

English and Mandarin native speakers' cue-weighting of lexical stress: Results from MMN and LDN

Zhen Zeng^{a,*}, Liqian Liu^{a,b}, Alba Tuninetti^c, Varghese Peter^d, Feng-Ming Tsao^e, Karen Mattock^a

^a Western Sydney University

^b University of Oslo

^c Bilkent University

^d University of the Sunshine Coast

^e National Taiwan University

ARTICLE INFO

Keywords:

Stress perception
Language rhythmicity
Cue-weighting
Mis-match negativity (MMN)
Late discriminative negativity (LDN)
Multi-feature paradigm
Iambic-trochaic law

ABSTRACT

Past research on how listeners weight stress cues such as pitch, duration and intensity has reported two inconsistent patterns: listeners' weighting conforms to 1) their native language experience (e.g., language rhythmicity, lexical tone), and 2) a general "iambic-trochaic law" (ITL), favouring innate sound groupings in cue perception. This study aims to tease apart the above effects by investigating the weighting of pitch, duration and intensity cues in stress-timed (Australian English) and non-stress-timed and tonal (Taiwan Mandarin) language speaking adults using a mismatch negativity (MMN) multi-feature paradigm. Results show effects that can be explained by language-specific rhythmic influence, but only partially by the ITL. Moreover, these findings revealed cross-linguistic differences indexed by both MMN and late discriminative negativity (LDN) responses at cue and syllable position levels, and thus call for more sophisticated perspectives for existing cue-weighting models.

1. Introduction

Sensitivity to speech rhythm permits the segmentation of words from fluent speech. In English, the perception of speech rhythm largely relies on analysing auditory cues pertinent to lexical stress (Cutler & Norris, 1988), including pitch, duration and intensity (Adams & Munro, 1978). Two theoretical frameworks have emerged regarding how these cues are weighted during stress perception. On the one hand, listeners are sensitive to and can segment speech based on the rhythmic properties of their native language(s) (e.g., stress- vs. syllable-timed) (Goyet, de Schonen, & Nazzi, 2010; Houston, Santelmann, & Jusczyk, 2004). On the other hand, listeners may also use what is considered a domain-general method – the iambic-trochaic law (ITL) – for sound grouping. The ITL refers to an innate tendency to group sounds that vary in pitch and intensity into trochaic (strong-weak grouping) patterns and sounds that vary in duration into an iambic (short-long grouping) pattern (Hayes, 1995; Nespor et al., 2008). To date, most comparative studies of these two frameworks have investigated languages with trichaic/iambic stress bias. For example, the prediction of language-specific influence

and the ITL in terms of pitch and intensity grouping bias overlap but the prediction for duration bias differs for languages like English and German. However, there is a paucity of studies investigating these effects in languages that are less conforming to the ITL, such as Mandarin, a tonal and syllable-timed language, where cues such as pitch and duration may highlight differences in ways that point to differences between language-specific influence and the ITL. The present study focuses on evaluating these effects by assessing the neural acoustic correlates of English lexical stress cues among native speakers of English and Mandarin. In the following, we review experimental research on stress cue-weighting from perspectives of language-specific and the ITL, and then assess lexical stress cue-weighting by Mandarin and English speakers using a mismatch negativity (MMN) multi-feature paradigm.

1.1. Language-specific influence on stress cue-weighting

Researchers have proposed three main categories of speech rhythmicity, characterised by the distinct unit of perceived isochronous recurrence: stress-, syllable-, or morae-timed languages (Abercrombie,

* Corresponding author at: Locked Bag 1797, Penrith 2750, NSW, Australia.
E-mail address: zh.zeng2015@gmail.com (Z. Zeng).

1964, 1967; Ladefoged & Johnson, 2014; Pike, 1945). Although contrastive stress placement (i.e., “PERmit” vs. “perMIT” in English) is common in stress-timed languages (Fry, 1958), native speakers of these languages show a trochaic bias in word segmentation relevant to stress patterns. For example, English speakers use a metrical segmentation strategy wherein disyllabic words are segmented faster and more reliably when the stress is at the initial syllable position (Cutler & Norris, 1988), since words with initial stress appear about three times more frequently than words without initial stress in English (Cutler & Carter, 1987).

Another way stress is analysed is through the perceptual cue-weighting. Acoustically, stress is characterised by stronger pitch, duration, and intensity measures (Adams & Munro, 1978). Some studies have attempted to map out the perceptual hierarchy of these cues using behavioural paradigms. Using a forced-choice auditory identification task, Chrabaszc, Winn, Lin, and Idsardi (2014) found that English speakers weighted pitch (height) as the most important cue for accurate judgment of stress placement (in disyllabic non-words), followed by intensity and then duration. The findings did not fully match the hypotheses derived from the literature, which state that pitch is the most important cue (e.g., Fry, 1958) but intensity is the least important (e.g., Mattys, 2000) for English speakers. Additionally, the assumption that all manipulated changes across cues are comparable ignores the role of cue salience (Ellis & Jones, 2009). It is difficult to gauge whether the identification was based on language-specific reliance on a particular cue or its salience (how acoustically easy it is to perceive a sound).

