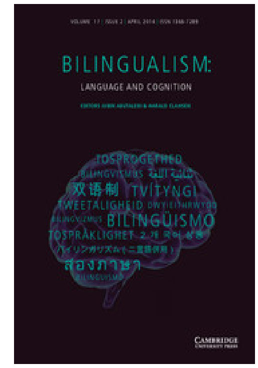

Stress processing in Mandarin and Korean second language learners of English

CANDISE Y. LIN, MIN WANG, WILLIAM J. IDSARDI and YI XU

Bilingualism: Language and Cognition / Volume 17 / Issue 02 / April 2014, pp 316 - 346
DOI: 10.1017/S1366728913000333, Published online: 02 August 2013

Link to this article: http://journals.cambridge.org/abstract_S1366728913000333

How to cite this article:
CANDISE Y. LIN, MIN WANG, WILLIAM J. IDSARDI and YI XU (2014). Stress processing in Mandarin and Korean second language learners of English . Bilingualism: Language and Cognition, 17, pp 316-346 doi:10.1017/S1366728913000333

Bilingualism: Language and Cognition 17 (2), 2014, 316–346 © Cambridge University Press 2013 doi:10.1017/S1366728913000333

Stress processing in Mandarin and Korean second language learners of English*

CANDISE Y. LIN
Department of Human Development and Quantitative Methodology, University of Maryland, College Park, USA
MIN WANG
Department of Human Development and Quantitative Methodology, University of Maryland, College Park, USA
WILLIAM J. IDSARDI
Department of Linguistics, University of Maryland, College Park, USA
YI XU
Department of Speech, Hearing and Phonetic Sciences, University College London, UK

(Received: December 27, 2012; final revision received: June 1, 2013; accepted: June 4, 2013; first published online 2 August 2013)

This study examined stress processing among Mandarin and Korean second language learners of English and English monolinguals. While both English and Mandarin have contrastive stress at the word-level, Korean does not. Consequently, Mandarin speakers may have an advantage over Korean speakers in English stress processing, even when matched for their general English proficiency. Experiment 1 assessed participants' stress encoding ability for nonwords in a short-term memory task. Experiment 2 examined the effect of stress in online word recognition in a lexical decision task by manipulating word frequency, stress location, and vowel quality. The results of both experiments support an advantage for English and Mandarin speakers over Korean speakers in stress processing of real words and nonwords. Only Korean speakers' lexical judgment of nonwords was modulated by word frequency, suggesting that they do not utilize stress in lexical access. Only English speakers' word recognition was facilitated by vowel quality changes. These results suggest that the abilities of non-native speakers to process stress in their L2 is influenced by the characteristics of the stress systems in their L1.

Keywords: stress processing, second language acquisition, cross-linguistic influence, prosody

Introduction

Adult second language (L2) learners' ability to perceive speech contrasts in the L2 may be heavily dependent on the phonological structures of their native languages (L1). L2 contrasts that do not exist in L1 are difficult to discriminate and acquire and such difficulties have been shown with both vowel and consonant contrasts (Best & Strange, 1992; Flege, Bohn & Jang, 1997; Flege & MacKay, 2004; Werker & Tees, 1984). For example, the English /r–l/ distinction is difficult for Japanese and Korean speakers but not for German speakers (Borden, Gerber & Milsark, 1983; Goto, 1971; Iverson, Kuhl, Akahane-Yamada, Diesch, Tohkura, Kettermann & Siebert, 2003; Miyawaki, Strange, Verbrugge, Liberman, Jenkins & Fujimura, 1975). Although the poor perception of non-native segmental contrasts is well documented, less attention has been paid to L2 learners' perception of non-native suprasegmental contrasts. Suprasegmenal (or prosodic) features, such as intonation and rhythm, are speech attributes that accompany consonants and vowels but which are not limited to single sounds and often extend over syllables, words or phrases (Crystal, 2003). One suprasegmental feature that may have important influence on L2 processing is stress. Stress is prominence given to a certain syllable in a word (Kager, 2007). Languages differ in whether they use word stress to indicate lexical contrast and in the phonetic attributes used to mark prominence. A learner whose L1 does not use stress to differentiate between lexical items may have difficulty recognizing words in an L2 that does (Dupoux, Peperkamp & Sebastián-Gallés, 2001). The purpose of the current study was to compare stress processing in two groups of English L2 learners: one group's L1 has lexically contrastive stress whereas the other group's L1 does not have stress at the word level.

* The first author was supported by the NSF IGERT award DGE-9791455 to the University of Maryland. We would like to thank the five native English speakers who recorded the auditory stimuli and the Mandarin and Korean speakers who completed the pre-experimental familiarity ratings. We thank the editor, Dr. David Green, and an anonymous reviewer for their valuable comments and suggestions.

Address for correspondence:
Candise Lin, Dept of Human Development, 3304 Benjamin Building, University of Maryland, College Park, MD 20742, USA
candisec@umd.edu

Stress typology

Stress is lexically contrastive in languages such as English, Russian, Spanish, and German. Such languages can include minimal pairs of words that differ only in stress location. For example, the American English (henceforth: English) words *trusty* and *trustee* share the same phonemes and they can only be distinguished from each other by stress location. The location of stress is generally less predictable in languages with lexically contrastive stress compared to those with no contrastive stress. In English, stress can fall on any syllable, although it is partially dependent on word class. In contrast, stress in French is always word- or phrase-final (Schane, 1968) whereas in Finnish, stress is always word-initial (Karlsson, 1999). When stress is not used to distinguish between lexical items, there are no minimal pairs of words that differ in stress location alone. Previous studies have consistently shown that speakers of non-contrastive stress languages (e.g., French) have greater difficulty discriminating minimal pairs of nonwords that differ only in the location of stress; this difficulty is termed "stress deafness" (Dupoux, Pallier, Sebastián-Gallés & Mehler, 1997; Dupoux et al., 2001; Peperkmap & Dupoux, 2002). "Stress deafness" was also observed in native French speakers with advanced L2 proficiency in Spanish (Dupoux, Sebastián-Gallés, Navarrete & Peperkamp, 2008), as well as in simultaneous French–Spanish bilinguals exposed to both languages from birth (Dupoux, Peperkamp & Sebastián-Gallés, 2010).

Theoretical framework

Several models in the L2 literature have been proposed to explain learning of non-native contrasts such as the Perceptual Assimilation Model (PAM) (Best, 1995; Best, McRoberts & Sithole, 1988) and the Speech Learning Model (SLM) (Flege, 1992, 1995; Flege, Schirru & MacKay, 2003). Since both of these models were originally developed to address the learning of segmental contrasts, they may not be generalizable to the learning of L2 suprasegmental contrasts in the current study. The Stress Parameter Model (SPM) (Peperkamp, 2004; Peperkamp & Dupoux, 2002), on the other hand, is concerned with the influence of typology on the encoding and representation of stress in phonological memory. Since SPM is directly related to stress processing, we built on this model for our investigations of the processing of stress contrasts by L2 learners with or without experiences with contrastive stress in their L1.

The Stress Parameter Model hypothesizes that learners' ability to discriminate minimal stress pairs is determined by whether they set the Stress Parameter to encode stress in their phonological representation early in life. Research has shown that prelexical infants are sensitive to the stress cues in their native languages (Jusczyk, Cutler & Redanz, 1993; Mehler, Jusczyk, Lambertz, Halsted, Bertoncini & Amiel-Tison, 1988) and the cues would help them determine whether stress is contrastive in their native language. If they decide that stress is contrastive, they will set their Stress Parameter to encode stress in the phonological representation. If stress is non-contrastive, infants would not develop any strategy for processing stress. Stress is lexically contrastive in Spanish but not in French. Using the head-turn preference paradigm, Skoruppa, Pons, Christophe, Bosch, Dupoux, Sebastián-Gallés, Limissuri & Peperkamp (2009) showed that nine-month-old infants from Spanish-speaking families spent more time listening to a stress pattern different from what they heard in the familiarization phase than infants raised in French-speaking families, and that infants from Spanish-speaking families were more sensitive to the difference between initial stress and medial stress than infants from French-speaking families.

It seems that once the Stress Parameter has been set to off in infancy, it cannot be reset (switched on again) because even French speakers who were advanced L2 learners of Spanish showed difficulty perceiving stress contrasts. Dupoux et al. (2008) compared three groups of French speakers learning Spanish after age 15. The groups had varying L2 proficiency (i.e., beginner, intermediate, and advanced). Regardless of their Spanish proficiency, the French late L2 learners' perception of stress contrasts was similar to that of the French monolinguals. The researchers concluded that "stress deafness" is a robust processing limitation that was not malleable through extensive exposure to a L2 with contrastive stress.

Even simultaneous French–Spanish bilinguals who were exposed to Spanish from birth did not perform as well as the monolingual Spanish speakers in stress perception in Dupoux et al. (2010). When the simultaneous bilinguals were divided into two groups based on language dominance, Spanish-dominant bilinguals performed similarly to Spanish monolinguals, whereas French-dominant bilinguals performed similarly to French late learners of Spanish. Dupoux and colleagues reasoned that this was probably because, in terms of phonological perception, simultaneous bilingual can only have one language that is processed at a native-like fashion. Thus, those simultaneous bilinguals who showed difficulty perceiving stress contrasts were probably less native-like in their Spanish processing. Overall, SPM predicts that the difficulty with non-native stress perception is largely determined by whether the learners' L1 has lexically contrastive stress.

Although the proposed Stress Parameter is binary, recent evidence has suggested that stress processing abilities are more dynamic and influenced by environmental input and the typological difference between the L1 and L2 prosodic systems. Tremblay (2008, 2009) found that Canadian French speakers are less stress deaf than European French speakers. In a cross-modal priming task, Tremblay (2008) examined the effect of stress in English word recognition in Canadian French speakers who were L2 English learners. The primes were the first syllable of the target words that either matched or mismatched the stress pattern of the target word (e.g., /ˈmɪs/ or /mɪs/ for *mystery*). A significant priming effect was found for the congruent stress primes. Since Canadian French has a greater contrast between strong and weak syllables than European French, Canadian French speakers' better perception could be attributed to this difference between the two French languages. In the current study, we conceptualized the mechanism for stress processing as a gradient construct which is influenced by the characteristics of the stress system in one's language.

Stress in Mandarin and Korean

The current study compared native Mandarin and Korean speakers' perception of L2 English stress contrasts. The reason for choosing these languages is that the typological difference between each L1–L2 combinations resulted in different levels of stress processing abilities as hypothesized by SPM. Although Mandarin is conventionally classified as a tonal language, lexically contrastive stress exists in standard Mandarin based on the Beijing dialect (hereafter: Mandarin) and it is realized with similar phonetic attributes as in English (Chen & Xu, 2006; Duanmu, 2007; M. Lin & Yan, 1980; T. Lin, 1985; T. Lin & Wang, 1984). As discussed above, in English there are minimal pairs of words that differ in stress alone (e.g., *trusty* /ˈtrʌsti/ and *trustee* /trʌˈsti/). The same holds for Mandarin. For instance, the minimal pair, /ˈtʊ̄ŋˈçī/ "East West" and /ˈtʊ̄ŋçi/ "thing", is written using the same characters 东西. The only difference between the two words is that in the former both syllables carry stress, while in the latter the second syllable is unstressed. In Mandarin, full syllables carry one of the four lexical tones and syllables with these tones are pronounced louder and have longer duration and greater amplitude than light syllables (Duanmu, 2007; M. Lin & Yan, 1980; T. Lin, 1985; T. Lin & Wang, 1984). In contrast, light syllables carry the neutral tone and show considerable vowel reduction (Chao, 1968) and weaker articulatory strength (Chen & Xu, 2006). Duanmu (2007) equated full syllables in Mandarin to stressed syllables in English, and light syllables in Mandarin to unstressed syllables in English (see also Chao, 1968). In English, stressed syllables are pronounced with greater intensity, higher pitch, longer duration (Fry, 1955, 1958) and greater articulatory strength (Xu & Xu, 2005). Like other languages with contrastive stress, the location of stress is generally unpredictable in Mandarin. There are three possible stress patterns including 1–0 (heavy–light, e.g., [paa.pa] "dad"), 1–2 (heavy–heavy where the first syllable has more stress, e.g., [tɕii.xwaa] "plans"), and 2–1 (heavy–heavy where the second syllable has more stress, e.g., [swuu.ʂɤɤ] "dorm") (Duanmu, 2007).[1]

In contrast, Korean does not have lexically contrastive stress (Jun, 2005; Sohn 1999). In Korean, there are no minimal word pairs differing in stress alone. Korean also does not have fixed stress at the word level (Jun, 1995) like Finnish or Turkish. According to Guion (2005), Korean prosody differs from English in two major ways. First, the basic building block of prosody in Korean is tone patterns whereas in English it is stress accent. Second, the domain of prosodic association for English stress is the lexical word whereas Korean tone patterns are associated with the accentual phrase. For example, stress falls on the initial syllable in English words such as *table* and *orchestra*. In Korean, an intonation phrase such as 첫교장선생님 (first principal) consists of three lexical words (i.e. 첫 'first', 교장 "headmaster", and 선생님 "teacher"). Such phrase is known as the Accentual Phrase (AP) in which the default pitch pattern in the Seoul Korean dialect is Low–High–Low–High (LHLH) (Jun 1998, 2000). When the AP-initial segment is either aspirated or tense, the AP begins with the H tone (Jun, 1998, 2000; Kim & Cho, 2009). However, the higher pitch in Korean does not signal any lexical shift whereas the higher pitch in English is one of the acoustic cues for lexical stress (Fry, 1958).

Hypotheses

Assuming SPM (Peperkamp & Dupoux, 2002), we hypothesized that Mandarin and English speakers would have better stress perception in English than Korean speakers. Mandarin infants would encode stress in their phonological representation since they can use acoustic cues to deduce that stress location is unpredictable in their L1. Native Mandarin speakers may be able to transfer stress encoding strategies when learning a contrastive stress L2 language such as English. In contrast, since stress is not part of the phonetic realization of the Korean prosodic system, Korean speakers may not have developed a processing mechanism for stress. When Korean speakers learn English as a L2, they do not have stress-encoding strategies available due to their L1 experience. Thus Korean speakers will exhibit similar difficulty with stress perception as the French learners of L2 Spanish as reported in previous studies (Dupoux et al., 2008, 2010).

SPM predicts that speakers' difficulty with stress perception would vary depending on the regularity of stress location in their L1 (Peperkamp, 2004; Peperkamp & Dupoux, 2002). Thus, this model may only account for

[1] Numbers denote stress levels.

languages with stress at the word level. Our investigation of stress perception in Korean L2 learners of English extends this model to suprasegmental learning and processing in L2 learning with an L1 that has no word level stress.

