of dimensions (Carrell *et al.*, 1981; Melara and Marks, 1990). If the baseline RT of one dimension is faster than that of another, it suggests that the acoustic differences between tokens along the faster-RT dimension are more discriminable. Carrell *et al.* (1981) systematically varied vowel quality and *f*0 values of syllables in speeded classification tasks and found that the symmetric interaction between vowel and *f*0 found by Miller (1978) was only replicated when the discriminatiblity between the two dimensions was the same. In cases of different relative discriminability, the more discriminable dimension (faster baseline RT) interfered more with the processing of the less discriminable dimension (slower baseline RT) than vice versa. Thus, in the absence of equal discriminability of dimensions, a finding of integrality may be difficult to interpret.

To determine whether the three dimensions of the stimuli in the present study were equally discriminable, a two-way repeated-measure analysis of variance (ANOVA) was used to compare the average RT of the baseline consonant, vowel, and tone identification conditions, with RT as the dependent variable, and Dimension (vowel, consonant, tone) and Group (Chinese in Chinese, Chinese in English, English in English) as two fixed factors. Results as shown in Fig. 1 showed a significant main effect for Dimension [$F(2, 39) = 19.48$, $p < 0.0001$] and a significant main effect for Group [$F(2, 39) = 4.70$, $p = 0.01$], but the interaction term was not significant [$F(4, 78) = 1.34$, $p = 0.26$]. These results suggest that there were differences in terms of the discriminability of the three dimensions and there were also differences among the three experimental groups.

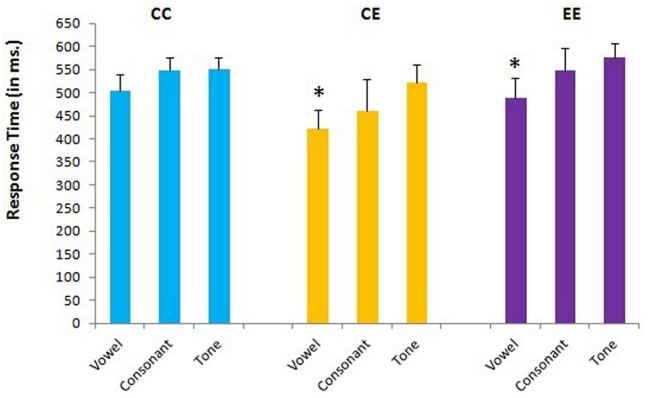

FIG. 1. (Color online) Mean RTs of the vowel, consonant, and tone classification in their respective baseline condition. (CC: Chinese-in-Chinese; CE: Chinese in English; EE: English in English.) Error bars indicate 95% confidence intervals.

Post hoc Tukey's HSD tests showed that for the Chinese in Chinese group, response times were equal between consonant (548.82 ms) and tone (549.22 ms) [$F(1, 78) = -0.02$, $p = 0.99$] between consonant and vowel (504.62 ms) [$F(1, 78) = 2.21$, $p = 0.41$] and between tone and vowel [$F(1, 78) = 2.23$, $p = 0.40$]. In the Chinese in English group, there was no RT difference between consonant (461.05 ms) and tone (522.18 ms) [$F(1, 78) = -2.53$, $p = 0.23$] or between consonant and vowel (421.85) [$F(1, 78) = 1.62$, $p = 0.78$]. However, the vowel baseline RT was much shorter than the tone baseline RT [$F(1, 78) = -4.51$, $p = 0.003$]. Finally, in the English in English group, RTs were equal between consonant (547.53 ms) and tone (575.12 ms) [$F(1, 78) = -1.33$, $p = 0.92$] and between consonant and vowel (487.48 ms) [$F(1, 78) = 2.90$, $p = 0.10$], but the vowel baseline RT was much shorter than the tone baseline RT [$F(1, 78) = -4.24$, $p = 0.002$].

The baseline RT comparison showed that vowels were significantly more discriminable than tones for the Chinese in English and the English in English groups. This was potentially problematic, because differences in dimensional discriminability might be confounded with the interference effect in the orthogonal condition (Carrell *et al.*, 1981; Melara and Marks, 1990). On the other hand, the discriminabilities of the consonantal and tonal dimensions were approximately the same for all three subject groups, validating cross-group comparison. Thus, the rest of this paper will focus on the dimensional interaction between consonant and tone.

