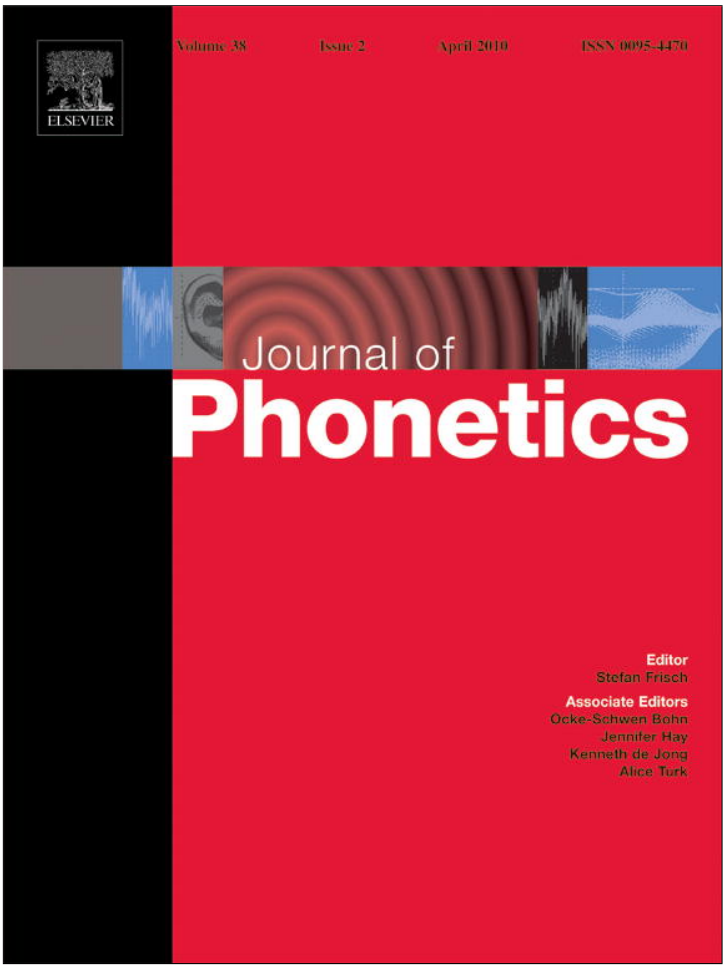

Journal of Phonetics 38 (2010) 260–271

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/phonetics

The weighting of vowel quality in native and non-native listeners' perception of English lexical stress

Yanhong Zhang [a,*], Alexander Francis [b]

[a] Program in Linguistics, Purdue University, West Lafayette, IN 47907, USA
[b] Program in Linguistics and Department of Speech, Language and Hearing Sciences, Purdue University, West Lafayette, IN 47907, USA

ARTICLE INFO

Article history:
Received 23 April 2009
Received in revised form
12 November 2009
Accepted 23 November 2009

ABSTRACT

Although vowel quality is an important cue to the perception of English lexical stress, few studies have examined the role this cue plays for non-native speakers. Previous research found that Mandarin speakers had problems using vowel reduction as a cue in English lexical stress production. Assuming native-like perception is a prerequisite to native-like production for non-native speech, this study compared the weight Mandarin learners and native speakers of English gave to vowel quality as a cue to English lexical stress in comparison to that given to f0, duration and intensity under natural and flat pitch contour conditions. Listeners judged lexical stress placement in synthesized tokens of the word *desert*, in which the first syllable *de-* was varied systematically in vowel quality and each of the other cues depending on the pair of cues in focus. Results showed that both English and Mandarin listeners consistently weighted vowel quality more than other cues. Vowel quality and duration were treated as a combinational cue by both language groups. However, Mandarin and English listeners showed different patterns for the processing of vowel quality and other prosodic cues, and Mandarin listeners' processing of vowel quality and pitch cues was influenced by the pitch contour conditions. These findings can be explained in terms of language-specific or cue-specific influence, and provide new insight into the relationship between production and perception in second language speech learning.

1. Introduction

Lexical stress has different properties in different languages. In languages like English and Spanish, stress is contrastive in that words can vary in only the location of stress, such as *contract* (first syllable stressed: verb; second syllable stressed: noun), whereas stress in other languages like French is positionally fixed (i.e., occurring always on the last non-schwa syllable within a word). Native experience with a particular stress pattern can result in difficulties when trying to learn the stress pattern of a different language. For example, in a series of stress studies, Dupoux and colleagues (Dupoux, Pallier, Sebastian, & Mehler, 1997; Dupoux, Peperkamp, & Sebastian-Galles, 2001) found that French listeners had difficulties in discriminating Spanish stress contrasts, and claimed that native listeners of languages with a positional stress system could experience "stress deafness" when exposed to a contrastive stress language. Subsequently, Dupoux, Sebastian-Galles, Navarrete, and Peperkamp (2008) suggested that non-native listeners' stress deafness might result from their inability to encode contrastive stress in their phonological representation. That is, the problem may be fundamentally linguistic, not perceptual. Nevertheless, perceptual factors may still play a significant role in explaining patterns of cross-language perception of stress because, in these studies, Dupoux and colleagues only varied pitch used to cue stress location. Considering that English stress is generally cued by multiple acoustic features such as intensity, fundamental frequency (f0), duration and vowel quality (Fry, 1955, 1958, 1965), it is not clear whether French speakers might have exhibited a different pattern of responses had other acoustic cues been included in the study.

A number of studies have suggested that listeners' non-native experience with a particular acoustic cue (rather than with a particular phonological category per se) might contribute to their difficulty in producing and/or perceiving prosodic units in a second language. For example, although English and Spanish both possess contrastive stress, vowel quality differences are associated with stress in English but not in Spanish, and native Spanish speakers have been found to have problems using vowel quality to signal English lexical stress (Flege & Bohn, 1989). One way to explore the possibility that experience with a *cue* matters more than experience with stress as such would be to investigate stress perception by native speakers of a language that uses specific acoustic features that may serve as cues to stress in the

* Corresponding author. Tel.: +1 609 423 0826.
E-mail addresses: yzhang648@gmail.com (Y. Zhang), francisa@purdue.edu (A. Francis).

0095-4470/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.wocn.2009.11.002

target language, but employs them as cues to other phonological categories than lexical stress. For example, Mandarin is a tonal language, and lexical tones in Mandarin are cued by pitch, duration and intensity, but not by vowel quality. If experience with specific cues matters more than experience with stress in and of itself, this would predict that Mandarin learners of English should exhibit difficulty in using vowel quality, but not pitch, duration or intensity, as cues to the perception of English lexical stress. The purpose of the present study is to examine the function of vowel quality in native Mandarin speakers' perception of English stress contrasts.

English lexical stress is acoustically manifest in multiple dimensions, especially pitch (f0), vowel duration, intensity, and vowel quality (Beckman, 1986; Fry, 1955, 1958, 1965; Sluijter & van Heuven, 1996a, 1996b). Inappropriate manipulation of any or all of these correlates can lead to poor discrimination of the stress contrast (e.g., Zhang, Nissen, & Francis, 2008). In particular, non-native speakers of English often have problems with vowel reduction related to production of unstressed syllables, which affects intelligibility and may contribute to the perception of foreign accent (Flege & Bohn, 1989; Fokes & Bond, 1989; Hammond, 1986; Zhang et al., 2008). In Zhang et al. (2008), native Mandarin speakers were asked to produce two-syllable English words differing only in stress position (first or second syllable), e.g. *contract*, *record*. Results indicated that native Mandarin speakers either did not reduce the vowel or did not use an appropriate reduced vowel in many unstressed syllables, although they were relatively good at manipulating the other acoustic correlates of stress (duration, intensity and average f0). It is possible that Mandarin listeners' difficulties with producing reduced vowels to cue the English stress contrast may result, in large part, from inappropriate perception, a factor that has been suggested for other cases of foreign accent (e.g., Flege, Bohn, & Jang, 1997; Rochet, 1995). In this case, we would expect native Mandarin speakers' perception of cues to English lexical stress to reflect their difficulties in production: They should have more problems using cues related to vowel quality as compared to other cues, such as duration, intensity and f0.

Alternatively, there is a possibility that Mandarin speakers' problems with reducing vowels in unstressed syllables might lie in production alone. It has been well documented that native language affects adult learners' production of non-native vowels (Flege et al., 1997; Flege & Hillenbrand, 1987; Rochet, 1995). Mandarin speakers may not produce English vowels that do not exist in Mandarin Chinese in a native-like manner (Chen, Robb, Gilbert, & Lerman, 2001; Zhang et al., 2008). For example, Zhang et al. (2008) observed that Mandarin speakers apparently aimed to use the high front lax vowel [ɪ] for the unstressed syllable *de-* in *desert* (verb), but their production of [ɪ] was not very close to that of their native English counterparts. That is, although their production was acoustically most similar to a canonical native English [ɪ] (of all of the English vowels, the closest one to this production in acoustic space was [ɪ]), it was still not sufficiently close to [ɪ] to be clearly identified as such by native listeners. Previous studies suggest that vowel quality plays a greater role in English stress contrasts than it does in other stress languages (e.g., Dutch), and native English speakers rely heavily on vowel quality in their perception of lexical stress (Campbell & Beckman, 1997; Cutler, 1986; Cutler & van Donselaar, 2001; Sluijter & van Heuven, 1996b).