To gain a clearer view of cue-weighting and perceptual salience, studies have increasingly focused on neural paradigms, including measuring event-related potentials (ERPs) and, in particular, the mismatch negativity (MMN) response and the late discriminative negativity (LDN) response. The MMN response, with peaks around 100–250 ms after the change onset in adults, has been extensively used to investigate auditory discrimination, although the precise neural function has been debated (see Fitzgerald & Todd, 2020 for a review). It is elicited by violations in the regular aspects of the auditory stimulation, and is larger when the difference between the standard and deviant is more marked (for a review, see Paavilainen, 2013; Fitzgerald & Todd, 2020). Occasionally, MMN is followed by another component, LDN, around 300–600 ms after the change onset (e.g., Korpilahti, Krause, Holopainen, & Lang, 2001; Martynova, Kirjavainen, & Cheour, 2003). Although the functional significance of the LDN remains inconclusive, evidence suggests that it can be (1) language-specific (Korpilahti et al., 2001; Yu, Shafer, & Sussman, 2017), (2) the establishment of more permanent internal representations of recurring regularities in the incoming auditory stream (Zachau et al., 2005), and/or (3) attention-related, representing involuntary reorienting of attention back to the main task following distraction (Escera, Alho, Schröger, & Winkler, 2000; Shestakova, Huotilainen, Čeponien, & Cheour, 2003), or is a manifestation of attention to a recently presented stimulus, indicating surprising or difficult processing (Cheour, Korpilahti, Martynova, & Lang, 2001; Courchesne, 1977; 1978).

In a study examining cue salience in a passive oddball paradigm, Zora, Schwarz, and Heldner (2015) recorded English-speaking adults' MMN responses to changes in pitch and intensity for words and pseudo-words. Both pitch and intensity deviants elicited MMN responses, with larger amplitude in words than pseudo-words and with pitch eliciting larger MMN responses than intensity. In addition, this study found early and late responses after the change onset for real words, whereas only a late response was found for the pseudo-words, which indicates attention to changes but difficulty in processing according to the authors. It was concluded that, in terms of cue-weighting, intensity is an important prosodic cue for determining word stress and contributes to lexical access, even though it is not as salient as pitch. In another study, duration as a stress cue was manipulated in an attempt to elicit MMN responses to stress in both the speech and the music contexts in English-speaking adults (Peter, McArthur, & Thompson, 2012). Results showed that

long-short but not short-long speech stimuli elicited MMN responses, which was interpreted as a stronger brain sensitivity to familiar stress patterns relative to listeners' native language. LDN responses for both long-short and short-long stress patterns in music (but not speech) stimuli were observed, possibly due to higher processing cost in music than in language. Interestingly, although a number of studies found a novelty effect of LDN (e.g., Chen, Peter, Wijnen, Schnack, & Burnham, 2018; Peter et al., 2012; Zora et al., 2015), there has also been interpretation of LDN as a language-specific familiarity effect. In a lexical tone study, English and Mandarin adults' MMN responses to (Mandarin) distinct tone contour contrast (T3-T1) and less distinct tone contour contrast (T2-T3) was investigated in a multi-feature paradigm (Yu et al., 2017). Results showed overall larger LDN in the native language (Mandarin) group than in the non-native language (English) group and larger LDN for the acoustically more distinct contrast than the less distinct contrast.

While all listeners rely on cues such as pitch, duration and intensity to segment words, the cue-weighting mechanism of the target language is dependent on the rhythmic class, bias and acoustic attributes of prosodic contrasts of their native language(s) (Goyet et al., 2010; Houston et al., 2004; Jusczyk, Cutler, & Redanz, 1993). As most comparative studies (e.g., stress-timed vs. syllable-timed) in this field have been conducted with European languages, it is not well-understood how native speakers of tonal languages such as Mandarin weight English stress, due to a lack of research. Nevertheless, (Taiwan) Mandarin is an ideal candidate for stress cue-weighting due to its multifaceted nature. Firstly, Mandarin can be classified as a syllable-timed language by acoustic analysis (Mok, 2009), and there is no consistent claim that Mandarin has a iambic/trochaic bias at the word level (see Lavin, 2002). Secondly, it is a tonal language where pitch contour variations are lexical, but duration variations are not obvious (Qin, Chien, & Tremblay, 2017). These mean that although Mandarin adults are sensitive to pitch cues, they may not be able to segment English speech efficiently because 1) Mandarin speakers may not be familiar with the metrical segmentation strategy and thus might not have a general trochaic bias like their English-speaking counterparts; 2) contrastive stress placement within a word in Mandarin is less common compared to English; and 3) they don't assign all three cues of pitch, duration and intensity to stress. As a result, Mandarin speakers might exhibit diminished sensitivity to some of the stress cues or to stress in general.