Korean speakers' difficulty with stress perception was hypothesized on the basis that they do not utilize stress when accessing or encoding lexical items in short-term memory, rather than their inability to hear the phonetic prominence in stressed syllables in English. In fact, in languages without contrastive stress such as Korean and French, not every syllable is given equal prominence and speakers of these languages may actually be sensitive to these prominence distinctions. For example, the Accentual Phrase in standard Korean is marked by a phrase-final High tone (Jun, 1998), which is realized through a rise in fundamental frequency. Research has shown that native Korean speakers could use this change in pitch to identify word boundaries (Kim, Broersma & Cho, 2012). Similarly, native French speakers also use this prominence distinction in speech segmentation (Christophe, Peperkamp, Pallier, Block & Mehler, 2004; Tremblay, Coughlin, Bahler & Gaillard, 2012; Tyler & Cutler, 2009). Thus, it is important to note that the hypothesized difficulty of Korean speakers may stem from problems with encoding stress at the lexical level rather than perceiving syllable prominence at the acoustic level.

The current study

Stress processing was examined using a sequence recall task and a lexical decision task, both adapted from Dupoux et al. (2008, 2010). The sequence recall task involved two minimal pairs of nonwords including a phoneme contrast (e.g. /p/ vs. /t/: /ˈkupi/ /ˈkuti/) and a stress contrast (/ˈmipa/ /miˈpa/). Since both /p/ and /t/ are legal phonemes in all three languages, there should not be any substantial group difference in processing the phoneme contrast. The absence of group differences in the phoneme condition would suggest that differences in the stress condition cannot be attributed to differences in participants' short-term memory capacity. During the training phase of the experiment, all participants learned to associate the number key [1] with the first nonword in a minimal pair and [2] with the second nonword. In the familiarization phase, participants heard one token of each nonword and had to press the corresponding key [1] or [2]. Once participants had accumulated seven consecutive correct responses, they entered the testing phase in which they heard nonword sequences varying in length from two to six nonwords, and had to press the keys in the corresponding order. For example, the correct response for the sequence /ˈmipa/ /miˈpa/ /ˈmipa/ /miˈpa/ is 1212. The auditory stimuli in the experiment were recorded by one male and one female native English speaker, thus increasing phonetic variability. In addition, the inter-stimulus interval between each nonword was only 80 ms, which placed a high demand on short-term memory and prevented the use of contrasting acoustic stress cues (e.g., 1 represents the word in which initial syllable has higher pitch than final syllable and 2 represents the word in which final syllable has higher pitch than initial syllable) to memorize the sequences. In other words, participants' performance depended on their ability to encode the input into its phonological representation as soon as the input for each nonword was available.

Participants in the current study included three language groups, Korean and Mandarin L2 learners of English, and native English speakers. We hypothesized that, assumning SPM, both Mandarin and English speakers would outperform Korean speakers in the stress condition since stress is lexically contrastive in Mandarin and English but not in Korean.

In the lexical decision task, stimuli were real English words as well as nonwords obtained from changing the location of stress in the real words (e.g., *between* /bɪˈtwin/ has final stress, and its nonword counterpart /ˈbɪtwin/ has initial stress). Word–nonword minimal pairs differing in one phoneme were used as controls (e.g., /əˈweɪ/ *away* vs. /əˈmeɪ/). In addition to stress location, we also manipulated word frequency in the real words and vowel quality change in the nonwords. We hypothesized that both native and non-native speakers would process words with higher frequency faster and more accurately than words with lower frequency. Based on SPM, it was hypothesized that both English and Mandarin speakers would outperform Korean speakers in rejecting the nonwords with stress change. Since vowel reduction is not part of the phonetic realization in the Korean prosody (Lee, Guion & Harada, 2006), SPM would also predict that Mandarin and English speakers, but not Korean speakers, would show better rejection of nonwords with vowel change than those with no vowel change.

Experiment 1: Sequence recall

Participants

There were 20 Mandarin L2 learners of English, 19 Korean L2 learners of English, and 21 English monolinguals (they will be referred to as Mandarin speakers, Korean speakers, and native English speakers, respectively, after this point). They were all current students at a mid-Atlantic university in the United States. English proficiency was measured by the Language Experience and Proficiency Questionnaire (LEAP-Q) (Marian, Blumenfeld & Kaushanskaya, 2007), a standardized test, and a cloze test (Bachman, 1982). Table 1 shows the descriptive statistics of the demographic information and proficiency ratings collected from the

Table 1. *Participants' demographics.*

	Korean	Mandarin	English
Age of English acquisition	12.1 (2.40)	9.8 (2.22)	
Length of U.S. residence (months)	35.3 (33.4)	18.4 (16.33)	
Percentage of L2 use and exposure	50 (19.1)	47.0 (14.4)	
Self-rated English proficiency			
Speaking	5.11 (1.97)	6.0 (1.86)	8.9 (2.07)
Understanding spoken language	5.79 (1.84)	6.75 (1.71)	9.0 (2.07)
Reading	6.74 (1.41)	7.3 (1.56)	9.0 (1.79)
TOFEL (%)	82.1 (1.86)	84.9 (0.6)	
Cloze (%)	49.62 (0.11)	55.78 (0.17)	77.5 (0.09)

Note: The self-rating was based on an 11-point Likert scale with 0 = "none" and 10 = "perfect". The TOFEL and cloze test scores were reported in terms of percentage with a total possible of 100%.

questionnaire and tests. Student's *t*-tests (with equal variance not assumed) showed that Mandarin speakers began to learn English significantly earlier than Korean speakers ($t(35.2) = 3.169$, $p = .003$), although Korean speakers have lived in the United States marginally longer than Mandarin speakers ($t(23.6) = 1.969$, $p = .051$). The L2 learner groups did not differ significantly in their percentage of L2 use and exposure ($t(31.0) = .543$, $p = .591$). For the self-reported proficiency ratings, native English speakers rated their proficiency in all three areas – speaking, understanding, and reading – significantly higher than Mandarin speakers (all ps $< .003$) and Korean speakers (all ps $< .001$). Mandarin and Korean speakers did not differ in their self-rated proficiency in the three areas (all ps $> .1$). The standardized measure, the TOEFL (Test of English as a Foreign Language) assesses English competence in speaking, listening, reading, and writing. The TOEFL test includes paper-, computer-, and internet-based versions. Thus, to make the scores comparable across non-native participants, we calculated percentages by dividing raw scores from total possible scores for each test. *T*-test comparisons revealed no significant difference between Mandarin and Korean speakers in TOEFL scores ($t(36) = 1.189$, $p = .244$).

Finally, the cloze test (Bachman, 1982) is a 30-item measure of syntactic and lexical knowledge in which participants read passages and fill in missing words to complete sentences based on passage content. Participants earned one point for each correct response and a percentage score for total number correct was calculated for each participant. Native English speakers performed significantly better than Mandarin and Korean speakers (both ps $< .001$), while the non-native groups performed similarly ($t(33) = 1.273$, $p = .212$). Native English speakers' self-reported proficiency ratings and cloze test scores confirmed their status as native English speakers. The three proficiency measures provided convergent evidence that the two non-native groups did not differ in their English proficiency. However, in light of the fact that the Mandarin speakers' means across all three assessments were slightly higher than those of the Korean speakers, in subsequent analyses we entered cloze test scores in the mixed-effects models as a proxy covariate to account for the variance in the proficiency measures.

Measures

There were three minimal contrasts in the sequence recall task (Dupoux et al., 2001, 2008, 2010) including the phoneme contrast (/ˈkupi/ /ˈkuti/) and the stress contrast (/ˈmipa/ /miˈpa/). The linguistic stimuli were recorded by two native speakers of American English (with Northeastern dialect), one male and one female, who were naïve to the hypotheses of the current study. The target nonwords were embedded in the sentence frame "I think ____ may not come" in which the focus of the sentence falls on the negation to prevent additional emphasis on the nonwords. Each speaker repeated the sentence for each target word 10 times. Three tokens were randomly selected for each nonword from each speaker. The tokens were randomly selected to construct the sequences, but no token was used consecutively in a sequence. Acoustic analyses of the stress contrasts revealed that the pitch of the stressed syllables was 77.51 Hz higher than that of the unstressed syllables ($t(46) = 4.530$, $p < .001$). The intensity of the stressed syllables was 2.917db higher than that of the unstressed syllables ($t(46) = 2.112$, $p = .04$). The duration of the stressed syllables was 64 ms longer than that of the unstressed syllables ($t(46) = 3.634$, $p = .001$). These results indicate that the acoustic cues of stress are unambiguous for these words.

Procedures

Participants always completed the phoneme contrast first and then the stress contrast. For both contrasts, the procedure was the same. First, in the learning phase, participants learned to associate the pairs of nonwords with numbers [1] and [2] on a keyboard (e.g., match [1] with /ˈkupi/ and [2] with /ˈkuti/). Next, they completed a checking phase in which they heard one nonword at a time, and pressed the corresponding number labeled on the computer keyboard. Feedback was provided following each key press. When the participants had accumulated seven consecutive correct responses, the computer program (E-prime, Psychology Software Inc., Pittsburgh, PA) proceeded to the testing phase. The testing phase consisted of five sub-blocks arranged in incremental order of difficulty. The first sub-block contained two-word sequences (e.g., /ˈkupi/ /ˈkuti/) and the second sub-block contained three-word sequences (e.g., /ˈkupi/ /kuti/ /ˈkupi/). The third sub-block contained four-word sequences (e.g., /ˈkupi/ /ˈkuti/ /ˈkupi/ /ˈkuti/) and the fourth sub-block contained five-word sequences (e.g., /ˈkupi/ /ˈkuti/ /ˈkuti/ /ˈkupi/ /ˈkuti/). Finally, the fifth sub-block contained six-word sequences (e.g., /ˈkupi/ /ˈkupi/ /ˈkuti/ /ˈkupi/ /ˈkupi/ /ˈkuti/; see Appendix A for a list of sequences). There were eight randomly ordered trials in each sub-block for each participant.

In each trial, participants first saw a fixation sign "+" for 500 ms in the center of the screen. Then they heard the sequence, followed by the cue word "OK" presented auditorily, after which they could respond. The word "OK" was used to prevent participants from using echoic memory to recall sequences. A visual reminder (e.g., "2 words") regarding the number of words in each sequence was presented at the beginning of each new block. Participants must have made the same number of key presses corresponding to the length of the input strings to proceed to the next trial (e.g., for /miˈpa/ /miˈpa/ /ˈmipa/ /miˈpa/, there must be four key presses). No feedback was given during the real test trials. If an incorrect number of key presses was recorded, the trial would automatically terminate after 5000 ms. About 2.6% for the phoneme condition and 0.7% for the stress condition were excluded on this basis.

Results

For data analyses, only responses that were 100% correct transcriptions of the sequences were coded as correct. All other responses were coded as incorrect. Responses that were 100% incorrect transcriptions (e.g., the response was 12121 for a sequence of 21212) were secondarily coded as inversions. Participants who had more inversions than correct responses were excluded from data analyses in that contrast. The reason for exclusion was to rule out incorrect responses due to the potential transposition of directions (e.g., participants who mistakenly associated the number key [1] with the second item and vice versa). This criterion resulted in the exclusion of one Mandarin and two Korean speakers from the phoneme contrast and three native English speakers from the stress contrast. The data from these participants were only excluded from one section of the experiment but was retained for the other section.

Table 2. *Total accuracy and accuracy in individual sequence lengths in the two contrasts in the sequence recall task in Experiment 1.*

			Length of sequence (words)				
Contrast		Total	2	3	4	5	6
Stress	Mandarin	.465	.856	.625	.469	.175	.200
	Korean	.315	.717	.480	.237	.086	.053
	English	.401	.826	.681	.285	.160	.104
Phoneme	Mandarin	.499	.888	.699	.474	.283	.132
	Korean	.416	.779	.630	.382	.213	.081
	English	.506	.955	.792	.432	.193	.142

Table 2 and Figure 1 show the total accuracies across combined and individual sequence lengths in all three contrasts across language groups. All subsequent analyses were conducted in R studio, an open source software environment for statistical computing and graphics (R Development Core Team, 2008), using the lme4 package for generalized linear (binomial) mixed-effects regression (Bates & Maechler, 2010) of the unaveraged data for each participant. Figures were created in R using the ggplot2 (Wickham, 2009) package. In the model predicting accuracy, subjects and items were entered as random effects while language groups (English, Korean, and Mandarin) and conditions (phoneme and stress), and proficiency as fixed effects. Essentially, our model (Table 3) considers the effects of language groups, conditions, and proficiency on sequence recall accuracy after accounting for the variance among participants and items. Estimates represent the log odds of change in accuracy resulting from a change in language group, condition, proficiency, or a combination of two or more of the variables and p-values are based on the *Wald* z distribution. Data were analysed with two separate models. The initial model treated the English group as the intercept and this model compared the performance of the native speakers with that of the Mandarin and Korean speakers. Then the second model treated the Korean group as the intercept and this model compared the performance between the two L2 groups. Initially, the models include one three-way interaction among language group, condition, and proficiency and three two-way interactions (i.e., language × condition, language × proficiency, and condition × proficiency). The interaction terms were removed one at a time until we found the model with the best goodness-of-fit using the Chi-square test.

