THE EFFECT OF STRESS ON ENGLISH WORD RECOGNITION BY NATIVE SPEAKERS OF TYPOLOGICALLY DIFFERENT LANGUAGES

Shu-chen Ou & Zhe-chen Guo

National Sun Yat-sen Univeristy, Taiwan
sherryou@faculty.nsysu.edu.tw; m031020006@student.nsysu.edu.tw

ABSTRACT

The study investigates the effect of stress on English word recognition by speakers of lexical tone and lexical pitch accent languages. Two on-line tasks that involved Taiwan Mandarin and Japanese speakers as test subjects and English natives as controls were conducted: (i) a uni-modal lexical decision task and (ii) a cross-modal priming and lexical decision task. It was found that Mandarin and Japanese speakers were accurate in auditorily discerning words and non-words based on stress patterns (e.g., *PEOple* vs. **peoPLE*) insofar as the stressed syllable bore a high tone. This suggests that their representation of stress may be tonal in nature. Besides, Mandarin speakers responded faster when hearing a prime with primary stress and seeing a word with the first syllable stressed. The same was not found in Japanese speakers, suggesting that the use of stress in English word recognition was constrained by native lexical prosody.

Keywords: word recognition, English stress, tone and pitch accent language speakers

1. INTRODUCTION

Typologically referred to as lexical stress language, English uses stress to distinguish word meanings or part of speech (e.g., *PERmit* (n.) vs. *perMIT* (v.)). Despite this fact, native speakers of English may not exploit the cue of stress effectively in word recognition as such minimal pairs are arguably few in the English lexicon [3, 4]. In contrast, speakers of Dutch, a language typologically related to English, are reportedly able to employ stress cue in recognizing English words in the cross-modal priming and lexical decision task: they respond faster in stress matching than stress mismatching conditions (e.g., upon hearing a fragment like *MU-* and seeing a word, they identify the word faster if it is *music* than if it is *museum*). This is presumably because the function load of stress is heavier in Dutch than in English [2, 5]. Yet, the use of stress in word recognition by speakers whose native languages are typologically distant from English (e.g., lexical tone and lexical pitch accent languages) is less known.

Native speakers of Taiwan Mandarin, a lexical tone language, have been found to be partially deaf to English stress in off-line tasks based on the empirical evidence that they are able to identify stress when it is cued mainly by higher pitch (e.g., when the target word is embedded in the falling intonation, an unmarked pitch accent context) but not so when the main cue is longer duration (e.g., when the word is in the rising intonation, a marked pitch accent context) [7]. One explanation is that they tend to interpret stress as tonal distinction (i.e., strong stress = [+high] tone; weak stress = [-high] tone) [1]. It is then interesting to investigate whether the tonal interpretation of stress based on pitch would have an effect in more demanding on-line word recognition. A hypothesis is that strong and weak stress encoded as tonal contrasts would both facilitate Taiwanese speakers' responses in the stress matching condition of the tasks used in [2, 5].

Japanese, which is typologically classified as lexical pitch accent language, falls between a lexical tone language and a lexical stress language. It is similar to the former in the sense that a syllable bears a certain pitch; it resembles the latter because the nuclear high pitch falls on the accented syllable of a word. Studies have indicated that Japanese speakers show a good command of pitch, duration, and intensity in producing English stress [6]. It is therefore of interest to the study to investigate what would be the phonetic cues employed by Japanese speakers in on-line word recognition.

In sum, the question addressed concerns whether and how speakers of Mandarin and Japanese, whose native languages are typologically distant from English, employ the cue of lexical stress in on-line word recognition. Two experiments were conducted. One was a uni-modal lexical decision task in which an auditory stimulus was presented one at a time (e.g., *PEOple* or **peoPLE*) and a response regarding its word status (i.e., *yes* for real English word and *no* for nonword) was required. The words were pronounced in both the falling and the rising intonations. The other was a cross-modal priming lexical decision task in which subjects heard a word fragment (e.g., *CAM-*), saw a visual word (e.g. *campus, campaign,* or *yulf*), and responded whether the word was a real English word. Both accuracy rates and reaction times (RT) would be measured.

2. METHOD

2.1. Materials and design

Words used in the two experiments were all selected from a 7000-word list compiled for high school students by the Ministry of Education of Taiwan to guarantee that non-native participants were familiar with them. In addition, they had to be words which, according to the CELEX database, occur at least five times per million words. The designs of the two experiments are as follows.

