

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



Wei-Lun Chung<sup>a</sup>, Gavin M. Bidelman<sup>a,b,\*</sup>

<sup>a</sup> School of Communication Sciences and Disorders, University of Memphis, TN 38152, USA

<sup>b</sup> 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

### 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.

© 2016 Elsevier Inc. All rights reserved.

## 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, Streltnikov, Huottilainen, & 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](mailto:g.bidelman@memphis.edu) (G.M. Bidelman).

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 predominantly 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 (Chrabaszc, 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

observed behaviorally (Archibald, 1997; Chung & Jarmulowicz, submitted for publication) result from an impoverished long-term representation, weaker neural encoding and/or tracking of intensity variations in English, their second language.

## 2. Materials and methods

### 2.1. Participants

Ten adult monolingual speakers of American English (2 males) and ten adult native speakers of Mandarin Chinese (3 males) participated in the experiment. Participants were closely matched in age (English:  $M = 27.97$ ,  $SD = 3.60$  years; Mandarin:  $M = 29.89$ ,  $SD = 4.37$  years), years of formal education (English:  $M = 18.7$  years,  $SD = 2.53$ ; Mandarin:  $M = 19.0$ ,  $SD = 2.23$  years), and were strongly right handed ( $\geq 84\%$ ) as measured by the Edinburgh Handedness Inventory (Oldfield, 1971). Participants reported no history of hearing, speech, language, or neuropsychiatric disorders. Each participant also completed a language history questionnaire (Li, Sepanski, & Zhao, 2006). Consistent with our previous cross-language studies (Bidelman & Dexter, 2015; Bidelman & Lee, 2015), Mandarin participants were characterized as late bilinguals who spoke Mandarin as their first language and began learning English in childhood. They were born and raised in China or Taiwan, and their onset age of English (L2) instruction was age  $\sim 11$  ( $M = 11.9$ ,  $SD = 1.37$  years). According to the Common European Framework of Reference for Language: Learning, Teaching, Assessment (CEFR; Council of Europe, 2001), Mandarin speakers were classified as independent users with B2 upper-intermediate-level English proficiency because they needed to pass the B2 threshold to meet admissions requirements as international students in the USA. Participants gave written informed consent in compliance with a protocol approved by the University of Memphis Institutional Review Board and received a monetary compensation for their participation.

### 2.2. Stimuli

Stimuli consisted of audio recordings of two English pseudowords adopted from Wade-Woolley and Heggie (2015). Pseudowords were extracted from two sentences produced by a female native English speaker. Pseudowords were used to control lexical-semantic meaning. This allowed us to isolate neural mechanisms subserving stress-related processing and ensure stimuli were lexically neutral for both groups. The standard pseudoword (*nocTICity*) consisted of a legal stress pattern whereas the deviant pseudoword (*NOcticity*) contained an illegal stress pattern. The two tokens differed in the change of primary stress placement from the second to the first syllable, which we expected to elicit two consecutive MMNs (Honbolygó & Csépe, 2013; Honbolygó et al., 2004; Ylinen et al., 2009). The acoustic properties of the pseudoword stimuli are shown in Figure 1.

The two audio recordings were converted to mono sound files and were matched in sampling rate (48,828 Hz), duration (828 ms), and overall RMS amplitude. While overall duration was closely matched, our stimuli were natural productions. Consequently, the first and second syllable durations differed slightly between tokens (*nocTICity*: 245 ms and 188 ms; *NOcticity*: 292 ms and 167 ms), maximal amplitude (*nocTICity*: 80.22 dB and 80.71 dB; *NOcticity*: 84.07 dB and 70.65 dB), and maximal pitch (*nocTICity*: 218 Hz and 232 Hz; *NOcticity*: 196 Hz and 81 Hz). Nevertheless, the most salient acoustic cue differentiating the two tokens' stress was amplitude variations (Fig. 1). Mismatch responses were recorded in a passive auditory oddball paradigm (standard: *nocTICity*; oddball (deviant): *NOcticity*) with an inter-

stimulus interval of 750 ms and standard/deviant ratio of 85/15%, respectively. Standards (680 trials) and deviants (120 trials) were presented according to a pseudo-random schedule such that at least two standard tokens intervened between subsequent deviants (e.g., Bidelman & Dexter, 2015).

### 2.3. Data acquisition and preprocessing

Electrophysiological recordings followed typical procedures used in our laboratory (Bidelman, 2015; Bidelman & Grall, 2014). Participants reclined comfortably in an electro-acoustically shielded booth to facilitate recording of neurophysiologic responses. They were instructed to relax and refrain from extraneous body movement (to minimize myogenic artifacts), ignore the sounds they hear (to divert attention away from the auditory stimuli), and were allowed to watch a muted subtitled movie to maintain a calm yet wakeful state. MMNs were recorded under a passive listening paradigm for  $\sim 30$ –45 min in order to collect responses to the 800 total stimulus trials. An additional 1.5 h of EEG data were acquired for experiments examining pitch processing of tonal sequences and are reported in our previous report (Bidelman & Chung, 2015). Stimulus presentation was controlled by a MATLAB routed to a TDT RP2 interface (Tucker-Davis Technologies) and delivered binaurally at an intensity of 82 dB SPL through insert earphones (ER-2; Etymotic Research). Stimulus intensity was calibrated using a Larson-Davis SPL meter (Model LxT) measured in a 2-cc coupler (IEC 60126).