Although vowel quality clearly plays a role in identifying English stress, there is as yet no consensus about the exact weight of vowel quality in comparison to other acoustic parameters of stress such as duration, intensity and f0. Previous studies by Beckman (1986) and Sluijter and van Heuven (1996b) agree that the relative strength of vowel quality is less than duration but greater than intensity, but they disagree about the ranking of average f0 with respect to the other cues. According to Fry (1958) and Beckman (1986), f0 is the most reliable stress cue and was ranked higher than all other cues. Sluijter and van Heuven (1996b), however, argued that prior studies by Beckman (1986) and Fry (1958) co-varied stress and accent, and that average f0 functions as an important correlate for accent instead of for lexical stress. According to Sluijter and van Heuven (1996b), stress refers to the different degree of prominence of each syllable within a word, while accent specifies the focused word in a sentence or a phrase due to the need of the specific language communication, usually the stressed syllable of that focused word becomes accented. Sluijter and van Heuven (1996b) emphasized that average f0 is the least reliable acoustic correlate in comparison to duration, vowel quality and even intensity for English lexical stress. According to Sluijter and van Heuven (1996b), vowel quality was a stronger cue to stress than f0.

Further evidence for the importance of vowel reduction from a psycholinguistic perspective is found in the work of Cutler (1986), who used a cross-modal lexical priming task to examine how lexical prosody (stress) alone constrains native English speakers' access to the lexicon independently of segmental factors (i.e., vowel reduction). She focused on English heteronyms, such as *FOREbear* (*meaning ancestor*) vs. *forBEAR* (*meaning withhold*), in which stress is contrastive but there are no segmental differences (no difference in vowels because there is no vowel reduction in either word). Results showed that English speakers could not use prosodic stress information alone to accomplish lexical recognition. Cutler (1986) attributed this finding to the fact that English has only a very small number of true minimal stress pairs in which vowel reduction does not apply. Cutler and van Donselaar (2001) replicated the experiment for Dutch, which has an equally small number of stress heteronym pairs, and found the opposite result. They argued that this opposite result might indicate that Dutch speakers used prosodic information to identify lexical items. Combining these two different findings, their interpretation was that Dutch and English differ in the amount of vowel reduction (Dutch has much less vowel reduction than English), and vowel quality outweighs other prosodic information in recognition of stress contrasts in English but not in Dutch.

Unlike English, Mandarin is a tonal language. Lexical tones in Mandarin are acoustically instantiated primarily in terms of pitch, then duration and intensity (Fu, Zeng, Shannon, & Soli, 1998; Gandour, 1978, 1983; Howie, 1976; Liu & Samuel, 2004; Whalen & Xu, 1992). Vowel quality is not an acoustic cue to Mandarin tones. Following the line that acoustic features used in the L2 phonological system but not in the L1 might be underattended (Francis & Nusbaum, 2002; Guion & Pederson, 2007; Iverson et al., 2003), it is possible that native Mandarin listeners might not attend at all to vowel quality as a stress cue or they may attend less to it than to other acoustic cues that they are more familiar with from the tonal contrasts of their first language experience. On the other hand, Mandarin listeners might be able to use vowel quality as a cue to English lexical stress, because they may actually be quite sensitive to subtle vowel differences both in Mandarin and in English. According to the Perceptual Assimilation Model proposed by Best (1995), the difficulty of discrimination on non-native vowel contrasts depends on the similarity or dissimilarity of each individual vowel to listeners' native vowel categories. Best (1995) predicts six possible discrimination patterns in light of vowel assimilations between L1 and L2. For example, if two non-native vowels are assimilated, respectively, to two different native categories (Two-Category Assimilation), then the discrimination of these vowel contrasts should be excellent. On the other hand, if two vowels are assimilated to

the same native category, then the discrimination could be poor or good depending on the degree of divergence of each vowel from that native category. Jia, Strange, Wu, and Collado (2006) examined English vowel discrimination by native Mandarin speakers. They found that experienced Mandarin learners of English were very accurate at discriminating all six pairs of vowel contrasts in English (i.e., /i-ɪ/, /i-e/, /ɛ-æ/, /æ-ɑ/, /ɑ-ʌ/, /u-ɑ/). Mandarin speakers who have been in the US for 2–5 years scored 96.2% correct vowel identification (ranging from 91.4% to 99.7%) while even those had been in the US for less than 2 years scored 94.5% (ranging from 85.2% to 99.7%). Given their sensitivity to different vowel categories in English, Mandarin listeners might be able to distinguish stressed syllables from unstressed ones by identifying different vowel categories, although they might still present a different pattern from their native English counterparts, possibly corresponding to differences in their production (Zhang et al., 2008). To our knowledge, no studies have examined the function of vowel quality in the perception of English stress by non-native listeners. To investigate this, in the present study we compared the weight given to vowel quality with that to other acoustic cues such as duration, pitch and intensity in the perception of English lexical stress across native English listeners and native Mandarin listeners.

The second goal of this study was to examine the role of the pitch contour in stress identification. It is possible that lexical-level acoustic information in a phrase or sentence with a flat pitch contour may be more easily accessed than in sentences with a natural pitch contour (Viswanathan & Russell, 1985). We may hypothesize that the flattened pitch contour might reduce listeners' reliance on f0 and/or increase their reliance on other cues such as vowel quality or duration, in both native and non-native listeners' stress identification, as compared with the weight given in utterances with normal f0 contour.

In sum, the current study attempts to address the following questions: How do native and non-native listeners weight vowel quality in comparison to other prosodic cues to stress perception? Are there any differences between English and Mandarin listeners with regard to cues to English lexical stress? Do pitch contour conditions (natural vs. flat) interact with the use of specific acoustic cues? These issues were investigated in three experiments. Experiment 1 examined the weighting of vowel quality in comparison to f0 as cues to stress perception under natural and flat pitch contour conditions by Mandarin listeners and English listeners. Experiment 2 constituted a similar comparison between vowel quality and duration, and Experiment 3 examined the same relationship between vowel quality and intensity. Specifically, identical participants participated in each of three experiments. Participants consist of 24 native English listeners (12 men and 12 women) and 24 native Mandarin listeners (12 men and 12 women). The presentation order of these three experiments was counterbalanced across language backgrounds of listeners and across listeners' gender. Also, within each of 12 listeners of the same gender as well as the same language background, presentations of natural vs. flat pitch contour conditions for each experiment were counterbalanced, that is, half of the participants were presented with natural pitch contour condition first and the other half with flat pitch contour condition first across each experiment.

2. Experiment 1

In this experiment, the perceptual weighting of vowel quality vs. f0 in two different pitch contour conditions was examined for native and non-native listeners. Listeners participated in two different sessions: One session is for vowel quality vs. f0 in natural pitch contour condition, and the other session for vowel quality vs. f0 in flat pitch contour condition.

2.1. Listeners

Two groups of listeners were recruited from the Purdue University community. The first group consisted of 24 native English listeners, 12 females and 12 males aged 18–26 (M=21). The second group consisted of 24 native Mandarin Chinese listeners, 12 females and 12 males aged 28–35 (M=31.5), all graduate students at the time of participation. All of the native Mandarin Chinese listeners were born in and grew up in mainland China with the exception of one male listener who was born and grew up in Taiwan. Purdue University was the first stop for all of them upon their arrival in the United States. The duration of their time in the US ranged from 2 months to 7 years (M=3.7). Prior to entering the US, no Chinese listeners had any experience being immersed in a native English language environment, although 11 of them had a native English teacher for English classes at the college level in China. This was only for one semester in all cases. Moreover, none of them specialized in English language/literature studies in college. All Chinese listeners were asked to evaluate their comprehensive English proficiency on a 10-point scale (1 is worst, 10 is excellent), and the average self-evaluation of English proficiency was 7, ranging from 6 to 8. All participants passed a brief hearing screening (20 dB HL at 500 Hz, 15 dB at 1000, 2000 and 4000 Hz) to be qualified for their participation in this experiment. No participant reported any prior history of speech or language problems. All participants also answered a questionnaire regarding their language background.

2.2. Stimuli

A pair of utterances of *DEsert* (noun) and *deSERT* (verb) produced by a female native speaker of American English (originally from the Midwest of the US, 20 years old) was selected to be used as an exemplar in the current stress perception experiments. This token was chosen from the utterances recorded for, and reported by Zhang et al. (2008). This specific token was chosen for two reasons: first, only one syllable (i.e., *de-*), instead of two syllables, in the *desert* pair involves vowel reduction, which simplifies the re-synthesis process; second, this pair of recordings was acoustically clear enough to display identifiable pitch and intensity contours, which can guarantee the technical manipulation for the current perception stimuli.

