

## Perception of English Lexical Stress by English, Beijing Dialect, and Guangzhou Cantonese Speakers

Xing-rong GUO<sup>1,a</sup>, Xiao-xiang CHEN<sup>1,b,\*</sup> and Yi-ming GUO<sup>2,c</sup>

<sup>1</sup>College of Foreign Languages, Hunan University,  
Changsha, China

<sup>2</sup>School of Economics and Management, Shanghai Maritime University,  
Shanghai, China

<sup>a</sup>guoxingrong356@163.com, <sup>b</sup>huda319@163.com, <sup>c</sup>ymguo@shmtu.edu.cn

**Keywords:** English lexical stress, Perceptual cues, Perception, English, Chinese dialect.

**Abstract.** This study explored how listeners' L1 dialect affects their weighting of acoustic correlates in perceiving English lexical stress. 20 American English speakers, 20 Beijing dialect speakers, and 20 Guangzhou Cantonese speakers produced nonce disyllabic words. Results showed that the salience hierarchy for the perception of English lexical stress by English listeners were: vowel quality>duration>F0>intensity; Beijing listeners: vowel quality>F0>duration>intensity; Guangzhou listeners: F0>vowel quality>duration>intensity. Stress perception of the Beijing dialect listeners is more similar to that of the native English listeners than Guangzhou listeners. These findings reveal that L1 dialect may transfer to the perception of L2 English stress. Results suggest that explicitly taught of these cues in English class is necessary at the preliminary stage of learning English stress to avoid the effect of L1 dialect.

### Introduction

Perception of second language (L2) speech sounds is heavily affected by native language (L1) prosodic knowledge [1]. Such perceptual bias has been repeatedly found in second language acquisition. For instance, Japanese listeners have difficulty when discriminating the English /r/-/l/ contrast, since they perceived these contrast as variants of a phoneme in their L1 [2]. The L1-L2 interference not only occur in segmental contrasts but also in suprasegmental dimensions. Perception of lexical stress is influenced by L1 in two aspects: the stress pattern and the acoustic cues. Many research revealed that speakers of non-stress languages have difficulties with the perception of lexical stress [3, 4].

However, little is known about how L1 dialect acoustic system affect L2 perception. Previous research indicates that L2 speakers are adept at using acoustic cues in target language if these correlates are already actively applied for realizing prosodic contrasts in their L1. For example, [5] found the Standard Beijing dialect learners used duration more than the Taiwanese Beijing dialect learners in perceiving English stress, since the former uses duration together with F0 to realize lexical stress, but the latter does not have this stress distinction.

Evidence to date indicated that perception of stress is the result of the fully interaction of F0, duration, intensity and vowel quality [1]. Based on this opinion, this study investigates how such interaction of perceptual cues to nonnative lexical stress is affected by the prosodic characteristics of listeners' L1 dialect.

## Perceptual Cues to English Lexical Stress

The principle perceptual cues that correlate with lexical stress include F0, duration, intensity and vowel quality [6, 7]. There has been a lack of consensus concerning the relative weights of the acoustic correlates in English lexical stress. There were many different versions of the salience hierarchy of the prosodic cues of stress in English lexical stress, for instance, F0>duration>intensity [8, 9]; F0>intensity>duration [10]; duration>intensity>F0 [11]; duration>F0>intensity [12]; vowel quality>F0>intensity>duration [1]. All these studies confirmed that stressed syllables tend to have higher F0, greater intensity, longer duration and full vowel quality than unstressed ones.

## Present Study

Cross-linguistic studies of stress perception remain scarce, and few studies have compared F0, duration, intensity, and vowel quality simultaneously. The present study aims to provide further insights on perceptual correlates of English lexical stress by examining the relative weight of the four acoustic cues in stress perception by native speakers of American English (AE), Beijing dialect (BJ) and Guangzhou Cantonese (GZ). These languages were chosen in the present study because they are typologically diverse. There are some differences and similarities between English vs. Chinese dialects, Beijing dialect vs. Guangzhou Cantonese. For example, in English there are minimal pairs of words that only differ in stress. Similar stress pairs are also found in Beijing dialect. Whereas Cantonese does not have lexical stress contrast. Different languages relied on different prosodic cues.

Based on previous research, Table 1 hypothesizes relative weight of the four acoustic correlates in the three groups. This study predicted that, when exposed to the same material, speakers of different languages/dialects would attend to perceptual cues in a different fashion. If Beijing and Guangzhou dialect groups transferred their L1 perception strategies of lexical tone to L2 English, they might primarily attend to use F0 cue. Vowel quality was predicted to be the second strongest cue for all three groups. Intensity was hypothesized to be the weakest cue for the three groups; it was expected duration differences to be observed by Beijing and Guangzhou dialect listeners.