There has been behavioural and neural evidence that Mandarin speakers show poorer sensitivity to English stress cue-weighting compared to the native speakers of English. Using a behavioural forced-choice paradigm, Wang (2008) reported pitch as a decisive cue for stress perception in Mandarin learners of English when compared to English speakers who used all three cues in stress perception of English disyllabic non-words. The author suggested that Mandarin listeners are 'deaf' to duration and intensity contrasts (Wang, 2008). This finding was also replicated in other studies using similar behavioural paradigms (Ou, 2010; Vickie & Andruski, 2010). In Chrabaszc et al. (2014), English speakers showed overall better stress discrimination signaled by segmental and suprasegmental cues, although both English speakers and Mandarin speakers weighted pitch (height) as the most important cue for accurate judgment of stress placement (in disyllabic non-words), followed by intensity and then duration. In a similar vein, Chung and Bidelman (2016) found that while both English and Mandarin adults can track stress for the non-word pair “nocTICity” and “NOcticity”, English adults had larger ERPs than Mandarin adults, indicating better tracking of intensity changes.

In another behavioural study, however, Cantonese adults who were highly competent in English showed an advantage instead of a disadvantage in stress processing compared to English adults (Tong, Lee, Lee, & Burnham, 2015). Specifically, Cantonese (a syllable-timed and tonal language) adults speaking English and Cantonese had marginally higher accuracy than did English adults when they were asked to decide the syllabic position of stress in a range of disyllabic non-words.

Nevertheless, it is not clear whether the first syllable and the second syllable of these non-words were equally probable when eliciting stress perception in these adults. Thus, more carefully controlled studies are needed in this area to discern any task-specific performances from listeners' perceptual sensitivity.

Taken together, Mandarin participants' perception of English in both behavioural and neural paradigms has generally revealed diminished sensitivity to some of the stress cues, possibly due to their lower proficiency in English. Higher sensitivity to stress cues is possible when Mandarin participants also show high competency in English. However, these studies did not look into how their sensitivity to particular syllabic positions would compare to their English counterparts. Thus, it is still not clear whether Mandarin speaker also use the ITL.

1.2. Iambic-trochaic law on stress cue-weighting

While cross-language studies point to a language-specific influence on stress cue-weighting, the iambic-trochaic law (ITL) (Hayes, 1995) – a domain-general law of sound grouping – posits that sounds varying in pitch or intensity are grouped trochaically (or strong-weak pattern), and those varying in duration are grouped iambically (or weak-strong pattern) (see Crowhurst, 2020 for a review). There has been evidence for this perceptual grouping pattern not only in human, but also non-human species (Cook, Rouse, Wilson, & Reichmuth, 2013; Fitch, 2012; Gámez et al., 2018; Hasegawa, Okanoya, Hasegawa, & Seki, 2011; Patel, Iversen, Bregman, & Schulz, 2009; Schachner, Brady, Pepperberg, & Hauser, 2009). In human studies, there has been evidence for this in cross-modal studies (e.g., Peña, Mehler, & Nespors, 2011), and in auditory studies using non-speech (Bolton, 1894; Iversen, Patel, & Ohgushi, 2008; Jeon & Arvaniti, 2016; Kusumoto & Moreton, 1997; Woodrow, 1909, 1951) and speech stimuli (Bhatara, Boll-Avetisyan, Unger, Nazzi, & Höhle, 2013; Bion, Benavides-Varela, & Nespors, 2011; Crowhurst, 2016; Crowhurst & Olivares, 2014; Hay & Diehl, 2007; Nespors et al., 2008). For example, Hay and Diehl (2007) tested rhythmic grouping in native listeners of French and English using both speech and non-speech stimuli. In both cases, French and English listeners grouped intensity-varied sequences as trochees and duration-varied sequences as iambs. Similar findings have been replicated among Italian speakers for pitch and duration (Bion et al., 2011). Italian adults were better at remembering pairs of syllables that had prominence signaled by pitch in the first syllable, whereas they were better at remembering pairs of syllables with prominence signaled by duration in the second syllable. Recall that in Peter et al. (2012), English speakers showed larger MMN responses to long-short compared to short-long patterns, which was interpreted as a familiarity effect of MMN reflecting a trochaic bias for duration cues. Likely, such responses are instead effects from the ITL if the MMN was interpreted as a novelty effect. Thus, while the grouping bias for duration cues in stress-timed language speakers may differentiate language-specific influence from the ITL, it may be subject to interpretation if control conditions (pitch and intensity conditions where predictions of language specific influence and the ITL align) are not included.