The first step was to create a single version of the *desert* token such that it would be perceived ambiguously with respect to stress placement. This was done using Praat 4.5.12 (Boersma & Weenink, 1992–2009) following the Praat instruction manual for source-filter (PSOLA) re-synthesis. Relevant acoustic parameters for both syllables, including f0, intensity, duration, and vowel quality (F1–F3), were manipulated to have the same values as the averaged measures derived from the original stressed and unstressed syllables of the selected *desert* pair (Tables 1 and 2).

Table 1
Values of acoustic parameters of the syllable *de-* in *desert* in stressed and unstressed conditions produced by an American female speaker of English from the previous production study. The average of acoustic parameters is derived from the (stressed+unstressed)/2.

De	Dur (ms)	f0 (Hz)	Int (dB)	F1 (Hz)	F2 (Hz)	F3 (Hz)
Stressed	155	213	74	595	1918	2883
Unstressed	104	204	68	422	2008	2994
Average	130	209	71	508.5	1963	2939

Table 2
Values of acoustic parameters of the syllable *-sert* in *desert* in stressed and unstressed conditions produced by an American female speaker of English from the previous production study. The average of the acoustic parameters is derived from the (stressed+unstressed)/2.

Sert	Dur (ms)	f0 (Hz)	Int (dB)	F1 (Hz)	F2 (Hz)	F3 (Hz)
Stressed	502	193	71	592	1738	2301
Unstressed	487	185	68	640	1896	2499
Average	495	189	69.5	616	1817	2400

Then a pilot study was conducted to determine whether this stimulus was in fact ambiguous.

2.2.1. Pilot study

Five American women participated in the pilot study. All were native speakers of American English, studying at Purdue University as undergraduate or graduate students. Their average age was 24 years, ranging from 21 to 36 years. None were from the linguistics program and all self-reported normal hearing, speech and language abilities.

Stimuli in the pilot study consisted of the resynthesized ambiguous *desert* token, as well as six pairs of stress contrasts (including *contract*, *object*, *permit*, *rebel*, *record*, and *subject*) as distracters. These distracters were selected from utterances produced by different female native English speakers in the previous stress study by Zhang et al. (2008). Selection of distracter word pairs was made by choosing those tokens with the highest acceptability ratings across female English speakers in the previous study.

Because the re-synthesis process introduced a small amount of error into the resulting sound file, comparison with the original, un-resynthesized tokens of distracter words would have been confounded. Therefore, a small amount of additional noise was added to each distracter sound file, only for the purposes of pilot testing. Subsequent tests, which involved only resynthesized sound files, did not require added noise. Adding noise was accomplished as follows: First, for each sound file, a waveform consisting of Gaussian noise was synthesized with the same duration and sampling rate as the file to which it would be added using the Praat 'create sound from formula…' function. The intensity contour of the noise was then multiplied by the intensity tier of the target sound file, resulting in a noise file with approximately equal intensity over time. This was then added to the original sound. Adding the noise resulted in a drop of the harmonic-to-noise ratio (HNR) of the stressed vowel of randomly selected stimuli from approximately 13 dB (without added noise) to approximately 10 dB. As HNR of purely periodic sounds can serve as a proxy for the signal-to-noise ratio (SNR), we conclude that adding noise in this manner decreased the SNR by approximately 3 dB. All words were still clearly intelligible with the added noise, but minor impurities in the resynthesized signal were less easily detected.

The pilot experiment was run using E-prime version 1.1 (Schneider, Eschman, & Zuccolotto, 2002). All participants were seated at a computer (Dell Optiplex/Windows XP) wearing headphones (Sennheiser, HD 25-1) that played sound at a comfortable listening level of 60–65 dBA in a quiet lab. A model RB-620 response pad (Cedrus Corporation) was provided for making a response by pressing the left or right button. Prior to the experiment, the experimenter informed these five native English listeners of the stress shift rule in English disyllabic words, and they clearly claimed their understanding of the association of the stress location on the first or the second syllable of English disyllabic words to the noun–verb category. For each trial, listeners first heard a word, then were asked to identify whether what they heard was a noun or verb. 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. The response choices were displayed with capitals standing for the stressed syllables, and the participants were informed of this kind of display in the very beginning of the experiment. There was a total of 13 tokens, including one target ambiguous token *desert* and 6 other word pairs (noun and verb for each). Each distracter token was repeated 10 times and the target ambiguous *desert* token was repeated 20 times, so that a set of 140 stimuli were randomly presented to each subject. The pilot experiment took *ca.* 15–20 min for each subject.

Statistical analyses on the identification data of the ambiguous "*desert*" token by all five native listeners of American English showed an overall 47% noun and 53% verb judgment. It is possible that the relatively long length of the second syllable '*-sert*' may have influenced listeners such that they gave more verb than noun responses. In order to offset this, the second syllable '*-sert*' was shortened by 100 ms, resulting in a length of 395 ms for the final perception experiment.

2.2.2. Stimuli manipulation

Using Praat 4.5.12 (Boersma & Weenink, 1992–2009), the revised ambiguous token was manipulated to create two stimulus sets contrasting vowel quality vs. f0 in both the natural and flat pitch contour conditions, respectively, according to the following procedure: based on the originally recorded stressed values for the first three formant frequencies (F1–F3), seven versions of the syllable *de-* were resynthesized in a continuum ranging from stressed [ɛ] (F1=597 Hz; F2=1854 Hz; F3=2786 Hz) to unstressed [ɪ] (F1=462 Hz; F2=1964 Hz; and F3=2975 Hz). Then from each of the seven steps along the vowel quality continuum, a seven-step continuum of f0 was created ranging from the stressed value 213 Hz to the unstressed value 204 Hz. Flanagan (1957) stated that the difference limen (DL) for f0 is ± 1 cps (1 Hz), at least for values of f0 between 80 and 160 Hz. So the overall difference here is about 7 times the DL, and each step is separated by just about 1 DL assuming that sensitivity does not change much between 160 and 204 Hz. Duration and intensity were held constant at the mean value of the stressed and unstressed syllable measures. The second syllable *-sert* was held constant with the values of the baseline ambiguous stimulus. Thus, a total of 49 *desert* tokens with natural pitch contour were made. To make the stimulus set with the flat pitch contour, each of the 49 *desert* tokens created under the natural pitch contour as described above was replicated, and then the f0 contour of each syllable was set to its average value for the duration of the syllable.

2.3. Procedure

Experimental equipment and parameters were the same as for the pilot study. Before participants started the identification task, the experimenter explained the stress shift rule in disyllabic words, such as the *desert* pair. All of them including Mandarin listeners and English listeners stated that they understood this rule and knew the stress difference between the noun *DEsert* and verb *deSERT*. Generally, this English stress rule is taught in English class in Chinese middle school.

Each listener completed two sessions on two different days. Half of the listeners in each group heard stimuli with a natural pitch contour first, while another half first heard the flat pitch contour condition. For each session, 49 stimuli were repeated in random order in each of 10 blocks, yielding a total of 490 trials.

On each trial, listeners heard a stimulus accompanied by a written presentation of possible responses showing the noun/verb stress pair on the screen: *DEsert*(*noun*) and *deSERT*(*verb*). Listeners were asked to identify what they heard as being more likely a verb or a noun, making their judgment by pressing the corresponding button on the response pad, the left button for the noun and the right button for the verb. After making the identification, listeners were prompted to listen to the next stimulus to start a new trial. Although this experiment was self-paced with no limit on time to respond, listeners were instructed to respond as quickly and accurately as possible. The average running time for each session was around 30–40 min. In addition, the first block of identification results made by each listener was treated as familiarization and not scored, so a total of 441 identification results made by each of 48 listeners for each session were used for further analyses.

2.4. Analysis

Identification functions were used as the first step in evaluating categorization. Identification functions were calculated as the proportion of noun responses to all seven stimuli sharing a given vowel quality value and to all seven stimuli sharing a given f0 value.

Logistic regression has been used to rank the relative importance of relevant acoustic cues in speech perception studies (Benki, 2001; Clayards, Aslin, Tanenhaus, & Jacobs, 2007; Mayo & Turk, 2004; Morrison, 2005). According to Pampel (2000), the Wald statistics is used to test the significance of the individual coefficient of each independent variable, which is calculated as the squared ratio of the unstandardized coefficient to its standard error, following a chi-square distribution. The odds ratio derived from exponentiating the estimated coefficient represents the effect size, such that the more distant the odds ratio is from 1.0 in either direction (<1 or >1), the greater effect of that independent variable. Here, logistic regression analyses were adopted to compare the difference in weighting vowel quality and f0 by native and non-native English listeners in the perception of English lexical stress under difference pitch contour conditions (natural vs. flat).