C. Dimensional interference

The interference effect as shown in Fig. 2 was evaluated by comparing the average RT in the orthogonal condition with the corresponding baseline condition [e.g., the condition in which the consonant was the target dimension and tone varied irrelevantly (C/T) vs the consonantal dimension baseline (C/B) (See Table III for explanation of abbreviations)]. Any significant difference would indicate that the identification of the target dimension was affected by variation in the non-target dimension. A three-way repeated-

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 128.210.206.145 On: Tue, 18 Nov 2014 17:52:48

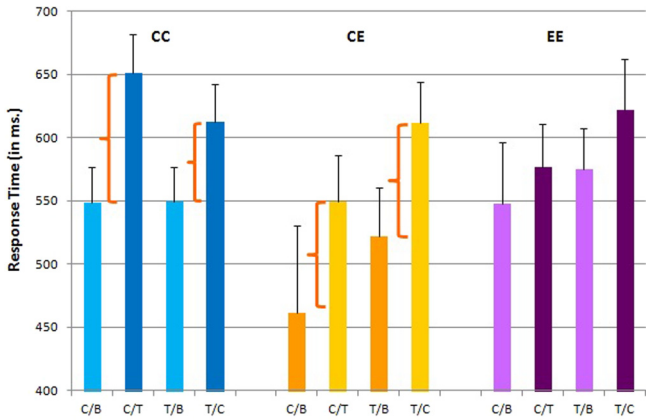

FIG. 2. (Color online) Dimensional interference effects are shown by comparing the RT between the orthogonal condition and the corresponding baseline condition (indicated by big bracket). The columns are grouped by the three experimental groups. (CC: Chinese-in-Chinese; CE: Chinese in English; EE: English in English; C/B: consonant classification in the baseline condition; C/T: consonant classification with tone changing in the orthogonal condition; T/B: tone classification in the baseline condition; T/C: tone classification with consonant varying in the orthogonal condition.) Error bars indicate 95% confidence intervals.

measure ANOVA was performed with RT as the dependent variable, and the three fixed factors of Condition (Baseline, Orthogonal), Task (C/T, T/C), and Group (Chinese in Chinese, Chinese in English, English in English). Results showed a significant main effect for Task [$F(1, 39) = 11.47$, $p = 0.002$] and for Condition [$F(1, 39) = 81.05$, $p < 0.0001$], but the main effect of Group was not significant [$F(2, 39) = 3.12$, $p = 0.0554$]. The three-way interaction term was not significant [$F(2, 39) = 1.35$, $p = 0.27$]. Among the two-way interaction terms, the effect of Group* Task was significant [$F(2, 39) = 9.72$, $p = 0.0004$] and that of Group*Condition was significant [$F(2, 39) = 4.48$, $p = 0.0177$], but the effect of Task *Condition was not significant [$F(1, 39) = 0.20$, $p = 0.65$]. These results suggest that there were differences among the experimental groups in terms of their RTs across conditions and task types. Thus, *post hoc* Tukey's HSD tests were performed to examine dimensional interference within each group.

Results showed that for the Chinese in Chinese group, there was significant T/C interference effect such that RT in T/C blocks (612.14 ms) was significantly longer than RT in T/B blocks (549.22 ms) [$F(1, 39) = 3.60$, $p = 0.04$], and a significant C/T interference effect such that RT in C/T blocks (651.16 ms) was significantly longer than RT in C/B blocks (548.82 ms) [$F(1, 39) = 5.85$, $p < 0.0001$]. In other words, tone judgment was affected by irrelevant variability of consonant, and consonant judgment was also affected by irrelevant variability of tone.

The Chinese in English group also showed a significant T/C interference effect, with significantly longer RT in T/C blocks (611.28 ms) than in T/B blocks (522.18 ms) [$F(1, 39) = 4.22$, $p = 0.007$], and a significant C/T interference effect with a significantly longer RT in C/T blocks (549.35 ms) than in C/B blocks (461.05 ms) [$F(1, 39) = 4.18$, $p = 0.008$]. Like the Chinese in Chinese group, Chinese in English listeners also exhibited mutual interference between the consonant and tone dimension.

