Acoustic characteristics of English lexical stress produced by native Mandarin speakers

Yanhong Zhang
Program in Linguistics, Purdue University, West Lafayette, Indiana 47907

Shawn L. Nissen
Department of Communication Disorders, Brigham Young University, Provo, Utah 84602

Alexander L. Francis[a]
Program in Linguistics and Department of Speech, Language and Hearing Sciences, Purdue University, West Lafayette, Indiana 47907

Native speakers of Mandarin Chinese have difficulty producing native-like English stress contrasts. Acoustically, English lexical stress is multidimensional, involving manipulation of fundamental frequency (F0), duration, intensity and vowel quality. Errors in any or all of these correlates could interfere with perception of the stress contrast, but it is unknown which correlates are most problematic for Mandarin speakers. This study compares the use of these correlates in the production of lexical stress contrasts by 10 Mandarin and 10 native English speakers. Results showed that Mandarin speakers produced significantly less native-like stress patterns, although they did use all four acoustic correlates to distinguish stressed from unstressed syllables. Mandarin and English speakers' use of amplitude and duration were comparable for both stressed and unstressed syllables, but Mandarin speakers produced stressed syllables with a higher F0 than English speakers. There were also significant differences in formant patterns across groups, such that Mandarin speakers produced English-like vowel reduction in certain unstressed syllables, but not in others. Results suggest that Mandarin speakers' production of lexical stress contrasts in English is influenced partly by native-language experience with Mandarin lexical tones, and partly by similarities and differences between Mandarin and English vowel inventories.
© 2008 Acoustical Society of America. [DOI: 10.1121/1.2902165]

I. INTRODUCTION

Adults who learn a second language (L2) are seldom able to speak that language without accent. Although the degree of an accent is related to many factors such as age and language environment, the primary influence on the nature of an individual's accent is the sound system of their native language (L1) (Flege and Hillenbrand, 1987; Lord, 2005; Piske *et al.*, 2001; Tahta and Wood, 1981). The interference of native phonetics and phonology on the acquisition of non-native vowels and consonants has been studied extensively, and results typically suggest that L2 learners have relatively greater difficulty perceiving and producing non-native contrasts that involve phonetic features dissimilar to those used in their native language. Similar difficulties in L2 acquisition have been identified in suprasegmental domains as well. For example, native Mandarin speakers learning English as a second language have been repeatedly shown to have difficulties producing English lexical and/or sentential stress, and it has been argued that this difficulty may result in large part from the influence of native suprasegmental (tonal) categories (Archibald, 1997; Chen *et al.*, 2001a; Juffs, 1990; Hung, 1993). However, most research in this area has focused on impressionistic observations rather than acoustic analysis (with the notable exception of Chen *et al.*, 2001a) and often confounds the phonological issue of stress placement with the phonetic problem of native-like stress production.

Here we attempt to dissociate the question of whether, or to what degree, non-native speakers are able to apply phonological rules of stress placement, in order to focus on the question of whether they are able to correctly produce the phonetic properties that correlate with the English stress contrast under conditions in which they know unambiguously where stress is to be placed. Thus, we ask whether Mandarin speakers are capable of producing native-like patterns of fundamental frequency, intensity, duration and vowel formant frequencies associated with English stressed and unstressed syllables when there is no question of their knowing *where* to place stress. An inability to correctly produce these acoustic correlates of English stress under these circumstances would suggest that their native language experience with producing (and possibly perceiving) the specific acoustic cue patterns related to Mandarin phonetic categories (tonal and/or segmental) interferes with their ability to produce qualitatively different patterns of these same cues in the service of producing English lexical stress distinctions.

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

A. English stress

A number of studies have explored the acoustic correlates of lexical stress in American English (Beckman, 1986; Bolinger, 1958; Campbell and Beckman, 1997; Fry, 1955, 1958, 1965; Lieberman, 1960, 1975; Sluijter and van Heuven, 1996; Sluijter *et al.*, 1997). Most of these studies focused on lexical stress in English disyllabic words in which the location of stress on the first or second syllable led the word to be identified as either a noun or a verb, respectively. Results of these studies consistently indicate that the acoustic correlates of average fundamental frequency (F0), intensity, syllable duration, and vowel quality are associated with the perception and production of English lexical stress: Stressed syllables have higher F0, greater intensity, and longer duration than unstressed syllables. Moreover, recent research suggests that the *alignment* of F0 events with respect to segments within a syllable may play an important role in both tonal and intonational categories (For intonation, cf. Arvaniti and Gårding, 2007; Atterer and Ladd, 2004; Grabe *et al.*, 2000; Mennen, 2004. For tone, see Xu, 1998, 1999; Xu and Liu, 2007, 2006) and it may be worth investigating this property in the production of stress as well. Although, to our knowledge, pitch peak alignment has not been implicated as a specific cue to the placement of lexical stress, misalignment of a pitch peak in a stressed syllable might contribute to the perception of non-nativeness in L2 speakers.

The precise measure of computing intensity is debated. Fry (1955, 1958) and Beckman (1986) identified average intensity over the syllable as a possible acoustic correlate of stress differences, while others (Sluijter and van Heuven, 1996; Sluijter *et al.*, 1997) have argued that spectral tilt (differences in intensity over the frequency spectrum of a given vowel) is a more appropriate measure. Since both measures are associated with increased vocal effort (Liénard and Di Benedetto, 1999; Traunmüller, 1989), it is possible that either may serve as acceptable correlates of the English stress contrast. However, since measurement of spectral tilt is highly dependent on the height or location of the first formant (F1), it is not possible to compare spectral tilt across vowels differing in quality (formant frequencies), as between reduced (unstressed) and unreduced (stressed) versions of the same vowel, so in the current study only average intensity was used.

Finally, the process of vowel reduction has been consistently identified as a correlate of the English lexical stress contrast. Although this feature has not been extensively examined in cross-language studies, many researchers have discussed its importance in general terms. For example, nonnative speakers' use of unreduced vowels in unstressed syllables has been argued to "contribute importantly to foreign accent" (Flege and Bohn, 1989) and is an "extremely typical" phenomenon in Spanish-accented English (Hammond, 1986). Fokes *et al.* (1984, 1989) and Flege and Bohn (1989) also concluded that the inability of L2 speakers to perform appropriate vowel reduction contributed to their non-native-like production of English, although the two articles differed in their assessment of the relative importance of vowel reduction in cuing the perception of native-like stress. Fokes *et al.* (1984) suggested that the inability of L2 learners to reduce the vowel in unstressed syllables could influence their ability to manipulate other phonetic correlates of English lexical stress, resulting in poorer performance on lexical stress production tasks. In contrast, Flege and Bohn (1989) argued that L2 learners of English first learn to produce stressed vs unstressed syllables contrasting in duration and intensity, and only subsequently learn (or fail to learn) to correctly reduce the vowels in unstressed syllables. Either way, vowel quality is clearly an important acoustic correlate of stress (Beckman, 1986; Fry, 1965) and failure to appropriately reduce unstressed vowels may contribute to the perception of a non-native accent (Fokes *et al.*, 1984; Flege and Bohn, 1989; Lee *et al.*, 2006).

B. Mandarin lexical tone

Unlike English, Mandarin is a tonal language. There are four lexical tones in Mandarin: tone 1 (high-level), tone 2 (high-rising), tone 3 (dipping), and tone 4 (high-falling). Tone, like stress in English, can distinguish word meaning independently of segmental properties. Some scholars have argued that Mandarin exhibits linguistic characteristics that are similar to lexical stress. For instance, syllables carrying the so-called neutral tone, which is usually found in syntactic particles within lexical units of two or more syllables, have been found to be less prominent than syllables carrying the four basic lexical tones (Chao, 1968; Chen and Xu, 2006).

Many studies have focused on the acoustic examination of Mandarin tones (Howie, 1976; Fu *et al.*, 1998; Gandour, 1978, 1983; Liu and Samuel, 2004; Whalen and Xu, 1992). In general, these studies have demonstrated that F0 is the primary acoustic cue for Mandarin tones, but that syllable duration and amplitude contour vary consistently across lexical tone categories. For example, the falling tone (tone 4) is typically much shorter than the other tones, especially the first tone (high level) which is typically quite long. Similarly, the third (dipping) tone is long, but also exhibits a mid-syllable decrease in amplitude. Perceptual research has shown that these non-pitch cues can also function as acoustic cues to Mandarin tones in the absence of F0 information (Fu *et al.*, 1998; Liu and Samuel, 2004; Whalen and Xu, 1992). Thus, based on their experience with controlling the F0, duration, and intensity of individual syllables to express lexical tone distinctions, from a purely phonetic perspective, it is possible that Mandarin speakers may be able to control these same acoustic properties to produce native English-like lexical stress contrasts.

This seems unlikely, however, as research on cross-language perception and L2 production of speech sounds clearly indicates a strong influence of the native *phonological* system on the perception and production of non-native sounds, and only some Mandarin tones map clearly onto English intonational patterns (see Francis *et al.*, 2008, for discussion of cross-language mapping between Mandarin and Cantonese tones and English intonational categories). Interestingly, the specific nature of L1 category influence on L2 perception and production (in terms of facilitation or interference) also appears to depend in large part on the relative

J. Acoust. Soc. Am., Vol. 123, No. 6, June 2008

Zhang *et al.*: Mandarin English lexical stress   4499

degree of (phonetic featural) similarity between the native and non-native categories (Best, 1995; Flege, 1995; Flege and Davidian, 1985). For example, according to Flege's Speech Learning Model (SLM), the presence of one or more native categories that are phonetically similar to a non-native category may interfere with the perception and production or acquisition of that L2 category. In contrast, Best's Perceptual Assimilation Model (PAM) would predict improved perception of an L2 contrast if each sound is sufficiently similar to a different native category. Such a situation would result in *two-category* assimilation, whereby each sound in a non-native contrast is assimilated to a different native category. Even if both sounds of the L2 contrast are assimilated to the same native category, PAM predicts improved perception of the contrast if one of the two is more successfully assimilated (a case of a *category goodness* contrast).

More interestingly, according to PAM non-native sounds that are *uncategorizable* to any native phoneme category may be easy to discriminate perceptually (perhaps even more easily than for native speakers), while still being extremely difficult to produce in a native-like manner (Best *et al.*, 2001; Best *et al.*, 1988). However, this last possibility seems unlikely in the case of F0 patterns, since these, unlike clicks (the typical example of uncategorizable sounds) can easily be recognized as speech sounds. Still, depending on which theory one adopts, and, more importantly, on the specific degree of similarity between the native and the L2 category or categories, one might expect either an increase or decrease in ease of acquisition when an L2 category is determined to be similar to a non-native one along one or more phonetic dimensions. Although the SLM and PAM have traditionally been applied to production and perception of segmental phonemes, there is nothing about the models themselves that would necessarily restrict their predictions to the segmental domain, and either may be able to account for the acquisition of suprasegmental aspects of speech, such as intonation or stress.

