Brain & Language 155–156 (2016) 49–57

Contents lists available at ScienceDirect

Brain & Language

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

Cortical encoding and neurophysiological tracking of intensity and pitch cues signaling English stress patterns in native and nonnative speakers

Wei-Lun Chung [a], Gavin M. Bidelman [a,b,*]

[a] School of Communication Sciences and Disorders, University of Memphis, TN 38152, USA
[b] Institute for Intelligent System, University of Memphis, TN 38152, USA

ARTICLE INFO

Article history:
Received 21 October 2015
Revised 4 April 2016
Accepted 6 April 2016
Available online 30 April 2016

Keywords:
Suprasegmental phonology
Prosody
Auditory event-related brain potentials (ERPs)
Mismatch negativity (MMN)
Mandarin Chinese

ABSTRACT

We examined cross-language differences in neural encoding and tracking of intensity and pitch cues signaling English stress patterns. Auditory mismatch negativities (MMNs) were recorded in English and Mandarin listeners in response to contrastive English pseudowords whose primary stress occurred either on the first or second syllable (i.e., "nocTICity" vs. "NOCticity"). The contrastive syllable stress elicited two consecutive MMNs in both language groups, but English speakers demonstrated larger responses to stress patterns than Mandarin speakers. Correlations between the amplitude of ERPs and continuous changes in the running intensity and pitch of speech assessed how well each language group's brain activity tracked these salient acoustic features of lexical stress. We found that English speakers' neural responses tracked intensity changes in speech more closely than Mandarin speakers (higher brain–acoustic correlation). Findings demonstrate more robust and precise processing of English stress (intensity) patterns in early auditory cortical responses of native relative to nonnative speakers.

1. Introduction

Nonnative speakers of English need to accurately perceive and produce word stress patterns for proper, native-like communication. Stress is the relative emphasis (i.e., weighting) that occurs in speech which provides a differential acoustic weighting between at least two syllables. Acoustically, a stressed syllable may have higher fundamental frequency, higher intensity, and longer duration relative to an unstressed syllable (Kehoe, Stoel-Gammon, & Buder, 1995). Listeners attend to these acoustic features signaling English stress patterns and then perceive primary stress based on the relative weightings and interaction between these acoustic cues (Flege & Bohn, 1989). However, nonnative speakers who use a different prosodic system (e.g., lexical tone) in their first language would have difficulties exploring the novel interactions between multi-dimensional acoustic features signaling English stress patterns. For example, Mandarin speakers who use lexical tone varying in fundamental frequency (Howie, 1976) in their native prosodic system might perceive English stress patterns relying on fundamental frequency, but not other acoustic cues (e.g., intensity) non-specific to Mandarin tone. In this study, we aimed to examine whether native and nonnative speakers of American English differed in their *neurophysiological* processing of English stress patterns.

To study how the human brain controls prosodic information, the mismatch negativity (MMN) has been used to examine the pre-attentive detection of violations in legal stress patterns among multiple words (Honbolygó & Csépe, 2013; Honbolygó, Csépe, & Ragó, 2004; Ylinen, Strelnikov, Huotilainen, & Näätänen, 2009). The MMN is advantageous for examining cross-language differences in speech processing as it reflects neural processing related not only to early auditory deviance detection, but also the pre-attentive detection of regularity violations (Winkler, Denham, & Nelken, 2009). MMNs are sensitive to long-term representations of linguistic rules (Näätänen, 2001) and also index the neuroplastic effects of long-term linguistic experience on speech processing (Bidelman & Dexter, 2015; Chandrasekaran, Krishnan, & Gandour, 2007; Näätänen et al., 1997). As such, these cortical responses offer a unique window into the early auditory cortical processing of both acoustic and linguistic information and how language experience modulates brain activity coding important stimulus features.