Finally, analysis of the English in English group showed no T/C interference effect: RT in T/C blocks (621.61 ms) was not significantly longer than in T/B blocks (575.12 ms) [$F(1, 39) = 2.57$, $p = 0.33$]. Similarly, there was no significant C/T interference: RT in C/T blocks (575.36 ms) was not statistically different from RT in C/B blocks (547.53 ms) [$F(1, 39) = 1.60$, $p = 0.90$]. That is to say, in strong contrast to the two Chinese groups, when English listeners were making decisions about *f*0 pattern, they were able to ignore the concurrent but irrelevant variation of the consonant dimension and vice versa, suggesting that a change of *f*0 contour, even when it was linguistically meaningful, did not slow down consonant processing.

D. Dimensional integrality

Since both Chinese groups exhibited bi-directional interference between consonantal and tonal dimensions, it was important to determine whether the degree of integrality was equal, that is, whether the magnitude of consonant interfering with tone was the same as that of tone interfering with consonant. Since the English participants did not show any integrality, their data are not analyzed in this section. A two-way repeated-measure ANOVA was performed to measure dimensional integrality. The dependent variable, denoted as RT_{diff}, was measured by subtracting the RT in the baseline condition from the RT in the orthogonal condition (e.g., $RT_{C/T} - RT_{C/B}$, and $RT_{T/C} - RT_{T/B}$). The two fixed factors were Task (C/T, T/C) and Group (Chinese in Chinese, Chinese in English). Results showed no significant main effect for Task [$F(1, 25) = 1.12$, $p = 0.30$] and no significant main effect for Group [$F(1, 25) = 0.11$, $p = 0.74$], and the interaction term was likewise not significant [$F(1, 25) = 1.21$, $p = 0.28$]. These results indicate that both Chinese groups demonstrated symmetric integrality between consonant and vowel: The magnitude of consonant interfering with tone did not differ significantly from the magnitude of tone interfering with consonant. In addition, the magnitude of consonant interfering with tone was the same for both groups, and so was the magnitude of tone interfering with consonant, suggesting that language expectation did not have an effect on the Chinese late bilinguals' processing of consonantal and tonal information.

IV. DISCUSSION

The goal of this study was to investigate processing interactions between consonants and tones by Mandarin/English bilinguals in comparison with native English listeners. Results showed that the Chinese listeners processed consonant and tone with mutual integrality regardless of language expectation, but the English listeners showed no dimensional interaction between consonant and intonation.

A. Baseline relative discriminability

The average baseline RTs for judging consonant and tone were not significantly different regardless of experimental group. This suggests that, in the absence of interference from other dimensions, both Chinese and English listeners

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 128.210.206.145 On: Tue, 18 Nov 2014 17:52:48

show similar sensitivity toward consonant ([p^h], [t^h]) and tone (Mandarin Tone 2 and Tone 4), at least as implemented in the stimuli used here. Interestingly, the observation that Chinese listeners are equally sensitive to consonantal and tonal differences conforms to results found in previous studies, but the observation that English listeners also show equivalent sensitivity to consonantal and tonal dimensions does not. Both Repp and Lin (1990) and Lee and Nusbaum (1993) previously found unequal discriminability between consonant and tone among their English participants, who were more sensitive to consonant than to tone.

This seeming discrepancy between the present study and previous ones may be accounted for by the linguistic relevance of the tonal dimension in the present study, in which the English listeners were specifically asked to classify tonal differences as intonation patterns. It is possible that the salience of same pitch information may be enhanced when English listeners are treating the *f*0 contour as intonation rather than as an abstract, non-linguistic property of the stimulus. The possibility that linguistic meaningfulness can amplify the discriminability of pitch variation by non-tonal language listeners is supported by Braun and Johnson (2011), who reported that Dutch listeners were more attentive to the same pitch movement of Mandarin Tone 2 and Tone 4 when it had linguistic relevance than when it did not (i.e., when the Mandarin tones were interpreted as the intonational contrasts between declarative statements versus echo questions in Dutch rather than as simple *f*0 contour differences).