C. Mandarin speakers' production of English stress

There is evidence that native Mandarin speakers have difficulty producing L2 English stress contrasts in a native-like manner. While it is possible that this difficulty arises from interference from the Mandarin sentential stress (intonational) system, existing evidence currently seems to suggest a strong interference from the Mandarin *tonal* system.[1] For example, Juffs (1990) reported errors made by native Chinese speakers who were college students and had little or no experience with spoken English outside the classroom. Many of these speakers' errors consisted of mistakes in stress placement, suggesting that they simply did not know what syllables required stress in the utterances they were asked to produce. However, even when stress was produced on the appropriate syllable, they showed evidence of difficulty with the phonetic manipulation of specific correlates of stress. For example, some speakers tended to use a falling tone to signal an English stressed syllable. The use of a falling tone, with its overall lower average F0, for a stressed syllable suggests that these speakers were not aware of the general association

between English stress and higher (average) F0, but may instead have been overextending the English tendency to use a sharply falling F0 contour for strongly emphatic stress (as in "*Yes*, I *do*") (Chao, 1972). Alternatively, it is possible that they were correctly recognizing that the English stressed syllable should be produced with a higher *initial* F0 value—in other words, they were focusing on the location of a pitch peak, rather than on an average syllable value (see discussion of F0 peak location, below). In contrast, other speakers did achieve an overall higher average pitch in stressed syllables, but also lengthened these syllables much more than was necessary. This suggests that these speakers simply superimposed *all* properties of the Mandarin high tone onto the English stressed syllable (including its association with very long syllable duration), rather than simply producing an overall higher average F0. Taken together, these results suggest that, even when Mandarin speakers know which syllable to stress, they may do so by transferring production patterns from their native tonal inventory.

To control for Mandarin speakers' lack of knowledge about where stress is to be placed, Chen *et al.* (2001a) examined the production of English sentence stress under conditions in which the speaker was clearly aware of the proper location of stress. They found that native Mandarin speakers employed many of the same acoustic correlates of stress as English speakers, including duration, amplitude, and fundamental frequency, but their use of these correlates was significantly different from American speakers. For example, Mandarin speakers produced stressed words with higher F0 compared to English speakers. Chen *et al.* (2001a) argued that this was a result of Mandarin speakers' native language experience, since Mandarin typically exhibits a much greater range of pitch fluctuation during the course of a sentence than does English. Thus, Mandarin speakers are used to producing high pitches at a higher point in their average pitch range than are English speakers, and this tendency transfers to the L2 as well. Although their results regarding F0, duration and intensity are very informative, Chen *et al.* (2001a) did not examine the possible influence of native phonology (whether tonal or segmental) on the production of L2 vowel quality as a cue to English stress.

The investigation of vowel quality is central to the present study, unlike previous work, because it is in this domain that we may begin to distinguish between interference that results from the fundamental difference between tone and stress *systems* and interference that arises from incomplete or inaccurate acquisition of individual lexical items. Interference of a systematic origin should be relatively uniform across lexical items, for example, leading to a uniform lack of vowel reduction or, conversely, a tendency to overgeneralize a principle of vowel reduction in unstressed syllables. Interference that arises on an item-by-item basis should, in contrast, be much more variable across items (Flege and Bohn, 1989).

The present study focused on three factors involved in the production of stress: (1) the acoustic correlates used by Mandarin and English speakers to indicate lexical stress placement in English, including F0, duration, intensity and vowel quality; (2) differences between the two groups in

terms of their use of these features; (3) the degree to which Mandarin speakers' pattern of acoustic correlate production can be explained by the structure of their native language phonology (both suprasegmental and segmental).

II. METHODS

A. Subjects

Two groups of speakers participated in this experiment: ten native speakers of American English (five women, five men) and ten native speakers of Mandarin Chinese (five women, five men). English participants ranged in age from 21 to 28 years of age ($M = 25$), while Mandarin speakers were 26–35 years of age ($M = 32$). The English speakers were all native residents of the United States (U.S.), while the Mandarin speakers were all originally from the People's Republic of China (PRC) and had lived in the U.S. for three to four years prior to participating in the experiment. All participants were recruited from within Purdue University community (West Lafayette, IN) and had normal hearing, speech, and language ability by self-report.

None of the Mandarin speakers had any English-immersion experience before arriving at Purdue University; all of their prior English experience was obtained in class while in China. None was enrolled in an English language department or school in China, although eight reported having had native English speakers as college English teachers at some point in their education. Since coming to the U.S., all Mandarin speakers had been exposed primarily to Midwestern dialects of American English. Of the American English speakers, seven were from the central Midwest (six from Indiana and one from Ohio, one of whom also spoke American-African English). There was also one American English speaker from each of California, New York, and Louisiana.

B. Stimuli

Seven pairs of disyllabic words were selected following the methodology of Beckman (1986) and Fry (1955, 1958). Each word pair consisted of a noun and a verb that had identical spelling forms and differed only in terms of stress placement (noun: stress on initial syllable; verb: stress on final syllable). These stimulus pairs were formed from the following corpus of word forms: *contract*, *desert*, *object*, *permit*, *rebel*, *record*, and *subject*. Each target word was elicited in isolation and in the semantically neutral frame sentence *I said __ this time* and was accompanied by associated context sentences created specifically for each word, which are shown in Table I.

Based on the work of Peterson and Barney (1952), ten familiar English words (*beat*, *bit*, *bet*, *bat*, *bought*, *father*, *bird*, *butt*, *put*, *boot*) were used to map English vowel spaces by native English speakers of America and by native Mandarin speakers of English. Similarly, a list of Chinese characters was selected for mapping the Mandarin speakers' Mandarin vowel space, as shown in Table II.

TABLE I. Stimuli and context sentences to aid in establishing the stressed syllable.

Target word	Noun/verb	Context sentence
Contract	noun	Mr. Smith has finally agreed to sign the new *contract*.
	verb	Will steel *contract* when it is cooled?
Desert	noun	They got lost in the *desert*.
	verb	Will he *desert* his team?
Object	noun	What is the *object* on the table?
	verb	They won't *object* to your decision.
Permit	noun	In order to park here, you need a *permit*.
	verb	Would you *permit* her request?
Rebel	noun	The *rebel* army did this.
	verb	They *rebelled* at this unwelcome suggestion.
Record	noun	Can I get a copy of my health *record*?
	verb	She *recorded* all songs her daughter sang yesterday.
Subject	noun	What is the *subject* of this sentence?
	verb	Must you *subject* me to this boring twaddle?

C. Procedure

Prior to recording, participants were asked to fill out a language background questionnaire. All recordings took place in a single-walled sound-attenuated booth and were made using a digital audio recorder (SONY DAT, TCD-D8), Studio V3 amplifier, and a unidirectional Hypercardiod dynamic microphone (Audio-Technica D1000HE).

The microphone was placed approximately 20 cm from the speaker's lips at an angle of 45° (horizontal) during recording. The speech tokens were sampled at a rate of 44.1 kHz with a quantization of 16 bits and low-pass filtered at 22.05 kHz. Each token was then saved as an individual sound file and normalized to a RMS amplitude of 70 dB using Praat 4.3 (Boersma and Weenink, 2004).

All stimuli were presented to speakers on individual file cards organized into three sets. One set of cards showed each word (target or distracter) at the top with the corresponding context sentence and frame sentence below. The second set of cards showed only target words and corresponding context sentences. The third set of cards showed only the English words and Chinese words for mapping vowel spaces.

Speakers were instructed to speak naturally at a typical rate and loudness level. Each speaker first read the first set of cards, context sentence first then the frame sentence, twice for each card. Before the next reading, the experimenter explained to the speaker the rule that stress needs to be shifted between syllables when some English words shift from noun to verb (e.g., CONtract vs conTRACT). The need for this type of stress shift to differentiate noun from verb for some English words should be familiar to the participants, because it is part of the standard middle school English class curriculum in the PRC. For the second set of recordings, speakers read only the target words in isolation. Target word pronunciation was indicated by referring to the context sentence. Again, each card was read twice. This elicitation procedure yielded 1120 tokens (14 words × 2 contexts × 2 repetitions × 20 subjects). Only the 560 tokens produced in isolation were used in subsequent analyses (both instrumental and perceptual) since each production is assumed to represent the speaker's best attempt to produce stress on the appropriate

TABLE II. All monophthong and diphthong vowel phonemes involved in this experiment, including corresponding Chinese characters and English words used in the vowel space mapping task. Note: (ü) indicates that this transcription is used when vowel is produced in isolation. Transcriptions based on those in Duanmu (2000) with the substitution of the IPA symbol [ɐ] for [A].

Mandarin vowels (IPA)	Chinese character	PinYin romanization	Tone	English vowels (IPA)	English word
i	一	yi	55	i	beat
iɛ	爷	ye	25	*	
yɛ	约	yue	55	*	
iɐ	鸭	ya	55	*	
y	迂	yu (ü)	55	*	
*	*	*	*	I	bit
ei	悲	bei	55	ei	
*	*	*	*	ɛ	bet
*	*	*	*	æ	bat
ai	哀	ai	55	ai	
au	熬	ao	25	*	
*	*	*	*	ɑ	father
*	*	*	*	ʌ	butt
*	*	*	*	ɔ	bought
*	*	*	*	ɝ	bird
ɤ	鹅	e	25	*	
o	播	bo	55	*	
ou	欧	ou	55	ou	
*	*	*	*	ʊ	put
u	屋	wu	55	u	boot
uo	窝	wo	55	*	
uɐ	挖	wa	55	*	
ɐ	阿	a	55	*	

syllable (initial for nouns, final for verbs). Moreover, one production could not be analyzed, leaving a total of 559 stress-contrasting tokens. Finally, all speakers read the list of English vowel space-mapping words, and Mandarin speakers read the list of Chinese characters.