In more recent years, we have seen mounting evidence of MMNs sensitive to long-term representations of language-specific stress rules. In Finnish and Hungarian, primary stress is always placed on the first syllable of a disyllabic word. The change of stress from the first syllable to the second syllable elicits two consecutive MMNs in disyllabic words in native speakers of Finnish (Ylinen et al., 2009) and Hungarian (Honbolygó & Csépe,

* Corresponding author at: School of Communication Sciences & Disorders, University of Memphis, 4055 North Park Loop, Memphis, TN 38152, USA.
E-mail address: g.bidelman@memphis.edu (G.M. Bidelman).

http://dx.doi.org/10.1016/j.bandl.2016.04.004
0093-934X/

2013; Honbolygó et al., 2004). Such consecutive, multi-response MMNs reflect the fact that native speakers of Finnish and Hungarian could detect shifts in the stress patterns of disyllabic words. It is noteworthy that deviations with illegal stress patterns elicited two successive MMNs, whereas deviations with legal stress patterns did not elicit two MMNs (Honbolygó & Csépe, 2013). This suggests that only deviations with illegal stress patterns violating the long-term representation of language-specific stress rules elicit two consecutive MMNs. Hence, Honbolygó and Csépe (2013) proposed that the process of word stress is dependent on both a short-term memory trace for acoustic speech sounds and a long-term representation of language-specific stress rules (e.g., strong–weak stress patterns dominate Finnish and Hungarian). Additionally, amplitudes of the P2 (~150–250 ms) component of the auditory event-related potentials (ERPs) were enhanced in response to stressed relative to unstressed syllables (Cunillera, Gomila, & Rodríguez-Fornells, 2008; Cunillera, Toro, Sebastián-Gallés, & Rodríguez-Fornells, 2006). Collectively, these findings suggest that word stress might be regarded as a linguistic category or abstract integrated representation of multi-dimensional acoustic features that is captured by serial components of the ERPs including the MMN and P2 responses (Paavilainen, Arajärvi, & Takegata, 2007; Phillips et al., 2000; Saarinen, Paavilainen, Schöger, Tervaniemi, & Näätänen, 1992; Shestakova et al., 2002).

Furthermore, peak amplitudes of MMNs are considered as a measure of discrimination accuracy. Larger MMN amplitudes have been found to be related to listeners' better pre-attentive detection of speech sounds (Kujala, Kallio, Tervaniemi, & Näätänen, 2001; Novitski, Tervaniemi, Huotilainen, & Näätänen, 2004; Näätänen, 2001; Näätänen, Schröger, Karakas, Tervaniemi, & Paavilainen, 1993; Tiitinen, May, Reinikainen, & Näätänen, 1994). Enhanced MMN amplitudes are also observed following intensive training for speech (Kraus et al., 1995; Menning, Imaizumi, Zwitserlood, & Pantev, 2002) and non-speech sounds (Menning, Roberts, & Pantev, 2000; Näätänen et al., 1993). Additionally, MMN is a neurophysiological marker of language-specific sound features. Larger MMN amplitudes have been found in response to pairs of vowel categories in a native language relative to those in a nonnative language (Winkler et al., 1999). Enhanced MMN amplitudes were also observed in Finnish learning voicing contrasts not used in Finnish fricative sounds (Tamminen, Peltola, Kujala, & Näätänen, 2015). It is worth noting that MMN amplitude may also index discrimination accuracy of second language learning. For example, Finnish-speaking L2 learners of French demonstrated enhanced amplitudes of MMNs in response to French sounds during learning (Shestakova, Huotilainen, Ceponiene, & Cheour, 2003). Russian-speaking L2 learners of Finnish showed smaller amplitudes of MMNs in response to Finnish sounds varying in duration than did native Finnish speakers (Nenonen, Shestakova, Huotilainen, & Näätänen, 2003). Taken together, these findings suggest that language experience (even short-term) enhances the brain's automatic processing of relevant stress cues depending on how they are exploited in a given language. Here, we investigated if similar language-specific differences exist between native English and Mandarin listeners' processing of lexical stress, as indexed by the MMN.