B. Dimensional integrality between Chinese listeners

The Chinese listeners exhibited symmetric processing interaction between the consonantal and tonal information, in both Mandarin and English contexts. Notwithstanding the ambient language, Mandarin listeners do not seem to be able to divert their attention from tonal variation while identifying consonants, nor can they neglect consonant variation while making judgment about tones, so the processing time of either dimension is lengthened by interference from variability in the other. This suggests that consonant and tone are equally important in the combined percept. The close association between consonant and tone in Mandarin context is predicted, because both dimensions are critical for lexical decision. However, it is worth noting that Mandarin listeners demonstrated the same behavior when listening in their native tonal language and in the context of a non-native, non-tonal language.

One possible explanation for this lack of a context effect is that the Chinese listeners may have transferred processing strategies from their native language into the second language. Chinese listeners presumably have developed categorical representations of pitch contours due to long-term experience with tones (Xu *et al.*, 2006), and therefore, their pitch processing may continue to be influenced by these representations irrespective of the surrounding language. This may also relate to the idea of "equivalence classification" in the Speech Learning Model (SLM) (Flege, 1995). According to this argument, because the two English intonational patterns are sufficiently similar to identifiable counterparts in Mandarin (Tone 2 and Tone 4), Mandarin listeners might assimilate the English intonational patterns to those similar Mandarin tones, particularly when the domain of the English intonations is restricted to a syllable. This possibility may be further examined in future studies that directly address the patterns of cross-language tonal assimilation by instructing listeners to make decisions about the similarity of L2 tokens to L1 categories (Flege *et al.*, 1997).

An alternative explanation for the homogeneous behavior of the two Chinese groups might be the fact that the present stimuli were synthesized based on Mandarin pronunciations of the eight syllables. It is true that the sound inventories of Mandarin and English share the two aspirated stops [p^h] and [t^h], and the *f*0 contours are generally similar between Mandarin Tone 2 and 4 and English rising (L-H%) and falling (H-L%) intonation patterns. Nonetheless, there still exist fine-grained sub-phonemic and acoustic differences between these sounds in the two languages, and Mandarin/English bilinguals may be sensitive to them. For example, Chao and Chen (2008) reported that the VOT for Mandarin aspirated stops ([p^h] (mean 82 ms, range 35–147 ms) and [t^h] (mean 81 ms, range 45–123 ms) is significantly longer than their English counterparts ([p^h], 62 ms; [t^h], 73 ms). In the stimuli of the present study, VOT was 120 ms for [p^h] and 105 ms for [t^h], which are both closer to the prototypes of Mandarin than English. Although these differences might seem small, studies have shown that listeners are sensitive to a similar degree of within-category variance in VOT in word recognition tasks (McMurray *et al.*, 2008). In addition, the *f*0 contours of the present stimuli also reflect typical acoustic properties of Mandarin Tone 2 and Tone 4: Tone 2 features a small-scale *f*0 dip in the first 70 ms following the burst before the subsequent large rise, while Tone 4 maintains a plateau of high *f*0 value during the first 95 ms before a drastic drop of *f*0 contour. It is possible that the Chinese listeners in the English experiment were cued by these prototypically Chinese acoustic properties, which might have triggered an unintended switch of language mode from English to Mandarin, leading them to process English words in a similar fashion as in Mandarin (see similar arguments by Flege and Eefting, 1987). However, the results of the present study cannot be taken as conclusive evidence of this hypothesis.

Yet another issue to settle is why the present study failed to find asymmetric integrality between consonant and tone among Chinese listeners as previously found by Tong *et al.* (2008). Although both studies used Mandarin Tone 2 and 4, in the stimuli of Tong *et al.* (2008), Tone 2 began at 109 Hz and Tone 4 at 138 Hz (29 Hz difference). In the present study, Tone 2 started at 130 Hz and Tone 4 at 187 Hz (57 Hz difference). Thus, the two tones may have been just enough more discriminable in the present study to make them resistant to interference from consonants. This hypothesis is further supported by observations by Cutler and Chen (1997), in which Cantonese listeners showed the most conspicuous decrease in sensitivity to tonal information when the tones in question were inherently difficult to distinguish (i.e., they had similar *f*0 values at the beginning of the syllable). In contrast, in stimuli including a more psychoacoustically distinguishable

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 128.210.206.145 On: Tue, 18 Nov 2014 17:52:48

pair of tones (i.e., one with large differences in *f*0 values at the beginning of the syllable), listeners were equally sensitive to tonal and segmental distinctions. However, no significant differences in discriminability were found in the baseline conditions between the tonal and consonantal tokens in the present study, so an explanation based purely in terms of relative discriminability seems unlikely.