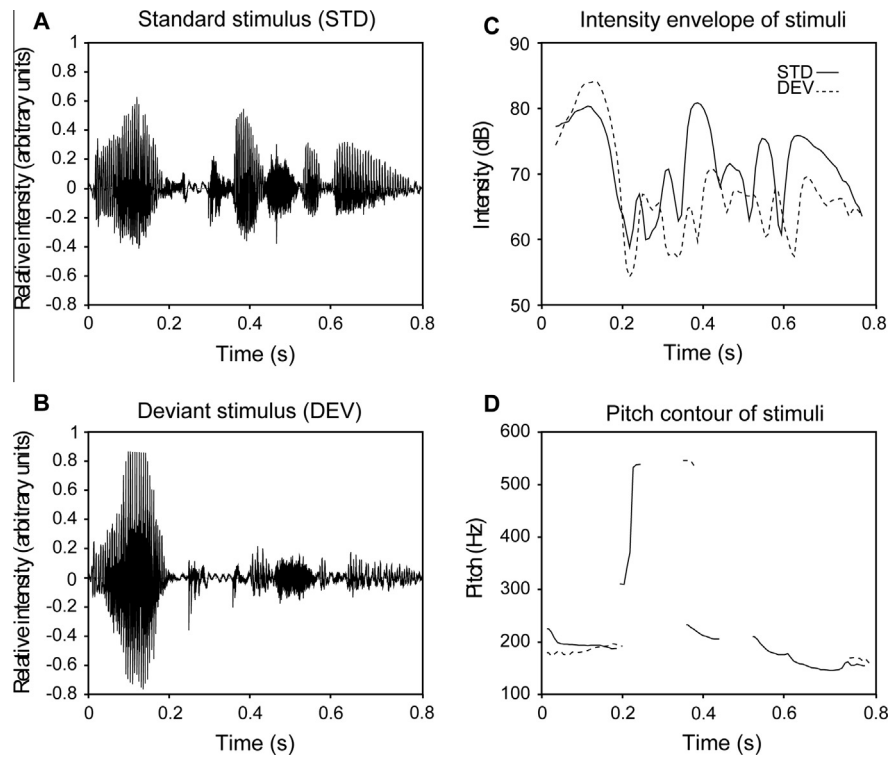
Neuroelectric activity was recorded from 64 electrodes at standard 10–10 locations around the scalp (Oostenveld & Praamstra, 2001). EEGs were digitized using a sampling rate of 500 Hz (SynAmps RT amplifiers; Compumedics Neuroscan) using an online passband of DC–200 Hz. Responses were then stored to disk for offline analysis. Electrodes placed on the outer canthi of the eyes and the superior and inferior orbit were used to monitor ocular activity. During online acquisition, all electrodes were referenced to an additional sensor placed  $\sim 1$  cm posterior to Cz. However, data were re-referenced off-line to a common average reference of all channels. Contact impedances were maintained  $< 10$  k $\Omega$  throughout the duration of the experiment.

Subsequent preprocessing was performed in Curry 7 (Compumedics Neuroscan) and custom routines coded in MATLAB<sup>®</sup> 2014b (The MathWorks, Inc.). Data visualization and scalp topographies were computed using EEG/ERPLAB (Delorme & Makeig, 2004; Lopez-Calderon & Luck, 2014). Ocular artifacts (saccades and blink artifacts) were then corrected in the continuous EEG using a principal component analysis (PCA) (Wallstrom, Kass, Miller, Cohn, & Fox, 2004). The PCA decomposition provided a set of independent components which best explained the topography of the blink/saccadic artifacts. The scalp projection of the first two PCA loadings was subtracted from the continuous EEG traces to nullify ocular contamination in the final ERPs. Cleaned EEGs were then digitally filtered (1–20 Hz; zero-phase), epoched ( $-200$  to 1200 ms), baseline-corrected to the pre-stimulus period, and subsequently averaged in the time domain to obtain ERPs for each stress condition per participant.

### 2.4. ERP response analyses

#### 2.4.1. Mismatch negativities (MMNs)

MMNs for English stress patterns in derived pseudowords were computed as the difference between ERPs recorded for the deviant pseudowords *NOcticity* and those recorded for the standard pseudoword *nocTICity* (i.e., deviant – standard). The global field power (GFP) was then calculated for each group's MMN waveforms (Lehmann & Skrandies, 1980). GFP quantifies the overall amount of neuroelectric activity at each time sample from the aggregate

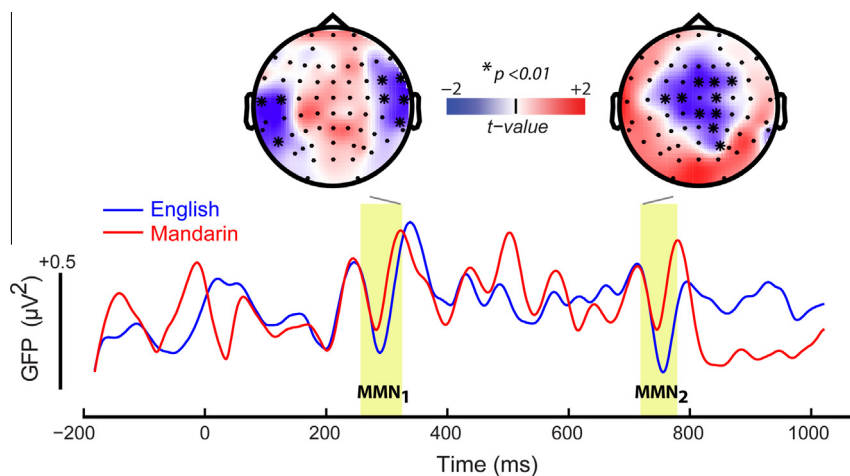


**Fig. 1.** Acoustic characteristics of pseudoword stimuli. Sound waveforms for the standard (A) and deviant (B) speech tokens with primary stress on the second and first syllable. (C) Intensity envelope and (D) pitch (i.e., fundamental frequency) contours of standard (solid lines) and deviant (dotted) tokens.

multichannel evoked potential recordings and allowed us to assess group differences in MMN without bias of selecting particular electrode(s). Two MMNs were apparent in listeners' responses corresponding to the changes in stress at the first and second syllables of our stimuli. Thus, MMNs were analyzed in two separate 40-ms time windows, computed as the maximum deflection in the GFP at the first and second MMN waves (MMN<sub>1</sub>: ~250 ms; MMN<sub>2</sub>: ~750 ms) (see Fig. 2).