Although the ITL has been replicated in various studies, the language-specific influence seems to influence results in studies using paradigms designed to test ITL. For example, Iversen et al. (2008; see also Kusumoto & Moreton, 1997) tested the ITL by assessing English and Japanese adults' grouping of streams of complex tones where every second tone varied in either intensity or duration. Although English listeners showed higher accuracy for an iambic grouping of duration cues, Japanese listeners showed less consistent grouping patterns for duration cues, with a bias toward a trochaic grouping. Similarly, in Bhatara et al. (2013), the language-specific influence of French was found in addition to the ITL, when both French and German adults were required to group sequences of continuous sequences of segmentally varying coarticulated syllables alternating in intensity, duration or neither. Both French and German speakers showed grouping preferences as predicted by the ITL such that sequences with intensity variation were

judged to be trochaic and sequences with duration variation were judged to be iambic above chance level. However, French participants showed this effect less consistently and needed stronger acoustic contrasts (i.e., larger differences between syllables in intensity and duration) than German speakers.

1.3. The present study

The mechanisms underlying the perceptual weighting of pitch, intensity and duration cues in speech are not yet fully understood: most studies that have examined the ITL have found some language-specific effects, and studies that looked specifically into native vs. non-native language processing yielded inconsistent findings and did not take general sound processing mechanisms (such as the ITL) into consideration. Moreover, a confound between the domain-specific and domain-general accounts of stress cue-weighting still exists: as the language-specific influence for stress-timed language speakers is based off trochaic grouping, it can be argued as just a part of the ITL. So is iambic grouping with a syllable-timed language – French. To differentiate the language-specific influence from the ITL, the language spoken by the participants should have features that are less conforming to the ITL in theory, such as Mandarin. The present study investigated language-specific influence and the ITL in English and Mandarin adults, assessing their weighting of the stress cues of pitch, intensity and duration. We adopted an MMN paradigm, with /dede/ as the standard sound for the paradigm. Deviants were created by manipulating pitch, duration or intensity at either the first or the second syllable of the disyllabic non-word. The deviants with (incremental) changes at the first syllable were considered as stress-initial, or trochaic, and the ones with (incremental) changes at the second syllable were considered stress-final, or iambic.

We used the MMN multi-feature paradigm as used by multiple studies with adults (Honbolygó, Kolozsvári, & Csépe, 2017; Partanen, Vainio, Kujala, & Huotilainen, 2011; Sorokin, Alku, & Kujala, 2010). The multi-feature paradigm was firstly proposed in Näätänen, Pakarinen, Rinne, and Takegata (2004) to optimize the traditional oddball paradigm, such that multiple deviants (i.e., unfamiliar or non-repeated sounds) could be incorporated in one EEG recording session, compared to the traditional oddball paradigm where only one deviant is included. Results recorded from the multi-feature paradigm were similar to those recorded in the traditional oddball paradigm (Pakarinen et al., 2009). In addition to replicating previous results about English adults, this study anticipated both effects of language-specific and the ITL effects in Mandarin adults. We expected discrimination effects either as a novelty effect, indexed by a larger MMN to non-native linguistic features compared to native features (e.g., Honbolygó & Csépe, 2013), or as a familiarity effect (Peter et al., 2012), indexed by a larger MMN to native linguistic features compared to non-native features. In terms of the LDN responses, the present study also expected either a novelty effect (Chen et al., 2018; Peter et al., 2012; a marker of language-related processing difficulty) or a language specific familiarity effect (Yu et al., 2017). Specifically,

- (1) If the participants' stress cue-weighting was predicted by language-specific influence, English adults would show a trochaic grouping preference. Mandarin adults, speaking a syllable-timed language without obvious stress, should have no iambic/trochaic bias across the syllable positions for each of the three cues.
- (2) If participants' stress cue-weighting was predicted by the ITL, both English adults and Mandarin adults would exhibit a trochaic grouping preference for pitch and intensity cues but an iambic grouping preference for duration cues.
- (3) In terms of the perceptual hierarchy, English adults may weight pitch as the most important cue of the three cues and intensity the least important, whereas Mandarin adults would weight pitch cues as the most important and duration the least important.

2. Methods

2.1. Participants

The current sample consisted of 20 Mandarin adult listeners (age range: 19–35 years; mean = 21.5; 10 females) recruited at National Taiwan University and 18 Australian English listeners (age range: 21–32 years, mean = 24.3; 7 females) recruited at Western Sydney University. Informed consent was obtained from all participants. The sample size of the present study is comparable to those in [Partanen et al. \(2011\)](#) and [Pakarinen et al. \(2009\)](#), where multiple cue-weighting data from 14 to 15 participants were collected in a multi-feature paradigm. Additionally, power analysis showed that a minimum of 14 participants per language group would be sufficient for a small-to-medium effect ($f = 0.25$) ([Faul, Erdfelder, Lang, & Buchner, 2007](#)). Participants reported normal hearing and normal or corrected-to-normal vision. According to the Language Experience and Proficiency Questionnaire (LEAP-Q) ([Marian, Blumenfeld, & Kaushanskaya, 2007](#)) completed, all Mandarin adults had L2 English classes in standard school education (year7-year 12), did not speak English in their daily life, and had never lived in an English-speaking country at the time of testing. While their main language exposure was Mandarin, some of these adults were also exposed to a spoken Chinese dialect (Southern Min = 5). Their exposure to the dialect(s) ranged from zero to 30% as measured by the LEAP-Q. All English adult participants were monolingual, although some ($N = 2$) of their parents also speak a language in addition to English (Egyptian). None of the participants was musicians or reported receiving consistent/formal music training. This research project was approved by the ethics committee for human research at Western Sydney University (Approval number: H13161).