D. Acceptability rating

Subjective ratings of acceptability or accentedness are commonly used in the evaluation of a speaker's foreign ac-

cent (Flege, 1984, 1988; Southwood and Flege, 1999). Such ratings are obtained by asking native listeners to assign a numeric value to a segment of speech based on its perceived quality (Francis and Nusbaum, 1999; Schmidt-Nielsen, 1995). To determine the acceptability of each recorded token, a listening evaluation test was conducted. Five native English-speaking graduate students in the linguistics or English as a second language program of Purdue University served as paid consultants. Linguistically trained listeners

were selected because of the increased likelihood that they would be able to focus on stress characteristics alone, ignoring other possible non-native (segmental) pronunciations in the speech samples. Each listener evaluated the acceptability of each of the 559 tokens on five separate occasions over a two-week period. Words were presented randomly but blocked by speaker gender.

For each token, listeners first heard the word and were asked to determine which word was said. Both possible choices for each word (e.g., conTRACT or CONtract) were displayed on the screen prior to playing the sound and remained on the screen until a choice was made. After listeners identified the token a new screen appeared showing their choice (e.g., either CONtract or conTRACT) and asked them to provide a rating of acceptability on a scale from 1 (poor) to 5 (excellent). The sound was repeated after this second screen was displayed, but the screen did not clear until a choice had been selected. Token presentation and data collection was carried out using *E-prime* version 1.1 (Schneider *et al.*, 2002).

E. Acoustic measurements

Using Praat acoustic analysis software (Boersma and Weenink, 2004), the following acoustic parameters were measured for each token: syllable duration (in ms); average intensity (in dB); average fundamental frequency (F0, in Hz); time of F0 peak and the first and second formant frequencies (F1 and F2, in Hz). The parameters related to intensity and F0 were measured within a syllable, and the formant frequencies were measured within the vowel. Only F1 and F2 measures were used to map the speakers' vowel space.

Syllable and vowel boundaries were segmented according to the following criteria: (1) word/syllable 1 onset: The first upward-going zero crossing at the beginning of the waveform; (2) word/syllable 2 offset: The ending point of the sound waveform at the last downward-going zero crossing; (3) syllable 1 offset/syllable 2 onset: In words with a stop consonant as the onset of the second syllable (such as *rebel, contract, object, subject, record*), this was defined at the beginning of the silence of the stop gap. In words with no medial stop consonant (*permit, desert*), then the boundary was marked as the transition between the acoustic (spectrographic) pattern of the initial consonant of the second syllable and the segment immediately preceding it. Segmentation criteria were based on both waveform and spectrogram cues as described by Peterson and Lehiste (1960). Based on these segmentations, syllable and vowel durations were calculated in millisecond increments. In addition, for diphthongal vowels (i.e., in Mandarin), formant frequencies were measured twice, once for the initial vocalic portion and once for the final portion. For this purpose, the transition point between the two vowel segments was visually identified as the midpoint of the transition between the two steady states, or the midpoint between the initial formant frequencies and the final ones, in the absence of any steady state. Average formant values were calculated between the onset of the vowel and this midpoint (for the initial vocalic portion) and between this midpoint and the end of the vowel (for the final vocalic portion).

The average intensity measure was calculated as the mean of multiple intensity values extracted and smoothed over the number of time points necessary to capture the minimum predicted pitch of each individual participant. F0 measures were measured as the average value over the entire syllable, and were computed using a Hanning analysis window and the autocorrelation method described in Boersma (1993). When measuring F0, the pitch range for female talkers was set to 100–500 Hz and 75–300 Hz for male talkers, as recommended in the Praat manual. The time of the F0 peak was identified automatically from the F0 contour, and subsequently converted to a proportion of the syllable by reference to the syllable duration. F0 was remeasured manually (as the reciprocal of each manually identified period of the syllable's acoustic waveform) when the pitch contour was absent, or displayed incompletely or intermittently through the syllable, and when displayed F0 values were suspiciously high or low compared to the rest of that talker's utterances. In most cases, these display problems were due to the presence of glottalization, especially in unstressed syllables produced by female American English and male Chinese speakers.

A linear predictive coding (LPC) based tracking algorithm was used to determine formant calculations for the entire vocalic segment of interest [as implemented in the Praat Sound to LPC (burg) method]. The LPC analysis employed a 25 ms Gaussian window with +6 dB pre-emphasis over 50 Hz. These computed formant frequencies were then averaged across the entire vowel, or, in the case of the diphthong, across the initial or final portion of the diphthong, respectively. In order to quantify the property of vowel quality, we used two measures derived from the center frequencies of the first and second formants (F1 and F2) as described by Blomgren *et al.* (1998). The statistic compact-diffuse (C-D), calculated as the difference between F1 and F2 (F2 − F1), is correlated with the phonetic property of tongue height. High vowels such as [i] and [u] typically have a relatively large C-D value, while low vowels such as [a] have a smaller C-D value. The statistic grave-acute (G-A), calculated as the arithmetic mean of F1 and F2 [(F1 + F2)/2], is correlated with the phonetic dimension of tongue advancement (front/back), such that front vowels such as [i] or [æ] typically have a relatively small value of G-A, while back vowels such as [u] or [o] typically have relatively large values.

III. RESULTS

A. Acceptability ratings

Listeners correctly identified the majority of tokens produced by both English and Mandarin speakers. The five tokens that were identified incorrectly by more than two listeners were excluded from further analysis. The mean acceptability rating for each of the remaining tokens was then calculated only across raters who correctly identified the token (all but 11 tokens were correctly identified by all listeners), as shown in Table III. Raters were relatively uniform

TABLE III. Results of perceptual evaluation of productions by American English and Mandarin speakers. Note: Accuracy = proportion of correct identifications; Avg = mean acceptability rating across 3–5 raters (see text) on a five-point scale where 1 = poor and 5 = excellent; s.d. = standard deviation for each mean rating.

| | English Speakers' Productions | | | Mandarin Speakers' Productions | | |
| | Identification | Rating | | Identification | Rating | |
Word	Accuracy	Avg	s.d.	Accuracy	Avg	s.d.
Contract N	0.99	4.60	0.18	1.00	3.47	0.38
Contract V	0.99	4.43	0.51	0.98	2.77	0.74
Desert N	1.00	4.29	0.47	0.95	2.88	0.78
Desert V	0.99	4.51	0.19	0.99	3.19	0.60
Object N	1.00	4.44	0.29	0.97	3.17	0.59
Object V	0.99	4.30	0.45	0.97	3.22	0.45
Permit N	0.94	4.15	0.59	0.94	2.72	0.62
Permit V	0.96	3.97	0.53	1.00	2.67	0.47
Rebel N	1.00	4.64	0.22	0.98	1.88	1.14
Rebel V	0.97	4.32	0.67	0.99	3.14	0.27
Record N	0.99	4.59	0.30	0.98	2.89	0.85
Record V	0.93	4.19	0.97	0.99	3.14	0.32
Subject N	0.98	4.21	0.49	0.99	3.61	0.60
Subject V	0.97	4.07	0.51	0.98	2.98	0.43
Average	0.98	4.34	0.46	0.98	2.98	0.59

in their assessment of both the English and Mandarin utterances. The mean range between the lowest and highest acceptability rating for a given word was 1.8 overall (1.6 for English productions, 1.9 for Mandarin). The mean rating score for correctly identified words produced by Mandarin speakers was 2.98 ($SD=0.74$, $Mdn=3.04$), while for the American group it was 4.34 ($SD=0.53$, $Mdn=4.49$). Most Mandarin speakers' productions were rated less than 3.5 (204 out of 277 tokens), but the majority of English speakers' productions scored higher than 4 (256 out of 276 tokens). A t-test showed that the rating difference between the two language groups was statistically significant, $t(551)=24.97$, $p<0.001$.

B. Acoustic analyses

To confirm the reliability of our acoustic measurements, 10% of all tokens (56) were selected for independent re-analysis by a second judge who was naive to the purpose of the experiment. Across raters, mean formant values differed by at most 25 Hz for F2 and 12 Hz for F1, mean F0 measures differed by no more than 3 Hz, and mean vowel and syllable durations differed by no more than 16 ms. Pearson's product moment correlation analysis of the two sets of measurements showed a strong correlation of at least $r=0.95$ and $p<0.001$ for all measures except the duration of the second syllable ($r=0.77$, $p<0.001$) and the location of the F0 peak within the second syllable ($r=0.88$, $p<0.001$). The comparatively poor correlation for measures involving the duration of the second syllable appears to derive from differences in the identification of the end of the syllable in cases in which the burst release was difficult to differentiate from background noise.

Using the originally measured values for each acoustic variable, a mixed factorial analysis of variance (ANOVA) was performed with native language and gender as between-subjects variables and stress (stressed or unstressed) as the within-subjects factor. All post hoc (Tukey HSD) tests were performed with a critical p value of 0.05. Means for each measure for each group, gender, and stress condition are given in Table IV.

1. Average F0

Results of the analysis of average F0 showed significant main effects of stress [$F(1,16)=148.19$, $p=0.001$], native language [$F(1,16)=15.73$, $p=0.001$], and gender [$F(1,16)=164.23$, $p<0.001$]. There were significant interactions between stress and language [$F(1,16)=12.42$, $p=0.003$] and gender and stress [$F(1,16)=21.09$, $p<0.001$]. The three-way interaction was not significant [$F(1,16)=0.41$, $p=0.53$]. The significant effect of gender was expected: The mean average F0 was 229 Hz for females and 176 Hz for males. Post hoc (Tukey HSD) tests showed that, for each language group, the F0 of the stressed syllables, averaged across males and females, was significantly higher than that of the unstressed syllables (Mandarin: stressed = 198 Hz, unstressed = 163 Hz; American: stressed = 164 Hz, unstressed = 145 Hz). In addition, in stressed syllables Mandarin speakers produced significantly higher F0 than English speakers, but not in unstressed syllables (averaged across genders: Mandarin: stressed = 198 Hz; American: stressed 164 Hz). Thus, the language-group difference (Mandarin > American English) is purely the result of Mandarin speakers producing stressed syllables with significantly higher F0 than do American English speakers.