In addition, ID function slopes and midpoints were compared. Following Benki (2001), Morrison (2005), Nissen, Harris, Jennings, Eggett, and Buck (2005), and Kondaurova and Francis (2008), logistic regression was run on each listener's identification data, the estimated coefficients of each stress cue were extracted, and the 50% cross-over midpoint was calculated according to the equation $x=(\log(.5/1-.5)-a)/b$ (where a is the coefficient of interception and b is the regression slope). The estimated coefficients and the scores of the 50% cross-over midpoint were separately submitted to one-way ANOVAs with language group as the independent factor to test the difference between language groups in both pitch contour conditions. Also, the estimated coefficients were submitted to a series of one-way ANOVAs with contour condition as the independent factor to examine the possible difference between natural and flat pitch contour condition for each language group.

Finally, following Flege et al. (1997), Bohn (1995) and Escudero and Boersma (2004), the comparison between different language groups with respect to the use of acoustic cues was made by calculating the "effect score". In the present experiment, for each listener, the "spectral effect" (vowel quality) scores along /ɛ/-/ɪ/ were derived by subtracting the percentage of noun responses given to the /ɪ/ endpoint averaging over seven f0 steps from the percentage of noun responses given to the /ɛ/ endpoint. A similar calculation was done to get the f0 effect. Then, these effect scores were separately submitted to one-way ANOVAs examining the possible effect of language background and the effect of the pitch contour condition, respectively.

2.5. Results

Fig. 1 shows the proportion of noun identifications along the seven-step vowel quality and seven-step f0 continua averaged across subjects for English and Mandarin listeners in the natural (Fig. 1a) and the flat (Fig. 1b) pitch contour conditions. The percentage of noun responses (stressed *de-*) decreased systematically as vowel quality changed from /ɛ/ to /ɪ/ and as f0 decreased. However, the slopes of the f0 ID curves were less steep than those for vowel quality, suggesting that native and non-native English listeners weighted vowel quality heavier than f0.

Results of logistic regression analyses showed a significant effect of vowel quality (English: *Wald's* $\chi^2(1)=481.72$, $p<.001$; Mandarin: *Wald's* $\chi^2(1)=523.27$, $p<.001$) and f0 (English: *Wald's* $\chi^2(1)=16.0$, $p<.001$; Mandarin: *Wald's* $\chi^2(1)=62.94$, $p<.001$) on stress identification by English and Mandarin listeners in the natural pitch contour condition. The odds ratio of vowel quality was also more distant from 1 than that of f0 for English listeners (vowel quality: .53; f0: .89) and for Mandarin listeners (vowel quality: .51; f0: .81), indicating that both groups of listeners responded more to vowel quality than to f0 in this task. In addition, Mandarin listeners showed a significant interaction between vowel quality and f0: *Wald's* $\chi^2(1)=6.07$, $p=.01$, and the odds ratio of the interaction is close to 1.0, ranked lower than vowel quality and f0 alone. These results indicated that Mandarin listeners tended to identify synthesized *desert* sounds with their first syllable *de-* having both full vowel and higher pitch, as nouns. In other words, Mandarin listeners treated vowel quality and pitch in a combined way. In contrast, English listeners did not show a significant interaction: *Wald's* $\chi^2(1)=.41$, $p=.52$, indicating that English listeners identified synthesized *desert* sounds with their first syllable *de-* having either full vowel or higher pitch, as nouns. In other words, English listeners treated pitch and vowel quality independently.

For the flat pitch contour condition, results of logistic regression showed significant effects of vowel quality (English: *Wald's* $\chi^2(1)=543.12$, $p<.001$; Mandarin: *Wald's* $\chi^2(1)=504.03$, $p<.001$) and pitch (English: *Wald's* $\chi^2(1)=4.10$, $p=.04$; Mandarin: *Wald's* $\chi^2(1)=6.89$, $p=.01$) for both groups. Similarly, English and Mandarin listeners weighted vowel quality heavier than f0 in terms of odds ratios (English: vowel quality: .48; f0: .94; Mandarin: vowel quality: .49; f0: .92). There was no significant interaction between vowel quality and f0 for either English or Mandarin listeners (English: *Wald's* $\chi^2(1)=1.1$, $p=.30$; Mandarin: *Wald's* $\chi^2(1)=2.4$, $p=.12$), indicating that both groups of listeners treated vowel quality and f0 independently when f0 was flat.

In terms of the slope comparison along vowel quality or f0 between English and Mandarin listeners in each pitch contour condition, there was no significant difference between English and Mandarin listeners with regard to the use of vowel quality in stress identification in either pitch contour condition (natural: $F(1,36)=.26$, $p=.61$; flat: $F(1,40)=.56$, $p=.46$), or with regard to the use of f0 (natural: $F(1,16)=2.03$, $p=.17$; flat: $F(1,2)=.31$, $p=.64$), indicating that Mandarin listeners treated both vowel quality and f0 in a similar manner to English listeners. In addition, the comparison of the 50% cross-over point along the vowel quality continuum between Mandarin and English listeners was not significant in either pitch contour condition (natural: $F(1,35)=1.31$, $p=.26$; flat: $F(1,40)=.85$, $p=.36$), and the comparison along f0 was not significant (natural: $F(1,16)=.05$, $p=.82$; flat: $F(1,2)=1.16$, $p=.39$).

The comparison of effect scores along the vowel quality continuum between English and Mandarin listeners was not

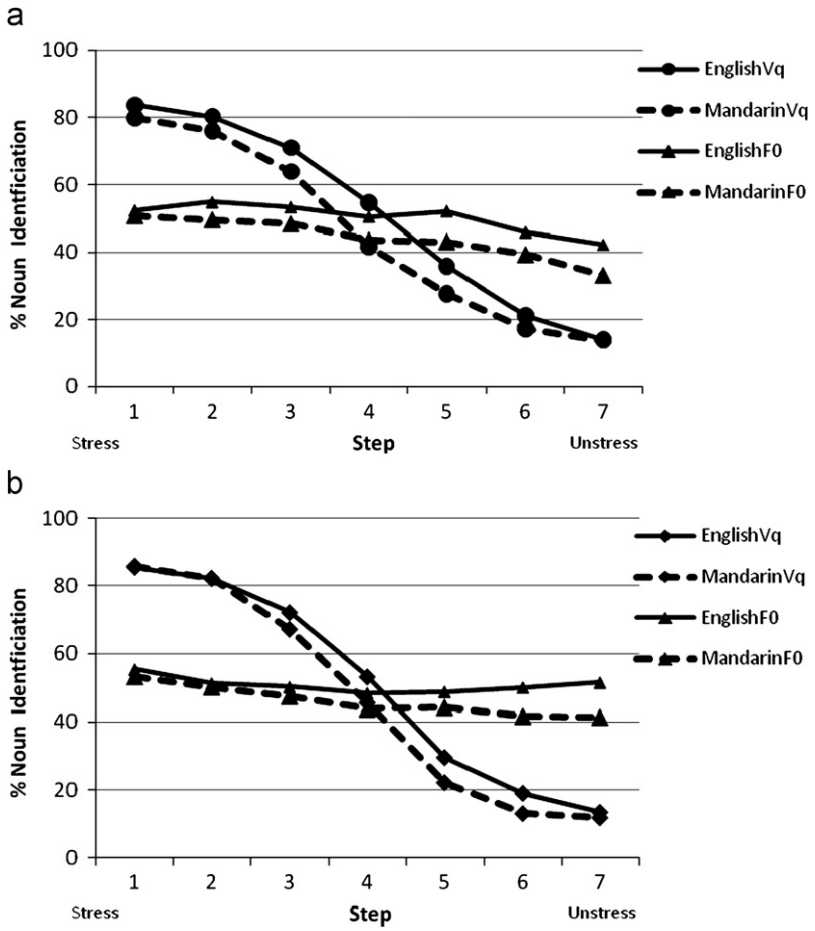

Fig. 1. Mean proportion of noun responses along vowel quality continuum and along the f0 continuum in natural and flat pitch contour conditions by English listeners and by Mandarin listeners. *EnglishVq*: vowel quality response by native English listeners; *MandarinVq*: vowel quality response by native Mandarin listeners; *EnglishF0*: f0 response by native English listeners; *MandarinF0*: f0 response by native Mandarin listeners. Left panel: Natural pitch contour condition; Right panel: Flat pitch contour condition: (a) natural pitch contour and (b) flat pitch contour conditions.

significant in either pitch contour condition (natural: $F(1,46)=.22$, $p=.64$; flat: $F(1,46)=.04$, $p=.85$). However, the language group difference was significant for f0 in both pitch contour conditions (natural: $F(1,46)=4.18$, $p=.046$; flat: $F(1,46)=9.53$, $p=.003$), such that Mandarin listeners had greater f0 effect scores than native English listeners in both pitch contour conditions (natural: English: .10, Mandarin: .18; flat: English: .04, Mandarin: .12).