2.2. Stimuli

To ensure that there would be no lexicality effect in the MMN paradigm (e.g., [Shtyrov & Pulvermüller, 2002](#)), we selected a stimulus that was not a word in either English or Mandarin. A syllable /de/ (180 ms; 65 dB; 190 Hz) with a neutral pitch contour was produced by a monolingual Australian English female speaker and recorded using Adobe Audition in a sound-proof booth in MARCS Institute Phonetics Lab. The syllable was then concatenated to form a bisyllabic non-word (in both English and Mandarin) /dede/ with 20 ms silence gap between the syllables (to ensure the naturalness of the stimuli). This allowed us to control the acoustic features between the first and the second syllable. The disyllabic non-word with an intensity of 65 dB, F0 at 190 Hz, and a duration of 380 ms served as the 'standard' stimulus in the MMN paradigm. Stress deviants were manipulated in Praat software ([Boersma & Weenink, 2013](#)) with pitch, intensity and duration changes at either the initial or final syllable position (see [Table 1](#)). Specifically, pitch and intensity incremental changes were on the whole syllable, but duration incremental changes are made on the vowel only. Please note that, as this study is carried out together with a mirrored study collecting data from infants aged 7–8 months and 10–11 months, these magnitudes of acoustic change were determined by English monolingual infants' threshold of sound change perception in a behavioural paradigm. The manipulated syllable is either 10% higher in fundamental frequency (from results of a pilot study to determine the discriminatory threshold

Table 1

Manipulation of Deviants. [~dede]: pitch deviant at syllable 2, [de ~ de]: pitch deviant at syllable 2; [de:de]: duration deviant at syllable 1, [dede:]: duration deviant at syllable 2; [ˈdede]: intensity deviant at syllable 1; [deˈde]: intensity deviant at syllable 2.

	Pitch	Duration	Intensity
Stress on syllable 1	[~dede]	[de:de]	[ˈdede]
Stress on syllable 2	[de ~ de]	[dede:]	[deˈde]

of pitch changes in infants), 6 dB louder, or 33% longer than the un-manipulated syllable ([Bull, Eilers, & Oller, 1984](#)).

2.3. Procedure

An MMN multi-feature paradigm was employed. During the experimental session, the standard sounds alternated with each of the six randomized deviants (in the oddball block). The standard/deviant probability ratio was 50/50 ([Näätänen et al., 2004](#)). A control block consisting of six sub-blocks was also included, with each sub-block repeating one of the six deviants for 100 trials and the order presentation of the six sub-blocks randomised. There were 1800 stimuli in total, with 600 deviants for the control block, 600 standards and 600 deviants for the oddball block. The stimuli were presented with an inter-stimulus interval of 500 ms at a constant intensity of 65 dB SPL. The total duration of the experiment was 30 min.

Participants were seated in a sound-attenuated room ~ 0.5 m from a Genelec 8010A speaker where the sound stimuli were played. While listening to the stimuli presented through Presentation software (Version 21.0; Neurobehavioral Systems, Inc.), participants were fitted with a 32-channel EEG cap (ActiCAP Slim, Brain Products) attached with gel and electrodes. The continuous EEG signal was recorded at a sampling rate of 500 Hz with the reference electrode at FCz. Electrode impedances were kept below 50 kΩ at the start of the recording. Participants' EEG was recorded using LiveAmp amplifier and BrainVision Recorder. To maintain engagement during the study, a movie was played silently on a screen next to the speaker with subtitles in their native language. Importantly, thanks to the portability of the LiveAmp system (Brain Products), the same experimental equipment and procedures were adhered to by the same research personnel across the Sydney and Taipei Labs.

2.4. EEG pre-processing

The EEG was analysed using Fieldtrip Toolbox ([Oostenveld, Fries, Maris, & Schoffelen, 2011](#)) in MATLAB 2019a. The continuous EEG was firstly band-pass filtered between 0.1 and 30 Hz using a windowed sinc finite impulse response filter and then divided into epochs between -100 and 800 ms relative to sound onsets. Epochs were then baseline corrected between -100 and 0 ms. The EEG was then subjected to Independent Component Analysis. Components with stereotypical features of eye blinks and eye movements were removed from the EEG. The EEG signal was then re-referenced to the average of the mastoids. Trials exceeding $\pm 100 \mu\text{V}$ were removed and then averaged separately for deviant and control to obtain the event-related potential (ERP) waves. Difference waves were computed by subtracting the ERP for the control stimulus (instead of the standard stimulus) from the deviant stimulus. In this way, ERPs to physically identical stimuli were compared for the calculation of MMN responses, as it reflects the brain response to a change as opposed to ERP effects due to physical differences between standard and deviant ([Jacobsen & Schröger, 2003](#); [Peter, McArthur, & Thompson, 2010](#)). Individual ERP waves were averaged to create grand-averaged ERPs.