Within the two latency windows (MMN<sub>1</sub> and MMN<sub>2</sub>), mismatch responses were compared between language groups via a clustered

permutation test (Maris & Oostenveld, 2007) using the FieldTrip toolbox (Oostenveld, Fries, Maris, & Schoffelen, 2011). For each electrode and time sample, an independent samples *t*-test was conducted contrasting English and Mandarin MMN responses. All samples exceeding a significance threshold of  $p = 0.01$  were then clustered on the basis of their temporal and spatial proximity. Significant differences between groups were then determined by generating and comparing surrogate clusters from 2500 resamples permuting between groups via Monte Carlo simulation (Oostenveld et al., 2011). Importantly, this cluster-based approach



**Fig. 2.** Cross-language differences in the neural differentiation of stress. Shown here are MMN responses evoked by changes in stress between the first and second syllable of English pseudowords (see Fig. 1). Grand mean MMN global field power (GFP) time waveforms show the overall neural differentiation of speech in each language group. Two distinct mismatch responses are visible corresponding to changes in stress at the first and second syllable, MMN<sub>1</sub> and MMN<sub>2</sub> respectively. (top) Scalp topographies (*t*-value maps) show the electrode sites that differentiate groups based on a clustered permutation test (Maris and Oostenveld, 2007). Red colors, Mandarin > English; blue colors, English > Mandarin. Starred electrodes (\*) denote channel clusters that show a significant difference between groups ( $p < 0.01$ ). English listeners show stronger (i.e., more negative) MMN responses at temporal electrode sites for the first syllable and at fronto-central sites for the second syllable.

corrects for multiple comparisons across the aggregate of all channel  $\times$  time ( $64 \times 701 = 44,864$ ) contrasts by controlling the family-wise error rate. Contiguous clusters at the  $p < 0.01$  level were deemed significant, i.e., scalp locations that distinguished each groups' ability to differentiate lexical stress patterns.

#### 2.4.2. P1–N1–P2 responses to standards

Subsequent ERP analysis focused on responses to standard tokens as they contained a more substantial number of trials (680) and hence a more favorable signal-to-noise ratio (SNR) than deviant and MMN waveforms. Analysis of standard responses also allowed us to further investigate where group differences in stress encoding appeared with regard to the obligatory waves of the auditory cortical ERPs (i.e., P1–N1–P2). For this analysis, peak amplitudes were measured for the prominent deflections of the ERPs (P1, N1, P2) per language group. Each syllable of the speech stimuli elicited a corresponding P1–N1–P2 signature “wavelet” (see Fig. 3). Therefore, we quantified P1 in three different time windows (denoted P1a, P1b, P1c), computed as the peak positive deflection within a 20 ms search window centered at  $\sim 65$ , 275, and 500 ms, respectively. Window times were chosen based on visual inspection of the grand mean ERPs and based on latency conventions of the auditory ERPs (Bidelman & Lee, 2015). Similarly, N1 and P2 were quantified for each syllable and group (N1a,b,c:  $\sim 100$ , 330, 550 ms; P2a,b,c:  $\sim 160$ , 430, 600 ms). Auditory cortical responses are maximal at the vertex and carry a prominent frontocentral scalp distribution (Picton et al., 1999; Woods, 1995). Hence, for data reduction purposes, P1–N1–P2 responses were quantified from the average potential at the Cz and FCz electrode sites (see Fig. 3).

#### 2.4.3. “Neural tracking” of speech amplitude envelope and pitch profiles

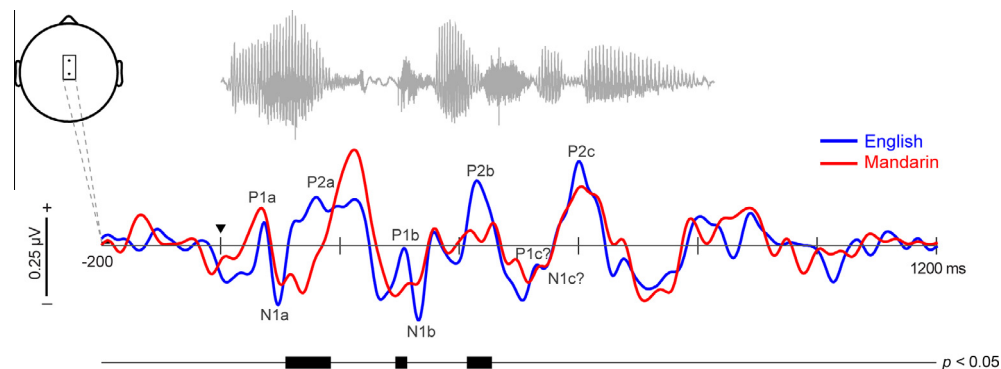
We next aimed to identify if native English speakers showed superior encoding of the ongoing stress patterns in English speech. To this end, we analyzed the degree to which each group's standard ERPs “tracked” changes in the running amplitude envelope of speech via cross-correlational analysis (Aiken & Picton, 2008). Acoustic envelopes were first extracted from the standard stimulus waveform (*noCTICity*) as the log magnitude of the Hilbert envelope (for details, see Fig. 1 of Aiken & Picton, 2006). The acoustic envelope was then lowpass filtered (1000 order, zero-phase FIR) at 20 Hz and decimated to match the bandwidth and sampling rate of the neural response waveforms. The stimulus and response amplitude profiles were then cross-correlated and the maximum cross-correlation between the acoustic amplitude envelope and

ERPs provided an estimate of how well each group's cortical responses tracked changes in the stimulus stress pattern (Aiken & Picton, 2008; Golumbic et al., 2013). Cross-correlations were used to account for the latency delay between the onset of the acoustic speech signal and the onset of the ERPs (Bidelman, Villafuerte, Moreno, & Alain, 2014; Galbraith & Brown, 1990). These stimulus-to-response cross-correlation coefficients were then compared between groups using an independent samples  $t$ -test. One-tailed tests were used given extensive prior work demonstrating cross-language differences in auditory processing between Mandarin and English listeners (Bidelman, Gandour, & Krishnan, 2011; Bidelman, Krishnan, & Gandour, 2011; Bidelman & Lee, 2015; Chandrasekaran, Krishnan, & Gandour, 2009; Krishnan, Gandour, & Bidelman, 2010; Krishnan, Gandour, & Suresh, 2015).

English stress consists of multidimensional features (e.g., intensity, pitch, duration). While our stimuli varied in stress primarily based on their intensity, pseudowords also contained some pitch variation (see Fig. 1). Additionally, previous ERP studies have shown superior pitch tracking in native speakers of tonal languages in the form of more faithful brainstem frequency-following responses to dynamic pitch contours (Bidelman, Gandour et al., 2011; Krishnan et al., 2010). Hence, we also investigated the degree to which modulations in ERP amplitudes varied with changes in the pitch patterns of pseudowords. To this end, stimulus pitch tracks (i.e., fundamental frequency contours) (Bidelman, Gandour, et al., 2011) were extracted from the standard speech waveform using an autocorrelation based pitch tracking algorithm as implemented in Praat (Boersma, 1993). Briefly, the algorithm works by sliding a 40 ms window in 10 ms increments over the time course of the stimulus. The autocorrelation function was computed for each 40 ms frame and the time lag corresponding to the maximum autocorrelation value within each frame was recorded. The reciprocal of this time lag (or pitch period) represents an estimate of the stimulus fundamental frequency ( $F_0$ ). The time lags associated with autocorrelation peaks from each frame were concatenated together to give a running  $F_0$ -pitch contour (see Fig. 1D). As with intensity profiles, stimulus-to-response cross-correlation coefficients were used to assess the degree to which changes in the running ERP amplitude reflected variations in pitch prosody.