The uni-modal lexical decision experiment required on-line auditory identification of correct stress and examined the prosodic information exploited by non-native speakers in this task. The experimental items were disyllabic word-nonword pairs, each including a real word and a nonword derived by altering the stress position of the real word (e.g., *MOney* vs. **moNEY*). The control items were word-nonword pairs involving segmental contrasts: the nonword of each pair was created by altering a phoneme of the real word (e.g., *system* vs. **syskem*). The number of the control and the experimental pairs was 15 and 28, respectively. All the words and nonwords were embedded in a statement and a yes/no question in the final position and recorded by a female native English speaker with the North American accent. The sentential contexts of the recorded items were then removed, leaving only the sentence-final words, which would be presented as the auditory stimuli to subjects. An acoustic analysis of these stimuli revealed that stressed syllables were more prominent than the unstressed ones in pitch, duration, and intensity in the falling intonation, but only in duration in the rising intonation. The differential stress cues in the two intonations would help identify the cue employed by the non-native subjects.

The cross-modal priming and lexical decision experiment was conducted to further investigate how lexical stress might be interpreted by the non-native subjects. The experimental items were word pairs with segmentally identical first syllables but different stress positions (e.g., *CAMpus* vs. *camPAIGN*), and 21 such pairs were selected. Each pair was then assigned a control item, a disyllabic word whose initial syllable had the same syllable structure as those of the pair but contained distinct segments. The results were 21 sets of words. All the words were carried in the final position of non-constraining sentences and recorded by the same English native speaker. Each recorded item was truncated to the first syllable of the last word, creating a fragment at the end to be used as a prime (e.g., *The word he said was CAM-*). For each set of words (e.g., *campus*, *campaign*, and *mountain*), the visual stimulus after priming was a nonword (e.g., *yulf*) or an experimental item of that set (e.g., *campus*). Permutation of types of priming fragments and visual words resulted in three trial conditions: (i) a test condition where the visual and auditory stimuli had segmentally identical first syllables but their stress might match (e.g., audio: *CAM-*; visual: *campus*) or mismatch (e.g., audio: *CAM-*; visual: *campaign*), (ii) a control condition where the two stimuli contained distinct segments (e.g., audio: *MOUN-*; visual: *campus*), and (iii) a nonword condition where the visual word was a nonword.

2.2 Procedure

Wearing headphones, subjects were tested individually in a sound-proof chamber with a desktop computer and informed that there were two listening experiments. In the unimodal lexical decision experiment, they were told that in each trial an auditory stimulus would be presented via the headphones and their task was to decide if the stimulus was a real English word or not by pressing *yes* or *no* buttons. This procedure was repeated until they completed the experiment. In the cross-modal priming experiment, they were told that in each trial an English sentence with the last word truncated would be heard and a word would be displayed on the screen. They were asked to determine whether the word on the screen was a real English word or not by pressing *yes* or *no* buttons. They repeated the above procedure until the end of the task.

2.3 Participants

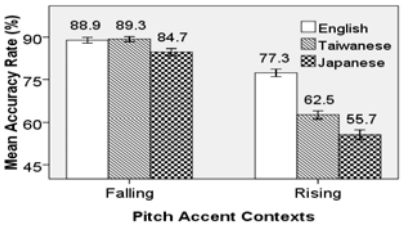
The participants were native speakers of English, Taiwan Mandarin, and Japanese, the number of subjects recruited being 18, 23, and 16, respectively. The Taiwanese subjects had to be English learners with a minimum TOEIC score of 800 and at least 10 years of learning experience. The requirements, however, were not obligatory for the Japanese subjects due to the availability of Japanese speakers.

3. RESULTS

3.1. Unimodal lexical decision task

In the unimodal lexical decision task, the mean accuracy rates of the three subject groups in the control (segmental) condition were all above 80% in both the pitch accent contexts, whereas their accuracy in the test condition varied as a function of their native language and the pitch accent context of the presented stimulus, as can be seen in Fig. 1.