The presence of an MMN/LDN response (by comparing the amplitude of the difference wave to zero) was tested using nonparametric cluster-based permutation statistics ([Maris & Oostenveld, 2007](#)). First, a series of t-tests were computed at each electrode and each time point, comparing the deviant and control waveforms. From this, clusters were formed by combining the sampling points where a significant effect was obtained ($p < .05$, two-tailed) based on temporal and spatial adjacency and polarity of the effect. Cluster-level statistics were then calculated by adding together all the t values within the cluster. To control for Type I errors, a permutation approach was used where the condition labels were randomly swapped, and the t-tests were repeated 2000 times to generate a data-driven null hypothesis distribution. The cluster-level statistics from the first step were considered significant if it fell in the

top 2.5 or bottom 2.5 percentile of the distribution.

2.5. Statistical analyses of amplitudes across cue and syllable positions

To facilitate maximum experimental validity for cross-linguistic comparison, EEG data were collected using the same experimental equipment and procedures by the same research personnel across labs. English and Mandarin adults' sensitivity to each cue across the syllable positions was measured by calculating the MMN and LDN amplitude values for each condition for the same analysis window for all speakers, and from five electrodes in the frontocentral region (Fz, FC1, Cz, FC2 and FCz), as MMN is usually seen at fronto-central electrodes (Näätänen, Paavilainen, Rinne, & Alho, 2007; Peter et al., 2010). This is conducted by averaging the amplitude in a 40 ms time window centred around the most negative peak within the significant time window. For the conditions that showed no significant MMN or LDN responses, we used the time windows that were comparable across the language groups. For example, there were significant MMN responses for D1 for English adults but a lack of significant MMN responses for D1 for Mandarin adults. In this case, D1 amplitude for Mandarin adults was calculated based on the same time window (40 ms around the most negative peak) for English adults for D1.

3. Results

3.1. Statistical analyses of the ERP components

The deviant-control difference waves for each cue for the two language groups are shown in Figs. 1 and 2. The difference waves showed negative peaks between 0 and 600 ms for cue position at the first syllable and between 200 and 700 ms for cue position at the second syllable for both language groups. Cluster-based permutation tests were then performed to confirm significant ERP components in the corresponding

time windows. Significant time windows are shown in Table 2 and are also marked in the corresponding conditions in the deviant-control waves in Fig. 1, accompanied by topography graphs at the peaks. Essentially, a number of significant time windows were identified as the MMN, as the negative clusters fits the expected time range and peaks around 100–250 ms after the deviance onset. The LDN time windows were determined as the negative clusters peak around 300–450 ms after change onset. Moreover, the topographical locations of these effects are fronto-central, where MMN and LDN responses are expected.

We first analysed the MMN amplitudes through a 2 × 3 × 2 ANOVA with language group as a between-subject variable and cue (3) and syllable positions (2) as within-subject variables. There was no significant main effect of language group for MMN amplitudes, $F(1,36) = 1.71$, $p = .20$; $\eta_p^2 = .05$, indicating that, overall, English adults had similar MMN responses to all cues and syllable positions compared to Mandarin adults. However, there was a significant main effect of cue, $F(2,72) = 13.37$, $p < .001$; $\eta_p^2 = .27$, and an interaction between language group and cue, $F(2,72) = 3.64$, $p = .03$, $\eta_p^2 = .09$. Table 3 demonstrates the MMN mean and SD for each cue for two language groups respectively. For English adults, a more prominent MMN was elicited for intensity cue compared to duration cue, $p' = .03$, while the MMN amplitudes for pitch and duration were similar, $p' = .45$. So were amplitudes for pitch and intensity, $p' = .93$. For Mandarin adults, MMN amplitudes were more prominent for pitch compared to duration, $p' < .001$, more prominent for intensity compared to duration. In comparison, MMN amplitudes for pitch and intensity were similar, $p' = .003$. Moreover, English adults exhibited similar responses compared to Mandarin adults when perceiving pitch ($p' = .56$) and intensity cues ($p' = .11$), but had more prominent MMN responses compared to Mandarin adults when perceiving duration cues ($p' = .01$). Thus, while English adults had the most prominent MMN for intensity cue, Mandarin adults had the most prominent MMN for pitch cue. Whereas both groups had similar MMN responses to pitch and intensity cues, English adults exhibited larger