With respect to the comparison between natural and flat pitch contour conditions, there were mainly two aspects. First, a slope comparison did not show a main effect of vowel quality: $F(1,36)=.07$, $p=.79$, or f0: $F(1,8)=.50$, $p=.50$, in stress perception by English listeners. For Mandarin listeners, there was also no significant difference found with vowel quality: $F(1,40)=.03$, $p=.87$, or f0: $F(1,10)=1.42$, $p=.26$. Second, in terms of the comparison of effect scores, Mandarin listeners did not show a significant effect of vowel quality: $F(1,46)=.75$, $p=.39$, or f0: $F(1,46)=2.7$, $p=.11$, and English listeners did not show a significant effect of vowel quality: $F(1,46)=.17$, $p=.68$, but did with f0, $F(1,46)=5.74$, $p=.02$. English listeners showed a larger effect score of f0 in natural than in flat pitch contour conditions (natural: .10; flat: .04).

2.6. Discussion

Native and non-native listeners used both vowel quality and f0 in the perception of English lexical stress with natural and flat pitch contour, but the weight of vowel quality was much stronger than f0. This finding is consistent with previous research by Sluijter and van Heuven (1996a, 1996b). No difference with regard to vowel quality was found between native and non-native listeners, suggesting that Mandarin listeners applied an English-like pattern in using vowel quality to cue English stress. Pitch contour condition also did not have an influence on the use of vowel quality for both language groups. However, native and non-native listeners showed a difference in the use of f0, and their use of f0 was also influenced by the pitch contour condition.

First, Mandarin listeners tended to treat vowel quality and f0 in a combined way in the natural pitch contour condition but in an independent way in the flat pitch contour condition, while English listeners treated these two cues independently in both pitch contour conditions. Several studies have investigated the processing interaction between segmental (e.g., vowel quality) and supra-segmental (e.g., pitch) features using the Garner speeded classification paradigm (Lee & Nusbaum, 1993; Miller, 1978; Repp & Lin, 1990; Tong, Francis, & Gandour, 2008; Wood, 1974, 1975). Studies by Wood (1974, 1975) indicated that English listeners processed vowel and pitch dimensions in a separable manner. A cross-linguistic study by Lee and Nusbaum (1993) compared English and Mandarin listeners' processing of segmental and supra-segmental information, and found that Mandarin listeners consistently perceived these dimensions across Mandarin and non-Mandarin stimuli sets in an integral fashion (cf. Garner, 1974), whereas English listeners showed an

integral pattern for Mandarin stimuli but not for non-Mandarin stimuli. However, Lee and Nusbaum (1993) also pointed out that the integral pattern shown in English listeners' processing of segmental and supra-segmental information for Mandarin stimuli was possibly caused by the relative discriminability of two dimensions. They concluded that Mandarin speakers' simultaneous processing of segmental and supra-segmental information derives from the phonemic role of Mandarin tones, since tone in Mandarin is lexically as important as segments and native Mandarin listeners learned to always attend to both dimensions simultaneously in word recognition. In contrast, tone is not lexically relevant in English, and English listeners have no motivation to manipulate vowel and pitch integrally. Hence, Mandarin listeners seem to apply their native strategy to process vowel quality and pitch in the perception of English lexical stress.

Nevertheless, Mandarin listeners did not show a combined pattern for these two dimensions in the flat pitch contour condition. The difference might be due to Mandarin listeners' heightened sensitivity to f0 contour within the syllable domain. Gandour (1983) found that Mandarin listeners placed more emphasis on f0 direction (contour) instead of f0 height as compared to English listeners in the perception of syllable-based tone. Xu (1997, 1998) also indicated that the syllable is the reference domain for tonal alignment. These previous studies suggest that Mandarin listeners are very sensitive to the dynamic f0 contour in the syllable domain. It is highly possible that this kind of sensitivity motivates Mandarin listeners' applying their native strategy to perceive English lexical stress under the natural pitch contour condition by combining vowel quality and pitch. On the other hand, as constant pitch is not as informative and natural as dynamic pitch, Mandarin listeners might have been less motivated to combine vowel quality and pitch in the flat pitch contour condition. Accordingly, Mandarin listeners' processing of pitch and vowel quality for stress identification is influenced by their native experience but this influence is conditioned by the naturalness of the stimuli in question.

Second, Mandarin listeners consistently showed greater reliance on f0 than English listeners in both pitch contour conditions in terms of f0 effect scores. It is known that the primary acoustic property of Mandarin lexical tones is f0 (e.g., Gandour, 1983). Mandarin listeners' greater reliance on f0 in stress perception might be derived from the influence of the significant power of f0 in cuing Mandarin tones. Of course, this assumption cannot be conclusively confirmed unless one was to examine listeners with no native tonal language background.

Finally, English listeners showed a greater f0 effect score in natural than flat pitch contour conditions. According to Gandour (1983), English listeners directed their attention almost exclusively to the height instead of the direction dimension with respect to the use of f0 in perception of syllable-based tones. Gandour's (1983) findings suggest that English listeners might not be readily responsive to the pitch contour in the syllable domain, whereas it is known that pitch is an active cue at the sentence level in English. Hence, such an f0 difference between natural and flat pitch contour conditions observed for English listeners in stress identification might have nothing to do with the f0 contour sensitivity within syllable domain. Considering the calculation of effect score is derived from the difference between the percentages of noun responses at two endpoints, in which the sound *desert* had the first syllable *de*-regularly stressed at step 1 and the first syllable *de*- regularly unstressed at step 7, the identification of the stressed and unstressed syllables, respectively, at steps 1 and 7 with the natural pitch contour might be more clear-cut and less confusable than with the flat pitch contour. This could account for the greater value of f0 effect score in natural than in flat pitch contour condition.

3. Experiment 2

In this experiment, the weight given to vowel quality was compared to duration in natural and flat pitch contour conditions.

3.1. Method

The procedure for manipulating stimuli set in this experiment was the same as in Experiment 1. In this experiment, the syllable *de*- varied along vowel quality and duration while f0 and intensity were kept constant at their average values. First, seven versions of the *de*- syllable were resynthesized in terms of vowel quality, then from each of the seven steps along vowel quality, a seven-step continuum of duration was created from the stressed value 155 ms to the unstressed value 104 ms. The syllable *–sert* was held constant with the values of the baseline ambiguous stimulus exactly as in Experiment 1. For the flat pitch contour condition, the natural pitch contour was flattened for each stimulus to have the average f0 value. Otherwise, the participants, experimental procedures, and data analyses were identical to those used in Experiment 1.

3.2. Results

Fig. 2 presents the identification results of vowel quality vs. duration in natural (Fig. 2a) and flat (Fig. 2b) pitch contour conditions. It can be observed that the ID curves for vowel quality were steeper than for duration, suggesting that the weight of vowel quality was heavier than duration.

Results of logistic regression analyses showed significant effects of vowel quality (English: *Wald's* $\chi^2(1)=750.70$, $p<.001$; Mandarin: *Wald's* $\chi^2(1)=589.47$, $p<.001$), and duration (English: *Wald's* $\chi^2(1)=201.28$, $p<.001$; Mandarin: *Wald's* $\chi^2(1)=84.04$, $p<.001$) in the natural pitch contour condition. In terms of odds ratio, the importance of vowel quality was greater than duration for both language groups (English: vowel quality .41, duration .65; Mandarin: vowel quality .47, duration .77). Both English and Mandarin listeners showed a significant interaction between vowel quality and duration (English: *Wald's* $\chi^2(1)=59.75$, $p<.001$; Mandarin: *Wald's* $\chi^2(1)=7.32$, $p<.001$). Similarly, for the flat pitch contour condition, logistic regression showed a significant effect of vowel quality (English: *Wald's* $\chi^2(1)=605.62$, $p<.001$; Mandarin: *Wald's* $\chi^2(1)=483.93$, $p<.001$), and duration (English: *Wald's* $\chi^2(1)=224.67$, $p<.001$; Mandarin: *Wald's* $\chi^2(1)=87.81$, $p<.001$), as well as an interaction between the two (English: *Wald's* $\chi^2(1)=50.56$, $p<.001$; Mandarin: *Wald's* $\chi^2(1)=5.19$, $p=.02$). The odds ratios also showed the greater importance of vowel quality as compared to duration for both language groups (English: vowel quality .48, duration .65; Mandarin: vowel quality .53, duration .77). Results suggest that both English and Mandarin listeners treated vowel quality and duration as a combined cue in the perception of English stress under natural and flat pitch contour conditions.