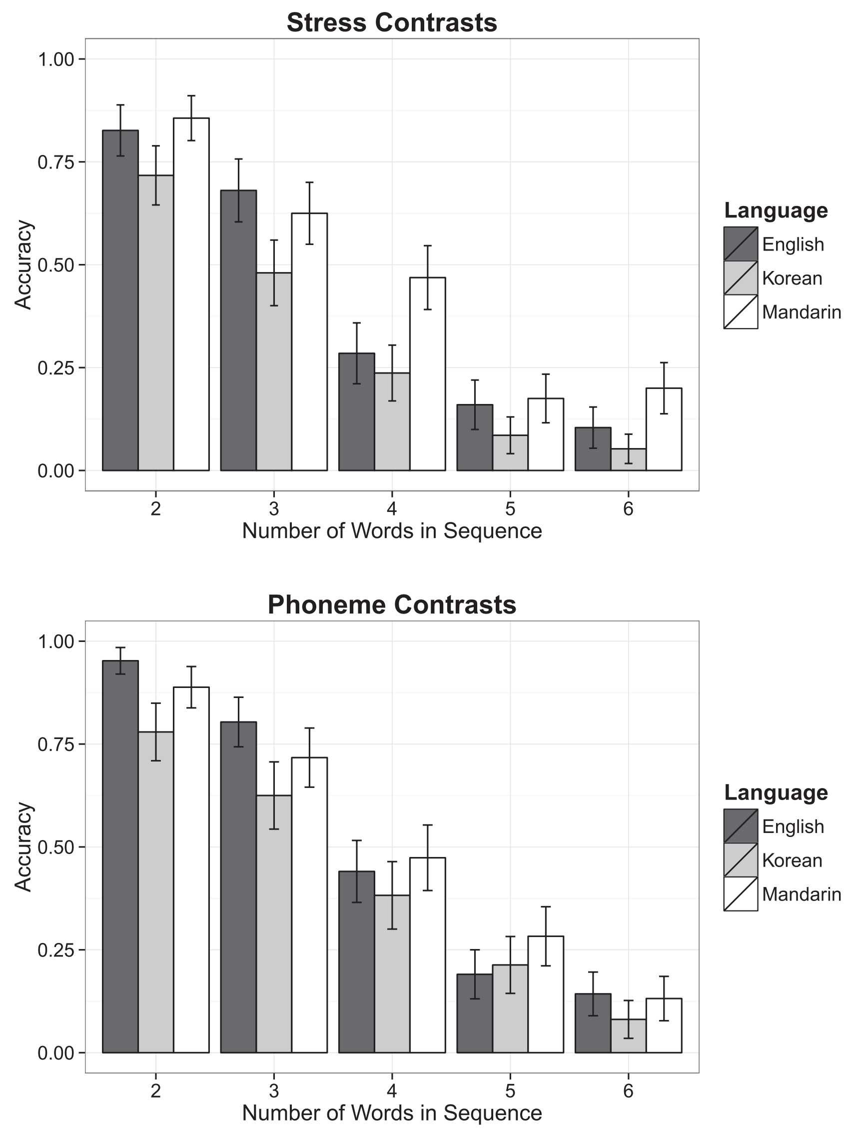

Figure 1. Accuracy in the sequence recall task in Experiment 1.

In the initial model with the English group as the baseline (upper part of Table 3), there was no significant difference in accuracy between the native English speakers and each of the L2 groups in the phoneme condition. The English speakers were significantly more accurate in the phoneme condition than in the stress condition. The interaction between language group and condition was significant when comparing the Mandarin and English groups but not significant when comparing the Korean and English groups. This is because the asymmetry in accuracy between the stress and phoneme conditions is present in the Korean speakers but not in the Mandarin speakers. In other words, the Mandarin speakers did not differ significantly in their performance in the phoneme vs. stress conditions. In the second model with the Korean group as the baseline (lower part of Table 3), the Korean and Mandarin speakers' accuracy in the phoneme condition did not differ significantly. The Korean speakers showed significantly better performance in the phoneme than in the stress condition. This asymmetry was not present in the Mandarin speakers, resulting in the marginally significant interaction between language group and condition. The effect of proficiency and its interaction with language group were not statistically significant.

Discussion

Generally, the results were consistent with our hypotheses. The Korean speakers had greater difficulty recalling the minimal stress pair (e.g., /'mipa/ /mi'pa/) compared to the minimal phoneme pair (e.g., /'kupi/ /'kuti/). This is probably due to having a less efficient automatic

Table 3. *Results from the generalized linear mixed-effects model predicting accuracy with language groups, conditions, and proficiency as fixed effects and participants and items as random effects. Estimates represent the log odds of change in accuracy in recalling the sequences resulting from a change in language group, condition, proficiency, or a combination of the variables.*

The intercept in this model ($\chi^2 = 2.861, p = .09, df = 3991$) estimated the English group's accuracy in the phoneme condition with low proficiency.

Fixed effects	Estimate (log odds)	SE	*z*-value	*p*-value
Intercept	−1.404	1.444	−.972	.331
Language Korean	−.249	1.614	−.154	.878
Language Mandarin	.933	1.551	.602	.547
Condition Stress	−.613	.144	−4.253	<.001
Proficiency	.018	.018	1.016	.309
Language Korean: Condition Stress	.043	.204	.209	.834
Language Mandarin: Condition Stress	.441	.201	2.197	.028
Language Korean: Proficiency	.004	.024	.149	.882
Language Mandarin: Proficiency	−.011	.021	−.538	.590

The intercept in this model ($\chi^2 = 2.761, p = .09, df = 2674$) estimated the Korean group's accuracy in the phoneme condition with low proficiency.

Fixed effects	Estimate (log odds)	SE	*z*-value	*p*-value
Intercept	−1.623	.866	−1.874	.061
Language Mandarin	1.151	1.072	1.074	.283
Condition stress	−.548	.142	−3.852	.0001
Proficiency	.021	.016	1.315	.188
Language Mandarin: Condition stress	.383	.198	1.936	.053
Language Mandarin: Proficiency	−.014	.020	−.724	.469

stress encoding mechanism as a result of having a non-contrastive stress L1. Since this asymmetry was not observed in the Mandarin group, Mandarin speakers' stress-encoding strategy developed from native language experience may have been transferred to stress processing in another language with contrastive stress, which gave them an advantage over Korean speakers in stress perception. The absence of significant group difference between the Mandarin and Korean speakers in the phoneme condition suggests that the Mandarin speakers' better recall of the stress contrasts is unlikely to be due to the Mandarin speakers having better short-term memory capacity than the Korean speakers. The fact that the sequence recall task used nonwords and L2 proficiency did not have a significant interaction with language group further demonstrated the robustness of cross-linguistic influence on L2 stress perception.

Interestingly, we found that both the Korean and English speakers had more difficulty encoding stress phonologically in short-term memory as compared to phonemes. Like the Korean speakers, the English speakers performed worse in the stress condition than in the phoneme condition. According to SPM, English speakers should have set their Stress Parameter to encode stress due to exposure to a language with unpredictable stress locations. Thus, it is unlikely that the English and Korean speakers' difficulty with the stress condition shares the same origin. The use of nonwords in the sequence recall task may be responsible for the English speakers' performance. In Experiment 1, the vowels in the unstressed syllables in the nonwords were not reduced. Thus, speakers could only rely on surpasegmental cues such as pitch, intensity, and duration to recognize stress. For the Mandarin speakers, these are all cues used in distinguishing between lexical tones, which could have facilitated their performance. Previous research has shown that English speakers do not heavily rely on suprasegmental information for lexical access due to the saliency of vowel reduction as a stress cue (Cutler, 1986). Cutler and van Donselaar (2001) suggested that there is a continuum of segmental consequences on stress processing in which English and Spanish represents the two extremes in this continuum. In Spanish, there is no vowel reduction and any syllable can be either stressed or unstressed. Spanish speakers must rely heavily on suprasegmental information to distinguish between minimal pairs. Native Spanish speakers showed lower error rates in the stress condition (19.6%) as compared to the phoneme condition (25%) in the sequence recall task in Dupoux et al. (2008). On the other hand, almost all unstressed syllables in English have vowel reduction. English speakers can make use of the change in vowel quality without relying on suprasegmental information to identify words. The absence of the vowel cue might have negatively impacted English speakers' performance. We tested this hypothesis in Experiment 2 by manipulation of vowel change in a lexical decision task.

Experiment 2: Lexical decision

In Experiment 2, we used a lexical decision task to extend the findings from Experiment 1 beyond speech perception by examining the influence of L1 stress characteristics on L2 online lexical access. We changed the location of stress in English words to make minimal pairs of real words and nonwords differing only in stress. Our predictions regarding language group differences were that based on the findings from Experiment 1, Korean speakers would show lower accuracy and slower response times than Mandarin and English speakers for the nonword pairs. For the real word pairs, we hypothesized that the two non-native English speaking groups would not differ significantly in accuracy or response time since they had similar levels of L2 proficiency; however, they would have lower accuracy level and greater response time than native English speakers.

Stimuli were created to vary in word frequency and in whether unstressed vowels were pronounced with vowel reduction in nonwords. Since the processing advantage for higher frequency words was a strong effect consistently found in native and non-native speakers (e.g., Grainger, 1990; Weber & Cutler, 2004), replication of this effect in the real words ensured that our participants were indeed using regular lexical retrieval strategies not contaminated by the manipulation of stress shift in the nonwords. In addition, if Korean speakers do not use stress for lexical access, then we may expect to observe the lexical frequency effect even in the nonwords.

The rationale for manipulating vowel reduction in the nonwords was to examine the use of segmental cues in relation to stress processing in native and non-native speakers. Most unstressed syllables in English undergo vowel reduction.[2] Due to the saliency of vowel reduction as a stress cue in English, native speakers' word recognition would be facilitated by the presence of vowel change in nonwords. However, it is unclear how vowel quality would influence lexical access in non-native speakers. We hypothesized that Korean speakers would show poorer performance than English speakers for words with vowel change than those with no vowel change since vowel reduction (i.e., a centralizing change in vowel quality) is not part of the phonetic realization in Korean prosody. On the other hand, because vowel reduction is an acoustic property of unstressed syllables in Mandarin (Duanmu, 2007; Shen, 1993), Mandarin speakers may benefit from vowel quality change in the nonword. Thus, we hypothesized that, similarly to English speakers, Mandarin speakers would perform better in the vowel reduction condition than in the no vowel change condition.

Participants

A total of 54 participants, 17 Mandarin speakers, 18 Korean speakers, and 19 English speakers from Experiment 1 participated in Experiment 2.

Materials

The current experiment was a mixed design with two categorical variables, lexicality and language groups, and a continuous variable, frequency. Lexicality and frequency were within-subject factors while language group was a between-subject factor. A total of 120 monomorphemic real words were used in the lexical decision task. Half of them were disyllabic and the other half trisyllabic. The words roughly fall into three frequency categories: high, medium, and low, with 20 in each (Appendix B). Frequency information was obtained from the English Lexicon Project (Balota, Yap, Cortese, Hutchison, Kessler, Loftis, Neely, Nelson, Simpson & Treiman, 2007) and was based on the Hyperspace Analogue to Language (HAL) log-transformed frequency norms (Lund & Burgess, 1996). We controlled for word length by matching the number of letters across the frequency categories. There was no significant difference in word length within the two- and three-syllable words (both $ps > .1$). For the disyllabic words, there were 29 words with trochaic stress patterns, and 31 words with iambic stress patterns. For the trisyllabic words, there were 28 initial-, 26 medial-, and 6 final-stressed words. Our stimuli reflect a typical distribution of stress patterns in English trisyllabic words. The English stress principles state that words with at least three syllables have antepenultimate stress if the penult is light; when the penult is heavy, it attracts stress (Hayes, 2009). Thus, in English there are a greater number of words with initial- and medial-stress than final-stress. In a database of 20,000 English words, Clopper (2002) identified 2619 trisyllabic words with initial stress, 1510 trisyllabic words with medial stress and 369 trisyllabic words with final stress.

For each real word, its nonword counterpart was obtained by changing the stress location. A total of 120

[2] There are many factors that affect vowel reduction, including syllable structure conditions and dialectal differences. These factors are beyond the scope of the current study.

nonwords were used, 60 disyllabic and 60 trisyllabic. Half of the words underwent vowel change as a result of stress shift. For example, when stress shifts to the second syllable for the real word e.g., *human* /ˈhyumən/, the schwa /ə/ is changed to a full vowel /æ/ and this makes the nonword /hyuˈmæn/. The vowel in the first syllable remains the same. For the real word e.g., *enough* /ɪˈnʌf/, the vowel quality of the stressed syllable is unchanged following stress location change; thus the pronunciation for the nonword is /ˈɪnʌf/.

In addition to the target stimuli of 120 minimal stress pairs, there were 60 minimal phonemic pairs serving as controls. Nonwords were obtained by changing one phoneme in real words while maintaining the same stress location. For example, the nonword version of *away* /əˈweɪ/ is /əˈmeɪ/. There were 30 disyllabic and 30 trisyllabic control words, with a representative range of frequency. The purpose of the control words was to prevent participants from developing stress-specific processing strategies. The fact that some nonwords differ from the real words in phonemes, while others differ from real words in stress would force participants to attend to both segmental and suprasegmental information in lexical judgment.

All of the auditory stimuli (including nonwords) were recorded by three native American English speakers (with a Northeastern dialect): two females and one male, without noticeable regional accents. These native speakers did not record the stimuli for Experiment 1. Each speaker recorded one-third of the 120 minimal stress pairs and 60 minimal phoneme pairs. The speakers used the International Phonetic Alphabet (IPA) for both real word and nonword pronunciation and practiced the pronunciation before recording. All stimuli were recorded inside a sound treated room using a microphone (Audio-Technica ATR 20) and the Cool Edit Pro software (Syntrillium Software, 2003) and the files were stored as uncompressed WAV, digitized at 44.1 kHz at 16 bits. Acoustic analysis of the minimal stress pair recordings revealed no significant difference between real words and nonwords in word-length ($t(238) = -.864$, $p = .388$). Pooling both real words and nonwords together, stressed syllables had significantly higher pitch (54.5 Hz mean difference, $t(235) = 6.435$, $p < .001$), longer duration (29.02 ms mean difference, $t(239) = 4.565$, $p < .001$), and greater intensity (4.78 db mean difference, $t(239) = 3.297$, $p = .001$) than unstressed syllables. There were no significant differences in stressed syllables between real words and nonwords in pitch ($t(238) = -.435$, $p = .664$), duration ($t(238) = 1.602$, $p = .111$), and intensity ($t(238) = -.690$, $p = .491$). There were no significant differences between real and nonwords in unstressed syllables in pitch ($t(234) = -.538$, $p = .591$) and intensity ($t(238) = -1.747$, $p = .082$). However, the unstressed syllables had significantly longer duration in nonwords (42.1 ms mean difference) than in real words ($t(238) = -2.952$, $p = .003$). This result suggested that the speakers did not reduce the duration in unstressed vowels in nonwords to the same degree as those in real words. In light of this result, we included syllable length as a covariate in the generalized linear mixed-effects models for nonwords.

Procedures

The random presentation of trials in the lexical decision task was implemented via E-prime (Psychology Software Inc., Pittsburgh, PA). Participants were asked to judge whether a sound they heard was a real English word. The instruction explicitly asked the participants to consider any mispronounced words as nonwords. Participants pressed the keyboard key labeled "YES" for real words and the key labeled "NO" for nonwords using the left and right index fingers. The experiment consisted of 180 trials (120 target stimuli and 60 fillers), half of which were real words and the other half nonwords, equaling the same number of positive and negative responses. In each trial, a fixation sign "+" was displayed in the center of the screen for 500 ms, followed by the auditory stimulus. Participants' response time was measured from the onset of the auditory stimulus. If participants did not respond within 5000 ms, the trial was terminated and recorded as an incorrect response. Participants received eight practice trials with feedback to familiarize them with the experimental procedure. Two lists were created so that participants heard only the real word or nonword within each minimal pair. For example, if the real word *until* /ʌnˈtɪl/ appeared in List 1, its nonword counterpart /ˈʌntɪl/ appeared in List 2. Participants were randomly assigned to either List 1 or 2.