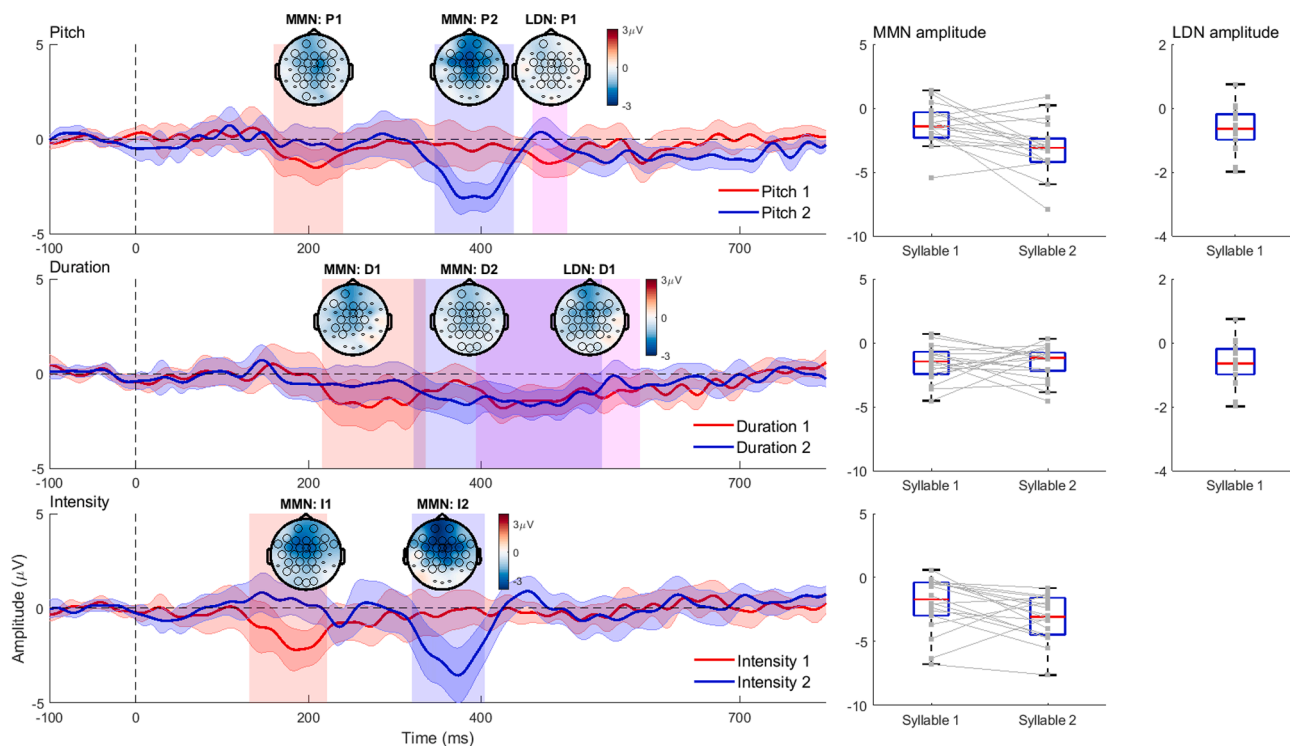


Fig. 1. MMN and LDN waveforms and scalp topography with highlighted significant time windows in English speaking adults, and boxplots of MMN and LDN peak amplitude comparison between stress on Syllable 1 and stress on Syllable 2 in English speaking adults. Significant time windows of MMN1, MMN2, LDN1 and LDN2 are respectively highlighted in Red, Blue, Magenta and Green. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