### 3. Results

Changes in the stress of English pseudowords from the second syllable to the first syllable elicited two consecutive MMN peaks which were prominent in the GFP waveforms (Fig. 2). Generally



**Fig. 3.** Grand average ERP responses to standard pseudowords with legal stress patterns. ERPs represent the average potential at a fronto-central electrode cluster (mean Cz and FCz; head inset). Arrow (▼) shows the onset of the time-locking speech stimulus. The speech waveform (gray) is shown above the neural traces for visual reference. Speech stimuli elicit three distinct neural signatures (i.e., P1–N1–P2 complex), corresponding to the repeated generation of the auditory cortical response to each consecutive syllable (labeled a–c). A permutation randomization test revealed group differences in the neural encoding of speech primarily in the latency range of the P2 wave (■ segments;  $p < 0.05$ ) during both the first and second syllables.

speaking, English speakers demonstrated larger MMNs to English stress patterns than did Mandarin speakers. Clustered based permutation tests (Maris & Oostenveld, 2007) revealed that these group differences ( $p < 0.01$ ) were attributable to English listeners' stronger MMNs at bilateral temporal scalp sites for the first ( $MMN_1$ ) and fronto-central locations for the second syllable ( $MMN_2$ ). More robust mismatch responses in the English group across both syllables indicate that native listeners show superior differentiation of the stress patterns in English speech relative to nonnative Mandarin listeners.

Having established that English listeners' cortical responses better differentiate stress patterns in speech, we next aimed to characterize if this group also showed superior encoding of English pseudowords. To this end, we analyzed ERP responses to standard speech tokens given the larger number of trials in this condition and hence more favorable SNR compared to participants' MMNs. ERPs to standard pseudowords with legal stress patterns are shown per language group in Fig. 3. Three wavelets (i.e., P1–N1–P2) were generated corresponding to the obligatory auditory cortical response repeatedly evoked at each consecutive syllable. In the English group, the P2a showed an extended/sustained response with quasi-bifid peaks occurring at 160 ms and another at ~220 ms. We have observed a similar sustained, bifid P2 in speech processing tasks in our previous studies (see Fig. 2f). However, it appears only the latter component (at 220 ms) was present in the Mandarin group. This wave is likely too early to be considered a traditional P300 response and was not observed consistently in all subjects. We posit that this response may reflect individual differences in post-perceptual demands of processing difficult speech material in the non-native listeners. Also apparent from Fig. 3, both the P2b and P2c appeared larger in English compared to Mandarin listeners. An initial permutation test between English and Mandarin responses ( $N = 200$  resamples) revealed group differences in speech encoding primarily in the latency range of the P2 responses (dark bands, Fig. 3).

This was confirmed by specific analysis of the P2 waves (Fig. 4). A two-way mixed model ANOVA (subjects as a random factor; group and wave number as fixed factors) revealed stronger encoding of stress in English speakers' P2 relative to Mandarin speakers [ $F_{1,18} = 6.09$ ,  $p = 0.024$ ] (Fig. 4C). The main effect of group with no group \* wave number interaction [ $F_{2,36} = 0.70$ ,  $p = 0.50$ ] indicates that English speakers had larger P2 amplitude across the board. No reliable group differences were observed for the more variable and less visible P1 [ $F_{1,18} = 0.56$ ,  $p = 0.47$ ] (Fig. 4A) and N1 deflections [ $F_{1,18} = 0.05$ ,  $p = 0.83$ ] (Fig. 4B). Collectively, these analyses reveal stronger encoding (raw ERP) and differentiation (MMN) of

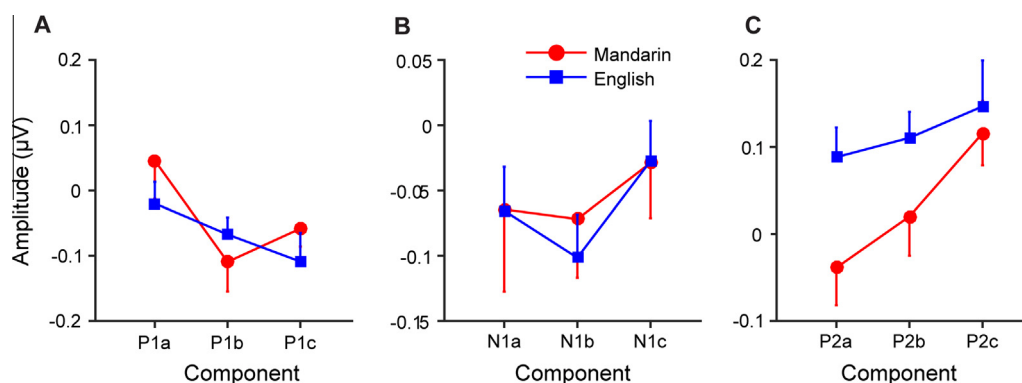
English-like stress patterns in native English listeners beginning ~150 ms after the onset of the speech stimulus.

The previous analyses focused on specific temporal windows that showed group differences in stress processing. To extend these results and determine if English listeners also show improved "neural tracking" of the running speech envelope, we compared modulations in ERP responses to the stimulus amplitude (e.g., Aiken & Picton, 2008). Stimulus-to-response envelope vs. ERP and pitch vs. ERP correlations are shown in Fig. 5 for each language group. Comparisons of cross-correlation coefficients revealed stronger stimulus–response correlations for intensity tracking in English listeners [ $t_{18} = 1.76$ ,  $p = 0.04$ ] (Fig. 5A). That is, English listeners' cortical responses tracked the continuous changes in the speech envelope more faithfully than in Mandarin speakers. As expected from the relative invariance in pitch of our stimuli (see Fig. 1), we found no group differences in terms of ERP pitch tracking [ $t_{18} = 0.52$ ,  $p = 0.30$ ] (Fig. 5B). Taken alongside our other analyses (MMN and P2 magnitudes), these results demonstrate that English speakers' early auditory cortical responses track stress patterns in speech based on intensity (but not pitch) cues with higher fidelity than in nonnative speakers. The fact these group effects are observed even during passive listening and for pseudowords indicates that superior stress processing in English listeners (i) occurs relatively automatically and (ii) in the absence of lexical–semantic meaning.