Table 1. Hypothesized hierarchy of auditory correlates in the three groups.

| Auditory cues | Language/Dialects group |                 |                     |
|---------------|-------------------------|-----------------|---------------------|
|               | America English         | Beijing Dialect | Guangzhou Cantonese |
| F0            | 3                       | 1               | 1                   |
| Intensity     | 4                       | 4               | 4                   |
| Duration      | 1                       | 3               | 3                   |
| Vowel quality | 2                       | 2               | 2                   |

Note: 1= most important, 4= least important

## Method

### Participants

Three groups of participants were recruited: 10 native American English speakers (10 for each gender; age range: 21-27, M=22.7), 10 Beijing dialect speakers (5 female; age range: 19-22, M=20.2), and 10 Cantonese speakers (5 female; age range: 19-23, M=18.8). The Chinese candidates' English proficiency was determined by the Language Experience and Proficiency Questionnaire (LEAP-Q) [13]. Chinese speakers at the intermediate English proficiency levels were invited for this experiment.

## Stimuli

In order to avoid segmental influences on perceptual judgments, the target stimuli were nonce words. Stimuli with  $C_1V_1C_2V_2$  structure were English disyllable non-words that differ in stress placement: /'mabə/ vs. /mə'ba/, which similar to [1]. F0, duration, intensity and vowel quality in both syllables were all manipulated by Praat (V 5.3.76) according to the report of [1]. Therefore,  $2 (F0) \times 2 (duration) \times 2 (intensity) \times 2 (vowel\ quality)$  manipulation on each syllable, resulting in 256 tokens in all.

## Procedure

The manipulated syllables were concatenated to construct the target stimuli. The participants were asked to listen to the different versions of the word and judge the position of lexical stress. All participants were tested individually by using the E-prime 2.0. The participants were asked to identify the stress placement. Each stimuli were heard three times. The average running time for the experiment was 40 min.

## Analysis and Results

Figure 1 shows the accuracy response of the four cues made by the three groups.

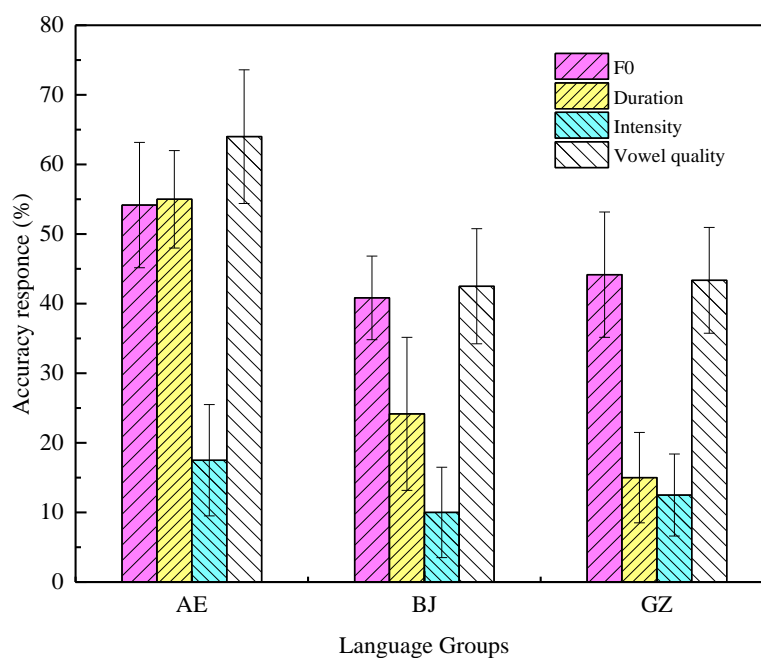


Figure 1. Accuracy response of the four cues made by the three groups

Figure 1 shows that F0 has significant difference in three groups. Factorial ANOVA revealed significant main effect difference in F0 [ $F(2,27) = 6.807, p = 0.004$ ], duration [ $F(2,27) = 59.073, p < 0.001$ ], and vowel quality [ $F(2,27) = 20.484, p < 0.001$ ], but not intensity [ $F(2,27) = 2.983, p = 0.068$ ]. Turkey HSD tests demonstrated AE speakers use F0 significantly different with BJ ( $p = 0.004$ ) and GZ speakers ( $p = 0.034$ ), but no significant difference were found among BJ and GZ groups ( $p = 0.654$ ); AE speakers used significantly different duration with BJ ( $p < 0.001$ ) and GZ speakers ( $p < 0.001$ ), but not BJ and GZ groups ( $p = 0.062$ ). There was no significant intensity difference between AE and BJ ( $p = 0.059$ ), AE and GZ speakers ( $p = 0.263$ ), as well as BJ and GZ groups ( $p = 0.706$ ). AE speakers used significantly different vowel quality with BJ ( $p < 0.001$ ) and GZ speakers ( $p < 0.001$ ), but there is no significant difference between BJ and GZ groups.