Another alternative explanation for this discrepancy has been suggested by Liu and Samuel (2007), who argued that with sufficient contextual information (such as in sentences or idioms), the segmental advantage over tone reported in earlier studies may give way to a more prominent role for the tonal dimension. Thus, when stimuli are more contextually rich than those used by Tong *et al*. (2008), we might expect less asymmetric or even fully symmetric integrality of tones and consonants. However, given that both Tong *et al*. (2008) and the present study used common, well-known monosyllabic words as stimuli, it is not clear whether any additional context might have been provided in the present study.

C. Dimensional integrality between Chinese and English listeners

In the present study, English listeners showed no dimensional interaction between consonant and intonation: The processing of *f*0 information was not slowed down by consonant variability, nor was the processing of consonant affected by *f*0 variation. Baseline judgment tasks revealed that all three groups of listeners exhibited the same degree of ability to make judgments based on the two dimensions in these stimuli, so the different patterns of processing interference between Chinese and English listeners cannot originate from inherent characteristics of the stimuli. A possible explanation hinges on processing cost under conditions of limited resources. Specifically, in the orthogonal condition where consonant and tone compete for attentional resources, Mandarin listeners employed both dimensions for lexical identification because ignoring either dimension would result in a perceptually high cost in a tonal language. English listeners, on the other hand, were able to ignore *f*0 variation when judging consonant and were able to ignore consonants when judging intonation, because *f*0 is less important than consonants for making lexical decisions in English and consonantal information is not relevant for making intonational decisions. Thus, ignoring one source of information when judging the other incurs little cost, and might even benefit processing efficiency.

Surprisingly, the performance of English listeners in the present experiment was inconsistent with the findings of Repp and Lin (1990) and Lee and Nusbaum (1993), both of whom found symmetric processing interaction between consonant and tone. It also appears to be inconsistent with Wood (1974), who reported asymmetric integration between consonant and *f*0 in English listeners, such that *f*0 variability interfered with consonant decisions but not vice versa. In all of these cases, integrality between *f*0 and consonants may be explained in terms of the role that onset *f*0 plays as a potentially significant cue to voicing of consonants in syllable-initial position (Kingston and Diehl, 1994; Llanos *et al*., 2013). According to this argument, *f*0 should be perceived integrally with other consonantal properties because it serves as an important cue to obstruent voicing. Thus, the primary question to be addressed here is why English speakers did not show any effect of *f*0 variability on consonantal processing in the present study. The answer may possibly be found in the acoustic features of the stimuli used here. In the present study, the consonants were voiceless aspirated stops [p^h] and [t^h], which are characterized by relatively long VOT, while previous studies used voiceless unaspirated stops [p] and [k], which have a short VOT. Moreover, the burst properties of the present stimuli may have been acoustically stronger, having been modeled on aspirated stops in which this is typically the case (Kingston and Diehl, 1994). Therefore, it is possible that the long VOT and comparatively powerful bursts in our stimuli may have provided our English listeners with sufficient acoustic information related to voicing that they had no need to incorporate *f*0 information into their decision.

Interestingly, the use of voiceless stops does not seem to have affected the processing interaction among the Mandarin Chinese listeners in this study. The present results for Mandarin listeners were comparable to the findings of earlier studies, which might be attributable to a difference in acoustic cue weighting strategies by listeners of tonal and non-tonal languages making obstruent voicing decisions. Research on the relationship between VOT and onset *f*0 suggests that while non-tonal language listeners tend to employ VOT and onset *f*0 as the primary and secondary cues for voiceless stop identification, tone language speakers suppress the use of onset *f*0 as a cue to stop consonant voicing, because they tend to prioritize *f*0 information for tonal identification (Francis *et al*., 2006; Xu and Xu, 2003). Consequently, in all the cross-linguistic studies on processing interaction between segments and suprasegmentals using either voiceless unaspirated or aspirated stops, including the present study, Mandarin Chinese listeners have demonstrated mutual interference between tone and consonant, irrespective of the consonant choice.