Figure 1: The mean accuracy rates of the three groups in the experimental condition, with standard error (SE) bars



All their responses to the experimental trials were explored with mixed-effects logistic regression. Being the dependent variable, each response was either correct (coded as 1) or incorrect (coded as 0). The predictors included Group (with the English group being the reference level) and two binary factors, each with two levels coded -0.5 and 0.5: Intonation, which was falling (-0.5) vs. rising intonation (0.5), and Position, which was trochaic (-0.5) vs. iambic (0.5) stress. A model selection procedure beginning with the full model yielded one that contained the predicators and their interactions as fixed-effects and by-subject and by-item intercepts as random effects. It then found a significant effect of Group, indicating that the English group responded more accurately than the Taiwanese (β = -.336, SE(β) = .150, z = -2.24, p < .05) and the Japanese (β = -.812, SE(β) = .155, z = -5.22, p < .001) groups. Intonation was also shown to be a crucial factor: when the intonation changed from falling to rising, accuracy would deteriorate (β = -1.050, SE(β) = .143, z = -7.37, p < .001). Such an effect of Intonation was found to interact with Group: compared with that of the English subjects, the accuracy of the Taiwanese (β = -.932, SE(β) = .197, z = -4.75, p < .001) and Japanese (β = -.581, SE(β) = .189, z = -3.07, p < .01) subjects was more adversely affected by the rising intonation.

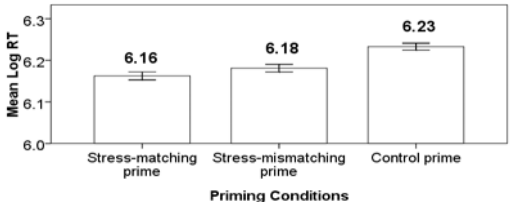
While further inspection of RTs might lend more insight to these finding, it was limited by non-native subjects' high error rates in the rising intonation condition. Thus, only the RTs of correct responses in the falling intonation were subjected to further mixed-effects regression analyses. Here RTs longer than 1100ms or shorter than 200ms were first eliminated as outliers, which accounted for 12.4% of the observations. The remaining ones were log-transformed and included as the dependent variable. The model for the RT data was the same as the previous one except that Intonation was excluded. It revealed only a marginally significant main effect suggesting that the Japanese group responded slower than the English one (β = .176, SE(β) = .098, t = -1.79, p < .10). In the unmarked pitch accent context, the response latencies to stress of Taiwanese group were native-like. Yet, their performance in the other context revealed that their encoding of stress was still dependent on pitch, a crucial cue to Mandarin and Japanese prosodic contrasts. A question then arose as to how stress contrasts based on pitch could be equated by the non-native subjects to their native pitch distinctions. To address the question, the results of the priming task were examined.

3.2. Cross-modal priming task

All responding with high accuracy (mean accuracy rate > 80%) to all the trial types, the three groups of subjects were examined for RTs of correct responses in the experimental and control conditions. Data from six participants were discarded for their low accuracy rates on the control trials (i.e., two standard deviations below the overall mean). Removal of RT outliers based on the criteria of the previous experiment excluded of 7.1 % of all the observations and the remaining valid RTs were log-transformed.

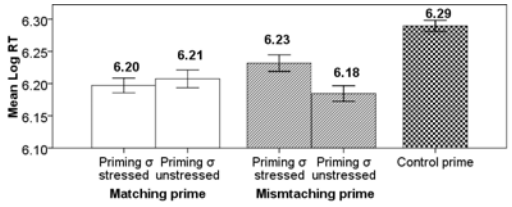
To examine the interpretation of stress by the non-native subjects' in word recognition, the RTs of the English group were first explored using mixed-effects regression. The predictors comprised the prime conditions (i.e., stress-matching, stress-mismatching, and control conditions, with the first one being the reference level) and other factors (e.g., word frequency). Drop-one model selection yielded a full model that included as fixed-effects all the main effects and their interactions and as random effects a by-subject random intercept and two by-item random intercepts for the priming fragment and the visual word. The results indicated that the RTs in the control condition were significantly longer than those in the other two (β = .100, SE(β) = .023, t = 4.33, p < .001), revealing the facilitation effects of matching segments (see Fig. 2). Yet, the English subjects' RTs in the two experimental conditions did not differ significantly. Although it appeared that the stress primes did not facilitate their word recognition, this might be attributed to the English subjects' sufficient familiarity with the visual words. As native speakers' word familiarity was not relevant to the inquiry of this study, subsequent analyses of the non-native subjects' RTs were conducted independently.