Forty-two participants completed Experiments 1 and 2 in separate sessions. There was a three-month interval between each session. Twelve participants (nine Korean and three English speakers) completed both Experiments 1 and 2 in a single session (always Experiment 1 first). No significant difference was shown in accuracy rates for either the real words or nonwords regardless of whether they completed the tasks in one or two sessions (both ps $> .1$). Thus, we combined both groups of participants for subsequent analyses.

Results

Response time (RT) data for incorrect responses were excluded from data analysis. Overall, 14.7% of the real word RT data (2.4% English, 8.4% Korean, and 6.6% Mandarin) and 38.6% of the nonword data (6.2% English, 26.5% Korean, and 12.8% Mandarin) were rejected. All RT data were log-transformed to improve normality. Table 4 shows the means of accuracy and RT for real words and nonwords across the three language groups.

Table 4. *Accuracy and response time for real words, nonwords with stress change, and nonwords with phoneme change in the lexical decision task in Experiment 2.*

	Real words					
	English		Korean		Mandarin	
	Accuracy	RT (ms)	Accuracy	RT (ms)	Accuracy	RT (ms)
	.927	1401.48	.826	1397.413	.794	1463.28
	Nonwords with stress change					
	English		Korean		Mandarin	
Vowel Change	Accuracy	RT (ms)	Accuracy	RT (ms)	Accuracy	RT (ms)
No	.778	1541.95	.405	1755.05	.597	1802.12
Yes	.854	1476.76	.399	1837.54	.605	1766.92
	Nonwords with phoneme change					
	English		Korean		Mandarin	
	Accuracy	RT (ms)	Accuracy	RT (ms)	Accuracy	RT (ms)
	.886	1571.3	.632	1868.54	.593	1846.61

As in Experiment 1, data were modeled in generalized linear (binomial, log-linear) mixed-effects regressions using the lme4 package (Bates & Maechler, 2010) in R studio (R Development Core Team, 2008). For all models, participant and items were entered as random effects. To predict accuracy and log RT for real words, language group, frequency, and proficiency were entered as fixed effects, to predict accuracy and log RT for nonwords, language group, vowel change, proficiency, and syllable length were entered as fixed effects. For accuracy results, estimates represent the log odds of change in accuracy resulting from a change in language group, vowel quality, proficiency, or a combination of two or more of the variables, and *p*-values are based on *Wald z* distributions. For RT results, estimates represent a change in RT (log ms) resulting from a change in language group, vowel quality, proficiency, or a combination of two or more of the variables, and *p*-values are based on MCMC sampling (Baayen, 2008; Bates, 2005). As in Experiment 1, data were analysed in two separate models in which the first model used English as the intercept while the second used Korean. We used the same model comparison procedure as in Experiment 1 and only the models with the best goodness-of-fit based on the Chi-square test were reported.

Real words

Table 4 shows the mean accuracy and response times for real words, nonwords with phoneme change and nonwords with stress change for all language groups.

Accuracy

In the initial model with the English group as the baseline (upper part of Table 5), the English speakers were significantly more accurate than the Korean and Mandarin speakers. There was a significant effect of frequency; the lexical judgment of the English speakers was more accurate for higher frequency words compared to lower frequency words. There was a significant interaction between frequency and language group for both the comparison between English and Mandarin and the comparison between English and Korean. The interaction was due to the fact that word frequency had a greater effect on lexical judgment for the Korean and Mandarin speakers in comparison to the English speakers. As can be seen in Figure 2, the slope of the regression lines for the Korean and Mandarin groups are more slanted than that of the regression line for the English group.

In the second model with the Korean group as the baseline (lower part of Table 5), both L2 groups showed a significant frequency effect in which they were more accurate judging high frequency than low frequency real words (inferred from the non-significant interaction between language group and frequency). The Korean speakers were significantly more accurate than the Mandarin speakers when making lexical judgment for real words.

Response times

The log-transformed RT data were subject to a similar analytical procedure as the accuracy. In the initial model (upper part of Table 6), the English speakers showed a

Table 5. *Results from the generalized linear mixed-effects model predicting real word accuracy with language groups, frequency, and proficiency as fixed effects and participants and items as random effects. Estimates represent the log odds of change in accuracy in lexical judgments resulting from a change in language group, frequency, proficiency, or a combination of the variables.*

The intercept in this model ($\chi^2 = 2.873$, $p = .09$, $df = 3170$) estimated English speakers' accuracy with low frequency words and low proficiency.

Fixed effects	Estimate (log odds)	SE	*z*-value	*p*-value
Intercept	−1.719	2.396	−.717	.473
Language Korean	−3.543	1.951	−1.816	.069
Language Mandarin	−5.972	1.856	−3.217	.001
Frequency	.491	.212	2.314	.021
Proficiency	.044	.030	1.486	.137
Language Korean: Frequency	.311	.108	2.883	.004
Language Mandarin: Frequency	.449	.102	4.402	<.001
Language Korean: Proficiency	−.001	.025	−.056	.956
Language Mandarin: Proficiency	.016	.022	.707	.479
Frequency: Proficiency	−.004	.002	−1.682	.093

The intercept in this model ($\chi^2 = 3.822$, $p = .05$, $df = 2093$) estimated Korean speakers' accuracy with low frequency words and low proficiency.

Fixed effects	Estimate (log odds)	SE	*z*-value	*p*-value
Intercept	−5.868	1.588	−3.694	.0002
Language Mandarin	−2.578	1.239	−2.081	.037
Frequency	.881	.166	5.314	<.001
Proficiency	.053	.028	1.913	.056
Language Mandarin: Frequency	.149	.082	1.815	.069
Language Mandarin: Proficiency	.017	.019	.891	.373
Frequency: Proficiency	−.005	.003	1.931	.054

significant effect of frequency. There was a significant interaction between language group and frequency for both the English and Korean comparison and the English and Mandarin comparison (Figure 2). In the second model (lower part of Table 6), the L2 groups did not differ significantly in response times. There was a significant two-way interaction between language group and frequency. The frequency effect was stronger in the Mandarin group than in the Korean group (Figure 2).

Nonwords

Accuracy

In the initial model with the English group as the baseline (upper part of Table 7), there was a significant interaction between language group and vowel change for both the English and Korean comparison and the English and Mandarin comparison. While the English speakers were more accurate rejecting nonwords with vowel change than those with no vowel change, both the Korean and Mandarin speakers' accuracy did not differ significantly between vowel change and no vowel change.

In the second model with the Korean group as the baseline (lower part of Table 7), the Mandarin speakers were significantly more accurate than the Korean speakers rejecting nonwords with no vowel change as well as nonwords with vowel change (inferred from the non-significant interaction between language and vowel change) (Figure 3). The significant interaction between language group and proficiency for nonwords with no vowel change: the difference in accuracy between the two L2 groups was smaller when they had higher L2 proficiency (Figure 4). The Korean speakers' accuracy in nonword rejection improved with increasing proficiency for both vowel change and no vowel change conditions whereas the Mandarin speakers showed lower accuracy rates with increasing proficiency in the no vowel change condition. The three-way interaction among language,

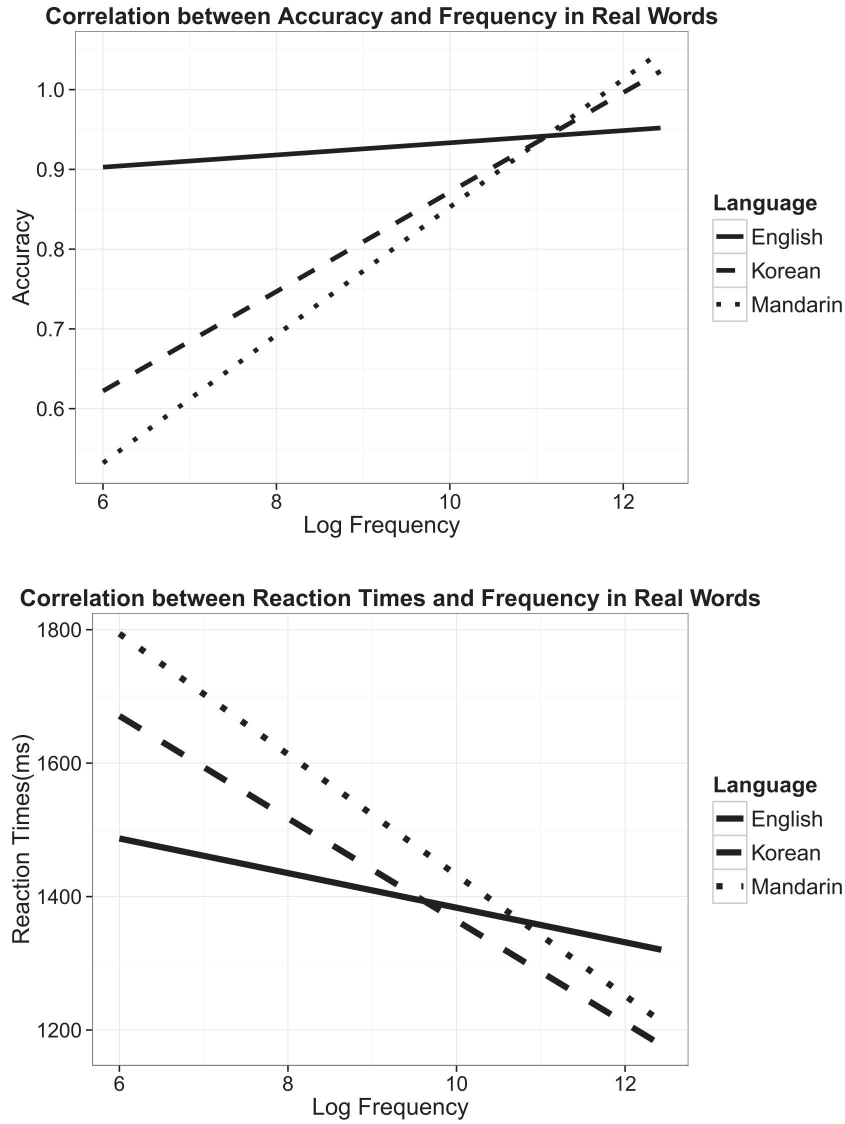

Figure 2. Interaction between language groups and log frequency in predicting accuracy and reaction times for real word lexical judgments in Experiment 2.

vowel change, and proficiency was not significant (this term was removed in the procedure of model comparison and thus not shown in the final model).

Stress errors

To help us understand whether speakers were more sensitive to stress errors in a particular stress location, additional analysis was conducted to explore the effect of stress error location on accuracy in each language group. We analyzed the disyllabic and trisyllabic words separately since the variable of stress error has two levels in disyllabic words and three levels in trisyllabic words. For a disyllabic nonword such as *human* /ˈhyumən/, the error is located in the final syllable. For a trisyllabic nonword such as *library* /laɪˈbrɛrɪ/, the error is located in the medial syllable. In the generalized mixed-effects log-linear models predicting nonword accuracy, stress error and proficiency were entered as fixed effects while participant and item were entered as random effects (see Appendix C for complete results). In general, the location of stress error did not have a significant effect on accuracy other than for trisyllabic medially-stressed nonwords for English speakers. They were more accurate rejecting nonwords with errors in the medial-stressed syllable compared to those with errors in the initial-stressed syllable.

Response times

Given participants' high error rates in their lexical decision of nonwords, the RT data may not be reliable. Thus, the following report on the RT results should be interpreted with some caution. In the initial model with

Table 6. *Results from the generalized linear mixed-effects model predicting real word reaction times with language groups, frequency, and proficiency as fixed effects and participants and items as random effects. Estimates represent the log odds of change in response times (in log units) in lexical judgments resulting from a change in language group, frequency, proficiency or a combination of two of the variables.*

The intercept in this model ($\chi^2 = 2.273$, $p = .13$, $df = 2697$) estimated English speakers' response times with low frequency words and low proficiency.

Fixed effects	Estimate (log ms)	SE	*t*-value	*p*-value
Intercept	3.263	.127	25.669	<.001
Language Korean	.196	.129	1.511	.080
Language Mandarin	.207	.124	1.677	.054
Frequency	−.018	.007	−2.633	.012
Proficiency	<−.001	.001	−.411	.633
Language Korean: Frequency	−.012	.003	−3.654	<.001
Language Mandarin: Frequency	−.014	.003	−4.984	<.001
Language Korean: Proficiency	−.001	.002	−.720	.396
Language Mandarin: Proficiency	<−.001	.002	−.437	.611
Frequency: Proficiency	<.001	<.001	1.507	.139

The intercept in this model ($\chi^2 = 3.412$, $p = .06$, $df = 1694$) estimated Korean speakers' response times with low frequency words and low proficiency.

Fixed effects	Estimate (log ms)	SE	*t*-value	*p*-value
Intercept	3.587	.096	37.18	<.001
Language Mandarin	−.169	.127	−1.34	.158
Frequency	−.043	.008	−5.26	<.001
Proficiency	−.004	.002	−2.46	.011
Language Mandarin: Frequency	.016	.011	1.5	<.001
Language Mandarin: Proficiency	.004	.002	1.76	.066
Frequency: Proficiency	<.001	<.001	2.48	.015
Language Mandarin: Frequency: Proficiency	<−.001	<.001	−1.84	.068

the English group as the baseline (upper part of Table 8), English speakers were significantly faster than the Korean but not the Mandarin speakers. The effect of vowel change was significant. The English speakers' lexical judgment was faster when there was a vowel change in the nonwords. There was a significant interaction between vowel change and the Korean and English groups but not between the Mandarin and English groups. Both the English and Mandarin speakers were faster to reject nonwords with vowel change than those with no vowel change whereas the opposite pattern was found in the Korean speakers.

In the second model with the Korean group as the baseline (lower part of Table 8), the Mandarin speakers were significantly faster than the Korean speakers to reject nonwords with no vowel change and nonwords with vowel change (inferred from the non-significant interaction between language and vowel change). There was a significant two-way interaction between language group and proficiency for nonwords with no vowel change. The difference in response times between the L2 groups was smaller when they have higher L2 proficiency. Similarly to the accuracy data, the three-way interaction among language, vowel change, and proficiency was not significant. The difference in RT between the Mandarin and Korean speakers with higher L2 proficiency was also smaller for nonwords with vowel change.