The comparison of ID function slopes showed no significant difference between English and Mandarin listeners either with the use of vowel quality (natural: $F(1,44)=.13$, $p=.72$; flat: $F(1,42)=.10$, $p=.75$) or with the use of duration (natural: $F(1,23)=.87$, $p=.36$; flat: $F(1,30)=.35$, $p=.56$) under either pitch contour condition. Also, the comparison of the 50% cross-over points did not show a significant difference between English and Mandarin listeners under either pitch contour condition for either vowel quality (natural: $F(1,44)=.67$, $p=.42$; flat: $F(1,42)=.02$, $p=.88$) or duration (natural: $F(1,23)=.90$, $p=.35$; flat: $F(1,30)=1.42$, $p=.24$). In terms of the comparison of effect scores, there was no significant difference between English and Mandarin

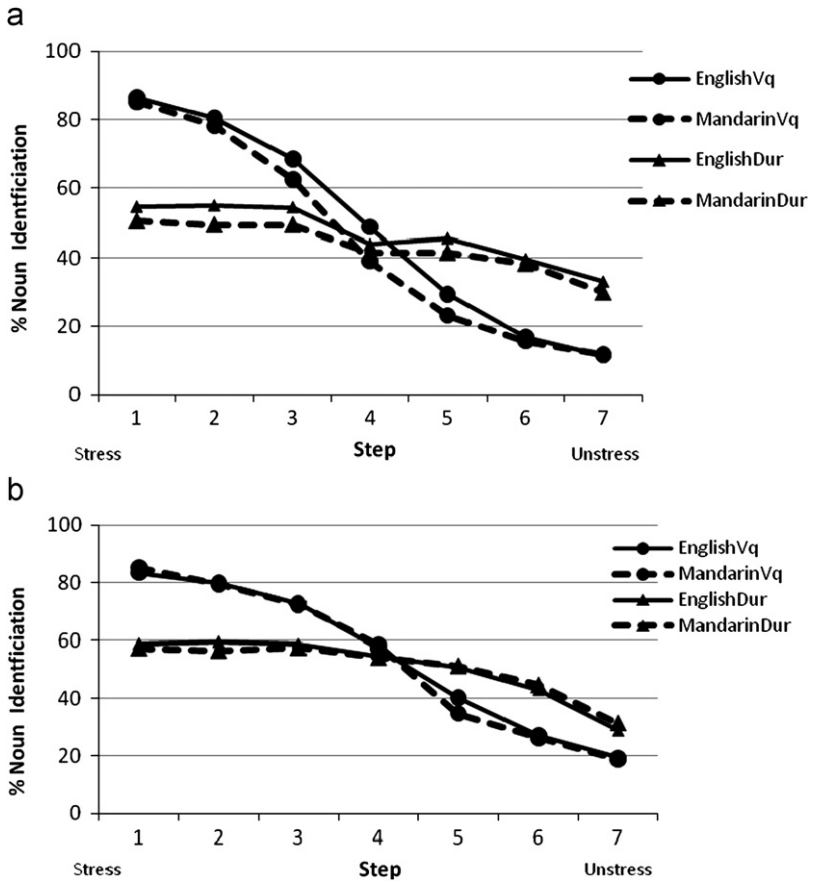

Fig. 2. Mean proportion of noun responses along vowel quality continuum and along the duration continuum in natural and flat pitch contour conditions by English listeners and by Mandarin listeners. *EnglishVq*: vowel quality response by native English listeners; *MandarinVq*: vowel quality response by native Mandarin listeners; *EnglishDur*: duration response by native English listeners; *MandarinDur*: duration response by native Mandarin listeners: (a) natural pitch contour condition and (b) flat pitch contour condition.

listeners in both pitch contour condition, either with regard to vowel quality (natural: $F(1,46)=.006$, $p=.94$; flat: $F(1,46)=.09$, $p=.77$) or with regard to duration (natural: $F(1,46)=.032$, $p=.86$; flat: $F(1,46)=.29$, $p=.60$).

With regard to the comparison between natural and flat pitch contour conditions, slope comparisons showed no significant difference between the two conditions in either the use of vowel quality or duration by English listeners (vowel quality: $F(1,44)=.02$, $p=.89$; duration : $F(1,31)=.84$, $p=.37$) and also not for Mandarin listeners (vowel quality: $F(1,42)=.017$, $p=.90$; duration : $F(1,22)=.43$, $p=.52$). The comparisons of effect scores also showed no significant difference in uses of these two cues by English (vowel quality: $F(1,46)=1.92$, $p=.17$; duration: $F(1,46)=1.6$, $p=.21$) or by Mandarin listeners (vowel quality: $F(1,46)=.55$, $p=.46$; duration: $F(1,46)=.67$, $p=.42$).

3.3. Discussion

Native and non-native listeners used both vowel quality and duration to perceive English lexical stress, and the weight they gave to vowel quality was greater than duration. Moreover, both groups of listeners treated vowel quality and duration as a combined cue for stress identification under both pitch contour conditions. No significant difference was found between Mandarin and English listeners with regard to vowel quality and duration in either pitch contour condition, indicating that Mandarin listeners had a native-like pattern when using vowel quality and duration for the perception of English lexical stress. In addition, no difference was found between natural and flat pitch contour conditions with respect to the use of these cues by either English or Mandarin listeners, suggesting that pitch contour condition has no influence on use of these cues in stress perception.

The combinational use of vowel quality and duration means that both cues were equally referred to for stress judgment. Specifically, to be identified as a noun, a *desert* sound has to have noun-like cue properties in terms of both longer duration and full vowel in its *de*-syllable. Previous studies agreed that more cues are better than fewer cues for speech perception when these multiple cues are not in conflict, and cooperating multiple cues, but not conflicting cues, can generally enhance phonetic clarity and increase the sensitivity of phonetic discrimination (Beckman, 1986; Groenen, Crul, Maassen, & van Bon, 1996; Hillenbrand, Getty, Clark, & Wheeler, 1995; Lea, 1977; Seidl, 2007). Considering that various acoustic cues contribute to different aspects of the same speech category, the combinational use of the cooperating part of these cues should produce more accurate stress identification. One possible explanation for the combined use of vowel quality and duration by English and Mandarin listeners might lie in that duration and vowel quality are both salient cues to English lexical stress, and combining these two cues provides better (more conservative) identification. Listeners who rely on the combination of two cues are less likely to treat an unstressed syllable as stressed. Alternatively, according to Seidl (2007), some cue might be necessary without being sufficient, meaning that the cue must be present but needs to be heard in combination with other cue(s) for a given decision to be made.

Following this line, the explanation for the combinational use of vowel quality and duration could be described like this: Both vowel quality and duration might be necessary cues for stress identification, but further examinations are needed to determine whether either is a sufficient cue for either English or Mandarin listeners for perceiving English lexical stress.

4. Experiment 3

In this experiment, the weighting of vowel quality in comparison to intensity in stress perception was examined under natural and flat pitch contour conditions by English and Mandarin listeners.

4.1. Method

Stimuli in this experiment were generated following the same procedure used in Experiments 1 and 2, except that stimuli varied in vowel quality and intensity while f0 and duration were kept constant at their average values. Specifically, seven versions of the syllable *de-* were resynthesized into a seven-step vowel-quality continuum exactly as in Experiments 1 and 2, then from each of the seven steps along vowel quality, a seven-step continuum of intensity was created from the stressed value 74 dB to the unstressed value 68 dB, while f0 and duration for all *de-* syllables were kept constant at their average values. The syllable *–sert* was held constant with the values of the baseline ambiguous stimulus exactly as in Experiments 1 and 2. For the flat pitch contour condition, the natural pitch contour was flattened for each stimulus along the average f0 value. Otherwise, the participants, experimental procedures, and data analyses were identical to those used in the prior two experiments.

4.2. Results

Fig. 3 presents identification results of vowel quality vs. intensity in natural (Fig. 3a) and flat (Fig. 3b) pitch contour conditions. These figures show that the ID curves of vowel quality are steeper than those of intensity, suggesting that the weight of vowel quality is heavier than intensity in both pitch contour conditions for both English and Mandarin listeners.

Logistic regression analyses for English listeners showed significant effects of vowel quality and intensity on stress identification under natural and flat pitch contour conditions (natural: vowel quality: *Wald's* $\chi^2(1)=479.80$, $p<.001$, intensity: *Wald's* $\chi^2(1)=12.52$, $p<.001$; flat: vowel quality: *Wald's* $\chi^2(1)=453.85$, $p<.001$, intensity: *Wald's* $\chi^2(1)=11.67$, $p<.001$). In terms of odds ratio, the importance of vowel quality was greater than intensity in both pitch contour conditions (natural: vowel quality .54, intensity .91; flat: vowel quality .54, intensity .91), but there was no significant interaction between vowel quality and intensity in either pitch contour condition (natural: *Wald's* $\chi^2(1)=3.46$, $p=.06$; flat: *Wald's* $\chi^2(1)=1.63$, $p=.20$). This suggests that English listeners treated vowel quality and intensity