Figure 2: The mean log RTs (with SE bars) of the English group in different priming conditions



The model used for the English group was applied to the non-native subjects' RTs, but with modifications. The change was that the three prime conditions were re-coded into three binary factors: Stress, which was priming fragment with (0.5) vs. without (-0.5) primary stress, Match, which was stress matching (0.5) vs. mismatching (-0.5) conditions, and Segment, which was whether the priming fragment and displayed word contained segmentally identical first syllables (0.5) or not (-0.5). The factors and their interactions were the fixed-effects. Fitting the Taiwanese subjects' RTs with this model indicated that segmentally matching primes significantly facilitated their responses (β = -.084, SE(β) = .012, t = -7.27, p < .001). Interestingly, a significant interaction was obtained between Match and Stress (β = -.064, SE(β) = .030, t = -2.11, p < .05): stress-matching primes facilitated lexical access only when the priming fragment bore primary stress, as shown in Fig. 3. Stressed and unstressed syllables might thus be interpreted differently by the lexical tone language speakers.

Figure 3: The mean log RTs (with SE bars) of the Taiwanese group in different priming conditions



Applying the model to the Japanese subjects' RTs showed that no factors or interactions had significant effects except for segmental priming (β = -.081, SE(β) = .014, t = -5.62, p < .001).

4. DISCUSSION

Except that it involves on-line word recognition, the uni-modal lexical decision experiment replicates the findings of [7]: Taiwanese participants are accurate in discerning words and non-words only in the context where strong or primary stress was cued by higher pitch. The same is observed for Japanese participants. This indicated that stress information, albeit included in the non-native learners' English lexical storage, is interpreted as lexical pitch distinction in their languages and that such interpretation extends to on-line word recognition.

It is then expected that weak and strong stress would both facilitate non-native learners' responses in stress matching condition of the cross-modal priming task as they may be interpreted as, for example, a [+high] tone and a [-high] tone. The results, however, only partially supports this hypothesis as the responses of Taiwanese subjects are faster in stress-matching condition only when the priming fragment has primary stress. The lack of the same facilitating effect from unstressed primes may be attributed to the fact that on-line and off-line tasks involve rather different modes of processing.

In contrast, the same is not found for the Japanese group. A preliminary account is that Japanese lexical pitch accent is encoded at the word/phrase level, requiring comparison of multiple syllables within a lexical item for differentiation of lexical meanings. Support for this account comes from [8], which suggests that due to differential pitch functionality among different first languages, Japanese listeners are more sensitive to Thai tones than Korean (no lexical pitch contrasts) ones but less so than Mandarin (contrastive tones on syllables) ones. In this view, Japanese speakers would not benefit from the auditory primes (as Mandarin speakers do), for the stimuli are monosyllabic and may not suffice to elicit priming effects on lexical access for them. While an alternative explanation would be that the Japanese subjects were selected without setting a threshold of English proficiency, it can be further investigated how exactly English stress is interpreted by native speakers of Japanese.

5. ACKNOWLEDGEMENTS

We thank Ministry of Science and Technology, Taiwan, for support of this research project (MOST102-2410-H-110-014-MY2).

6. REFERENCES

[1] Cheng, C.-C. (1968). English stresses and Chinese tones in Chinese sentences. *Phonetica, 18*, 77-88.

[2] Cooper, N., Cutler, A., & Wales, R. (2002). Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners. *Language and speech*, *45*, 207-228.

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

[4] Cutler, A., & Pasveer, D. (2006). Explaining cross-linguistic differences in effects of lexical stress on spoken-word recognition. *Proc. 3rd International Conference on Speech Prosody* Dresden (pp. 250-254).

[5] Cutler, A., Wales, R., Cooper, N., & Janssen, J. (2007). Dutch listeners' use of suprasegmental cues to English stress. *Proc. 16th ICPhS* Saarbrücken (pp. 1913-1916).

[6] Kondo, M. (2009). Is acquisition of L2 phonemes difficult? Production of English stress by Japanese speakers. *Proc. 10th GASLA* Champaign, 105-112.

[7] Ou, S.-C. (2010). Taiwanese EFL learners' perception of English word stress. *Concentric: Studies in Linguistics*, *36*, 1-23.

[8] Schaefer, V., & Darcy, I. (2014). *Laboratory Phonology*, *5*, 489-522.