Frequency effect in nonwords

If the Korean speakers' low accuracy in the nonwords condition was because they did not consider stress in lexical access and accepted stress-shifted nonwords (i.e., /ˈbɪtwin/) as real words (i.e., *between* /bɪˈtwin/), they should show a frequency effect in the nonwords condition. In contrast, this frequency effect should be absent in the Mandarin and English speakers since they would be able to differentiate between real words and nonwords with the wrong stress location. To examine these possibilities, we first constructed generalized linear mixed-effect models

Table 7. *Results from the generalized linear mixed-effects models predicting nonword accuracy with language groups, vowel change, proficiency, and syllable length as fixed effects and participants and items as random effects. Estimates represent the log odds of change in accuracy in lexical judgments resulting from a change in language group, vowel quality, proficiency, syllable length or a combination of any of the two variables.*

The intercept in this model ($\chi^2 = 5.182$, $p = .07$, $df = 3170$) estimated English speakers' accuracy with no vowel change, short syllable length, and low proficiency.

Fixed effects	Estimate (log odds)	SE	*z*-value	*p*-value
Intercept	3.055	1.604	1.905	.057
Language Korean	−5.185	1.778	−2.916	.003
Language Mandarin	−1.978	1.706	−1.16	.246
Vowel change	.678	.266	2.547	.011
Proficiency	−.023	.020	−1.135	.256
Length	.001	.001	1.42	.155
Language Korean: Vowel change	−.683	.231	−2.955	.003
Language Mandarin: Vowel change	−.568	.232	−2.452	.014
Language Korean: Proficiency	.049	.026	1.904	.057
Language Mandarin: Proficiency	.007	.023	.291	.771

The intercept in this model ($\chi^2 = 4.460$, $p = .035$, $df = 2093$) estimated Korean speakers' accuracy with no vowel change, short syllable length, and low proficiency.

Fixed effects	Estimate (log odds)	SE	*z*-value	*p*-value
Intercept	−2.070	.902	−2.295	.022
Language Mandarin	3.376	1.094	3.086	.002
Vowel change	−.005	.259	−.02	.984
Proficiency	.027	.016	1.646	.099
Length	<.001	.001	.789	.430
Language Mandarin: Vowel change	.079	.207	.382	.703
Language Mandarin: Proficiency	−.045	.020	−2.199	.028

with language group (English, Korean, and Mandarin), lexicality (real words or nonwords) and frequency as the fixed effects and participants and items as the random effects to predict accuracy and log RT in the lexical decision task. Nonwords' frequency was based on that of their real word counterparts. The three-way interaction among language group, lexicality, and frequency was significant for the comparison between the English and Korean and between the English and Mandarin (Table 9), suggesting that the frequency effect was different between real words and nonwords for all three language groups.

More importantly, in the models predicting lexical judgment of nonwords only (Table 10), there was a significant interaction between frequency and language group (in accuracy) when comparing the English and Korean speakers but not when comparing the English and Mandarin speakers. While the Korean speakers were significantly less accurate rejecting nonwords with higher frequency than those with lower frequency, this frequency effect was not significant for the English and Mandarin speakers (Figure 5). The Korean speakers' frequency effect is consistent for both real words and nonwords. The higher the frequency, the higher acceptance rate is for both real and nonce words. In the lexical decision task, the Korean speakers were more likely to press "YES" for high frequency words (e.g.,„*human*) than low frequency words (i.e., *potion*) and this resulted in correct responses in the real words condition. Since the Korean speakers had difficulty differentiating between real words and nonwords with stress change, they were also more likely to press "YES" for high frequency nonwords (e.g.,„/ˈhyumən/) than low frequency nonwords (e.g.„ /pəʊˈʃɛn/) and this resulted in incorrect responses in the nonwords condition.

The control words were constructed to prevent participants from developing stress-specific processing strategies and examine any possible difference among the language groups in phonemic processing. In the generalized linear mixed-effects regressions predicting

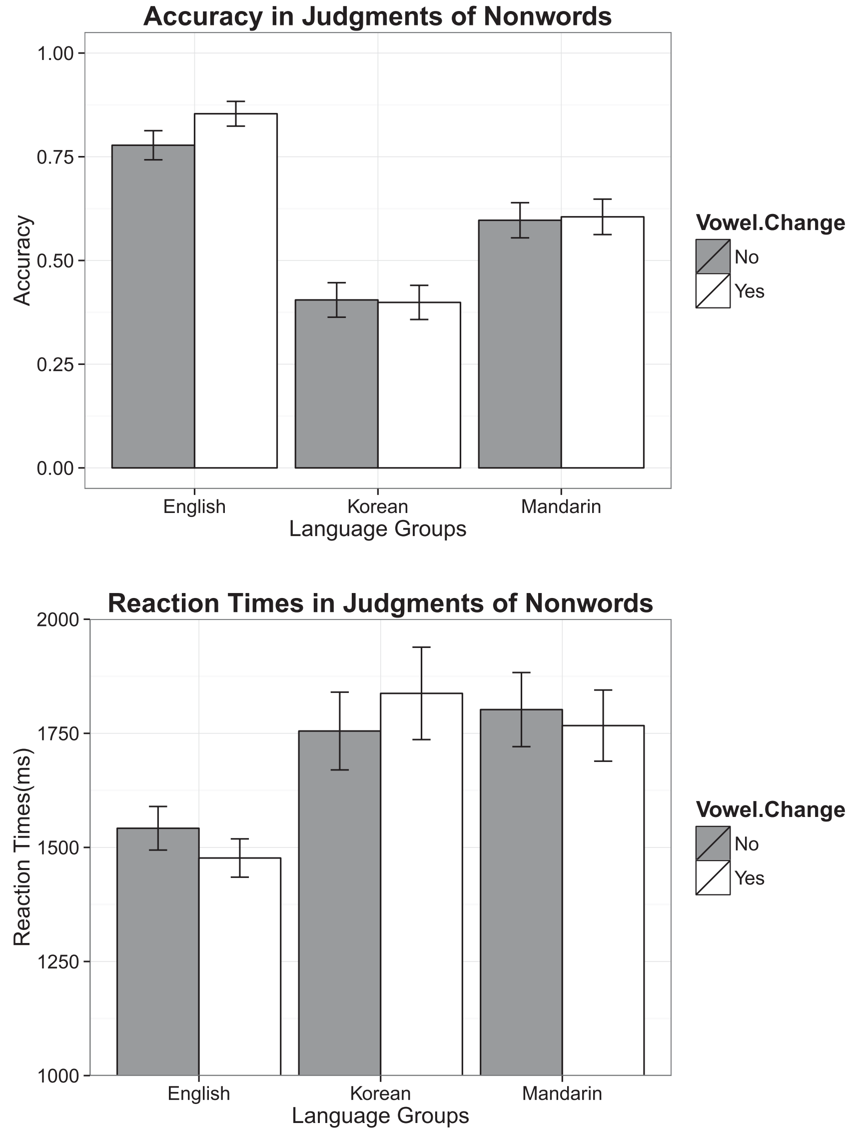

Figure 3. Accuracy and reaction times for the lexical judgment of nonwords in Experiment 2.

accuracy and log RT in rejecting nonwords with a phoneme change, language group and proficiency were entered in the model as fixed effects while participant and item were entered as random effects. The English speakers were significantly more accurate and faster than the Korean speakers (accuracy: $z = -5.41$, $p < .001$; RT: $t = 1.74$, $p = .062$) and the Mandarin speakers (accuracy: $z = -6.946$, $p < .001$; RT: $t = 1.84$, $p = .046$). However, the effect of language group in accuracy and RT was not significant in the model comparing the Korean and Mandarin speakers (accuracy: $z = -.285$, $p = .348$; RT: $t = -.075$, $p = .997$). The effect of proficiency and its interaction with language group were not significant in these models (all *p*s $> .1$).

We also conducted a separate set of analyses using *d*-prime scores (see Appendix D). The results were generally consistent with the analyses done by separating real words and nonwords.

Discussion

Both the Mandarin and Korean speakers were less accurate than the native English speakers in rejecting nonwords with the wrong stress location and with the wrong phoneme. More importantly, the L2 groups perform comparably when rejecting nonwords with a phoneme change whereas the Korean speakers were worse than the Mandarin speakers when rejecting nonwords with the incorrect stress pattern. In addition, a significant frequency effect for nonwords was observed for the Korean speakers, but not for the Mandarin or English speakers, suggesting that the Korean speakers treated

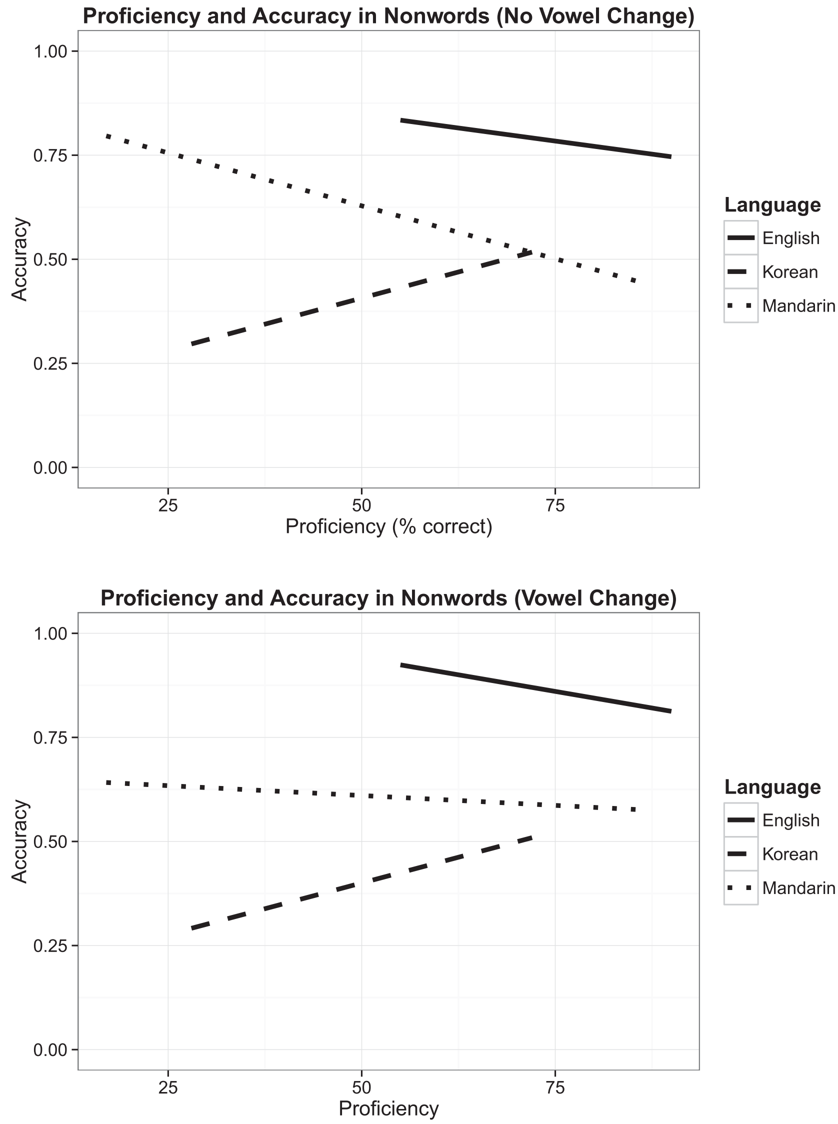

Figure 4. Interaction between proficiency and vowel change on lexical judgment accuracy in Experiment 2.

nonwords with the wrong stress location as real words. These results are consistent with those from Experiment 1. The Korean speakers showed an asymmetry in their performance in which they had more difficulty with items involving stress than those involving phonemes. The converging findings from both experiments suggest that the Korean speakers' lack of phonological representation for stress not only affected their encoding of minimal stress pairs of nonwords in short-term memory, but also impacted their online retrieval of real word stress patterns. Since the two L2 groups did not differ significantly in the phoneme conditions and we have taken into account of the variance of proficiency in the generalized linear mixed-effects models by the inclusion of the cloze scores as a fixed effect, the difference in stress processing between the Mandarin and Korean speakers may not be attributed to variation in L2 proficiency or segmental processing. However, there is preliminary evidence suggesting that Korean speakers' stress processing skill improves with better L2 proficiency.

These findings supported the hypotheses assuming SPM, suggesting that Korean speakers' difficulty with stress processing may be a result of the lack of efficient stress encoding strategies. Furthermore, the finding that Mandarin speakers were less accurate than English speakers in rejecting nonwords with stress shift indicated that although Mandarin speakers have abstract representation of stress and can discriminate minimal stress pairs, their L2 stress processing related to lexical retrieval was not as efficient as native speakers'. Alternatively, it is also possible that Mandarin speakers were less familiar with some of the words used in the

Table 8. *Results from the generalized linear mixed-effects models predicting nonword reaction times with language groups, vowel change, proficiency, and syllable length as fixed effects and participants and items as random effects. Estimates represent a change in response times (in log units) resulting from a change in language group, vowel quality, syllable length, or a combination of any of the two variables.*

The intercept in this model ($\chi^2 = 5.503$, $p = .06$, $df = 1918$) estimated English speakers' response times with no vowel change, short syllable length, and low proficiency.

Fixed effects	Estimate (log ms)	SE	*t*-value	*p*-value
Intercept	3.11	.132	23.601	.0001
Language Korean	.288	.152	1.903	.032
Language Mandarin	.070	.144	.488	.572
Vowel change	−.019	.009	−1.985	.039
Proficiency	.001	.002	.648	.442
Length	<−.001	<.001	−2.567	.006
Language Korean: Vowel change	.027	.013	2.087	.033
Language Mandarin: Vowel change	.014	.012	.961	.346
Language Korean: Proficiency	−.004	.002	−1.685	.053
Language Mandarin: Proficiency	<.001	.002	.128	.859

The intercept in this model ($\chi^2 = 4.481$, $p = .034$, $df = 1040$) estimated Korean speakers' response times with no vowel change, short syllable length, and low proficiency.

Fixed effects	Estimate (log ms)	SE	*t*-value	*p*-value
Intercept	3.40	.084	40.76	.0001
Language Mandarin	−.223	.105	−2.12	.017
Vowel change	.008	.014	.56	.459
Proficiency	−.003	.002	−1.7	.053
Length	<−.001	<.001	−1.67	.071
Language Mandarin: Vowel	−.017	.016	−1.1	.232
Language Mandarin: Proficiency	.004	.002	2.08	.017

lexical decision, particularly words in the low frequency condition (e.g., *cocoon*, *adamant*, *calvary*, *vendetta*, etc.)