( $p=0.974$ ). Table 2 presents the observed weighting of the four cues for each language group.

The results revealed the three groups reached significant difference for each language group in terms of F0, duration, vowel quality, but not intensity. Vowel quality was the dominant cue for AE and BJ groups. Whereas the GZ group weighed the F0 cue heavily. The AE listeners used all the four cues to a greater extent than the BJ and GZ listeners.

Figure 2 shows the interaction of the acoustic cues and stress pattern in stress perception by the three groups. Stress cues were contrast coded so as that each cue signal either trochaic or iambic.

Table 2. Observed hierarchy of auditory correlates in three groups.

| Auditory cues | Language/Dialects group |                 |                     |
|---------------|-------------------------|-----------------|---------------------|
|               | America English         | Beijing Dialect | Guangzhou Cantonese |
| F0            | 3                       | 2               | 1                   |
| Intensity     | 4                       | 4               | 4                   |
| Duration      | 2                       | 3               | 3                   |
| Vowel quality | 1                       | 1               | 2                   |

Note: 1 = most important, 4= least important.

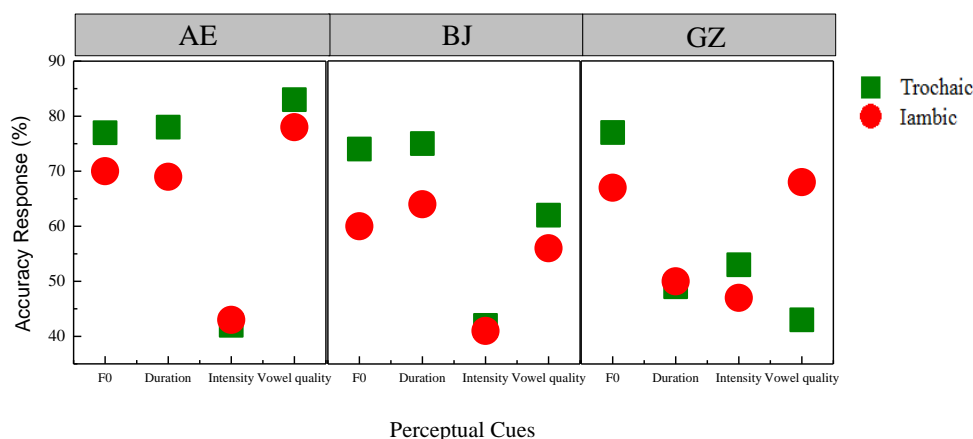


Figure 2. Stress identification for trochaic and iambic patterns by the three groups

Figure 2 shows that AE and BJ groups were more likely to identify syllables as trochaic than iambic relying on the four cues. For GZ group, duration and vowel quality for the iambic stress pattern were higher than the trochaic pattern. This means that both AE and BJ groups prefer to perceive disyllabic nonce words as trochaic pattern, but such preference was not so obvious in GZ group. BJ group showed more native-like pattern than GZ group in using F0, duration, intensity and vowel quality to identify trochaic and iambic stress patterns. Interesting, vowel quality manipulation has not created a significant effect between the trochaic and iambic in the AE group and BJ group ( $p > 0.05$ ) except GZ group ( $p < 0.05$ ).

## Discussion

In contrast to the predicted hierarchy before, the results indicated that there were more similarities between AE and BJ groups than between AE and GZ groups. For the AE and BJ groups, vowel quality was the strongest perceptual cue, whereas F0 was the strongest for the GZ group. Intensity was the least for the three groups, indicating that intensity was far less influential than the other cues.

For AE group, vowel quality, duration and F0 were significant cues in stress perception, whereas intensity cues were far less influential. For BJ group, vowel quality, duration and F0 were significant cues, whereas intensity was the least influential cue. But GZ listeners demonstrated a different pattern. F0 and vowel quality were also important perceptual cue for the GZ group, duration and intensity cues was less important. These results for the AE group showed slight differences from the hypothesized hierarchy of auditory correlates before. Vowel quality was the most robust cue, duration was a stronger cue than F0 and intensity.

For BJ group, L2 speakers transfer their L1 stress perception strategies to L2 tasks[5], BJ learners primarily relying on those auditory cues that have been actively explored for stress contrasts in their L1.