TABLE IV. Mean scores and standard deviations for all acoustic parameters for English stressed and unstressed syllable produced by native Mandarin and English speakers. Note: STR = stressed, UNSTR = unstressed. Each cell contains mean value with standard deviation in parentheses.

| | Mandarin | | | | English | | | |
| | Male | | Female | | Male | | Female | |
	STR	UNSTR	STR	UNSTR	STR	UNSTR	STR	UNSTR
F0 (Hz)	145	121	252	205	122	111	206	178
	(13)	(15)	(14)	(13)	(18)	(22)	(17)	(12)
Peak F0 loc. (%)	47	38	45	30	47	45	42	39
	(4)	(8)	(6)	(2)	(8)	(7)	(6)	(4)
Intensity (dB)	65	60	65	60	68	63	65	61
	(2)	(3)	(1)	(2)	(1)	(1)	(1)	(1)
Syllable duration (ms)	337	267	365	287	291	216	367	283
	(23)	(15)	(51)	(44)	(26)	(19)	(50)	(41)

2. Peak F0 location

There were significant effects of stress [$F(1,16)=18.18$, $p<0.001$] and of gender [$F(1,16)=10.38$, $p=0.005$], but not of language [$F(1,16)=3.45$, $p=0.079$]. There was a significant interaction between stress and native language [$F(1,16)=5.09$, $p=0.038$], but not between stress and gender [$F(1,16)=0.92$, $p=0.35$], or native language and gender [$F(1,16)=0.05$, $p=0.81$], and the three-way interaction was also not significant [$F(1,16)=0.32$, $p=0.58$]. *Post hoc* tests showed that for Mandarin speakers the location of peak F0 in stressed syllables was significantly different from the location of peak F0 in unstressed syllable ($p=0.003$, with the stressed location at 46% of the syllable and unstressed at 34%). In other words, Mandarin speakers produced the F0 peak location significantly earlier in unstressed syllables than that in stressed ones. For American English speakers, the difference in F0 peak location between stressed and unstressed syllables was not significant.

In addition, the F0 peak location of the stressed syllable in trochaic (strong-weak pattern) and in iambic (weak-strong) structure was also compared, because it was shown that English speakers tended to produce the peak F0 earlier in the stressed syllable in iambic words than in trochaic words (Munson *et al.*, 2003). A mixed factorial ANOVA was performed with native language and gender as between-subjects variables and with structure (trochee or iamb) as within-subject factor, and the F0 peak location of the stressed syllable as the dependent variable. Results showed a significant effect of structure [$F(1,16)=63.93$, $p<0.001$], but no significant effect of native language [$F(1,16)=0.66$, $p=0.43$], or gender [$F(1,16)=2.31$, $p=0.15$]. There was a significant interaction between native language and structure [$F(1,16)=12.5$, $p=0.003$], as well as between gender and structure [$F(1,16)=5.591$, $p=0.03$], but there was no significant three-way interaction [$F(1,16)=0.61$, $p=0.45$]. *Post hoc* tests showed that both Mandarin and American English speakers produced the F0 peak of the stressed syllable earlier in iambic words than that in trochaic words (Mandarin: trochaic=61%, iambic=32%; English: trochaic=50%, iambic=39%).

3. Intensity

Analysis of average intensity showed a significant effect of stress [$F(1,16)=259.85$, $p<0.001$], and language group [$F(1,16)=10.19$, $p=0.006$]. Gender did not show a main effect [$F(1,16)=2.29$, $p=0.149$], and none of the interactions were shown to be significant. *Post hoc* tests showed that for both language groups, stressed syllables (Mandarin: 65 dB; American: 67 dB) had a significant higher intensity than unstressed syllables (Mandarin: 60 dB; American: 62 dB). Interestingly, although the main effect of language group was significant, indicating that the intensity of speech produced by American English speakers was, on average, two dB higher than Mandarin speakers, *post hoc* analysis showed no significant difference between the intensities of either Mandarin and American English stressed syllables or those of unstressed syllables.

4. Duration

Results of the analyses of syllable durations showed significant effects of stress [$F(1,16)=380.68$, $p<0.01$] and gender [$F(1,16)=9.2$, $p=0.008$], but no effect of language [$F(1,16)=2.48$, $p=0.135$], and no significant interactions. Men produced syllables averaging 277 ms, while women's syllables averaged 325 ms. *Post hoc* tests showed that for both language groups stressed syllables had a significantly longer duration (Mandarin: 351 ms; American: 329 ms) than unstressed syllables (Mandarin: 277 ms; American: 250 ms).

5. Vowel space

Figure 1 shows the English, Mandarin and Mandarin-English vowel spaces, averaged across both male and female talkers (only peripheral vowels are shown). Both the Mandarin English and American English vowel spaces are roughly quadrilateral, consistent with the results of Chen *et al.*, 2001b. However, there are slight differences in the location of specific vowels between the two groups of speakers. In particular, the production of English [u] by native Mandarin speakers is farther "back" (in the sense of having lower F2) compared to the American English [u]. It has been documented that the American English production of [u] is often characterized by a higher F2 than similar phoneme produc-

J. Acoust. Soc. Am., Vol. 123, No. 6, June 2008

Zhang *et al.*: Mandarin English lexical stress   4505

FIG. 1. Comparison of three vowel spaces of American English, Mandarin Chinese, and Mandarin English.

tions in many other languages (for examples, compare vowel charts for various languages presented in IPA, 1999), which may be the result of a more advanced tongue placement during articulation. Such production differences may be the process of historical change, as suggested by Hillenbrand *et al.*'s (1995) comparison of their data with those of Peterson and Barney (1952). This hypothesis is supported by the observation that the present measurements of the F2 of [u] are even higher in frequency (more fronted) than those of Hillenbrand *et al.* (1995), which are in turn higher than those of Peterson and Barney (1952): 1406, 1051, and 910 Hz, respectively, averaged over men and women. Of course, some of this difference may be due to the much smaller number of participants in the present study leading to greater potential influence of inter-individual differences in absolute vocal tract size, as well as possible dialect differences between the participants in the Peterson and Barney (East coast of the U.S.), Hillenbrand *et al.* (central Michigan) and current (central In-

diana) studies (see also Harigawa, 1997, for data from Southern Californian English, with even more evidence of fronting of back vowels). Still, Fig. 1 shows that Mandarin speakers are attempting to approximate the more fronted American English [u] (as compared with their native [u]), although they do not achieve it perfectly. Despite these minor differences, the observation that the Mandarin English and American English vowel spaces share an overall similar structure suggests that Mandarin speakers' native vowel system does not interfere very much with production of English-like *stressed* vowels, at least when words are produced in isolation. Further analyses were carried out to investigate the production of stressed and unstressed vowels in the target disyllabic words.

6. Vowel reduction

For each syllable in each word, separate ANOVAs were conducted for both the C-D (Compact-Diffuse, related to the phonological feature contrast high/low) and G-A (Grave-Acute, related to the phonological feature contrast front/back) variables with three factors: gender, language, and stress. Significant results are shown in Table V, with a bold font indicating the significance of the *post-hoc* (Tukey HSD) test at the $p < 0.05$ level. In addition, F1 and F2 values, averaged across male and female participants, are shown in Table VI.

For most syllables, stressed vowels did not show a significant difference between Mandarin and American English speakers. The exceptions were *-mit (permit)* and *re-* (*rebel*). The main differences between Mandarin and English speakers were found in their productions of unstressed syllables, and most of these differences appeared in the C-D (Compact-Diffuse) feature with the exception of the word *rebel* in which the unstressed versions of both the initial syllable *re-* and the final syllable *-bel* showed significant differences be-

TABLE V. Statistically significant pairwise comparisons between formant measures for stressed and unstressed vowels, by syllable. Note: C-D refers to the compact-diffuse dimension (F2-F1); G-A refers to grave-acute (arithmetic mean of F1 and F2). S refers to stressed syllables, U to unstressed, AE to American English speakers' productions, and M to Mandarin speakers'. Thus, for example, E<M indicates that English speakers' productions of a given syllable showed smaller mean values of a given acoustic feature than did Mandarin speakers'.

| | Stressed/Unstressed | | | | American English/Mandarin | | | |
| | AE | | M | | STRESSED | | UNSTRESSED | |
Syllable	C-D	G-A	C-D	G-A	C-D	G-A	C-D	G-A
con-	**S < U**						**AE > M**	
-tract			**S < U**					
de-	**S < U**		**S < U**				**AE < M**	
-sert								
ob-	**S < U**						**AE > M**	
-ject								
per-								
-mit			**S > U**		**AE < M**		**AE < M**	
re-		**S > U**			**AE < M**		**AE < M**	**AE < M**
-bel	**S > U**	**S > U**		**S > U**			**AE < M**	**AE < M**
re-		**S > U**	**S < U**				**AE < M**	
-cord	**S < U**	**S < U**	**S < U**				**AE > M**	
sub-								
-ject	**S < U**		**S < U**					

Zhang *et al.*: Mandarin English lexical stress

TABLE VI. Average F1 and F2 values in Hz across male and female native speakers of Mandarin Chinese and American English.

| | Stressed | | | | Unstressed | | | |
| | English | | Mandarin | | English | | Mandarin | |
Syllable	F1	F2	F1	F2	F1	F2	F1	F2
con-	844	1495	828	1423	564	1819	694	1519
-tract	817	1768	789	1726	773	1768	750	1756
de-	610	1807	638	1880	452	1938	412	2159
-sert	583	1691	594	1644	582	1765	599	1687
ob-	873	1396	797	1368	676	1583	706	1346
-ject	680	1886	690	1867	627	1884	625	1884
per-	710	1496	670	1462	661	1507	610	1480
-mit	633	1997	480	2325	659	1925	543	2162
re-	694	1587	583	1812	530	1632	516	1905
-bel	717	1618	725	1735	624	1235	683	1559
re-	716	1723	674	1736	524	1774	470	1898
-cord	622	1378	711	1227	617	1761	665	1471
sub-	687	1551	762	1547	617	1654	620	1528
-ject	685	1893	683	1864	619	1912	600	1908

tween Mandarin and English speakers in terms of both the C-D (Compact-Diffuse) and G-A (Grave-Acute) features. Overall, five general patterns can be distinguished:

Type 1. *Correct non-reduction*. Neither English nor Mandarin speakers reduced the vowel in the following unstressed syllables (no significant differences were found for either C-D or G-A): *per- (permit), -sert (desert), sub- (subject)*, and *-ject (object)*.

Type 2. *Unexpected reduction*. Unlike English speakers, Mandarin speakers significantly reduced unstressed vowels (in terms of either C-D or G-A) in the following words: *-tract (contract)* and *-mit (permit)*.

Type 3. *Incorrect reduction*. In these syllables, both English and Mandarin speakers showed significant differences between stressed and unstressed vowels, but the unstressed vowel used by Mandarin speakers was in each case significantly different (in terms of either C-D or G-A, or both) from its English counterpart. These syllables include: *de- (desert), -bel (rebel), re- (record)*, and *-cord (record)*.

Type 4. *Lack of reduction*. Unlike the English speakers, Mandarin speakers did not show a significant change in either the C-D or G-A features from stressed to unstressed versions of the following syllables: *con- (contract), ob- (object)*, and *re- (rebel)*.

Type 5. *Correct reduction*. The only syllable in which both American and Mandarin speakers appear to show a similar degree and quality of vowel reduction is the syllable *-ject (subject)*.