Vowel quality change in nonwords facilitated the native English speakers' lexical judgment, but the non-native speakers did not benefit from the vowel cue to the same degree as the native speakers did. The native English speakers' lexical judgment was more accurate when there was a change in vowel quality in the nonwords. Since vowels are systematically reduced in unstressed syllables in real words, the absence of vowel reduction in unstressed syllables in nonwords apparently made the identification of stress location more difficult, resulting in higher errors in lexical judgment. Consistent with Cutler and van Donselaar (2001), our finding indicated that English speakers rely on segmental cues of stress in lexical access. The effect of vowel change also interacts with the location of stress errors in which English speakers were more sensitive to initial-stress errors when vowel quality was changed in the nonwords. Cutler and Carter (1987) have reported that 90% of the content words begin with a stressed syllable in English. Our results provided evidence to support that English speakers are sensitive to the predominant stress pattern in their native language.

The Korean speakers did not show sensitivity to any particular error patterns or change in vowel quality. The Korean speakers' difficulty rejecting nonwords with vowel reduction may be interpreted in two ways. First, Korean speakers did not use vowel reduction as a cue to stress. Second, Korean speakers did not perceive the difference between reduced vowels and non-reduced vowels due to the absence of reduced vowels in Korean (Lee et al., 2006), which is independent of stress.

In contrast, although reduced vowels exist in Mandarin, vowel quality change did not appear to aid the Mandarin speakers' lexical judgment. Shen (1993) conducted acoustic analyses of native Mandarin speakers' production of Chinese words. She observed that the unstressed low vowel /a/ shifts toward schwa, which is the mid vowel /ə/, in unstressed syllables. For example, the word 着 /ʈʂáu/ "to touch" is pronounced as /ʈʂə/ "-ing" when unstressed.

Table 9. *Results from the generalized linear mixed-effects models predicting lexical judgment accuracy and response times with language groups, frequency, and lexicality as fixed effects and participants and items as random effects. Estimates represent a change in accuracy (in log odds) or response times (in log units) resulting from a change in language group, frequency, lexicality, or a combination of any of the two or more variables.*

The intercept in this model ($\chi^2 = 260.21$, $p < .001$, $df = 6348$) estimated English speakers' accuracy with low frequency real words.

Fixed effects	Estimate (log odds)	SE	*z*-value	*p*-value
Intercept	1.074	.649	1.654	.098
Language Korean	−4.613	.734	−6.286	<.001
Language Mandarin	−5.791	.744	−7.782	<.001
Frequency	.207	.071	2.935	.003
Lexicality	1.448	.815	1.777	.075
Language Korean: Frequency	.410	.082	4.982	<.001
Language Mandarin: Frequency	.522	.0837	6.240	<.001
Language Korean: Lexicality	3.961	.889	4.450	<.001
Language Mandarin: Lexicality	5.189	.901	5.756	<.001
Frequency: Lexicality	−.286	.088	−3.217	.001
Language Korean: Frequency: Lexicality	−.579	.099	−5.832	<.001
Language Mandarin: Frequency: Lexicality	−.591	.101	−5.878	<.001

The intercept in this model ($\chi^2 = 139.54$, $p < .001$, $df = 4620$) estimated English speakers' response times with low frequency real words.

Fixed effects	Estimate (log ms)	SE	*t*-value	*p*-value
Intercept	3.198	.024	134.37	<.001
Language Korean	.145	.031	4.75	<.001
Language Mandarin	.171	.032	5.41	<.001
Frequency	−.007	.002	−3.44	.006
Lexicality	−.004	.028	−.15	.877
Language Korean: Frequency	−.015	.003	−5.84	<.001
Language Mandarin: Frequency	−.016	.003	−6.10	<.001
Language Korean: Lexicality	−.092	.038	−2.40	.016
Language Mandarin: Lexicality	−.130	.037	−3.54	.004
Frequency: Lexicality	.004	.003	1.29	.196
Language Korean: Frequency: Lexicality	.017	.004	4.26	<.001
Language Mandarin: Frequency: Lexicality	.018	.004	4.77	<.001

Despite their use of vowel reduction as a stress cue in their L1 production, the Mandarin speakers did not benefit from this cue when processing L2 words. One possible explanation for this finding is the difference in frequency and phonological environment associated with vowel reduction between the two languages. In English, vowel reduction can occur in any position depending on stress location. For example, in words such as *support* /səˈpɔrt/, *catalog* /ˈkætəˌlɒg/, and *museum* /mjuːˈzɪəm/ vowel reduction appears in the initial, medial, or final syllable. In contrast, since unstressed syllables can never appear in word-initial positions in Mandarin (Duanmu, 2007), vowel reduction can only occur in the medial and final syllable (e.g., both medial and final syllables in出去了 /tʂʰūtɕʰylə/ "have gone out", final syllable in哭着 /kūʈʂə/ "crying"). Our study also showed that the Mandarin speakers had lower accuracy rates in the lexical judgment of nonwords with no vowel change with increasing L2 English proficiency (see Figure 4). It is possible that as Mandarin speakers become more proficient in L2 English, they realized that stress placement is strongly correlated with vowel quality

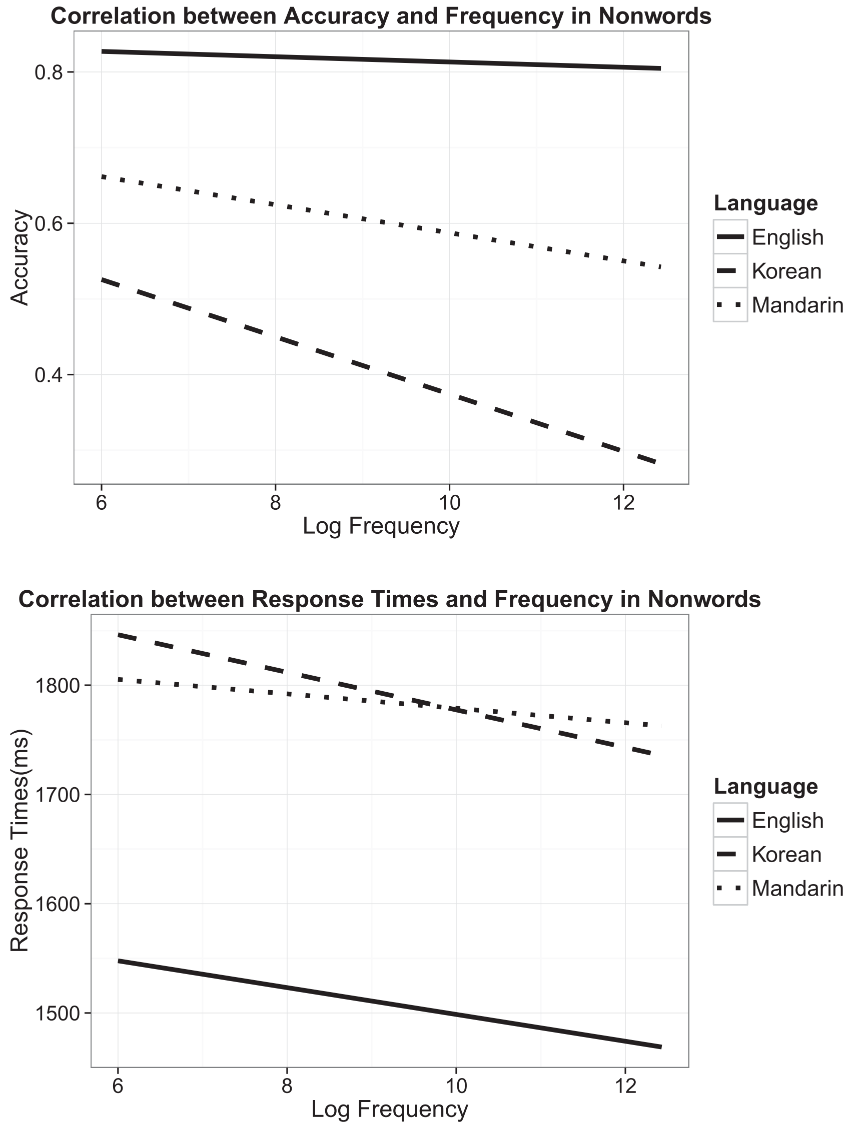

Figure 5. Correlation between frequency and performance for the lexical judgment of nonwords in Experiment 2.

in English stress. Thus, Mandarin speakers with more advanced L2 proficiency performed less well when stress is not cued by vowel quality.

General discussion

The current study examined stress encoding in minimal pairs and the influence of stress on lexical access among native English speakers, and native Korean and native Mandarin speakers who are L2 learners of English. The comparison between Korean and Mandarin speakers allowed us to test the hypothesis as guided by SPM (Peperkamp & Dupoux, 2002). We hypothesized that Mandarin speakers would show better stress perception than Korean speakers. Results from Experiments 1 and 2 were consistent with the hypothesis. Korean speakers have not developed efficient stress encoding strategies as a result of the absence of contrastive stress in their L1. This has led to difficulty with stress processing in L2 English, evident in their poorer performance in both encoding minimal stress pairs of nonwords in sequence recall and rejecting words with the wrong stress patterns in the lexical decision task, relative to Mandarin and native English speakers.

The role of L1 in stress processing

The current results are consistent with the Dupoux et al. (2008) study, in which the French speakers who learned Spanish as their L2 also showed profound stress processing difficulties despite their high proficiency in Spanish. In the current study, the Korean speakers'

Table 10. *Results from the generalized linear mixed-effects models predicting nonword accuracy and response times with language groups and frequency as fixed effects and participants and items as random effects. Estimates represent a change in accuracy (in log odds) or response times (in log units) resulting from a change in language group, frequency, or a combination of the two variables.*

The intercept in this model ($\chi^2 = 20.732$, $p = .007$, $df = 3174$) estimated English speakers' accuracy with low frequency nonwords.

Fixed effects	Estimate (log odds)	SE	*z*-value	*p*-value
Intercept	2.005	.649	3.090	.002
Language Korean	−.635	.601	−1.057	.291
Language Mandarin	−.527	.609	−.866	.387
Frequency	−.011	.066	−.160	.873
Language Korean: Frequency	−.194	.056	−3.049	.002
Language Mandarin: Frequency	−.089	.056	−1.238	.216

The intercept in this model ($\chi^2 = 6370.6$, $p < .001$, $df = 1922$) estimated English speakers' response times with low frequency nonwords.

Fixed effects	Estimate (log ms)	SE	*t*-value	*p*-value
Intercept	3.192	.028	112.05	<.001
Language Korean	.055	.038	1.43	.129
Language Mandarin	.047	.036	1.29	.176
Frequency	−.003	.002	−1.22	.191
Language Korean: Frequency	.003	.003	.79	.438
Language Mandarin: Frequency	.002	.003	.83	.409

performance in the stress contrast in sequence recall was similar (31.4% accuracy across sequence lengths) to the French speakers in Dupoux et al. (33.9% accuracy in four-word sequences only in beginner L2 learners, 21.2% in intermediate, 25.8% in advanced). The Mandarin and native English speakers (46.5% and 40.1%, respectively) in the current study, however, did not perform as well as the Spanish speakers in Dupoux et al. (80.4%). All three language groups in the current study (Mandarin = 49.9%, Korean = 41.6%, and English = 50.6%) had lower accuracies in the phoneme contrast than the Spanish and French speakers in Dupoux et al. (accuracy range: 74–79%).

One possible reason for these differences is that the sequence recall task in the current study was more difficult than that in Dupoux et al. (2008). Our sequence recall task consisted of five different sequence lengths with the number of words to be recalled ranging from two to six. In contrast, the sequence recall task in Dupoux et al. included only four-word sequences. Another possible reason why the native English speakers in our study did not perform as well as the Spanish speakers in the stress contrast is that native English speakers rely more on segmental information than Spanish speakers in stress processing (Soto-Faraco, Sebastián-Gallés & Cutler, 2001). In the stress contrast of the sequence recall task, there was no vowel reduction in the unstressed syllables. Previous research has shown that real word minimal stress pairs that do not involve change in vowel quality such as *trusty* /ˈtrʌsti/ and *trustee* /trʌˈsti/ are stored in the mental lexicon by native English speakers as homophones (Cutler, 1986). The lack of vowel change in nonwords like /ˈmipa/–/miˈpa/ may result in native English speakers' less efficient encoding of the stress contrast, explaining their lower accuracy as compared to the Spanish speakers in Dupoux et al. (2008).

Regarding the lexical decision task in Experiment 2, since Spanish words were used in Dupoux et al. (2008) while English words were used in the current study, the following comparisons should be interpreted with caution. For nonwords, the Korean speakers in the present study (40% accuracy) performed similarly to the French speakers with intermediate L2 proficiency (39%). The Mandarin speakers' performance (60% accuracy) was better than that of the French speakers with advanced L2 proficiency (45%) but lower than that of the native Spanish speakers (95%). Although Mandarin speakers have abstract representation of stress as a result of having contrastive stress in their L1, their lexical access is constrained by L2 proficiency since they are L1 dominant.

Thus, they are better than speakers whose languages lack contrastive stress (i.e., Korean and French); yet, their stress processing skills are not as good as those of native speakers of the language in which the task was performed (i.e., English and Spanish). Note that Spanish stress is more phonologically regular and predictable than English. Also, English, but not Spanish, uses vowel reduction as a segmental cue to stress. These linguistic differences between the two languages make it difficult to draw any firm conclusions from the above comparisons.

"Stress deafness" as a gradient phenomenon

Dupoux et al. (2001, 2008, 2010) used the term "stress deafness" to describe the difficulty in stress processing as observed in the French speakers in their studies and the Korean speakers in the current study. However, findings from our current study showed that "stress deafness" may not be a dichotomous construct dependent on the setting of the Stress Parameter as postulated by SPM. Instead, speakers may have various degrees of difficulty depending on the characteristics of stress in their native languages. Peperkamp and Duxpoux (2002) found that performance on the stress contrast of their sequence recall task varied with the regularity of stress location. Speakers of languages with the most regular stress assignment, such as Turkish and Finnish, in which stress always falls on the final and initial syllable, respectively, showed the poorest performance in stress encoding. Speakers of languages where stress placement can vary across words depending on phonological and morphological factors (e.g., syllable weight, affixation, etc.) and where segmental information is not correlated with stress (i.e., absence of vowel reduction) showed the best performance. There are languages with a few exceptions in an otherwise regular stress system, such as Polish, in which a monosyllabic word receives stress if it is at the end of an utterance; otherwise stress always falls on the penultimate syllable. The performance of Polish speakers in the sequence recall task is superior to that of French and Finnish speakers but poorer than that of Spanish speakers (Peperkamp & Duxpoux, 2002).

In the current study, the Korean speakers' accuracy rates in both the sequence recall and the lexical decision tasks were not close to floor level, suggesting that Korean speakers do have some stress processing capabilities, although not as efficient as those of speakers whose L1 has lexically contrastive stress such as Mandarin. Findings from Experiment 2 suggest that Korean speakers' stress processing skills may be improved with higher L2 proficiency. In light of these findings, it is worthwhile for future research to examine stress perception in other L2 groups using similar tasks to establish a comprehensive typology in both L1 and L2 stress processing.

