Brain & Language 138 (2014) 61–70

Contents lists available at ScienceDirect

Brain & Language

journal homepage: www.elsevier.com/locate/b&l

Neural correlates of acoustic cues of English lexical stress in Cantonese-speaking children

Xiuhong Tong [a], Catherine McBride [a,*], Juan Zhang [b], Kevin K.H. Chung [c], Chia-Ying Lee [d], Lan Shuai [e], Xiuli Tong [f]

[a] Psychology Department, The Chinese University of Hong Kong, Shatin, Hong Kong
[b] Faculty of Education, University of Macau, Macao
[c] Department of Special Education and Counselling, The Hong Kong Institute of Education, Hong Kong
[d] The Institute of Linguistics, Academia Sinica, Taiwan
[e] Department of Electrical and Computer Engineering, Johns Hopkins University, United States
[f] Division of Speech and Hearing Sciences, The University of Hong Kong, Pokfulam, Hong Kong

ARTICLE INFO

Article history:
Accepted 2 September 2014
Available online xxxx

Keywords:
English lexical stress processing
Prosody development
MMN
p-MMR

ABSTRACT

The present study investigated the temporal course of neural discriminations of acoustic cues of English lexical stress (i.e., pitch, intensity and duration) in Cantonese-speaking children. We used an event-related potential (ERP) measure with a multiple-deviant oddball paradigm to record auditory mismatch responses to four deviants, namely, a change in pitch, intensity, or duration, or a change in all three acoustic dimensions, of English lexical stress in familiar words. In the time window of 170–270 ms, we found that the pitch deviant elicited significant positive mismatch responses (p-MMRs) and that the duration deviant elicited a mismatch negativity (MMN) response as compared with the standard. In the time window of 270–400 ms, the intensity deviant elicited a significant p-MMR, whereas both the duration and the three-dimension changed deviants elicited significant MMNs. These results suggest that Cantonese-speaking children are sensitive to either single or convergent acoustic cues of English words, and that the relative weighting of pitch, intensity and duration in stress processing may correlate with different ERP components at different time windows in Cantonese second graders.

1. Introduction

Research on speech has typically focused on how phonetic segments such as vowels and consonants are encoded during speech perception (e.g., Mesgarani, Cheung, Johnson, & Chang, 2014). There has been little work on the discrimination of suprasegmental features of speech, such as lexical stress in English. Lexical stress refers to the relative emphasis or prominence of syllables within words or of words in sentences, such as *PREsent*[1] (\'pre-zənt\; gift) and preSENT (\pri-'zent\) (Fry, 1955, 1958; Selkirk, 1980). Although behavioral research on native English-speaking adults' perception and production of English lexical stress has suggested that stress is acoustically related to pitch (i.e., fundamental frequency [F0]), duration, and intensity (e.g., Crystal, 1969; Kehoe, Stoel-Gammon, & Buder, 1995), the neural correlates of encoding pitch, intensity and duration during English lexical stress processing in children remains poorly understood. In particular, no study has yet examined the neural discriminations of acoustic cues of English lexical stress in children whose first language is a tonal language, such as Cantonese speakers learning English as a second language. In this study, we thus used an event-related potential (ERP) measure to explore neural discriminations of English lexical stress cues (i.e., pitch, intensity and duration) in Cantonese-speaking children. We focused on whether Cantonese-speaking second graders acquiring English as a second language can use these three acoustic cues in English stress perception; and to what extent the weight of each cue varies with unfolding of stress perception in those second graders, as well as what neural markers would be associated with each acoustic cue.

Within the last decades, researchers have become more interested, both theoretically and in empirical work, in stress perception and production in both native and non-native speakers. Empirical evidence on perception of lexical stress in adult native speakers of English suggests that F0, duration, and intensity are the main acoustic correlates of English stress perception (e.g., Fry, 1958; Kehoe et al., 1995; Mol & Uhlenbeck, 1955; Morton & Jassem, 1965). For example, Fry (1958) found that among the three

* Corresponding author.
E-mail address: cmcbride@psy.cuhk.edu.hk (C. McBride).

[1] The capitalized letters represent stress syllables.

http://dx.doi.org/10.1016/j.bandl.2014.09.004
0093-934X/

cues, F0 is the most important cue for English stress perception, followed by duration and intensity. Bolinger (1965) also argued that F0 is the strongest cue in English stress perception, and that both duration and intensity are only secondary. In addition, stressed syllables are characterized as having increased magnitudes of F0, longer duration, and greater intensity relative to unstressed syllables (e.g., Klatt, 1976; Lieberman, 1960).

There is increasing interest in perception of non-native lexical stress contrasts in adult listeners (e.g., Frost, 2011; Peperkamp & Dupoux, 2002; Wang, 2008). For example, Peperkamp and Dupoux (2002) reported that French adult speakers showed stress "deafness" in English lexical stress discrimination, because French is a language with predictable stress while English has an unpredictable stress pattern. In related work, Peperkamp, Vendelin and Dupoux (2010) showed that adult speakers of Standard French, Southeastern French, Finnish, and Hungarian, all of which have fixed stress patterns, had difficulties in perceiving stress contrasts. In contrast, there was no such "stress deafness" found in adult Spanish speakers whose native language has unpredictable stress. Frost (2011) argued that French and English native speakers may not process stress in the same way. These studies of "stress deafness" have been focused on adult speakers of languages with predictable stress such as French versus adult speakers of language with unpredictable stress, such as Spanish. Little is known about whether tone language speakers, whose L1 is a non-stressed language, such as Cantonese, are sensitive to English lexical stress, in particular, to the different acoustic dimensions including pitch, intensity, and duration. Thus, we move one step further by investigating whether Cantonese-speaking children are sensitive to three different acoustic correlates of English lexical stress (i.e., pitch, intensity and duration), and whether a similar order of perceived relative importance (F0–duration–intensity) would be observed in young Cantonese-speaking children.