In order to evaluate possible strategies Mandarin speakers may have used in the production of the English unstressed vowels, the average formant values for each vowel were converted to Bark scale values. These values were used to compute Euclidean distances for each stressed or unstressed vowel produced in the experimental words and those from the vowel space mapping task (mapping vowels). These distance measures are shown in Tables VII–X.

Although these tables are quite complex, a few general patterns may be observed from them. Table VII shows which of the Mandarin speakers' mapping vowels are closest to the vowel in a given syllable, while Table IX does the same for English speakers. Comparing the stressed syllables in these two tables shows that Mandarin and English speakers employed approximately the same vowel categories for stressed syllables in many cases. For example, both groups' productions of the vowel in the stressed syllable *con- (contract)* and *ob- (object)* syllables were closest to their productions of [ɑ] in the mapping task, and both produced *de- (desert)* with a vowel most similar to [ɛ].

Comparison of the distance between Mandarin speakers productions and *English* speakers' mapping vowels (Table VIII) with that in Table VII also helps elucidate some more ambiguous cases, such as the nearly equivalent distance between Mandarin speakers' productions of the stressed *-ject* syllable and their mapping vowels [ɛ] and [æ]. Given the overall similarity of these vowels and the very small difference between the two distances, such productions may still be acceptable, and, indeed, as shown in Table VIII, Mandarin speakers' productions of *-ject* are clearly closest to English speakers' [ɛ] mapping vowel which suggests that this syllable is being produced with a vowel that would be clearly identifiable to English speakers as [ɛ] rather than [æ].

With respect to unstressed vowels, the situation is more complex. In some cases, such as the unstressed syllable *de- (desert)*, Mandarin speakers' productions were closest to a vowel in their own English mapping vowel productions (Table VII) that corresponded to the English speakers' mapping vowel closest to English speakers' productions of this syllable ([ɪ]). However, the Mandarin production of this vowel was significantly different from that of native English speakers as shown in Table V, column 7, suggesting that the two mapping task vowels must have been quite different (see also the greater magnitude of the distance between the Mandarin speakers' production of this syllable and the English speakers' mapping vowel [ɪ] as shown in Table VIII).

J. Acoust. Soc. Am., Vol. 123, No. 6, June 2008

Zhang *et al.*: Mandarin English lexical stress   4507

TABLE VII. Euclidean distance in F1×F2 space between Mandarin speakers' stressed and unstressed vowels in the word production task and Mandarin speakers' productions of English vowels in the vowel space mapping task. Note: Smallest distance indicated in bold.

		i	I	ɛ	æ	ɑ	ɔ	ʌ	ɝ	ʊ	u
					English vowels (Mandarin speakers)						
con-	S	5.26	4.09	2.22	1.92	**0.23**	2.09	1.07	2.13	3.52	3.90
	U	4.26	3.07	1.39	1.40	0.82	2.12	0.10	1.41	2.88	3.39
-tract	S	4.33	3.23	1.19	**0.66**	1.21	3.10	1.19	2.47	3.93	4.47
	U	4.04	2.93	0.90	**0.43**	1.36	3.14	1.15	2.37	3.81	4.38
de-	S	3.08	1.99	**0.09**	0.71	2.13	3.51	1.55	2.36	3.63	4.28
	U	0.95	**0.35**	2.20	2.81	4.11	4.89	3.34	3.40	4.01	4.71
-sert	S	3.34	2.15	**0.89**	1.36	1.75	2.65	0.94	1.40	2.69	3.33
	U	3.25	2.06	**0.71**	1.22	1.81	2.82	1.03	1.58	2.86	3.50
ob-	S	5.27	4.08	2.32	2.11	**0.32**	1.75	1.00	1.88	3.23	3.59
	U	4.89	3.69	2.17	2.17	0.78	1.36	**0.72**	1.23	2.58	2.97
-ject	S	3.46	2.39	0.38	**0.31**	1.89	3.48	1.46	2.46	3.82	4.44
	U	2.98	1.89	0.17	0.81	2.21	3.55	1.60	2.36	3.61	4.26
per-	S	4.31	3.12	1.60	1.69	0.92	1.84	**0.21**	1.12	2.58	3.09
	U	3.94	2.75	1.52	1.80	1.40	1.95	**0.59**	0.82	2.26	2.83
-mit	S	1.41	**1.03**	2.01	2.50	4.04	5.13	3.37	3.71	4.55	5.25
	U	2.01	**1.19**	1.30	1.78	3.33	4.53	2.69	3.18	4.18	4.87
re-	S	2.85	1.69	**0.48**	1.16	2.21	3.31	1.49	2.02	3.20	3.87
	U	2.20	**1.03**	1.04	1.71	2.86	3.77	2.09	2.35	3.30	4.00
-bel	S	4.33	3.23	1.19	**0.66**	1.21	3.10	1.19	2.47	3.93	4.47
	U	4.04	2.93	0.90	**0.43**	1.36	3.14	1.15	2.37	3.81	4.38
re-	S	3.60	2.45	**0.51**	0.65	1.53	2.98	0.97	1.98	3.37	3.97
	U	1.92	**0.72**	1.43	2.11	3.16	3.89	2.36	2.41	3.19	3.89
-cord	S	5.37	4.18	2.76	2.76	1.19	**0.83**	1.31	1.38	2.49	2.74
	U	4.26	3.06	1.55	1.66	0.96	1.88	**0.20**	1.11	2.58	3.09
sub-	S	4.57	3.40	1.51	1.27	**0.54**	2.36	0.60	1.89	3.36	3.84
	U	3.83	2.64	1.30	1.58	1.36	2.15	**0.52**	1.04	2.46	3.04
-ject	S	3.41	2.34	**0.32**	0.37	1.91	3.46	1.44	2.43	3.78	4.40
	U	2.77	1.67	**0.38**	1.02	2.38	3.64	1.74	2.38	3.57	4.24

In other cases, Mandarin speakers' productions of unstressed vowels did not pattern with those of native American English speakers. For example, for the unstressed *con-* (contract), American English speakers recorded here used a vowel similar to [ɪ] as in *bit* (Table IX), but Mandarin subjects used [ʌ] as in *butt* (Table VII).[2] One possible explanation for this is that Mandarin speakers may have substituted a native short, central vowel, [ɤ], for the similar English [ʌ], and this argument is supported by the observation that, as shown in Table X, [ɤ] is indeed the closest Mandarin monophthong to the vowel in the unstressed *con-* syllable. However, this distance (1.18 Bark) is considerably larger than the distance between the vowel in the unstressed *con-* syllable and Mandarin speakers' production of English [ʌ] (0.10 Bark). This pattern of results is more consistent with the hypothesis that Mandarin speakers chose an English vowel as their target for this syllable, but, unlike the case of *de-* discussed above, the vowel that they chose was different from that chosen by the native speakers in this study (the possibility that this native production may have been nonstandard is discussed below).

Finally, sometimes Mandarin speakers seem to have tried but failed to produce sufficiently distinctive versions of stressed and unstressed vowels. For example, in the syllable *ob- (object)*, both Mandarin and American English speakers produced vowels similar to [ɑ] in stressed productions and [ʌ] in unstressed ones (Tables VII and IX). However, there was no significant difference in Mandarin speakers' stressed and unstressed vowels in terms of either the C-D or G-A dimensions (Table V). This pattern can be explained by examining the relative distance between the vowel in unstressed *ob-* and [ɑ], which was 0.78 for Mandarin speakers (Table VII), compared with 0.72 as a distance from [ʌ], and 1.51 for English speakers (Table IX), compared with 0.34 for [ʌ]. In other words, the Mandarin production of unstressed *ob-* was nearly equidistant between [ɑ] and [ʌ], while English speakers' productions were much closer to [ʌ] than to [ɑ], suggesting that Mandarin speakers were aware that they needed to produce a different vowel in the unstressed as compared to the stressed context, but were either not sure what that vowel should be or, perhaps, were simply unable to realize it to a sufficiently clear degree.

IV. DISCUSSION

Native Mandarin speakers were able to produce lexical stress contrasts that were correctly identified by linguistically trained native speakers of American English. Subsequent acoustic analyses indicated that both native English and native Mandarin speakers used the acoustic correlates of F0, intensity and duration in a similar manner: Both groups pro-

TABLE VIII. Euclidean distance in F1×F2 space between Mandarin speakers' stressed and unstressed vowels in the word production task and English speakers' productions of English vowels in the vowel space mapping task. Note: Smallest distance indicated in bold.

		English vowels (Mandarin speakers)									
		i	I	ϵ	æ	ɑ	ɔ	ʌ	ɝ	ʊ	u
con-	S	5.14	3.27	2.12	1.78	**0.33**	0.79	0.96	1.98	1.81	3.36
	U	4.26	2.38	1.48	1.76	1.23	1.41	**0.14**	0.94	0.77	2.46
-tract	S	4.00	2.20	0.91	**0.68**	1.63	2.04	0.96	1.69	1.56	3.37
	U	3.73	1.92	**0.64**	0.83	1.81	2.18	0.94	1.49	1.38	3.18
de-	S	2.84	0.98	**0.45**	1.58	2.59	2.87	1.47	1.25	1.25	2.77
	U	1.34	**1.24**	2.53	3.65	4.55	4.71	3.37	2.53	2.68	2.85
-sert	S	3.44	1.61	1.24	2.09	2.16	2.30	1.01	**0.29**	0.32	1.93
	U	3.30	1.45	1.09	1.99	2.24	2.41	1.07	**0.46**	0.49	2.06
ob-	S	5.23	3.35	2.28	2.07	**0.24**	0.51	0.96	1.90	1.73	3.15
	U	4.99	3.13	2.27	2.38	0.88	**0.78**	0.85	1.45	1.30	2.51
-ject	S	3.13	1.32	**0.11**	1.18	2.35	2.68	1.32	1.41	1.37	3.03
	U	2.75	0.89	**0.55**	1.68	2.67	2.94	1.53	1.24	1.26	2.71
per-	S	4.39	2.53	1.74	2.08	1.25	1.33	**0.44**	0.89	0.73	2.24
	U	4.13	2.31	1.78	2.35	1.72	1.74	0.81	0.51	**0.39**	1.79
-mit	S	0.93	**0.96**	2.23	3.25	4.50	4.74	3.35	2.70	2.81	3.41
	U	1.62	**0.27**	1.51	2.56	3.79	4.05	2.65	2.11	2.20	3.11
re-	S	2.81	0.96	0.93	2.02	2.66	2.87	1.49	**0.91**	0.97	2.28
	U	2.28	**0.65**	1.45	2.58	3.29	3.47	2.12	1.34	1.47	2.24
-bel	S	4.00	2.20	0.91	**0.68**	1.63	2.04	0.96	1.69	1.56	3.37
	U	3.73	1.92	**0.64**	0.83	1.81	2.18	0.94	1.49	1.38	3.18
re-	S	3.43	1.56	**0.61**	1.37	1.99	2.27	0.87	0.98	0.90	2.65
	U	2.21	**0.92**	1.85	2.98	3.58	3.72	2.42	1.53	1.68	2.07
-cord	S	5.54	3.70	2.86	2.91	1.04	**0.66**	1.45	1.94	1.81	2.68
	U	4.34	2.48	1.70	2.07	1.30	1.38	**0.44**	0.84	0.67	2.21
sub-	S	4.43	2.56	1.43	1.38	1.00	1.34	**0.37**	1.41	1.25	2.97
	U	3.96	2.13	1.55	2.16	1.73	1.81	0.70	0.41	**0.24**	1.92
-ject	S	3.10	1.28	**0.16**	1.23	2.37	2.69	1.32	1.37	1.33	2.98
	U	2.58	**0.70**	0.75	1.89	2.84	3.10	1.69	1.26	1.31	2.63

duced stressed syllables with a higher F0, longer duration and greater intensity than unstressed syllables.