Conclusion

In summary, the current study tested the predictions made by the Stress Parameter Model (Peperkamp & Dupoux, 2002) with regard to perception of non-native stress contrasts. The sequence recall and lexical decision tasks showed that the Mandarin and native English speakers outperformed the Korean speakers in the conditions involving stress. These results are partly consistent with SPM; Korean speakers may not have developed an efficient strategy for processing stress, resulting in their difficulty to perceive minimal stress pairs. In comparison to non-native speakers, native the English speakers' word recognition was facilitated by vowel quality change, suggesting that they rely more on segmental than suprasegmental information in spoken word recognition. The non-native speakers did not benefit from vowel change to the same degree as the native speakers. Taken together, these results suggest that the availability of stress representation in L1 phonological memory allows L2 learners to develop native-like stress processing abilities. However, when L2 learners need to process stress and perform lexical access simultaneously, they are constrained by their lower lexical proficiency and the fine-grained difference between the L1 and L2 prosodic systems.

Appendix A. Sequences used in Experiment 1

Two-word sequences	11, 12, 21, 22 (each repeated twice)
Three-word sequences	111, 112, 121, 122, 211, 212, 221, 222
Four-word sequences	1121, 1211, 1212, 1221, 2112, 2121, 2122, 2212
Five-word sequences	11211, 11212, 12211, 12212, 21112, 21211, 22112, 22121
Six-word sequences	112112, 112121, 122121, 122122, 211121, 212122, 221212, 222121

Appendix B. Stimuli used in Experiment 2

■■■	High frequency ($k = 20$)	Medium frequency ($k = 20$)	Low frequency ($k = 20$)
Range	12.43–10.18	9.40–7.98	7.90–6.00
Mean	11.52	8.53	7.05
Standard deviation	.744	.807	.736
Standard error	.023	.025	.023

Word	IPA	Nonword	Vowel change
under	'ʌndɚ	ʌn'dɛr	1
support	sə'pɔrt	'sʌpɔrt	1
program	'proʊgræm	proʊ'græm	0
between	bɪ'twin	'bitwɪn	1
every	'ɛvri	əv'ri	1
version	'vɜrʒən	vɚ'ʒɛn	1
money	'mʌni	mʌ'ni	0
always	'ɔlweɪz	ɔl'weɪz	0
order	'ɔrdɚ	ɔr'dɛr	1
enough	ɪ'nʌf	'ɪnʌf	0
human	'hyumən	hyu'mæn	1
until	ʌn'tɪl	'ʌntɪl	0
public	'pʌblɪk	pʌ'blɪk	0
control	kən'troʊl	'kɔntroʊl	1
design	dɪ'zaɪn	'dɪzaɪn	0
perhaps	pɚ'hæps	'pɛrhæps	1
instead	ɪn'stɛd	'instɛd	0
contact	'kɒntækt	kɒn'tækt	0
below	bɪ'ləʊ	'bɪləʊ	0
common	'kɒmən	kɒ'mɔn	1
reserve	rɪ'zɜːrv	'rɪzɚv	1
despite	dɪ'spaɪt	'dɪspaɪt	0
exact	ɪg'zækt	'ɪgzkæt	0
surface	'sɜːrfɪs	sɚ'fis	1
perform	pɚ'fɔːm	'pɛrfɔm	1
remain	rɪ'meɪn	'rɪmeɪn	0
platform	'plætfɔːrm	plæt'fɔːrm	0
solid	'sɒlɪd	sɒ'lɪd	0

Appendix B. Continued

Word	IPA	Nonword	Vowel change
campaign	kæm'peɪn	'kæmpeɪn	0
county	'kaʊntɪ	kaʊn'ti	1
domain	də'meɪn	'dəʊmeɪn	1
disease	dɪ'zi:z	'dɪzi:z	0
improve	ɪm'pru:v	'ɪmpruv	1
repeat	rɪ'pi:t	'ri:pɪt	1
argue	'ɑrgju:	ɑr'gju:	1
hotel	həʊ'tɛl	'həʊtɛl	0
detail	'di:teɪl	di'teɪl	1
profit	'prɒfɪt	prɒ'fɪt	0
quantum	'kwɒntəm	kwɒn'tʌm	1
defend	dɪ'fɛnd	'dɪfɛnd	0
confess	kən'fɛs	'kɔnfɛs	1
bracket	'brækɪt	bræ'kɪt	0
eclipse	ɪ'klɪps	'ɪklɪps	0
exile	'ɛgzaɪl	ɛg'zail	0
harvest	'hɑ:vɪst	hɑ'vɪst	1
donkey	'dɒŋkɪ	dɒŋ'ki	1
bury	'bɛrɪ	bɛ'rɪ	0
behold	bɪ'həʊld	'bɪhəʊld	0
contempt	kən'tɛmpt	'kɔntɛmpt	1
fossil	'fɒsəl	fɒ'sɛl	1
scandal	'skændəl	skæn'dɛl	1
potion	'pəʊʃən	pəʊ'ʃɛn	1
demise	dɪ'maɪz	'dɪmaɪz	0
cocoon	kə'ku:n	'kəʊkun	1
discreet	dɪ'skri:t	'dɪskrit	1
endorse	ɪn'dɔ:s	'indɔs	1
encore	'ɒŋkɔ:	ɒŋ'kɔ:	0
ensign	'ɛnsaɪn	ɛn'saɪn	0
distress	dɪ'strɛs	'dɪstrɛs	0
neglect	nɪ'glɛkt	'nɪglɛkt	0
possible	'pɒsɪbəl	pɒ'sɪbəl	0
already	ɔ:l'rɛdɪ	ɔ:lrɛ'dɪ	0
remember	rɪ'mɛmbɚ	'rɪmɛmbɚ	0
library	'laɪbrərɪ	laɪ'brɛrɪ	0
video	'vɪdɪˌəʊ	vɪ'dɪəʊ	0
understand	ˌʌndɚ'stænd	ʌn'dɛrstænd	1
together	tə'gɛðɚ	'tugɛðɚ	1
family	'fæmɪlɪ	fæ'mɪli	0
department	dɪ'pɑ:tmənt	dɪpɑt'mɛnt	1
magazine	ˌmægə'zi:n	mæ'gɛzin	1
document	'dɒkyəmənt	dɒkyə'mɛnt	1
industry	'ɪndəstrɪ	ɪn'dʌstrɪ	1
continue	kən'tɪnju:	'kɔntɪnju:	1
difficult	'dɪfɪkəlt	dɪ'fɪkəlt	0

Appendix B. Continued

Word	IPA	Nonword	Vowel change
monitor	ˈmɒnɪtɚ	mɒˈnɪtɚ	0
develop	dɪˈvɛləp	ˈdɪvɛləp	0
component	kəmˈpəʊnənt	ˈkɔmpəʊnənt	1
determine	dɪˈtɜːrmɪn	ˈdɪtɚmɪn	1
animal	ˈænɪməl	æˈnɪməl	0
represent	ˌrɛprɪˈzɛnt	ˌrɛˈprɪzənt	1
lieutenant	luːˈtɛnənt	ˈluːtɛnənt	0
precision	prɪˈsɪʒən	ˈprɪsɪʒən	0
merchandise	ˈmɜrtʃənˌdaɪz	mɜrtʃənˈdaɪz	0
catalogue	ˈkætəˌlɒg	kæˈtɛlɒg	1
distinguish	dɪˈstɪŋgwɪʃ	dɪstɪŋˈgwɪʃ	0
underneath	ˌʌndɚˈniːθ	ˈʌndɚniθ	1
emperor	ˈɛmpərɚ	ɛmˈpɛrɚ	1
admiral	ˈædmərəl	ædˈmɛrəl	1
obsolete	ɒbsəˈliːt	ɒbˈsɒlit	1
cylinder	ˈsɪlɪndɚ	sɪˈlɪndɚ	0
hurricane	ˈhʌrɪkeɪn	hʌˈrɪkeɪn	0
dilemma	dɪˈlɛmə	ˈdɪlɛmə	0
casino	kəˈsiːnəʊ	ˈkæsinəʊ	1
custody	ˈkʌstədɪ	kʌsˈtəʊdɪ	1
paradigm	ˈpærəˌdaɪm	pærəˈdaɪm	0
horizon	həˈraɪzən	ˈhɒraɪzən	1
placenta	pləˈsɛntə	ˈplæsɛntə	1
vitamin	ˈvaɪtəmɪn	vaɪtəˈmɪn	0
nirvanna	nɜːˈvɑːnə	ˈnɜːvɑnə	1
tsunami	tsʊˈnæmɪ	ˈtsʊnæmɪ	0
vendetta	vɛnˈdɛtə	ˈvɛndɛtə	0
requiem	ˈrɛkwɪˌɛm	rɛˈkwɪˌɛm	0
tequila	tɪˈkiːlə	ˈtɪkilə	1
magenta	məˈdʒɛntə	ˈmɛdʒɛntə	1
calvary	ˈkælvərɪ	kælˈværɪ	1
cartilage	ˈkɑːtɪlɪdʒ	kɑˈtɪlɪdʒ	1
tuxedo	tʌkˈsiːdəʊ	ˈtʌksidəʊ	1
pedestal	ˈpɛdɪstəl	pɛˈdɪstəl	0
adamant	ˈædəmənt	æˈdæmənt	1
omnibus	ˈɒmnɪˌbʌs	ɒmˈnɪˌbʌs	0
impetus	ˈɪmpɪtəs	ɪmˈpɪtəs	0
ovation	əʊˈveɪʃən	ˈəʊveɪʃən	0
ornament	ˈɔrnəmənt	ɔrˈnɛmənt	1
complexion	kəmˈplɛkʃən	ˈkɔmplɛkʃən	1
bionic	baɪˈɒnɪk	ˈbaɪɒnɪk	0
mnemonic	nɪˈmɒnɪk	ˈnɪmɒnɪk	0
improvise	ˈɪmprəˌvaɪz	ɪmˈprəʊvaɪz	1
atrophy	ˈætrəfɪ	æˈtrəʊfɪ	1
insipid	ɪnˈsɪpɪd	ˈɪnsɪpɪd	0
imprecise	ˌɪmprɪˈsaɪs	ɪmˈprɪsaɪs	0

Appendix C. Descriptive statistics for accuracy as a function of error stress location in the nonwords in each language group

Disyllabic error location	English Initial	English Final	Korean Initial	Korean Final	Mandarin Initial	Mandarin Final
Mean	.851	.841	.350	.358	.586	.596
SD	.356	.366	.478	.481	.493	.492
SE	.021	.023	.028	.030	.029	.032

Trisyllabic error	English Initial	English Medial	English Final	Korean Initial	Korean Medial	Korean Final	Mandarin Initial	Mandarin Medial	Mandarin Final
Mean	.727	.851	.685	.449	.473	.362	.627	.630	.444
SD	.447	.357	.469	.498	.500	.485	.485	.484	.503
SE	.030	.022	.064	.034	.031	.064	.034	.030	.075

Results from the exploratory analysis of stress location in Experiment 2. In each mixed-effect model, stress error location and proficiency were entered as fixed effects while participant and item were entered as random effects. The intercept in each model estimated accuracy for words with stress error in the initial syllable.

Disyllabic		Estimate (log odds)	SE	*z*-value	*p*-value
Mandarin ($df = 516$)	Intercept	1.477	.815	1.812	.07
	Error Medial Syllable	.896	.802	1.117	.264
	Proficiency	−.018	.014	−1.355	.175
	Error Medial Syllable: Proficiency	−.014	.013	−1.119	.263
Korean ($df = 544$)	Intercept	−2.346	.906	−2.59	.009
	Error Medial Syllable	1.469	.950	1.546	.122
	Proficiency	.030	.017	1.758	.079
	Error Medial Syllable: Proficiency	−.027	.017	−1.57	.116
English ($df = 545$)	Intercept	5.891	2.813	2.094	.036
	Error Medial Syllable	−2.706	2.593	−1.043	.297
	Proficiency	−.041	.035	−1.511	.249
	Error Medial Syllable: Proficiency	.036	.032	1.111	.266
Trisyllabic		**Estimate (log odds)**	**SE**	***z*-value**	***p*-value**
Mandarin ($df = 494$)	Intercept	.661	1.126	.587	.557
	Error Medial Syllable	1.144	.875	1.308	.191
	Error Final Syllable	.868	1.283	.677	.498
	Proficiency	.002	.019	.094	.925
	Error Medial Syllable: Proficiency	−.021	.014	−1.542	.123

	Error Final Syllable: Proficiency	−.032	.021	−1.509	.131
Korean ($df = 526$)	Intercept	−1.734	.944	−1.837	.066
	Error Medial Syllable	.553	1.063	−.52	.603
	Error Final Syllable	−3.949	2.426	−1.628	.104
	Proficiency	.029	.017	1.658	.097
	Error Medial Syllable: Proficiency	.015	.019	.767	.443
	Error Final Syllable: Proficiency	.064	.044	1.454	.146
English ($df = 525$)	Intercept	2.463	1.922	1.281	.200
	Error Medial Syllable	4.713	2.725	2.032	.038
	Error Final Syllable	1.182	3.228	.366	.714
	Proficiency	−.013	.024	−.534	.593
	Error Medial Syllable: Proficiency	−.044	.034	−1.321	.187
	Error Final Syllable: Proficiency	−.019	.040	−.468	.640

Appendix D. *d*-prime scores analysis in Experiment 2

Frequency was categorized into three conditions, high, intermediate, and low. For the stress condition, six *d*-prime scores were created for each participant as a result of a 2 (vowel change) × 3 (frequency) factorial design. For the phoneme condition, three *d*-prime scores were created for each participant for each of the three frequency conditions. We treated a correct response accepting a real word as a "hit" while an incorrect response rejecting a real word as a "miss". A correct response rejecting a nonword was labeled as a "correct rejection" whereas an incorrect response accepting a nonword was labeled as a "false alarm". For a hit rate of $H = 1.00$ or a false alarm rate of $FA = 0$, we used the conventional adjustment to set the minimum $p = 1/N$ where N is the number of trials used in the calculation of p and we set the maximum value for $p = (N-1)/N$. The formula we used to calculate *d*-prime scores was $z(H) - z(FA)$ (Swets, 1996).

Descriptive statistics for *d*-prime scores as a function of condition and frequency (and vowel change in the stress condition) in each language group.