Meanwhile, consonantal variability failed to interfere with intonational judgments among English listeners in our study, which was possibly because intonation in English usually spans over sequences of segments, and therefore the optimal strategy of identifying intonational patterns with restricted attentional resources is to ignore segmental information. The interference of consonant with *f*0 observed in Repp and Lin (1990) and Lee and Nusbaum (1993) may arise from an inherent artifact of their stimuli, as both reported that consonants were more discriminable than tones in their stimuli. As shown by Carrell *et al*. (1981), when two dimensions differ in baseline RT, in the orthogonal condition there will be asymmetric interference with the faster baseline dimension interfering more with the slower baseline dimension than vice versa (see Repp and Lin, 1990, for discussion).

V. CONCLUSION AND IMPLICATIONS FOR FUTURE STUDY

Focusing on the dimensional interaction between consonant and tone, the major finding of this study is that, given stimuli with equal discriminability and using voiceless

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 128.210.206.145 On: Tue, 18 Nov 2014 17:52:48

aspirated stops and Mandarin Tone 2 and Tone 4, Chinese listeners show symmetric interference between consonant and tone. In contrast, English listeners did not show such a tight perceptual integration of segmental and intonational dimensions: Consonant and intonation were symmetrically separate from one another. This outcome suggests that the dimensional interaction between segments and suprasegmentals is susceptible to language-specific processing constraints. Because of long-term linguistic experience with lexical tones, Mandarin Chinese listeners may always attend to *f*0. English listeners, on the other hand, may attend heavily to *f*0 patterns only when they stretch over phrases or sentences, dismissing *f*0 information when processing segments, especially those for which voicing is already easily determined by other cues. Nevertheless, it should be acknowledged that in the present study, the same Mandarin stimuli were used across experimental groups irrespective of ambient languages, and thus, the possibility cannot be excluded that these stimuli were perceived as Chinese syllables by the Mandarin listeners in the English context, and were perceived as non-prototypical English syllables by the English listeners. Subsequent studies may develop both English and Mandarin versions of the same stimuli set in order to further investigate the effect of language context and expectation.

More broadly, when considered along with the results of previous research, the present results suggest that dimensional processing interactions may be affected by the acoustic correlates of the segments and suprasegmentals in the experimental design, both in terms of which phonetic units are used and in terms of the sub-phonemic properties of the features that cue those units. In this regard, future research is needed to determine whether processing interactions may differ across different phonetic (segmental or suprasegmental) environments, i.e., whether processing interactions change as a function of the particular vowels, consonants or tones used in specific stimuli. Finally, the present results strongly suggest a need for future research to address the question of how acoustic properties of the signal may direct bilingual listeners' language expectation (and, if so, what cues may serve that function), and how, as suggested by McMurray and Jongman (2011), the use of incoming acoustic cues in phonetic decision-making may be affected by the perception of such language-specific cue properties.

ACKNOWLEDGMENTS

We would like to thank Fernando Llanos, Olga Dmitrieva and Rachel Chapman for assistance with stimulus generation, Samantha Berger and Audrey Bengert for assistance with running the English-context experiments, and Whitney Huang for assistance with statistical analyses. We would also like to thank Ben Munson for helpful comments on an earlier version of this manuscript.

[1]*F*0 functions as the primary perceptual cues of Mandarin tones, although other properties may also play a role, such as duration and amplitude contour (Blicher *et al*., 1990; Whalen and Xu, 1992).

[2]A shorter RT in the orthogonal as compared to the baseline condition is usually not observed because the task in the orthogonal condition is at least as cognitively demanding as the tasks in the baseline condition.

[3]In retrospect, the more frequent word *tie* could also have been used here. There was no principled basis for selecting *Thai* over *tie* except, perhaps, an unconscious preference for Thai cuisine over sartorial neckwear on the part of the second author.