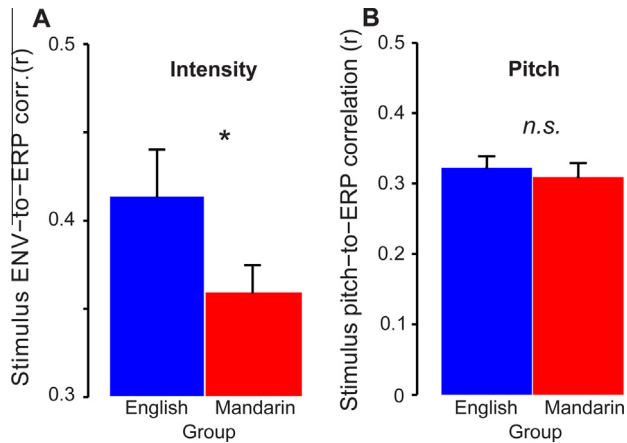
#### 4. Discussion

While considerable attention has been paid to the pre-attentive processing of stress, how native and nonnative speakers process aspects of English prosody in the brain has been largely unexplored. English is a stress/rhythmic language with variations in fundamental frequency, intensity, and duration (Kehoe et al., 1995), whereas Mandarin is a tone language dominated by prosodic variations in fundamental frequency rather than intensity (Howie, 1976). The present study compared the cortical processing of English-like stress patterns in derived pseudowords in English and Mandarin speakers. Our data reveal experience-dependent effects of native English experience on the neural encoding and tracking of intensity-based stress cues. Compared to Mandarin speakers, English listeners showed more robust encoding and faithful tracking of the amplitude envelope of running speech, consistent with their more pronounced experience with intensity-based stress cues in their native language.

In line with previous studies (Honbolygó & Csépe, 2013; Honbolygó et al., 2004; Ylinen et al., 2009), we found that a change



**Fig. 4.** Differences in P1, N1, and P2 component amplitude across speech syllables and language groups. No group differences were observed in the more variable (A) P1 and (B) N1 responses. (C) P2 was larger for English (■) compared to Mandarin (●) listeners indicating superior encoding of English pseudowords in native speakers ~150 ms after the stress event. errorbars =  $\pm 1$  s.e.m.



**Fig. 5.** English listeners show superior “neural tracking” of continuous stress patterns in speech. (A) Stimulus-to-response cross-correlation coefficients between the envelope of the stimulus waveform (standard token) and amplitude profile of the ERP response. (B) Correlations between the running pitch prosody and ERPs. Higher correlations denote that modulations in ERP amplitude more faithfully follow changes in the ongoing stimulus acoustics. English listeners’ cortical ERPs show superior tracking of the speech envelope relative to nonnative Mandarin listeners for intensity-based stress cues; no group differences are observed for pitch tracking suggesting that modulations in the cortical ERPs largely reflect changes in the amplitude profile of speech.  $p < 0.05$ , errorbars =  $\pm 1$  s.e.m.

of stress between syllables in derived pseudowords elicited two consecutive MMN responses in both English and Mandarin speakers. These findings indicate that cortical responses differentiated rhythmic patterns between standard and deviant speech tokens (i.e., *noCTiCity* vs. *NOcticity*) regardless of language group and corroborate the proposition that word stress is processed as a linguistic category represented by multi-dimensional acoustic features (Paavilainen et al., 2007; Phillips et al., 2000; Saarinen et al., 1992; Shestakova et al., 2002). Furthermore, we investigated the effect of language experience on the MMN components to understand the different functional properties of serial mismatch responses. Compared with Mandarin speakers, English speakers showed stronger responses in bilateral temporal scalp sites for the first MMN and fronto-central locations for the second MMN (Fig. 2). Tentatively, this change from temporal to more frontal MMN generators may reflect an early differentiation of stress patterns first based on short-term sensory memory traces from initial acoustic analysis (temporal scalp sites) and the latter recruitment of long-term neural representations of language-specific stress rules (fronto-central location) (Honbolygó & Csépe, 2013). This notion is further supported by our recent MMN study which demonstrated a differential recruitment of inferior frontal compared to temporal brain regions during degraded speech recognition in native compared to nonnative listeners (Bidelman & Dexter, 2015); higher-order brain regions are engaged only when speech sounds are linguistically relevant to the listener.

Compared to native English speakers, Mandarin speakers showed smaller MMN amplitudes elicited by a change in primary stress placement across multiple syllables (Fig. 2). Honbolygó and Csépe (2013) proposed that two consecutive MMNs are attributable to a short-term memory trace for acoustic information and a long-term representation of language-specific stress rules. This suggests that Mandarin speakers in the current study had poorer processing and an impoverished long-term representation of the stress patterns in English. Our electrophysiological findings here corroborate recent behavioral studies which demonstrate that Mandarin speakers are less proficient in exploiting suffix cues for English stress perception (Chung & Jarmulowicz, submitted for publication). We observed these group differences in relatively

early auditory cortical processing (MMN) suggesting that English speakers’ superior encoding of suffix cues occurs relatively automatically, and well before the integration with lexical-semantic knowledge representations as might be the case for Mandarin listeners (Archibald, 1997; Ullman, 2005). Such an explanation may account in part for the weaker MMN in Mandarin speakers (nonnative) relative to English (native) speakers.

Another possible explanation is how English and Mandarin speakers differ in exploiting acoustic cues to perceive English stress patterns. Indeed, Mandarin speakers demonstrated poorer processing of English stress patterns relative to native English speakers as evidenced by (i) smaller P2 amplitudes in response to primary stress events (i.e., the standard stimulus *noCTiCity* with a legal stress pattern) (Figs. 3 and 4) and (ii) smaller brain-acoustic correlations between ongoing amplitudes of ERP response and speech amplitude envelope. The sensitivity of the P2 to primary stress placement in the standard pseudoword supports the findings of previous studies that have shown that this component tags properties of word stress (Cunillera et al., 2006; Cunillera et al., 2008) and phonetic features of speech (Bidelman & Lee, 2015; Bidelman, Moreno, & Alain, 2013). Although both groups showed reliable encoding of English stress (indexed by P2), the weaker ERP amplitudes in the Mandarin group overall further indicates language dependent tuning of stress-related processing. This suggests that Mandarin speakers are not as sensitive to primary stress placement relative to native English speakers.