For the GZ group, the observed minimal role of duration for English stress perception is consistent with previous studies on the perception of L1 stress, but the fact that F0 and vowel quality influenced stress perception to a greater extent than duration. Cantonese has stronger syllable-timed rhythm than Mandarin, since the syllable structure of Cantonese is simple without word stress or any vowel reduction [14]. Neutral tones occur frequently in Beijing dialect with certain vowel reduction, so they were more sensitive than Cantonese listeners in using vowel quality.

Regarding to the affect of trochaic or iambic stress pattern, results suggest that both AE and BJ groups prefer to perceive disyllabic nonce words as trochaic, but such preference was not so obvious in GZ group. This pattern may have resulted from the dominance of trochaic words in English and Beijing dialect. In English, there are 70% of disyllabic content words are trochaic[15]. In Beijing dialect, the word only has strong-weak stress pattern since the unstressed syllables cannot occur in the first syllable, and vowel reduction occurs only in medial and final positions. This may account for the more stress sensitivity to the vowel cue in the trochaic stress pattern for Beijing dialect listeners. However, there is no neutral tone in Cantonese, so the listeners have not found any bias to trochaic contours as English and Beijing groups.

## **Conclusion**

In sum, a comparison of stress perception performance of the English, BJ and GZ listeners offers some interesting insights into the cross-linguistic and cross-dialect effect on the encoding of the L2 prosodic knowledge. Though there are prominent differences in different languages, similar performance can arise. Interesting, L1 listeners of different dialects with similar prosodic characteristics' would show quite dissimilar preference on L2 prosodic perception tasks. Although Beijing dialect and Guangzhou Cantonese listeners rely on different auditory cues for English stress perception. However, it could be infer that stress perception performance of the Beijing dialect listeners but not the Guangzhou Cantonese listeners is more similar to that of the native English listeners. Results suggest that the L1 dialect background is a potentially influential factor which may transfer from L1 to L2.

## **Acknowledgments**

This study was funded by National Social Science Foundation (14BYY144) and China Postdoctoral Science Foundation (2016M590348).

## References

- [1] A. Chrabaszcz, M. Winn, C.Y. Lin, W.J. Idsardi, Acoustic cues to perception of word stress by English, Mandarin, and Russian speakers. *Journal of speech, language, and hearing research*. 57(2014) 1468-1479.
- [2] K. Miyawaki, J.J. Jenkins, W. Strange, A.M. Liberman, R. Verbrugge, O. Fujimura. An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Attention Perception & Psychophysics*. 18 (1975) 331-340.
- [3] C.Y. Lin, Wang MIN, W.J. Idsardi, Y.I. Xu, Stress processing in Mandarin and Korean second language learners of English. *Bilingualism: Language and Cognition*. 17(2013) 316-46.
- [4] V.Y. Yu, J.E. Andruski, A cross-language study of perception of lexical stress in English. *Journal of psycholinguistic research*. 39 (2010) 323-344.
- [5] Z. Qin, A. Tremblay, Effects of native dialect on Mandarin listeners' use of prosodic cues to English stress. *The 7th Speech Prosody Conference 2014*, Trinity College, Dublin, Ireland.
- [6] D.B. Fry, Experiments in the perception of stress. *Language and Speech*. 1(1958):126-152.
- [7] M.E. Beckman, Stress and non-stress accent: Walter de Gruyter, 1986.
- [8] D.B. Fry, Duration and intensity as physical correlates of linguistic stress. *J. Acoust. Soc. Am*. 27(1955)765-768.
- [9] P. Ladefoged, *Phonetic data analysis: An introduction to fieldwork and instrumental techniques*: Wiley-Blackwell, 2003.
- [10] P. Lieberman, Some acoustic correlates of word stress in American English. *J. Acoust. Soc. Am*. 32(1960)451-454.
- [11] M.E. Beckman, J. Edwards, Articulatory evidence for differentiating stress categories. *Papers in laboratory phonology III: Phonological structure and phonetic form*. 1994, 7-33.
- [12] C. Adams, R.R. Munro, In search of the acoustic correlates of stress: fundamental frequency, amplitude, and duration in the connected utterance of some native and non-native speakers of English. *Phonetica*. 35(1978)125-156.
- [13] V. Marian, H.K. Blumenfeld, M. Kaushanskaya, The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*. 50(2007)940-967.
- [14] P. Mok, On the syllable-timing of Cantonese and Beijing Mandarin. *Chinese. Journal of Phonetics*. 2 (2009) 148-154.
- [15] A. Cutler, D.M. Carter, The predominance of strong initial syllables in the English vocabulary. *Computer Speech & Language*. 2(1987)133-142.