[4]Note that the experiment design included two blocks of a correlated condition in which the target dimension and the non-target dimension co-varied simultaneously. For instance, participants are presented with repetitive tokens of [p^hi], [t^hai] in random order and their task is to identify vowel. The correlated condition differs from the orthogonal condition in that the variation of the non-target dimension is relevant to the variation of the target dimension, and therefore any change in the value of the target dimension necessarily entails a change in value of the non-target dimension. In the Garner paradigm, a shorter response time in the correlated condition suggests that the processing of the target dimension is facilitated by the predictably redundant variability of the non-target dimension, indicating that the two dimensions are perceptually integrated. However, studies using the correlated condition to investigate speech stimuli (Eimas *et al*., 1978; Miller, 1978) found inconsistent results about the facilitation effect. Therefore, results of the correlated condition were not analyzed for the present project, and will not be discussed further.

Beckman, M. E., Hirschberg, J., and Shattuck-Hufnagel, S. (**2005**). "The original ToBI system and the evolution of the ToBI framework," in *Prosodic Typology: The Phonology of Intonation and Phrasing*, edited by S.-A. Jun (Oxford University Press, Oxford, UK), pp. 9–54.

Blicher, D. L., Diehl, R. L., and Cohen, L. B. (**1990**). "Effects of syllable duration on the perception of the Mandarin Tone 2/Tone 3 distinction: Evidence of auditory enhancement," J. Phonetics **18**, 37–49.

Boersma, P., and Weenink, D. (**2011**). "Praat: Doing phonetics by computer (Version 5.3.10) [Computer Program]," http://www.praat.org/ (Last viewed January 10, 2012).

Braun, E., and Johnson, E. K. (**2011**). "Question or tone 2? How language experience and linguistic function guide pitch processing," J. Phonetics **39**, 585–594.

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

Chao, K., and Chen, L. (**2008**). "A cross-linguistic study of voice onset time in stop consonant productions," Computational Linguistics and Chinese Language Processing **13**, 215–232.

Cutler, A., and Chen, H.-C. (**1997**). "Lexical tone in Cantonese spoken-word processing," Percept. Psychophys. **59**, 165–179.

Da, J. (**2010**). "Chinese text computing: Syllable frequency with tones," http://lingua.mtsu.edu/chinese-computing/phonology/syllabletone.php (Last viewed March 21, 2014).

Davis, M. (**2008**). "The corpus of contemporary American English: 450 million words, 1990–present," http://corpus.byu.edu/coca/ (Last viewed March 27, 2014).

Eimas, P. D., Tartter, V. C., Miller, J. L., and Keuthen, N. J. (**1978**). "Asymmetric dependencies in processing phonetic features," Percept. Psychophys. **23**, 12–20.

Elman, J. L., Diehl, R. L., and Buchwald, S. E. (**1977**). "Perceptual switching in bilinguals," J. Acoust. Soc. Am. **62**, 971–974.

Flege, J. (**1995**). "Speech learning in a second language," in *Phonological Development: Models, Research, and Implications*, edited by L. Menn and C. Stoel-Gammo (York Press, Timonium, MD), pp. 565–604.

Flege, J. E., Bohn, O.-S., and Jang, S. (**1997**). "Effects of experience on non-native speakers' production and perception of English vowels," J. Phonetics **25**, 437–470.

Flege, J. E., and Eefting, W. (**1987**). "Cross-language switching in stop consonant production and perception by Dutch speakers of English," Speech Commun. **6**, 185–202.

Francis, A. L., Ciocca, V. C., Ma, L., and Fenn, K. (**2008**). "Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers," J. Phonetics **36**, 268–294.

Francis, A. L., Ciocca, V. C., and Ng, B. K. C. (**2003**). "On the (non)categorical perception of lexical tones," Percept. Psychophys. **65**, 1029–1044.

Francis, A. L., Ciocca, V. C., Wong, V. K. M., and Chan, J. K. L. (**2006**). "Is fundamental frequency a cue to aspiration in initial stops?," J. Acoust. Soc. Am. **120**, 2884–2895.

Gandour, J. T., Dzemidzic, M., Wong, D., Lowen, M., Tong, Y., Hsieh, L., Satthamnuwong, N., and Lurito, J. (**2003**). "Temporal integration of speech prosody is shaped by language experience," Brain Language **84**, 318–336.