Additionally, brain-acoustic (Fig. 5) correlations demonstrated a more robust correspondence between the stimulus amplitude envelope and cortical activity in English relative to Mandarin speakers. In other words, Mandarin speakers’ responses were poorer at following changes in ongoing intensity cues that are critical for distinguishing the stress patterns in English speech. Intensity cues play a less prominent role in Mandarin prosody compared to English, where amplitude variations are more salient cue to signal stress. We infer that the stronger encoding and tracking of speech stress based on intensity is the result of the fact that this cue is a more perceptually salient to native speakers of English. Hence, the poorer neurophysiological encoding and tracking of the speech envelope may inhibit Mandarin speakers from fully distinguishing one word from another based on intensity variations. Consequently, our neuroimaging data may at least partially account for Mandarin listener’s difficulties in exploiting intensity cues for English stress perception observed behaviorally (Chrabaszcz et al., 2014).

Even though this study showed English and Mandarin speakers differed in neurophysiological tracking of English stress patterns, it has some limitations. First, durations of the first and second syllables between the standard and deviant stimuli were not identical. Hence, while stimuli differed in stress primarily based on amplitude cues (Fig. 1), we cannot rule out the possibility that at least some portion of the current MMNs reflect deviance processing for other acoustic correlates of stress (e.g., duration, voice quality). Future studies employing synthetic speech sounds are needed to address the independent role of amplitude and other acoustic cues (fundamental frequency, duration) in stress processing. Second, the current study employed only one pair of standard and deviant stress stimuli. While our stimulus design attempted to parallel previous studies (e.g., Honbolygó & Csépe, 2013), several tokens could be used to investigate natural variation for each word and examine how current findings generalize, for example, across different speakers. A more expansive stimulus set could be fruitful as previous work has shown that natural variability in standard items leads to a more reliable abstraction and salient auditory memory trace for acoustic stimuli (Phillips et al., 2000). Nevertheless, we observed robust group differences even in standard tokens. Thus, even if MMNs in the current study were in fact multidimensional,

it is clear that both deviance detection and ongoing encoding of lexical stress properties differs according to listeners' language background (Nenonen et al., 2003; Shestakova et al., 2003).

Future research could also examine the cortical processing of prosodic stimuli from a number of additional perspectives of prosody not examined in the present study. Several behavioral studies have revealed that Mandarin- and English-speaking children with reading disabilities (e.g., dyslexia) are impaired in prosodic perception compared to typically developing children (Goswami, Gerson, & Astruc, 2010; Goswami et al., 2013; Wang, Huss, Hämäläinen, & Goswami, 2012). In light of these previous suggestions and our current findings, it would be worthwhile to examine (1) whether children with dyslexia are more impaired in cortical processing of prosodic patterns relative to a control group and (2) which acoustic features are less reliable cues for prosodic perception in children with dyslexia. Lastly, the current study found that the pre-attentive detection of English stress patterns may involve a short-term memory trace for acoustic comparisons and a long-term representation of language-specific stress rules. Hence, an interesting avenue for future investigation might be to further examine the differential roles of acoustic and language-specific analyses in the perception of lexical stress.

In summary, findings of the current study indicate that both English (native) and Mandarin (nonnative) speakers tag English stress patterns at early cortical levels of auditory processing as evidenced by two consecutive MMNs elicited by a change in stress between adjacent word syllables. However, Mandarin speakers demonstrate poorer pre-attentive detection of English stress patterns than English speakers as revealed by less robust encoding (P2) of primary stress placement and poorer neural tracking of the speech amplitude envelope. Our ERP data are consistent with the notion that nonnative (Mandarin) listeners' poorer sensitivity to English stress patterns observed behaviorally (Archibald, 1997; Chung & Jarmulowicz, submitted for publication) result from (1) an impoverished long-term representation of the interaction between English derivational suffixes and primary stress placement, (2) a weaker neural encoding (P2) of primary stress placement, and (3) a poorer neural tracking of ongoing intensity variations in English, their second language. The paradigm used here might serve as an avenue for future research examining the cortical processing of prosody across normal and clinical populations.