However, these productions were still rated as significantly less acceptable than those of native English speakers, suggesting that the Mandarin speakers in this study produced stress contrasts with a discernable accent. Acoustically, differences between Mandarin and English speakers' production of stressed and unstressed syllables were noted, specifically in terms of the properties of average F0, F0 peak location, intensity, and vowel reduction. Mandarin speakers produced English stressed syllables with significantly higher F0 than did American speakers. Moreover, Mandarin speakers produced F0 peaks significantly earlier in the unstressed syllable than in stressed syllable, while English speakers showed no difference in F0 peak timing between stressed and unstressed syllables. In addition, Mandarin speakers were, on average, about 2 dB less intense, overall, than were English speakers, but it is unlikely that this difference, in itself, contributed significantly to the perception of non-nativeness in their production of the English stress contrast. Finally, Mandarin speakers showed a tendency to either not reduce, or incorrectly reduce vowels in unstressed syllables requiring vowel reduction. In general, these findings are consistent with the hypothesis that, although native Mandarin speakers are able to control certain acoustic correlates in an English-like manner to signal stress, they are not able to manage F0 and vowel quality in a strictly English-like manner due to interference from their native tonal system and vowel systems respectively.

With respect to the observed group differences in average F0, the present results are consistent with the results and conclusions of Chen et al. (2001a), who has argued that such behavior derives from tone language speakers' experience with using a larger proportion of their overall frequency range as compared to speakers of nontonal languages (Chen, 1974): Mandarin high tones are produced with an F0 at a much higher proportion of the talker's overall pitch range compared to English stress (see also Shen, 1989 and Adams and Munro 1978 for corroborative results). Therefore, although Mandarin speakers are able to transfer the use of F0 from the tonal domain to that of lexical stress, they are still strongly influenced by the native (tonal) domain within which they are used to manipulating this property. Thus, the acoustic property of F0 cannot be considered an independent feature to be manipulated at will, but rather must be controlled as part of the speakers' native language phonology.

Similarly, although analysis of the peak F0 location indicates that both American English and Mandarin speakers produced the peak F0 earlier in the stressed syllable in words with iambic stress than in words with trochaic stress, consistent with the findings of Munson et al. (2003), the two groups differed in terms of their location of peak F0 in

TABLE IX. Euclidean distance in F1×F2 space between English speakers' stressed and unstressed vowels in the word production task and English speakers' productions of English vowels in the vowel space mapping task. Note: Smallest distance indicated in bold.

		i	I	ɛ	æ	ɑ	ɔ	ʌ	ɝ	ʊ	u
						English vowels (Mandarin speakers)					
con-	S	4.96	3.11	1.89	1.45	**0.66**	1.13	0.95	2.00	1.83	3.49
	U	2.72	**0.90**	1.08	2.17	2.79	2.98	1.61	0.94	1.03	2.20
-tract	S	4.03	2.29	0.97	**0.43**	1.77	2.21	1.21	1.93	1.81	3.62
	U	3.81	2.02	0.72	**0.67**	1.81	2.21	1.04	1.65	1.54	3.34
de-	S	2.93	1.06	**0.74**	1.81	2.50	2.74	1.34	0.92	0.95	2.42
	U	2.06	**0.96**	2.01	3.15	3.79	3.93	2.63	1.73	1.88	2.17
-sert	S	3.23	1.40	1.17	2.11	2.35	2.50	1.19	**0.45**	0.52	1.97
	U	2.97	1.13	1.01	2.05	2.54	2.73	1.37	**0.73**	0.80	2.15
ob-	S	5.42	3.56	2.37	1.89	**0.36**	0.81	1.28	2.31	2.13	3.66
	U	3.95	2.07	1.21	1.67	1.51	1.72	**0.34**	0.77	0.61	2.40
-ject	S	3.03	1.22	**0.13**	1.26	2.45	2.77	1.40	1.42	1.39	3.02
	U	2.77	0.91	**0.54**	1.66	2.64	2.92	1.51	1.23	1.25	2.71
per-	S	4.40	2.53	1.58	1.77	1.07	1.27	**0.17**	1.09	0.91	2.56
	U	4.18	2.32	1.55	1.97	1.40	1.52	**0.39**	0.73	0.56	2.21
-mit	S	2.50	0.72	**0.60**	1.69	2.96	3.26	1.86	1.62	1.64	3.03
	U	2.82	1.01	**0.31**	1.44	2.64	2.95	1.57	1.46	1.45	2.99
re-	S	4.01	2.13	1.19	1.56	1.43	1.67	**0.26**	0.90	0.75	2.53
	U	3.31	1.62	1.66	2.60	2.58	2.65	1.49	**0.45**	0.62	1.48
-bel	S	4.03	2.29	0.97	**0.43**	1.77	2.21	1.21	1.93	1.81	3.62
	U	3.81	2.02	0.72	**0.67**	1.81	2.21	1.04	1.65	1.54	3.34
re-	S	3.66	1.81	**0.66**	1.11	1.79	2.11	0.77	1.21	1.11	2.90
	U	2.76	1.08	1.45	2.53	2.93	3.07	1.77	**0.89**	1.03	1.86
-cord	S	4.59	2.79	2.19	2.61	1.52	1.43	0.98	0.97	**0.86**	1.86
	U	3.11	1.24	0.79	1.78	2.34	2.56	1.17	**0.78**	0.79	2.35
sub-	S	4.11	2.24	1.34	1.70	1.35	1.56	**0.19**	0.86	0.70	2.44
	U	3.48	1.62	1.10	1.91	2.04	2.21	0.87	0.43	**0.39**	2.11
-ject	S	3.03	1.24	**0.09**	1.22	2.45	2.79	1.42	1.47	1.43	3.06
	U	2.66	0.79	**0.60**	1.73	2.76	3.03	1.63	1.31	1.34	2.74

stressed as compared to unstressed syllables. Mandarin speakers reached their peak F0 significantly earlier in unstressed syllables than in stressed syllables, while the American English speakers showed no difference in peak F0 timing between syllable types. Xu (1998, 1999; Xu and Liu, 2006) examined the peak F0 location in Chinese syllables across different lexical tones, finding a positive correlation between syllable duration and the location of the F0 peak. Longer syllables were found to have a later F0 peak relative to the syllable onset. In the present study, Mandarin speakers produced English unstressed syllables with significantly shorter durations than stressed syllables. This duration difference may have caused Mandarin speakers to incorrectly alter peak F0 timing.[3] Once again, it appears that, although Mandarin speakers are able to select F0 as a cue to be manipulated in the service of producing English lexical stress differences, they may only do so according to the linguistic conventions commonly used within their native language.

A. Vowel reduction

To examine vowel reduction, productions of vowels in stressed and unstressed syllables were referenced against productions of monosyllabic (stressed) vowels in the vowel space mapping task (Tables VII–X). Based on these comparisons, it appears that Mandarin speakers showed a great deal of similarity with English speakers in both their stressed and some unstressed vowel productions. In particular, for the majority of vowels used in the stressed syllable, Mandarin speakers employed approximately the same vowel categories as the English speakers. In agreement with this observation, the difference between most Mandarin and English stressed syllables was statistically insignificant (Table V, fifth and sixth columns), supporting the hypothesis that Mandarin speakers do not have significant difficulty learning to produce American English full (unreduced) monophthongal vowels.

Mandarin speakers' productions of unstressed vowels were also frequently comparable to those of English speakers. For example, in Type 1(*Correct nonreduction*) syllables such as *per- (permit)*, *-sert (desert)* and *sub- (subject)*, Mandarin speakers correctly did not reduce the vowel, just as American English speakers did not, while in Type 3(*Incorrect reduction*) syllables such as *de- (desert)* and *bel- (rebel)*, and in Type 5(*Correct reduction*) syllables such as *-ject (subject)*, Mandarin speakers reduced the vowel just as American English speakers did. However, in the Type 3(*Incorrect reduction*) cases (e.g., *de-* in the verb *desert*), although Mandarin speakers were not successful in achieving the English unstressed vowel quality, they did attain formant values that were comparable to their (accented) productions of the same vowels that were used by the native English speakers in the corresponding unstressed syllable. In other words, they ap-

TABLE X. Euclidean distance in F1×F2 space between Mandarin speakers' stressed and unstressed vowels in the word production task and Mandarin speakers' productions of Mandarin monophthongal vowels in the vowel space mapping task. Note: Smallest distance indicated in bold.