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 128.210.206.145 On: Tue, 18 Nov 2014 17:52:48

Garcia-Sierra, A., Diehl, R. L., and Champlin, C. (**2009**). "Testing the double phonemic boundary in bilinguals," Speech Commun. **51**, 369–378.
Garner, W. R. (**1974**). *The Processing of Information and Structure* (Erlbaum, Potomac, MD), 203 pp.
Garner, W. R. (**1976**). "Integration of stimulus dimensions in concept and choice processes," Cognit. Psychol. **8**, 98–123.
Grosjean, F. (**1998**). "Studying bilinguals: Methodological and conceptual issues," Bilingualism: Language Cognit. **1**, 131–149.
Halle, P. A., Chang, Y. C., and Best, C. T. (**2004**). "Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners," J. Phonetics **32**, 395–421.
Hazan, V. L., and Boulakia, G. (**1993**). "Perception and production of a voicing contrast by French-English bilinguals," Language Speech **36**, 17–38.
Hoeme, K. A., and Diehl, R. L. (**1994**). "Perception of vowel height: The role of *F*1-*F*0 distance," J. Acoust. Soc. Am. **96**, 661–674.
Kingston, J., and Diehl, R. L. (**1994**). "Phonetic knowledge," Language **70**, 419–494.
Klatt, D. H. (**1980**). "Software for a cascade/parallel formant synthesizer," J. Acoust. Soc. Am. **67**, 971–995.
Ladd, D. R., and Morton, R. (**1997**). "The perception of intonational emphasis: Continuous or categorical?," J. Phonetics **25**, 313–342.
Lee, L., and Nusbaum, H. C. (**1993**). "Processing interaction between segmental and suprasegmental information in native speakers of English and Mandarin Chinese," Percept. Psychophys. **53**, 157–165.
Liu, S., and Samuel, A. G. (**2007**). "The role of Mandarin tones in lexical access under different contextual conditions," Lang. Cogn. Process **22**, 566–594.
Llanos, F., Dmitrieva, O., Shultz, A. A., and Francis, A. L. (**2013**). "Auditory enhancement and second language experience in Spanish and English weighting of secondary voicing cues," J. Acoust. Soc. Am. **134**, 2213–2224.
McMurray, B., Aslin, R., Tanenhaus, M., Spivey, M., and Subik, D. (**2008**). "Gradient sensitivity to within-category variation in speech: Implications for categorical perception," J. Exp. Psychol.: Hum. Percept. Perform. **34**, 1609–1631.
McMurray, B., and Jongman, A. (**2011**). "What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations," Psychol. Rev. **118**, 219–246.
Melara, R. D., and Marks, L. E. (**1990**). "Perceptual primacy of dimensions: Support for a model of dimensional interaction," J. Exp. Psychol.: Hum. Percept. Perform. **16**, 398–414.
Miller, J. L. (**1978**). "Interactions in processing segmental and supra-segmental features of speech," Percept. Psychophys. **24**, 175–180.
Repp, B. H., and Lin, H. B. (**1990**). "Integration of segmental and tonal information in speech perception: A cross-linguistic study," J. Phonetics **18**, 481–495.
Tomiak, G. R., Mullennix, J. W., and Sawusch, J. R. (**1987**). "Integral processing of phoneme: Evidence for a phonetic mode of perception," J. Acoust. Soc. Am. **81**, 755–764.
Tong, Y., Francis, A. L., and Gandour, J. T. (**2008**). "Processing dependencies between segmental and suprasegmental features in Mandarin Chinese," Lang. Cogn. Process **23**, 689–708.
Whalen, D. H., and Levitt, A. G. (**1995**). "The universality of intrinsic Fo of vowels," J. Phonetics **23**, 349–366.
Whalen, D. H., and Xu, Y. (**1992**). "Information for Mandarin tones in the amplitude contour and in brief segments," Phonetica **49**, 25–47.
Wood, C. C. (**1974**). "Parallel processing of auditory and phonetic information in speech discrimination," Percept. Psychophys. **15**, 501–508.
Xu, C. X., and Xu, Y. (**2003**). "Effect of consonant aspiration on Mandarin tones," J. Int. Phonetics Assoc. **33**, 165–181.
Xu, Y., Gandour, J. T., and Francis, A. L. (**2006**). "Effects of language experience and stimulus complexity on the categorical perception of pitch direction," J. Acoust. Soc. Am. **120**, 1063–1074.

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 128.210.206.145 On: Tue, 18 Nov 2014 17:52:48