## References

- Aiken, S. J., & Picton, T. W. (2006). Envelope following responses to natural vowels. *Audiology & Neuro-Otology*, 11(4), 213–232.
- Aiken, S. J., & Picton, T. W. (2008). Envelope and spectral frequency-following responses to vowel sounds. *Hearing Research*, 245(1–2), 35–47.
- Archibald, J. (1997). The acquisition of English stress by speakers of tone languages: Lexical storage versus computation. *Linguistics*, 35, 167–181.
- Bidelman, G. M. (2015). Multichannel recordings of the human brainstem frequency-following response: Scalp topography, source generators, and distinctions from the transient ABR. *Hearing Research*, 323, 68–80.
- Bidelman, G. M., & Chung, W.-L. (2015). Tone-language speakers show hemispheric specialization and differential cortical processing of contour and interval cues for pitch. *Neuroscience*, 305, 384–392.
- Bidelman, G. M., & Dexter, L. (2015). Bilinguals at the “cocktail party”: Dissociable neural activity in auditory-linguistic brain regions reveals neurobiological basis for nonnative listeners' speech-in-noise recognition deficits. *Brain and Language*, 143, 32–41.
- Bidelman, G. M., Gandour, J. T., & Krishnan, A. (2011). Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *Journal of Cognitive Neuroscience*, 23(2), 425–434.
- Bidelman, G. M., & Grall, J. (2014). Functional organization for musical consonance and tonal pitch hierarchy in human auditory cortex. *NeuroImage*, 101, 204–214.
- Bidelman, G. M., Krishnan, A., & Gandour, J. T. (2011). Enhanced brainstem encoding predicts musicians' perceptual advantages with pitch. *The European Journal of Neuroscience*, 33(3), 530–538.
- Bidelman, G. M., & Lee, C.-C. (2015). Effects of language experience and stimulus context on the neural organization and categorical perception of speech. *NeuroImage*, 120, 191–200.
- Bidelman, G. M., Moreno, S., & Alain, C. (2013). Tracing the emergence of categorical speech perception in the human auditory system. *NeuroImage*, 79, 201–212.
- Bidelman, G. M., Villafuerte, J. W., Moreno, S., & Alain, C. (2014). Age-related changes in the subcortical-cortical encoding and categorical perception of speech. *Neurobiology of Aging*, 35(11), 2526–2540.
- Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proceedings of the Institute of Phonetic Sciences*, 17(1193), 97–110.
- Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2007). Mismatch negativity to pitch contours is influenced by language experience. *Brain Research*, 1128(1), 148–156.
- Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2009). Relative influence of musical and linguistic experience on early cortical processing of pitch contours. *Brain and Language*, 108(1), 1–9.
- Choi, J.-Y., Hasegawa-Johnson, M., & Cole, J. (2005). Finding intonational boundaries using acoustic cues related to the voice source. *The Journal of the Acoustical Society of America*, 118(4), 2579–2587.
- Chrabaszcz, A., Winn, M., Lin, C. Y., & Idsardi, W. J. (2014). Acoustic cues to perception of word stress by English, Mandarin, and Russian Speakers. *Journal of Speech, Language, and Hearing Research: JSLHR*, 57(4), 1468–1479.
- Chung, W.-L., & Jarmulowicz, L. (submitted for publication). Stress perception and production in English derivation, and word reading in adult Mandarin speaking English learners.
- Cunillera, T., Gomila, A., & Rodríguez-Fornells, A. (2008). Beneficial effects of word final stress in segmenting a new language: Evidence from ERPs. *BMC Neuroscience*, 9, 23.
- Cunillera, T., Toro, J. M., Sebastián-Gallés, N., & Rodríguez-Fornells, A. (2006). The effects of stress and statistical cues on continuous speech segmentation: An event-related brain potential study. *Brain Research*, 1123(1), 168–178.
- Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1), 9–21.
- Elder, C., Golombek, P., Nguyen, T. T. A., & Ingram, J. (2005). Vietnamese acquisition of English word stress. *TESOL Quarterly*, 39(2), 309–319.
- Flege, J. E., & Bohn, O.-S. (1989). An instrumental study of vowel reduction and stress placement in Spanish-accented English. *Studies in Second Language Acquisition*, 11(01), 35–62.
- Frost, D. (2011). Stress and cues to relative prominence in English and French: A perceptual study. *Journal of the International Phonetic Association*, 41(01), 67–84.
- Galbraith, G. C., & Brown, W. S. (1990). Cross-correlation and latency compensation analysis of click-evoked and frequency-following brain-stem responses in man. *Electroencephalography and Clinical Neurophysiology*, 77(4), 295–308.
- Golumbic, E. M. Z., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., ... Schroeder, C. E. (2013). Mechanisms underlying selective neuronal tracking of attended speech at a “cocktail party”. *Neuron*, 77(5), 980–991.
- Goswami, U., Gerson, D., & Astruc, L. (2010). Amplitude envelope perception, phonology and prosodic sensitivity in children with developmental dyslexia. *Reading and Writing: An Interdisciplinary Journal*, 23(8), 995–1019.
- Goswami, U., & Leong, V. (2013). Speech rhythm and temporal structure: Converging perspectives. *Laboratory Phonology*, 4, 67–92.
- Goswami, U., Mead, N., Fosker, T., Huss, M., Barnes, L., & Leong, V. (2013). Impaired perception of syllable stress in children with dyslexia: A longitudinal study. *Journal of Memory and Language*, 69(1), 1–17.
- Honbolygó, F., & Csépe, V. (2013). Saliency or template? ERP evidence for long-term representation of word stress. *International Journal of Psychophysiology*, 87(2), 165–172.
- Honbolygó, F., Csépe, V., & Ragó, A. (2004). Suprasegmental speech cues are automatically processed by the human brain: A mismatch negativity study. *Neuroscience Letters*, 363(1), 84–88.
- Howie, J. M. (1976). *Acoustical studies of Mandarin vowels and tones*. Cambridge: Cambridge University Press.
- Jarmulowicz, L., & Taran, V. L. (2013). Lexical morphology: Structure, process, and development. *Topics in Language Disorders*, 33(1), 57–72.
- Jarmulowicz, L. (2016). Stress production in derived English words as a developmental window. In J. Thomson & L. Jarmulowicz (Eds.), *Linguistic Rhythm and Literacy*. Amsterdam, Netherlands: TiLAR, John Benjamins Publishing Co.
- Kehoe, M., Stoel-Gammon, C., & Buder, E. H. (1995). Acoustic correlates of stress in young children's speech. *Journal of Speech and Hearing Research*, 38(2), 338–350.
- Kochanski, G., Grabe, E., Coleman, J., & Rosner, B. (2005). Loudness predicts prominence: Fundamental frequency lends little. *The Journal of the Acoustical Society of America*, 118(2), 1038–1054.
- Kraus, N., McGee, T., Carrell, T. D., King, C., Tremblay, K., & Nicol, T. (1995). Central auditory system plasticity associated with speech discrimination training. *Journal of Cognitive Neuroscience*, 7(1), 25–32.
- Krishnan, A., Gandour, J. T., & Bidelman, G. M. (2010). The effects of tone language experience on pitch processing in the brainstem. *Journal of Neurolinguistics*, 23(1), 81–95.
- Krishnan, A., Gandour, J. T., & Suresh, C. H. (2015). Pitch processing of dynamic lexical tones in the auditory cortex is influenced by sensory and extrasensory processes. *The European Journal of Neuroscience*, 41(11), 1496–1504.
- Kujala, T., Kallio, J., Tervaniemi, M., & Näätänen, R. (2001). The mismatch negativity as an index of temporal processing in audition. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, 112(9), 1712–1719.