In a recent behavioral study, we explored cross-language differences in the perception of primary stress cues between native English and Mandarin-speaking L2 learners of English (Chung & Jarmulowicz, submitted for publication; see also Wade-Woolley & Heggie, 2015). In the task, participants were asked to determine which one of two derived pseudowords varying in primary stress placement (e.g., *NOCticity* versus *nocTICity*) sounded like a real English word. English derived pseudowords with non-neutral suffixes (e.g., *-ity*) require primary stress placed on the syllable before suffixation (e.g., *nocTICity*) to satisfy legal stress patterns (Jarmulowicz, 2016; Jarmulowicz & Taran, 2013). Overall, we found that native English speakers showed better behavioral identification of legal stress patterns compared to Mandarin listeners (Chung & Jarmulowicz, submitted for publication). This suggests that nonnative speakers have greater challenge in using suffix cues to determine which syllable is stressed than do native listeners. Here, we extend these results by examining the neural basis of these cross-language differences in stress processing.

It is conceivable that speakers' native prosodic system might play an influential role in processing nonnative prosodic cues in their L2. For example, native French speakers who use lexical stress in a predictable way have difficulties discriminating stress patterns in pseudowords (Peperkamp & Dupoux, 2002; Peperkamp, Vendelin, & Dupoux, 2010). Frost (2011) further argued that the difficulties native French speakers encountered in English stress perception are attributable to different acoustic features used to signal French and English prosodic systems. By extension, explanation may account for the difficulties Mandarin speakers have in perceiving patterns of stressed and unstressed syllables in English (Chung & Jarmulowicz, submitted for publication). English and Mandarin prosodic systems share a common acoustic feature: *pitch*—the psychological correlate of fundamental frequency. Pitch is one of three acoustic features representing English stress patterns (Kehoe et al., 1995), and the primary acoustic feature for Mandarin tone perception (Howie, 1976). Several behavioral studies demonstrate that native Mandarin speakers tend to use pitch as a cue for perceptually distinguishing (Ou, 2010; Yu & Andruski, 2010) and producing English stress patterns (Zhang, Nissen, & Francis, 2008). This suggests a "prosodic transfer" whereby tone language speakers—who use pitch cues more predominantly in their native language—might rely more heavily on pitch-based cues in a stress/rhythmic language like English (Elder, Golombek, Nguyen, & Ingram, 2005; Pennington & Ellis, 2000). On the other hand, continuous variations in intensity appear to be critical for accurate stress perception (Goswami & Leong, 2013) and intensity may be a more important acoustic feature signaling English stress relative to pitch cues (Choi, Hasegawa-Johnson, & Cole, 2005; Kochanski, Grabe, Coleman, & Rosner, 2005). Hence, Mandarin speakers who use pitch predominately in their first language might be less efficient in perceiving English stress patterns based on intensity cues. Indeed, behavioral studies have shown that intensity is a less reliable cue for Mandarin listener's perception of English stress (Chrabaszcz, Winn, Lin, & Idsardi, 2014). Here, we hypothesized that Mandarin listeners might show poorer neural encoding and tracking of ongoing variations in the intensity envelope of English (L2) speech given the lesser importance of this cue in their native language (Mandarin).

The aim of the current study was to examine cross-language differences in the early auditory cortical processing of English stress patterns between native (English) and nonnative (Mandarin) speakers (as indexed by the MMN). To this end, we recorded mismatch negativity potentials in English and Mandarin listeners in response to English pseudowords that included occasional violations in primary stress placement. We used pseudowords to remove the lexical–semantic meaning from speech stimuli and thus, examine stress-related brain processing in the absence of lexical information, *per se*. The presence of two consecutive MMNs would provide further evidence that word stress is a linguistic category or abstract integrated representation of multi-dimensional acoustic features (Paavilainen et al., 2007; Phillips et al., 2000; Saarinen et al., 1992; Shestakova et al., 2002). In addition, we hypothesized that native English speakers would show superior cortical encoding and neurophysiological tracking of the running intensity profile of speech (i.e., amplitude envelope) compared to nonnative speakers. These findings would support the notion that nonnative listeners' poorer sensitivity to English stress patterns