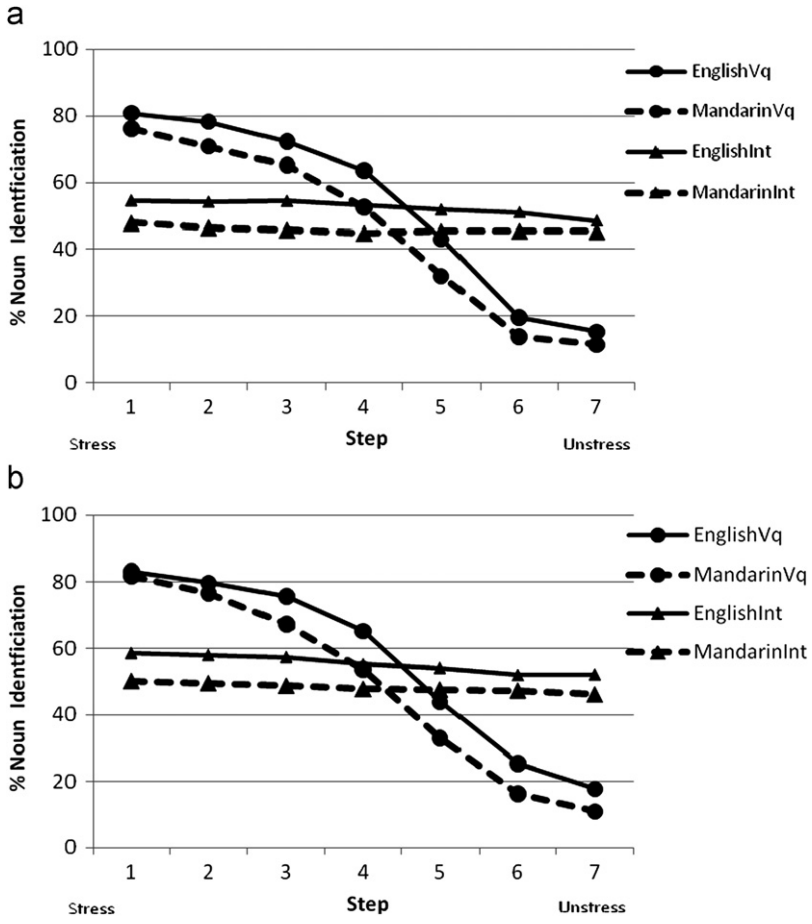

Fig. 3. Mean proportion of noun responses along vowel quality continuum and along the intensity continuum in natural and flat pitch contour conditions by English listeners and by Mandarin listeners. *EnglishVq*: vowel quality response by native English listeners; *MandarinVq*: vowel quality response by native Mandarin listeners; *EnglishInt*: intensity response by native English listeners; *MandarinInt*: intensity response by native Mandarin listeners: (a) natural pitch contour condition and (b) flat pitch contour condition.

independently. For native Mandarin listeners, results showed a significant effect of vowel quality under both pitch contour conditions (natural: *Wald's* $\chi^2(1)=398.18$, $p<.001$; flat: *Wald's* $\chi^2(1)=488.94$, $p<.001$), but no main effect of intensity (natural: *Wald's* $\chi^2(1)=.07$, $p=.79$; flat: *Wald's* $\chi^2(1)=3.17$, $p=.08$) as well as no significant interaction between vowel quality and intensity (natural: *Wald's* $\chi^2(1)=1.22$, $p=.27$; flat: *Wald's* $\chi^2(1)=.33$, $p=.57$). These results suggest that Mandarin listeners used only vowel quality to perceive English lexical stress when intensity was the only other cue available, while English listeners used both vowel quality and intensity independently.

In terms of slope comparison, there was no significant difference between English and Mandarin listeners' use of vowel quality in either pitch contour condition (natural: $F(1,39)=.002$, $p=.96$; flat: $F(1,39)=.03$, $p=.87$). Similarly, the comparison of the 50% cross-over midpoints also showed no significant difference between English and Mandarin listeners' use of vowel quality in either pitch contour condition (natural: $F(1,39)=2.11$, $p=.15$; flat: $F(1,39)=.23$, $p=.63$). Finally, the comparison of effect scores also showed no difference between English and Mandarin listeners' use of vowel quality in either pitch contour condition (natural: $F(1,46)=.001$, $p=.98$; flat: $F(1,46)=.35$, $p=.58$).

Mandarin listeners showed no difference between the natural and flat pitch contour conditions in terms of either vowel quality slopes ($F(1,38)=.002$, $p=.96$) or intensity slopes ($F(1,4)=1.36$, $p=.31$); Nor was there a significant effect of pitch contour condition on the slopes of English listeners' ID curves, either for vowel quality ($F(1,40)=.02$, $p=.89$) or intensity ($F(1,3)=.002$, $p=.97$). Similarly, comparison of English listeners' effect scores showed no significant difference between natural and flat pitch contour conditions for either vowel quality ($F(1,46)=.003$, $p=.96$) or intensity ($F(1,46)=.006$, $p=.94$) and results were similar for Mandarin listeners (vowel quality: $F(1,46)=.32$, $p=.57$; intensity: $F(1,46)=.25$, $p=.62$).

4.3. Discussion

Native English listeners used vowel quality and intensity independently in the perception of English lexical stress. They weighted vowel quality much heavier than intensity in natural and flat pitch contour conditions. This finding is consistent with the results of previous studies by Sluijter and van Heuven (1996a, 1996b) and Beckman (1986), in that the intensity was ranked much lower than vowel quality in cuing English lexical stress. In contrast, Mandarin listeners used only vowel quality and ignored intensity when making decisions about English lexical stress under both natural and flat pitch contour conditions. In other words, when vowel quality was available, intensity was not used by native Mandarin listeners. It is known that intensity is the least important cue for English lexical stress (e.g., Beckman, 1986), and intensity is also the secondary cue for Mandarin lexical stress (e.g., Whalen & Xu, 1992). Mandarin listeners' inability to use intensity in English lexical stress identification might result from the less important role that intensity plays in cuing both lexical stress in English and lexical tone in Mandarin.

5. General discussion

Although previous studies by Dupoux et al. (2008) suggested that non-native listeners' stress deafness is not necessarily a perceptual problem, results from a serious of identification tests in the present study with focus on cue weighting indicated that native and non-native listeners showed different perceptual patterns when only vowel quality and pitch, and vowel quality and intensity, are available for stress identification. However, a similar pattern was also found in using vowel quality and duration in the perception of English lexical stress by native and non-native listeners.

Native English listeners used all four acoustic cues to perceive English lexical stress in both natural and flat pitch contour conditions. In most cases, vowel quality was more salient than other cues. This finding is consistent with previous stress studies by Sluijter and van Heuven (1996a, 1996b) and Cutler (1986). Similarly, Mandarin listeners used vowel quality as a cue to perceive English lexical stress. Mandarin listeners categorized the target vowel contrasts ([ɛ] and [ɪ]) just like English listeners, and Mandarin listeners also weighted vowel quality more than other prosodic cues consistently like English listeners.

Vowel quality is not a cue used for Mandarin lexical tones and on this basis we expected that Mandarin listeners might not use this cue at all in English, or at least might have problems with its use. However, considering vowel quality as a segmental property, we might make other predictions. For example, although the phonological inventory of Mandarin Chinese does not include the lax vowel [ɪ], it does have a high front unrounded [i] as well as an [ɛ]-like mid-vowel sound (although this segment always follows a vowel like [i] or [y] to form diphthongs) (Duanmu, 2002). Based on this inventory, English [ɪ] and [ɛ] have been classified as non-native or "unfamiliar" vowels to native Mandarin speakers in cross-linguistic production/perception studies (Chen et al., 2001; Flege et al., 1997). According to the assumption of two-category assimilation in PAM (Best, 1995), if each member of the non-native sound pair is assimilated to a different native sound category, the discrimination of one non-native sound from another would be expected to be excellent. Following this assumption, Mandarin listeners would be expected to clearly discriminate [ɛ] from [ɪ], which could therefore account for Mandarin listeners' ability to use vowel quality as a cue to English lexical stress: Rather than perceiving this as a prosodic difference, Mandarin listeners may be cuing in on the vowel quality difference and identifying the two words (the noun and verb) as different on this basis in the same way that they might distinguish between e.g., *bet* and *bit*. Alternatively, the native-like use of vowel quality by Mandarin listeners in the perception of English lexical stress might result from Mandarin listeners' application of L2 general perception strategy irrespective of their native Mandarin background. According to Bohn (1995), a particular acoustic cue, not necessarily a familiar cue to L2 learners which is used in their native language, might be used by L2 learners just because this cue is easily accessed compared to other acoustic cues in their speech perception. Since vowel quality is not associated with cuing tonal contrasts in Mandarin Chinese, Mandarin listeners seem not to transfer their native tonal perception strategy to English stress perception. It is possible that vowel quality is easily accessed for native Mandarin speakers to identify stress contrasts. In other words, when information conveyed by other cues, like intensity, f0 or duration, is insufficient for Mandarin listeners to differentiate English stress contrasts, vowel quality difference will be used for stress identification whether or not Mandarin listeners have native experience with vowel quality cuing prosody.

In addition, in combination with the results of the previous study of stress production by Zhang et al. (2008), the present results seem not to support the proposition that "accented production is perceptually motivated" as might be suggested by applying the findings of Rochet (1995) to the present problem. In the study of stress production, Mandarin speakers were found to use an English-like [ɛ] for the stressed syllable *de-* as in the noun *desert*, but a non-native-English-like [ɪ] for the unstressed syllable *de-* as in the verb *desert* (Zhang et al., 2008). Yet in the present experiment, Mandarin listeners were able to use vowel quality as

a perceptual cue in a manner comparable to native English listeners. This inconsistency indicates that Mandarin speakers' inappropriate production of vowel reduction does not derive from their incorrect perception. Rather, it might be a production problem alone. According to Davidson (2006), errors in non-native speech sequence production might be caused by the articulatory coordination of adjacent sounds, which is language-specific, not necessarily caused by problematic perception.