		\textipa{5}	o	γ	i	u	y
		Mandarin vowels (Mandarin speakers)					
con-	S	**0.63**	2.19	1.94	5.02	4.90	5.00
	U	1.65	2.18	**1.18**	4.07	4.70	3.98
-tract	S	**1.61**	3.18	2.24	3.97	5.78	4.12
	U	**1.86**	3.21	2.13	3.68	5.75	3.82
de-	S	2.77	3.56	**2.15**	2.74	5.84	2.87
	U	4.85	4.88	3.30	**0.87**	6.46	0.75
-sert	S	2.56	2.68	**1.19**	3.19	4.87	3.05
	U	2.60	2.85	**1.37**	3.07	5.05	2.97
ob-	S	**0.87**	1.85	1.71	5.06	4.57	4.98
	U	1.53	1.44	**1.07**	4.76	4.08	4.58
-ject	S	2.46	3.53	**2.24**	3.08	5.93	3.26
	U	2.86	3.59	**2.15**	2.65	5.83	2.77
per-	S	1.78	1.90	**0.89**	4.16	4.39	4.01
	U	2.25	1.98	**0.58**	3.85	4.27	3.63
-mit	S	4.70	5.15	3.58	**0.85**	6.97	1.38
	U	3.97	4.56	3.02	**1.56**	6.56	1.89
re-	S	2.95	3.34	**1.83**	2.61	5.47	2.60
	U	3.61	3.78	2.21	**2.01**	5.68	1.94
-bel	S	**1.61**	3.18	2.24	3.97	5.78	4.12
	U	**1.86**	3.21	2.13	3.68	5.75	3.82
re-	S	2.21	3.03	**1.75**	3.31	5.44	3.36
	U	3.94	3.89	2.30	1.84	5.61	**1.62**
-cord	S	1.71	**0.92**	1.32	5.28	3.64	5.05
	U	1.82	1.94	**0.88**	4.11	4.42	3.96
sub-	S	**1.24**	2.43	1.67	4.30	5.07	4.31
	U	2.21	2.18	**0.80**	3.71	4.49	3.53
-ject	S	2.49	3.51	**2.21**	3.04	5.90	3.21
	U	3.06	3.67	**2.19**	2.45	5.84	2.55

peared to be aiming for the appropriate reduced vowel target, but missed producing it with the expected F1 and F2 values in the same way that they missed producing that target vowel when it was the target in a stressed monosyllable (in the vowel space mapping condition). In other words, Mandarin speakers' poor performance on vowel reduction in the present experiment appears to be due to an inability to correctly produce *specific* reduced vowels, and some of this may be related to their incorrect production of those vowels even in stressed contexts (e.g., the vowel space mapping task).

One explanation for this difficulty is interference from the native vowel system or, more properly, the *lack* of a sufficiently similar vowel in the Mandarin system leading to particularly inaccurate productions in a manner consistent with the results of Flege *et al.* (1997), who found that Mandarin speakers showed the least spectral accuracy when producing English vowels, including [ɪ], that are not found in Mandarin. Similarly, Chen *et al.* (2001b) showed that [ɪ], an "unfamiliar vowel" to Mandarin speakers, was pronounced less accurately than other vowels that were familiar to Mandarin speakers (that is, acoustically more similar to native Mandarin vowels). In particular, as in the present study, Chen *et al.* (2001b) showed that female speakers of Mandarin produced [ɪ] with a lower F1 than that of female speakers

of American English, while male speakers of Mandarin produced [ɪ] with a higher F2 than that of male speakers of American English. Thus, difficulties with native-like production of [ɪ] seem to be characteristic of Mandarin speakers' production of English, in a manner independent of the issue of lexical (or sentential) stress production.

In other cases, Mandarin speakers seem to have chosen a different target vowel than did the English speakers, as in the case of unstressed *con-*, where Mandarin speakers produced a vowel very similar to their [ʌ] mapping vowel, but English speakers produced a vowel more similar to their mapping vowel productions of ([ɪ]). Since the first syllable of the verb *contract* is quite commonly produced with the vowel [ʌ] in many varieties of English, it is quite possible that the Mandarin speakers in this study were in fact successfully approximating a native-like pronunciation of this word, albeit one that differed from the native pronunciation in the local dialect. The degree to which Mandarin (or any other non-native) speakers' perceived non-nativeness may derive from their (successfully) attaining an English target appropriate to a different English dialect than that of their listeners is an interesting and important sociolinguistic question, and deserves further exploration although it is beyond the scope of the present study.

Again, however, the fact that Mandarin speakers produced clearly different vowel qualities in the stressed and unstressed versions of the same syllable supports the hypothesis that they are capable of employing vowel change as a cue to lexical stress, at least in some cases. On the other hand, in other cases, Mandarin speakers did not appear to reduce unstressed vowels significantly, even though English speakers did show clear vowel reduction (e.g., in the syllable *ob-* in the word *object*). As described above, the behavior of this syllable can be explained in terms of Mandarin speakers' failure to correctly produce the reduced vowel [ʌ]. Examination of the two groups' vowel spaces (Fig. 1) showed that Mandarin speaker's productions of English [ɑ] and [ʌ] were each quite close to the American English [ɑ] and [ʌ] when producing the (stressed) words *father* and *butt*, respectively. Thus, Mandarin speakers should in principle have been able to produce both the stressed and unstressed versions of *ob-* accurately, and clearly moved in the expected (native-like) direction, but may not have managed the change with sufficient clarity. Indeed, in all cases in which American English speakers showed significant differences in formant frequencies between stressed and unstressed syllables and Mandarin speakers did not [*con-*, *ob-*, and *re-*(*bel*), see Table V], there are still some small differences observable in Mandarin speakers' productions, at least in terms of there being a difference in mapping vowel that is closest to the stressed as compared to the unstressed vowel (Table VII). The appearance of unexpected reductions (significant differences between stressed and unstressed vowel formant patterns for Mandarin but not English speakers), as in the syllables *-tract* and *-mit*, further supports the hypothesis that Mandarin speakers are aware of, and attempt to make use of, formant frequency differences to cue lexical stress differences.

J. Acoust. Soc. Am., Vol. 123, No. 6, June 2008

Zhang *et al.*: Mandarin English lexical stress   4511

V. CONCLUSION

In conclusion, it appears that Mandarin speakers are able to successfully approximate English-like patterns of duration and intensity when producing stress contrasts, as well as some of the native-like patterns of F0 production. Moreover, when their pattern of performance on these cues diverged from that of native English speakers, it did so in a manner consistent with the transfer of properties characteristic of the Mandarin tonal system. In contrast, Mandarin speakers, although clearly aware of the importance of vowel reduction as a cue to stress, had much more difficulty with this cue, but the precise pattern of difficulty was not systematic, and appeared to vary across the linguistic context or vowel category. This observation is consistent with the proposal of Flege and Bohn (1989), who suggested that L2 learners acquire L1 stress patterns for individual words. For instance, the pattern for the noun *object* might be learned at a different time than that of the verb *object*. The present results suggest further that learners might acquire the individual *cues* to stress based on the lexical item or vowel category, at least with respect to the cue of vowel reduction. Since Mandarin speakers were successful at producing English-like cues for duration, intensity, and to a limited extent F0, it is difficult to determine whether they learned to produce these cues systematically, whether they have simply already learned these cues for the specific words examined here, or whether transfer from their native suprasegmental phonological system was sufficient to achieve native-like patterns in the L2. Further research is needed to investigate the contribution of the observed non-English-like F0 patterns, such as the stressed syllables produced at F0 values that are too high and with a different alignment of F0 peaks within the syllable, to the perception of foreign accent in Mandarin speakers of English. In addition, it would be of interest to examine the relative contribution of the various cues examined in this study to the perception of stress in English.

ACKNOWLEDGMENTS

[1]The study of Mandarin intonation is still in its infancy, and is complicated by its interaction with tone. While the general consensus seems to be that Mandarin does possess at least a minimal set of intonational patterns that are independent from, but interact with, the tonal properties of a given utterance, there is considerable disagreement about the nature of the proposed system and the quality and degree of its interaction with lexical tone (Chao, 1968; Ho, 1977; Gårding, 1984; Kratochvil, 1998; Shen, 1990; see Schack, 2000, for review). This topic is far beyond the scope of the present article.

[2]In fact, it appears that English speakers produced a vowel in this context that is more or less equidistant from [ɪ], [ɛ], [ɝ], and [ʊ], though marginally closer to [ɪ]. The presence of the following [n] and concomitant nasalization of the preceding vowel may have complicated measurement of this vowel, and dialectal differences may have skewed these measures

toward [ɪ] and away from the expected [ʌ]. However, the main point remains, namely that Mandarin speakers did not produce the same unstressed vowel as did native English speakers.

[3]It is not yet known whether this timing difference contributes to the perception of non-native accent in the Mandarin speakers' productions, though recent research on peak timing in Mandarin tone production and cross-dialectal differences in F0 peak timing suggests that it might (Arvaniti and Garding, 2007; Atterer and Ladd, 2004; Grabe *et al.*, 2000; Mennen, 2004). We are currently carrying out perceptual investigations to explore this issue.

Adams, C., and Munro, R. (**1978**). "In search of the acoustic correlates of stress: Fundamental frequency, amplitude, and duration in the connected utterance of some native and non-native speakers of English," Phonetica **35**, 125–156.

Archibald, J. (**1997**). "The acquisition of English stress by speakers of non-accentual languages: Lexical storage versus computation of stress," Linguistics **35**, 167–181.

Arvaniti, A., and Gårding, G. (**2007**). "Dialectal variation in the rising accents of American English," in edited by J. Cole and J. Hualde *Laboratory Phonology*, (Mouton de Gruyter, Berlin), Vol. **9**.

Atterer, M., and Ladd, D. R. (**2004**). "On the phonetics and phonology of 'segmental anchoring' of F0: Evidence from German," J. Phonetics **32**, 177–197.

Beckman, M. E. (**1986**). *Stress and Non-stress Accent* (Foris, Dordrecht).

Best, C. T. (**1995**). "A direct realistic view of cross-language speech perception," in *Speech Perception and Linguistic Experience: Issues in Cross-language Research*, edited by W. Strange (York, Baltimore), pp. 171–204.

Best, C. T., McRoberts, G. W., and Goodell, E. (**2001**). "American listeners' perception of nonnative consonant contrasts varying in perceptual assimilation to English phonology," J. Acoust. Soc. Am. **1097**, 775–794.

Best, C. T., McRoberts, G. W., and Sithole, N. M. (**1988**). "Examination of perceptual reorganization for non-native speech contrasts: Zulu click discrimination by English-speaking adults and infants," J. Exp. Psychol. Hum. Percept. Perform. **4**, 45–60.

Blomgren, M., Robb, M., and Chen, Y. (**1998**). "A note on vowel centralization in stuttering and nonstuttering individuals," J. Speech Lang. Hear. Res. **41**, 1042–1051.

Boersma, P. (**1993**). *Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. IFA Proceedings [Proceedings of the Institute of Phonetic Sciences, Amsterdam]* **17**, 97–110. Downloaded from http://www.fon.hum.uva.nl/Proceedings/IFA-Proceedings.html. Last accessed May 3, 2007.

Boersma, P., and Weenink, D. (**2004**). http://www.fon.hum.uva.nl/praat/. Last accessed March 26, 2007.

Bolinger, D. L. (**1958**). "A theory of pitch accent in English," Word **14**, 109–119.