- Lehmann, D., & Skrandies, W. (1980). Reference-free identification of components of checkerboard-evoked multichannel potential fields. *Electroencephalography and Clinical Neurophysiology*, 48(6), 609–621.
- Li, P., Sepanski, S., & Zhao, X. (2006). Language history questionnaire: A web-based interface for bilingual research. *Behavior Research Methods*, 38(2), 202–210.
- Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, 8, 213.
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164(1), 177–190.
- Menning, H., Imaizumi, S., Zwitserlood, P., & Pantev, C. (2002). Plasticity of the human auditory cortex induced by discrimination learning of non-native, moratimed contrasts of the Japanese language. *Learning & Memory (Cold Spring Harbor, N.Y.)*, 9(5), 253–267.
- Menning, H., Roberts, L. E., & Pantev, C. (2000). Plastic changes in the auditory cortex induced by intensive frequency discrimination training. *NeuroReport*, 11(4), 817–822.
- Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology*, 38(1), 1–21.
- Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., ... Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, 385(6615), 432–434.
- Näätänen, R., Schröger, E., Karakas, S., Tervaniemi, M., & Paavilainen, P. (1993). Development of a memory trace for a complex sound in the human brain. *NeuroReport*, 4(5), 503–506.
- Nononen, S., Shestakova, A., Huotilainen, M., & Näätänen, R. (2003). Linguistic relevance of duration within the native language determines the accuracy of speech-sound duration processing. *Cognitive Brain Research*, 16(3), 492–495.
- Novitski, N., Tervaniemi, M., Huotilainen, M., & Näätänen, R. (2004). Frequency discrimination at different frequency levels as indexed by electrophysiological and behavioral measures. *Cognitive Brain Research*, 20(1), 26–36.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9(1), 97–113.
- Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, 2011, 156869.
- Oostenveld, R., & Praamstra, P. (2001). The five percent electrode system for high-resolution EEG and ERP measurements. *Clinical Neurophysiology*, 112(4), 713–719.
- Ou, S.-C. (2010). Taiwanese EFL learners' perception of English word stress. *Concentric: Studies in Linguistics*, 36(1), 1–23.
- Paavilainen, P., Arajärvi, P., & Takegata, R. (2007). Preattentive detection of nonsalient contingencies between auditory features. *NeuroReport*, 18(2), 159–163.
- Pennington, M. C., & Ellis, N. C. (2000). Cantonese speakers' memory for English sentences with prosodic cues. *The Modern Language Journal*, 84(3), 372–389.
- Peperkamp, S., & Dupoux, E. (2002). A typological study of stress “deafness”. In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology VII* (pp. 203–240). Berlin: Mouton de Gruyter.
- Peperkamp, S., Vendelin, I., & Dupoux, E. (2010). Perception of predictable stress: A cross-linguistic investigation. *Journal of Phonetics*, 38(3), 422–430.
- Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., ... Roberts, T. (2000). Auditory cortex accesses phonological categories: An MEG mismatch study. *Journal of Cognitive Neuroscience*, 12(6), 1038–1055.
- Picton, T. W., Alain, C., Woods, D. L., John, M. S., Scherg, M., Valdes-Sosa, P., ... Trujillo, N. J. (1999). Intracerebral sources of human auditory-evoked potentials. *Audiology & Neuro-Otology*, 4(2), 64–79.
- Saari, J., Paavilainen, P., Schöger, E., Tervaniemi, M., & Näätänen, R. (1992). Representation of abstract attributes of auditory stimuli in the human brain. *NeuroReport*, 3(12), 1149–1151.
- Shestakova, A., Brattico, E., Huotilainen, M., Galunov, V., Soloviev, A., Sams, M., ... Näätänen, R. (2002). Abstract phoneme representations in the left temporal cortex: Magnetic mismatch negativity study. *NeuroReport*, 13(14), 1813–1816.
- Shestakova, A., Huotilainen, M., Ceponiene, R., & Cheour, M. (2003). Event-related potentials associated with second language learning in children. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, 114(8), 1507–1512.
- Tamminen, H., Peltola, M. S., Kujala, T., & Näätänen, R. (2015). Phonetic training and non-native speech perception—New memory traces evolve in just three days as indexed by the mismatch negativity (MMN) and behavioural measures. *International Journal of Psychophysiology*, 97(1), 23–29.
- Tiitinen, H., May, P., Reinikainen, K., & Näätänen, R. (1994). Attentive novelty detection in humans is governed by pre-attentive sensory memory. *Nature*, 372(6501), 90–92.
- Ullman, M. T. (2005). A cognitive neuroscience perspective on second language acquisition: The declarative/procedural model. In C. Sanz (Ed.), *Mind and context in adult second language acquisition: Methods, theory, and practice* (pp. 141–178). Washington, DC: Georgetown University Press.
- Wade-Woolley, L., & Heggie, L. (2015). Implicit knowledge of word stress and derivational morphology guides skilled readers' decoding of multisyllabic words. *Scientific Studies of Reading*, 19(1), 21–30.
- Wallstrom, G. L., Kass, R. E., Miller, A., Cohn, J. F., & Fox, N. A. (2004). Automatic correction of ocular artifacts in the EEG: A comparison of regression-based and component-based methods. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, 53(2), 105–119.
- Wang, H.-L. S., Huss, M., Hämäläinen, J. A., & Goswami, U. (2012). Basic auditory processing and developmental dyslexia in Chinese. *Reading and Writing*, 25(2), 509–536.
- Winkler, I., Denham, S. L., & Nelken, I. (2009). Modeling the auditory scene: Predictive regularity representations and perceptual objects. *Trends in Cognitive Sciences*, 13(12), 532–540.
- Winkler, I., Lehtokoski, A., Alku, P., Vainio, M., Czigler, I., Csépe, V., et al. (1999). Pre-attentive detection of vowel contrasts utilizes both phonetic and auditory memory representations. *Cognitive Brain Research*, 7(3), 357–369.
- Woods, D. L. (1995). The component structure of the N1 wave of the human auditory evoked potential. *Electroencephalography and Clinical Neurophysiology. Supplement*, 44, 102–109.
- Ylinen, S., Strelnikov, K., Huotilainen, M., & Näätänen, R. (2009). Effects of prosodic familiarity on the automatic processing of words in the human brain. *International Journal of Psychophysiology*, 73(3), 362–368.
- Yu, V. Y., & Andruski, J. E. (2010). A cross-language study of perception of lexical stress in English. *Journal of Psycholinguistic Research*, 39(4), 323–344.
- Zhang, Y., Nissen, S. L., & Francis, A. L. (2008). Acoustic characteristics of English lexical stress produced by native Mandarin speakers. *The Journal of the Acoustical Society of America*, 123(6), 4498–4513.