Of course, another possible explanation for the lack of a production–perception link across these two studies could be that different subjects participated in each of them. It is possible that individual differences in perception, production and language experience across the two groups of learners might have resulted in one group that was unable to either produce or perceive the correct vowel contrast (in the Zhang et al. (2008) study, in which only production was measured) and another group that was able to perceive and produce the contrast correctly (in the present study, in which only perception was measured). On the other hand, previous studies have already shown that accurate perception still does not always guarantee accurate production. For example, Bent (2005) reported that English-speaking participants with little Mandarin experience showed high perceptual sensitivity to all Mandarin tone pairs but had difficulty accurately producing some tone pairs. Similarly, Sheldon and Strange (1982) examined the relationship between the production and perception of English /r/ and /l/ by native Japanese adults learning English in the US and found, for some subjects, that the production of the /r-l/contrast was more accurate than perception in certain phonetic environments. Thus, future work in this area should focus on comparing perception and production within a single group of speakers.

References

Beckman, M. E. (1986). *Stress and non-stress accent*. Dordrecht: Foris.

Benki, J. R. (2001). Place of articulation and first formant transition pattern both affect perception of voicing in English. *Journal of Phonetics*, *29*, 1–22.

Bent, T. (2005). *Perception and Production of non-native prosodic categories*. Doctoral dissertation, Northwestern University.

Best, C. T. (1995). A direct realistic view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171–206). Baltimore: York Press.

Boersma, P., and Weenink, D. (1992–2009). Praat: Doing phonetics by computer [Computer program]. 〈http://www.praat.org/〉.

Bohn, O. S. (1995). Cross-language speech perception in adults: First language transfer doesn't tell it all. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 279–304). Baltimore, MD: York Press.

Campbell, N., & Beckman, M. (1997). Stress, prominence, and spectral tilt. In A. Botinis, G. Kouroupetroglou, & G. Carayiannis (Eds.), *Proceedings of the ESCA workshop on intonation: Theory, models and applications* (pp. 67–70), Athens.

Chen, Y., Robb, M. P., Gilbert, H. R., & Lerman, J. W. (2001). Vowel production by Mandarin speakers of English. *Clinical Linguistics and Phonetics*, *15*(6), 427–440.

Clayards, M., Aslin, R.N., Tanenhaus, M.K., & Jacobs, R.A. (2007). Within category phonetic variability affects perceptual uncertainty. In *Proceedings of international congress of phonetic sciences* (pp. 701–703), Saarbruken, Germany.

Cutler, A. (1986). Forbear is a homophone: Lexical prosody does not constrain lexical access. *Language and Speech*, *29*, 101–220.

Cutler, A., & van Donselaar, W. (2001). Voornaam is not a homophone: Lexical prosody and lexical access in Dutch. *Language and Speech*, *44*, 171–195.

Davidson, L. (2006). Phonology, phonetics or frequency influences on the production of non-native sequences. *Journal of Phonetics*, *34*(1), 104–137.

Duanmu, S. (2002). *The phonology of standard Chinese*. New York: Oxford University Press.

Dupoux, E., Pallier, C., Sebastian, N., & Mehler, J. (1997). A destressing "deafness" in French?. *Journal of Memory and Language*, *36*, 406–421.

Dupoux, E., Peperkamp, S., & Sebastian-Galles, S. (2001). A robust method to study stress 'deafness'. *Journal of the Acoustical Society of America*, *110*, 1606–1618.

Dupoux, E., Sebastian-Galles, N., Navarete, E., & Peperkamp, S. (2008). Persistent stress 'deafness': The case of French learners of Spanish. *Cognition*, *106*(2), 682–706.

Escudero, P., & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition*, *26*(4), 551–585.

Flanagan, J. L. (1957). Estimation of the maximum precision necessary in quantizing certain "dimensions" of vowel sounds. *Journal of the Acoustical Society of America*, *29*(4), 533–534.

Flege, J. E., & Bohn, O. S. (1989). An instrumental study of vowel reduction and stress placement in Spanish-accented English. *Studies of Second Langauge Acquisition*, *11*, 35–62.

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

Flege, J. E., & Hillenbrand, J. (1987). Limits on phonetic accuracy in foreignlanguage production. In G. Ioup, & S. Weinberger (Eds.), *Interlangugae phonology: The acquisition of a second language sound system* (pp. 176–201). Cambridge: Newbury House.

Fokes, J., & Bond, Z. S. (1989). The vowels of stressed and unstressed syllables in nonnative English. *Language learning*, *39*(3), 341–373.

Francis, A. L., & Nusbaum, H. C. (2002). Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 349–366.

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America*, *27*, 765–768.

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

Fry, D. B. (1965). The dependence of stress judgments on vowel formant structure. In E. Zwimer, & W. Bethge (Eds.), *Proceedings of the 5th international congress of phonetics sciences* (pp. 306–311). Basel: Karger.

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

Gandour, J. (1978). The perception of tone. In V. Fromkin (Ed.), *Tone: A linguistics survey* (pp. 41–76). New York: Academy.

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

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

Groenen, P., Crul, T., Maassen, B., & van Bon, W. (1996). Perception of voicing cues by children with early otitis media with and without language impairment. *Journal of Speech and Hearing Research*, *39*, 43–54.

Guion, S. G., & Pederson, E. (2007). Investigating the role of attention in phonetic learning. In O.-S. Bohn, & M. Munro (Eds.), *Second-language speech learning: The role of language experiencein speech perception and production: A festschrift in honor of James E. Flege* (pp. 57–77). Amsterdam: Benjamins.

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

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, *97*(5), 3099–3111.

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

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., & Kettermann, A., et al. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, *87*, B47–B57.

Jia, G., Strange, W., Wu, Y., & Collado, J. (2006). Perception and production of English vowels by Mandarin speakers: age-related differences vary with amount of L2 exposure. *Journal of the Acoustical Society of America*, *119*(2), 1118–1130.

Kondaurova, M. V., & Francis, A. L. (2008). The relationship between native allophonic experience with vowel duration and perception of the English tense/lax vowel contrast by Spanish and Russian listeners. *Journal of the Acoustical Society of America*, *124*(6), 3959–3971.

Lea, W. A. (1977). Acoustic correlations of stress and juncture. In Hyman, L. M (Eds.), *Studies in stress and accent (South California occasional papers in linguistics, 4)*.

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

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

Mayo, C., & Turk, A. (2004). Adult–child difference in acoustic cue weighting are influenced by segmental context: Children are not always perceptually biased toward transitions. *Journal of Acoustic Society of America*, *115*(6), 3184–3193.

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

Morrison, G. S. (2005). An appropriate metric for cue weighting in L2 speech perception: Response to Escudero and Boersma (2004). *Studies in Second Language Acquisition*, *27*, 597–606.

Nissen, S. L., Harris, R. W., Jennings, L., Eggett, D. L., & Buck, H. (2005). Psychometrically equivalent trisyllabic words for speech reception threshold testing in Mandarin. *International Journal of Audiology*, *44*(7), 391–399.

Pampel, F. C. (2000). *Logistic regression: a primer*. Thousand Oaks, CA: Sage Publications.

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

Rochet, B. L. (1995). Perception and production of second-language speech sounds by adults. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 379–410). Baltimore, MD: York Press.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). E-Prime user's guide. Pittsburgh: Psychology Software Tools Inc..

Seidl, A. (2007). Infants' use and weighting of prosodic cues in clause segmentation. *Journal of Memory and Language*, *57*, 24–48.

Sheldon, A., & Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics*, *3*, 243–261.

Sluijter, A. M.C., & van Heuven, V. J. (1996a). Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America*, *100*(4), 2471–2485.

Sluijter, A. M. C., & van Heuven, V. J. (1996b). Acoustic correlates of linguistic stress and accent in Dutch and American English. In *Proceedings of the fourth international conference on spoken language processing* (Vol. 2, pp. 630–633), Philadelphia.

Tong, Y., Francis, A. L., & Gandour, J. (2008). Processing dependencies between segmental and suprasegmental features in Mandarin Chinese. *Language and Cognitive Processes*, *23*(5), 689–708.

Viswanathan, V., & Russell, W. (1985). New objective measures for the evaluation of pitch extractors. In Acoustic, speech, and signal processing, IEEE international conference on ICASSP'85 (Vol. 10, pp. 411–414), April.

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

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

Wood, C. C. (1975). Auditory and phonetic levels of processing in speech perception: Neurophysiological and information—Processing analyses. *Journal of Experimental Psychology: Human Perception & Performance*, *104*, 3–20.

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

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

Zhang, Y., Nissen, S., & Francis, A. L. (2008). Acoustic characteristics of English lexical stress produced by native Mandarin speakers. *Journal of the Acoustical Society of America*, *123*(6), 4498–4513.