In the linear mixed-effects models predicting *d*-prime scores, language group, condition (stress vs. phoneme), and proficiency were entered as fixed effects while participant was entered as a random effect. The models did not include item as a random variable since *d*-prime scores were computed across items separately for each participant. We did not include frequency as a fixed variable since we wanted the *d*-prime analyses to be as comparable as possible to the original analyses with nonwords (i.e., Table 7). The estimate represents a change in *d*-prime scores when there was a change in language group, condition, proficiency, or a combination of two or more of the variables. The models with the best goodness-of-fit based on the Chi-square test were reported.

In the linear mixed-effects models predicting *d*-prime scores in the stress condition only, language group, vowel change, and proficiency were entered as fixed effects while participant was entered as a random effect (item was not entered as a random effect). The estimate represents a change in *d*-prime scores when there was a change in language group, vowel quality, proficiency, or a combination of two or more of the variables.

Phoneme	High frequency			Medium frequency			Low frequency		
Language	Mean	SE	SD	Mean	SE	SD	Mean	SE	SD
English	.464	.406	.096	2.384	.292	.069	.237	.376	.088
Korean	1.207	.601	.142	1.445	.829	.195	1.414	.633	.149
Mandarin	.977	.838	.203	1.224	.754	.183	.808	.504	.122

Stress Language	Vowel Change	High frequency			Medium frequency			Low frequency		
		Mean	SE	SD	Mean	SE	SD	Mean	SE	SD
English	No	2.163	.486	.114	1.857	.554	.131	1.871	.529	.125
	Yes	2.215	.397	.093	2.085	.372	.088	2.281	.413	.097
Korean	No	1.073	.574	.135	.479	.492	.116	.178	.545	.128
	Yes	.774	.620	.146	.476	.481	.113	.729	.808	.190
Mandarin	No	1.523	.669	.162	.934	.669	.162	.776	.533	.129
	Yes	1.556	.662	.160	.868	.680	.165	.909	.093	.219

The intercept in this model ($\chi^2 = 6.324$, $p = .04$, $df = 468$) estimated English speakers' *d*-prime scores with low proficiency in the phoneme condition.

Fixed effects	Estimate	SE	*t*-value	*p*-value
Intercept	2.505	.548	4.568	.0002
Language Korean	−1.766	.631	−2.8	.002
Language Mandarin	−1.527	.605	−2.524	.007
Condition Stress	.344	.112	3.074	.002
Proficiency	−.009	.007	−1.429	.120
Language Korean: Condition Stress	−1.08	.158	−6.84	.0001
Language Mandarin: Condition Stress	−.253	.161	−1.576	.125
Language Korean: Proficiency	.022	.009	2.441	.009
Language Mandarin: Proficiency	.010	.008	1.28	.162

The intercept in this model ($\chi^2 = 2.119$, $p = .14$, $df = 309$) estimated Korean speakers' *d*-prime scores with low proficiency in the phoneme condition.

Fixed effects	Estimate	SE	*t*-value	*p*-value
Intercept	.739	.357	2.071	.028
Language Mandarin	.239	.461	.518	.561
Condition Stress	−.738	.109	−6.742	.0001
Proficiency	.012	.007	1.804	.051
Language Mandarin: Condition Stress	.829	.157	5.277	.0001
Language Mandarin: Proficiency	−.012	.008	−1.423	.115

The intercept in this model ($\chi^2 = 5.027$, $p = .08$, $df = 309$) estimated English speakers' *d*-prime scores with low proficiency and no vowel change.

Fixed effects	Estimate	SE	*t*-value	*p*-value
Intercept	2.703	.743	3.638	.0004
Language Korean	−2.959	.851	−3.476	.0004
Language Mandarin	−1.650	.815	−2.024	.0196
Vowel Change	.229	.105	2.179	.0394
Proficiency	.009	.009	−1.005	.233
Language Korean: Vowel Change	−.149	.149	−1.003	.333
Language Mandarin: Vowel Change	−.196	.151	−1.297	.216
Language Korean: Proficiency	.026	.012	2.106	.011
Language Mandarin: Proficiency	.010	.011	.902	.285

The intercept in this model ($\chi^2 = 3.642$, $p = .05$, $df = 203$) estimated Korean speakers' *d*-prime scores with low proficiency and no vowel change.

Fixed effects	Estimate	SE	*t*-value	*p*-value
Intercept	.026	.485	.054	.944
Language Mandarin	1.343	.594	2.26	.009
Vowel Change	−.483	.318	−1.515	.15
Proficiency	.011	.009	1.168	.177
Language Mandarin: Vowel Change	−.116	.170	−.683	.512
Language Mandarin: Proficiency	−.016	.011	−1.481	.084
Vowel Change: Proficiency	.011	.006	1.897	.073

Results of the *d*-prime scores analyses are consistent with the original analyses separating real words and nonwords. In the initial model with the English group as the baseline, we found that native English speakers had significantly higher *d*-prime scores than both Mandarin and Korean speakers in the phoneme condition. There was a significant interaction between language group and condition when comparing English and Korean speakers. English speakers' *d*-prime scores were significantly higher in the stress condition than in the phoneme condition whereas Korean speakers' *d*-prime scores were significantly lower in the stress condition. This interaction was not significant comparing English and Mandarin speakers. In the second model with the Korean group as the baseline, Korean and Mandarin speakers' *d*-prime scores did not differ significantly in the phoneme condition. However, there was a significant interaction between condition and language group in which Korean speakers' *d*-prime scores were significantly lower in the stress condition than in the phoneme condition while this asymmetry was not present in the Mandarin group.

Focusing on the stress condition only, in the initial model, we found that English speakers' *d*-prime scores were significantly higher than both L2 groups in the no vowel change condition. There was a significant effect of vowel change for English speakers. In the second model, Mandarin speakers showed significantly higher *d*-prime scores than Korean speakers in the stress condition. Neither L2 group showed an effect of vowel change.

References

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics*. Cambridge: Cambridge University Press.

Bachman, L. F. (1982). The trait structure of cloze test scores. *TESOL Quarterly, 16,* 61–70.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English lexicon project. *Behavior Research Method, 39,* 445–459.

Bates, D. M. (2005). Fitting generalized linear mixed models in R: Using the lme4 package. *R News: The Newsletter of the R Project, 5,* 27–30.

Bates, D. [M.], & Maechler, M. (2010). lme4: Generalized linear mixed-effects models using S4 classes. R package version 0.999375-37 [software].

Best, C. T. (1995). A direct realistic perspective on cross-language speech perception. In W. Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research*, pp. 171–206. Timonium, MD: York Press.

Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization of nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 345–360.

Best, C. T., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Haskins Laboratories Status Report on Speech Research, 109/110,* 89–108.

Borden, G., Gerber, A., & Milsark, G. (1983). Production and perception of the /r/–/l/ contrast in Korean adults learning English. *Language Learning, 33,* 499–526.

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

Chen, Y., & Xu, Y. (2006). Production of weak elements in speech – evidence from F0 patterns of neutral tone in Standard Chinese. *Phonetica, 63,* 47–75.

Christophe, A., Peperkamp, S., Pallier, C., Block, E., & Mehler, J. (2004). Phonological phrase boundaries constrain lexical access I: Adult data. *Journal of Memory and Language, 51,* 523–547.

Clopper, C. (2002). Frequency of stress patterns in English: A computational analysis. IULC Working Papers Online 2. [Available at https://www.indiana.edu/∼iulcwp/pdfs/02-clopper02.pdf; retrieved Janurary 13, 2013.]

Crystal, D. (2003). *A dictionary of linguistics and phonetics* (2nd edn.). Malden, MA: Blackwell.

Cutler, A. (1986). Forbear is a homophone: Lexical prosody does not constrain lexical access. *Language and Speech, 29,* 201–220.

Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech & Language, 2,* 133–142.

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

Duanmu, S. (2007). *The phonology of standard Chinese* (2nd edn.). New York: Oxford University Press.

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

Dupoux, E., Peperkamp, S., & Sebastián-Gallés, N. (2001). A robust method to study stress “deafness”. *Journal of Acoustical Society of America, 110,* 1606–1618.

Dupoux, E., Peperkmap, S., & Sebastián-Gallés, N. (2010). Limits on bilingualism revisited: Stress “deafness” in simultaneous French–Spanish bilinguals. *Cognition, 114,* 266–275.

Dupoux, E., Sebastián-Gallés, N., Navarrete, E., & Peperkamp, S. (2008). Persistent stress “deafness”: The case of French learners of Spanish. *Cognition, 106,* 682–706.

Flege, J. E. (1992). Speech learning in a second language. In C. Ferguson, L. Menn & C. Stoel-Gammon (eds.), *Phonological development: Models, research, and implications*, pp. 565–604. Timonium, MD: York Press.

Flege, J. E. (1995). Second-language speech learning: Findings, and problems. In W. Strange (ed.), *Speech perception and linguistic experience: Theoretical and methodological issues*, pp. 233–273. Timonium, MD: York Press.

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., & MacKay, I. R. A. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition, 26,* 1–34.

Flege, J. E., Schirru, C., & MacKay, I. R. A. (2003). Interacction between the native and second language phonetic subsystems. *Speech Communication, 40,* 467–491.

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.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds “L” and “R”. *Neuropsychologia, 9,* 317–323.

Grainger, J. (1990). Word frequency and neighborhood effects in lexical decision and naming. *Journal of Memory and Language, 29,* 228–244.

Guion, S. (2005). Knowledge of English word stress patterns in early and late Korean–English bilinguals. *Studies in Second Language Acquisition, 27,* 503–533.

Hayes, B. (2009). *Introductory phonology*. Malden, MA: Blackwell.

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

Jun, S.-A. (1995). A phonetic study of stress in Korean. *Journal of Acoustical Society of America, 98,* 2893.

Jun, S.-A. (1998). The accentual phrase in the Korean prosodic hierarchy. *Phonology, 15,* 189–226.

Jun, S.-A. (2000). K-ToBI (Korean ToBI) labelling conventions (version 3.1). http://www.linguistics.ucla.edu/people/jun/ktobi/K-tobi.html, retrieved Janurary 13, 2013.

Jun, S.-A. (2005). Korean intonational phonology and prosodic transcription. In S.-A. Jun (ed.), *Prosodic typology: The phonology of intonation and phrasing*, pp. 201–229. Oxford: Oxford University Press.

Jusczyk, P. W., Cutler, A., & Redanz, N. (1993). Preference for the predominant stress pattern of English words. *Child Development, 64,* 675–687.

Kager, R. (2007). Feet and metrical stress. In P. de Lacy (ed.), *The Cambridge handbook of phonology*, pp. 195–227. Cambridge: Cambridge University Press.

Karlsson, F. (1999). *Finnish: An essential grammar*. London: Routledge.

Kim, S., Broersma, M., & Cho, T. (2012). The use of prosodic cues in processing an unfamiliar language. *Studies in Second Language Acquisition, 34,* 415–444.

Kim, S., & Cho, T. (2009). The use of phrase-level prosodic information in lexical segmentation: Evidence from word-spotting experiments in Korean. *Journal of the Acoustical Society of America, 125,* 3373–3386.

Lee, B., Guion, S. G., & Harada, T. (2006). Acoustic analyses of the production of unstressed English vowels by early and late Korean and Japanese bilinguals. *Studies in Second Language Acquisition, 28,* 287–513.

Lin, M., & Yan, J. (1980). Beijinghua qingsheng de shengxue xingzhi [Acoustic properties of Mandarin neutral tone]. *Dialect, 3,* 166–178.

Lin, T. (1985). Tantao Beijing hua qingsheng xingzhi de chubu shiyan [Preliminary experiments on the nature of Mandarin neutral tone]. In T. Lin & L. Wang (eds.), *Working Papers in Experimental Phonetics*, pp. 1–26. Beijing: Beijing University Press.

Lin, T., & Wang, W. (1984). Shengdiao ganzhi wenti [Perception of tones]. *Zhongguo Yuyan Xuebao* [Bulletin of Chinese Linguistics]*, 2,* 59–69.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers, 28,* 203–208.

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research, 50,* 940–967.

Mehler, J., Jusczyk, P. W., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition, 29,* 144–178.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A., Jenkins, J., & Fujimura, O. (1975). An effect of language experience: The discrimination of /r/ and /l/ by native speakers of Japanese and English. *Perception & Psychophysics, 18,* 331–340.

Peperkamp, S. (2004). Lexical exception in stress systems: Arguments from early language acquisition and adult speech perception. *Language, 80,* 98–126.

Peperkamp, S., & Dupoux, E. (2002). A typological study of stress “deafness”. In C. Gussenhoven & N. Warner (eds.), *Laboratory Phonology 7*, pp. 203–240. Berlin: Mouton de Gruyter.

R Development Core Team. (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org.

Schane, S. A. (1968). *French phonology and morphology*. Cambridge, MA: MIT Press.

Shen, X. S. (1993). Relative duration as a perceptual cue to stress in Mandarin. *Language and Speech, 36,* 415–433.

Skoruppa, K., Pons, F., Christophe, A., Bosch, L., Dupoux, E., Sebastián-Gallés, N., Limissuri, R. A., & Peperkamp, S. (2009). Language-specific stress perception by 9-month-old French and Spanish infants. *Developmental Science, 12,* 914–919.

Sohn, H.-M. (1999). *The Korean language*. Cambridge: Cambridge University Press.

Soto-Faraco, S., Sebastián-Gallés, N., & Cutler, A. (2001). Segmental and suprasegmental mismatch in lexical access. *Journal of Memory and Language, 45,* 412–432.

Swets, J. A. (1996). *Signal detection theory and ROC analysis in psychology and diagnostics: Collected papers*. Hillsdale, NJ: Lawrence Erlbaum.

Syntrillium Software (2003). http://cooledit-pro.software.informer.com/.

Tremblay, A. (2008). Is second language lexical access prosodically constrained? Processing of word stress by French Canadian second language learners of English. *Applied Psycholinguistics, 29,* 553–584.

Tremblay, A. (2009). Phonetic variability and the variable perception of L2 word stress by French Canadian listeners. *International Journal of Bilingualism, 13,* 35–62.

Tremblay, A., Coughlin, C. E., Bahler, C., & Gaillard, S. (2012). Differential contribution of prosodic cues in the native and non-native segmentation of French speech. *Journal of Laboratory Phonology, 3,* 385–423.

Tyler, M. D., & Cutler, A. (2009). Cross-language differences in cue use for speech segmentation. *Journal of the Acoustical Society of America, 126,* 367–376.

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken word recognition. *Journal of Memory and Language, 50,* 1–25.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7,* 49–63.

Wickham, H. (2009). *Ggplot2: Elegant graphics for data analysis*. New York: Springer.

Xu, Y., & Xu, C. X. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics, 33,* 159–197.