Campbell, N., and Beckman, M. (**1997**). "Stress, prominence, and spectral tilt," in *Proceedings of ESCA Workshop on Intonation: Theory, Models and Applications*, edited by A. Botinis, G. Kouroupetroglou, and G. Carayiannis, Athens, pp. 67–70.

Chao, Y. R. (**1968**). *A Grammar of Spoken Chinese* (University of California Press, Berkeley, CA).

Chao, Y. (**1972**). *Mandarin Primer* (Harvard University Press, Cambridge, MA).

Chen, G. T. (**1974**). "The pitch range of English and Chinese speakers," J. Chin. Linguist. **2**, 159–171.

Chen, Y., Robb, M. P., Gilbert, H. R., and Lerman, J. W. (**2001a**). "A study of sentence stress production in Mandarin speakers of American English," J. Acoust. Soc. Am. **4**, 1681–1690.

Chen, Y., Robb, M. P., Gilbert, H. R., and Lerman, J. W. (**2001b**). "Vowel production by Mandarin speakers of English," Clin. Linguist. Phonetics **6**, 427–440.

Chen, Y., and Xu, Y. (**2006**). "Production of weak elements in speech—evidence from f0 patterns of neutral tone in standard Chinese," Phonetica **63**, 47–75.

Duanmu, S. (**2000**) The Phonology of Standard Chinese, Oxford university Press, Oxford, England.

Flege, J. E. (**1984**). "The detection of French accent by American listeners," J. Acoust. Soc. Am. **3**, 692–707.

Flege, J. E. (**1988**). "Factors affecting degree of perceived foreign accent in English sentences," J. Acoust. Soc. Am. **1**, 70–79.

Flege, J. E. (**1995**). "Second language speech learning: Theory, findings, and problems," in *Speech Perception and Linguistic Experience: Issues in*

Cross-language Research, edited by W. Strange (York, Baltimore), pp. 233–277.

Flege, J. E., and Bohn, O. S. (**1989**). "An instrumental study of vowel reduction and stress placement in Spanish-accented English," Stud. Second Lang. Acquis. **11**, 35–62.

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 Davidian, R. (**1985**). "Transfer and developmental processes in adult foreign language speech production," Appl. Psycholinguist. **5**, 323–347.

Flege, J. E., and Hillenbrand, J. (**1987**). "Limits on phonetic accuracy in foreign language production," in *Interlanguae Phonology: the Acquisition of a Second Language Sound System*, edited by G. Ioup and S. Weinberger (Newbury House, Cambridge), pp. 176–201.

Fokes, J. E., Bond, Z. S., and Steinberg, M. (**1984**). "Patterns of word stress by native and non-native speakers," in *Proceedings of the Tenth International Congress of Phonetic Sciences*, edited by M. Van den Broecke and A. Cohen (Foris, Dordrecht), pp. 682–686.

Fokes, J., and Bond, Z. S. (**1989**). "The vowels of stressed and unstressed syllables in Nonnative English," Lang. Learn. **3**, 341–373.

Francis, A. L., and Nusbaum, H. C. (**1999**). "Evaluating the quality of synthetic speech," in *Human Factors and Voice Interactive Systems*, edited by D. Gardner-Bonneau (Kluwer, Boston), pp. 63–97.

Francis, A. L., Ciocca, V., Ma, L., and Fenn, K. (2008). "Perceptual learning of Cantonese lexical tones by tonal and non-tonal language speakers," J. Phonetics, published online 13 February, 2008.

Fry, D. B. (**1955**). "Duration and intensity as physical correlates of linguistic stress," J. Acoust. Soc. Am. **27**, 765–768.

Fry, D. B. (**1958**). "Experiments in the perception of stress," Lang Speech **1**, 126–152.

Fry, D. B. (**1965**). "The dependence of stress judgments on vowel formant structure," in *Proceedings of the 5th International Congress of Phonetics Sciences*, eds.X. Zwerner, and W. Bethge, Karger: Basel, pp. 306–311.

Fu, Q. J., Zeng, F. G., Shannon, R. V., and Soli, S. D. (**1998**). "Importance of tonal envelope cues in Chinese speech recognition," J. Acoust. Soc. Am. **1**, 505–510.

Gandour, J. (**1978**). "The perception of tone," in *Tone: A Linguistic Survey*, edited by V. Fromkin (Academy, New York), pp. 41–76.

Gandour, J. (**1983**). "Tone perception in far eastern languages," J. Phonetics **11**, 149–175.

Gårding, Eva. (**1984**). "Chinese and Swedish in a generative model of intonation," in *Nordic Prosody III, Papers from a Symposium* edited by C. C. Elert, I. Johansson, and E. Strangert (Almqvist and Wiksell, Stockholm), pp. 79–91.

Grabe, E., Post, B., Nolan, F., and Farrar, K. (**2000**). "Pitch accent realization in four varieties of British English," J. Phonetics **28**, 161–185.

Hammond, R. H. (**1986**). "Error analysis and the natural approach to teaching foreign languages," Lenguas Modernas **13**, 129–139.

Harigawa, R. (**1997**). "Dialect variation and formant frequency: The American English vowels revisited," J. Acoust. Soc. Am. **1**, 655–658.

Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **5**, 3099–3111.

Ho, Aichen T. (**1977**). "Intonation variation in a Mandarin sentence for three expressions: Interrogative, exclamatory and declarative," Phonetica **34**, 446–457.

Howie, J. (**1976**). *Acoustical Studies of Mandarin Vowels and Tones* (Cambridge University Press, Cambridge).

Hung, T. T. N. (**1993**). "The role of phonology in the teaching of pronunciation to bilingual students," Language, Culture and Curriculum **3**, 249–256.

International Phonetic Association (**1999**). *Handbook of the International Phonetic Association* (Cambridge University Press, Cambridge).

Juffs, A. (**1990**). "Tone, syllable structure and interlanguage phonology: Chinese learner's stress errors," Int. Rev. Appl. Linguistics **2**, 99–117.

Kratochvil, P. (**1998**). "Intonation in Beijing Chinese," in *Intonation Systems: A Survey of Twenty Languages*, edited by D. Hirst and A. DiCristo (Cambridge University Press, Cambridge, MA), pp. 417–431.

Lee, B., Guion, S. G., and Harada, T. (**2006**). "Acoustic analysis of the production of unstressed English vowels by early and late Korean and Japanese bilinguals," Stud. Second Lang. Acquis. **28**, 487–513.

Lieberman, P. (**1960**). "Some acoustic correlates of word stress in American English," J. Acoust. Soc. Am. **32**, 451–454.

Lieberman, P. (**1975**). *Intonation, Perception and Language* (M.I.T. Press, Cambridge, Massachusetts).

Liénard, J. S., and DiBenedetto, M. G. (**1999**). "Effects of vocal effort on spectral properties of vowels," J. Acoust. Soc. Am. **1**, 411–422.

Liu, S., and Samuel, A. G. (**2004**). "Perception of Mandarin lexical tones when f0 information is neutralized," Lang Speech **47**, 109–138.

Lord, G. (**2005**). "(How) can we teach foreign language pronunciation? On the effects of a Spanish phonetics course," Hispania–A journal devoted to the teaching of Spanish and Portuguese **3**, 557–567.

Mennen, I. (**2004**). "Bi-directional interference in the intonation of Dutch speakers of Greek," J. Phonetics **32**, 543–563.

Munson, B., Bjorum, E. M., and Windsor, J. (**2003**). "Acoustic and perceptual correlates of stress in nonwords produced by children with suspected developmental apraxia of speech and children with phonological disorder," J. Speech Lang. Hear. Res. **46**, 189–202.

Peterson, G. E., and Barney, H. L., (**1952**). "Control methods used in a study of the vowels," J. Acoust. Soc. Am. **24**, 175–184.

Peterson, G. E., and Lehiste, L. (**1960**). "Duration of syllable nuclei in English," J. Acoust. Soc. Am. **32**, 693–703.

Piske, T., MacKay, I. R. A., and Flege, J. E. (**2001**). "Factors affecting degree of foreign accent in an L2: A review," J. Phonetics **2**, 191–215.

Schack, K. (**2000**). "Comparison of intonation patterns in Mandarin and English for a particular speaker," in University of Rochester Working Papers in the Language Sciences, edited by K. M. Crosswhite and J. McDonough, Spring 2000, pp. 24–55. Available online at http://www.bcs.rochester.edu/cls/s2000n1/schack.pdf. Last accessed October 1, 2007.

Schmidt-Nielsen, A. (**1995**). "Intelligibility and acceptability testing for speech technology," in *Applied Speech Technology*, edited by A. K. Syrdal, R. W. Bennett, and S. L. Greenspan (CRC press, Boca Raton, FL), pp. 195–232.

Schneider, W., Eschman, A., and Zuccolotto, A. (**2002**). *E-Prime User's Guide*. (Psychology Software Tools Inc., Pittsburgh).

Shen, X. S. (**1989**). "Toward a register approach in teaching Mandarin tones," J. Chin. Lang. Teachers Assoc. **24**, 27–47.

Shen, X.-N. S. (**1990**). *The Prosody of Mandarin Chinese*, University of California Publications in Linguistics (University of California Press, Berkeley, CA), Vol. **118**.

Sluijter, A. M. C., and Heuven, V. J. (**1996**). "Spectral balance as an acoustic correlate of linguistic stress," J. Acoust. Soc. Am. **4**, 2471–2485.

Sluijter, A. M. C., Heuven, V. J., and Pacilly, J. J. A. (**1997**). "Spectral balance as a cue in the perception of linguistic stress," J. Acoust. Soc. Am. **1**, 503–513.

Southwood, M. H., and Flege, J. E. (**1999**). "Scaling foreign accent: Direct magnitude estimation versus interval scaling," Clin. Linguist. Phonetics **5**, 335–349.

Tahta, S., and Wood, M. (**1981**). "Foreign accents: Factors relating to transfer of accent from the first language to a second language," Lang Speech **3**, 265–272.

Traunmüller, H. (**1989**). "Articulatory dynamics of loud and normal speech," J. Acoust. Soc. Am. **85**, 295–312.

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

Xu, Y. (**1998**). "Consistency of tone-syllable alignment across different syllable structures and speaking rates," Phonetica **55**, 179–203.

Xu, Y. (**1999**). "Effects of tone and focus on the formation and alignment of F0 contours," J. Phonetics **27**, 55–105.

Xu, Y., and Liu, F. (2006). "Tonal alignment, syllable structure and coarticulation: Toward an integrated model," Italian J. Ling. **18**, 125–159.