There have been only a few empirical studies on English lexical stress perception in Chinese[2] learners of English (e.g., Chan, 2007; Wang, 2008). Wang (2008) evaluated the effects of F0, duration, and intensity on English stress perception in Mandarin Chinese learners of English and native English speakers. Results demonstrated that all three cues had a significant influence on English stress perception for native English speakers, but only F0 was found to be important for Chinese learners of English. Similar findings were obtained in adult Cantonese learners of English by Chan (2007) who found that Cantonese speakers used F0 as the primary cue in English stress perception, but the native English speakers used spectral balance (i.e., the distribution of intensity over the frequency spectrum) as the most important cue in stress perception.

The finding that Chinese learners of English rely more on F0 than other acoustic cues indicates some transfer of reliance on F0 from the L1 tonal language to L2 stress (e.g., Nguyen & Ingram, 2005; Pennington & Ellis, 2000). More specifically, perceptual studies of Chinese lexical tone suggest that F0 is the primary acoustic cue for Chinese tone perception (e.g., Khouw & Ciocca, 2007; Vance, 1976). Chinese speakers, therefore, may transfer the strategy in perceiving lexical tone to English stress perception (e.g., Wang, 2008). This possibility of transfer is also supported by studies on English stress production in Chinese speakers, revealing that Chinese speakers may adopt the strategies used in their native tone production task to produce English stress (Zhang, Nissen, & Francis, 2008). For example, Zhang et al. (2008) provided extensive acoustic analyses of English stress production by Mandarin Chinese speakers and English speakers and demonstrated that Mandarin Chinese speakers used the acoustic cues of F0, duration, and intensity in a similar manner as native English speakers in stress production. That is, both Chinese speakers and native English speakers produced stressed syllables with a higher F0, longer duration, and greater intensity than unstressed syllables. These findings suggest that F0, duration, and intensity are all implicated in English stress perception in L2 English learners. Moreover, the relative importance of these acoustic cues in stress perception in L2 learners may be influenced by the native tone languages. Studies of child learners of English lexical stress perception-particularly the more subtle perception of pitch, intensity, and duration-are needed to further explore these possibilities.

Another important issue yet to be examined to date is the neural markers of the acoustic cues of stress in stress perception. In particular, we know of no research that has systematically manipulated the three different acoustic correlates of English lexical stress including F0, intensity and duration and evaluated their effects on perception of English lexical stress in Cantonese-speaking children who are English learners. Therefore, in this study, we adopted an ERP measure to explore the neural discriminations of English stress and further evaluate the relative importance of three acoustic correlates of English stress (i.e., F0, duration, and intensity) in stress perception in Cantonese-speaking second graders acquiring English as a second language.

It is widely known that the ERP measure is an approach with a very fine temporal resolution; it can be used to represent the brain's response to either a passive or an eliciting input. In ERP studies of speech perception, the auditory passive oddball paradigm is often used to examine participants' discriminative ability in speech perception and production with either single or multiple deviants (e.g., for reviews see Cheour, Leppänen, & Kraus, 2000; Näätänen, 2001; Näätänen, Paavilainen, Alho, Reinikainen, & Sams, 1989; Näätänen, Pakarinen, Rinne, & Takegata, 2004). In the passive oddball paradigm, participants are usually presented with a stream of frequent stimuli (standard) and infrequent stimuli differing in some discriminable change (for reviews, see Cheour et al., 2000; Näätänen, Paavilainen, Rinne, & Alho, 2007). A specific ERP component, i.e., the mismatch negativity (MMN), is often observed in this paradigm by subtracting the ERP responses to frequent stimuli (standard) from those of infrequent stimuli (deviant) (e.g., Chandrasekaran, Gandour, & Krishnan, 2007; Cheour et al., 1997; Näätänen et al., 1989). The MMN is found to distribute over the fronto-central electrodes with a peak in the time window of between 150 ms and 250 ms from the change onset of the stimuli in adults, and it reflects automatic, pre-attentive cortical processing. The MMN is suggested to be an indicator of the participant's ability to discriminate between the standard and the deviant; the MMN has been found to become smaller or disappear as the degree of deviance between the standard and deviant is reduced (for a review, see Näätänen et al., 2007). The MMN, which can be obtained irrespective of participants' attention or the behavioral task administered, is a useful tool to use to examine auditory or speech perception in infants and children, who are limited in attention or motivation (e.g., Cheour et al., 2000; Kuhl, 1998; Lee et al., 2012; Morr, Shafer, Kreuzer, & Kurtzberg, 2002). However, previous research has predominantly focused on investigations of the segmental level of speech such as vowels and consonants (for a review, see Näätänen et al., 2007), so less is known about the brain responses to suprasegmental features, such as English lexical stress in Cantonese-speaking children who learn English as a second language.

There have been a few ERP studies on neural discriminations of German stress in German monolinguals (Weber, Hahne, Friedrich, & Friederici, 2004). For example, Weber et al. (2004) used an MMN paradigm to investigate German-speaking adults' and 4- and 5-month-old infants' ERP responses to trochaic (on the first syllable) and iambic (on the second syllable) stress patterns in two-syllable

[2] In the present study the word "Chinese" is used as a blanket term referring to the distinct languages of Mandarin and Cantonese.