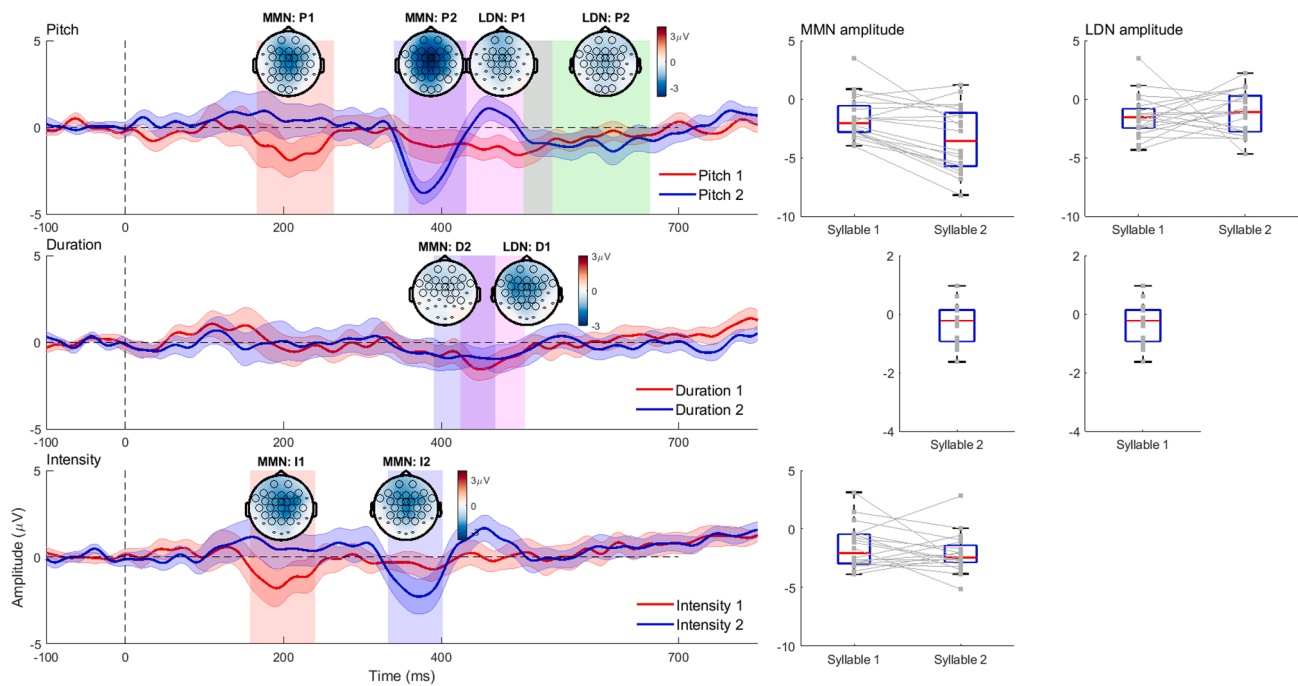


Fig. 2. MMN and LDN waveforms and scalp topography with highlighted significant time windows in Mandarin-speaking adults, and boxplots of MMN and LDN peak amplitude comparison between stress on Syllable 1 and stress on Syllable 2 in Mandarin-speaking adults. Significant time windows of MMN1, MMN2, LDN1 and LDN2 are respectively highlighted in Red, Blue, Magenta and Green. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2

Significant ERP time window for pitch (P), Duration (D) and Intensity (I) in first (1) and second (2) syllable positions for English and Mandarin native speakers.

	Stress on Syllable 1		Stress on Syllable 2	
	MMN1	LDN1	MMN2	LDN2
English	P1: 160–240 ms	P1: 460–500 ms	P2: 346–438 ms	P2: N/A
	D1: 216–336 ms	D1: 394–584 ms	D2: 322–540 ms	D2: N/A
	I1: 132–222 ms	I1: N/A	I2: 320–404 ms	I2: N/A
Mandarin	P1: 166–254 ms	P1: 358–540 ms	P2: 340–432 ms	P2: 504–664 ms
	D1: N/A	D1: 424–506 ms	D2: 390–468 ms	D2: N/A
	I1: 158–240 ms	I1: N/A	I2: 332–402 ms	I2: N/A

Table 3

Perceptual hierarchy of cue in English and Mandarin adults indexed by MMN amplitude mean and SD.

	English adults	Mandarin adults
Pitch	-2.19(.43)	-2.54(.34)
Duration	-1.59 (.21)	-0.58 (.23)
Intensity	-2.61(.39)	-1.82(.34)

MMN responses to duration cues compared to Mandarin adults.

We also found a significant main effect of syllable position, $F(1,36) = 21.40, p < .001; \eta_p^2 = .37$, as well as an interaction between cue and syllable position $F(2,36) = 5.18, p = .01; \eta_p^2 = .12$. Pairwise comparisons were performed to examine the cue and syllable position interaction, with Bonferroni adjustment. The MMN amplitudes across syllable positions were significantly different for pitch and intensity cues, with a larger MMN response for P2 compared to P1, $p' < .001$, and a larger

MMN response for I2 compared to I1, $p' = .03$. In contrast, D1 and D2 values were statistically similar, $p' = .42$. Table 4 presents the mean and SD of the MMN amplitude across cue and syllabic position in participants.

To interpret English adults' and Mandarin adults' sensitivity to language-specific influence and ITL based on our priori hypotheses, we calculated their MMN amplitudes to each condition. Pairwise comparison comparing amplitudes for syllable 1 and syllable 2 for each cue with Bonferroni score adjustment revealed that, for English adults, MMN responses were larger in amplitude for Syllable 2 compared to Syllable 1 for pitch ($p' = .01$), larger in amplitude for syllable 2 compared to syllable 1 for intensity changes ($p' = .046$), but were similar in amplitude for Syllable 1 and Syllable 2 for duration changes ($p' = .97$). For Mandarin adults, the MMN responses were larger for Syllable 2 compared to Syllable 1 for pitch changes ($p' < .001$), but were similar in amplitude for Syllable 1 and Syllable 2 for intensity ($p' = .27$) and similar in amplitude for Syllable 1 and Syllable 2 duration changes. Table 5 shows the mean and SD for each syllable position and cue within each language group.

As only pitch and duration cues at the first syllable elicited LDN responses for both language groups, we conducted analyses with these responses for cues at those positions for the two language groups. A 2×3 ANOVA with language group as between-subject variable and cue (3) as within-subjects did not show main effects of language group, $F(1,36) = .04, p = .85; \eta_p^2 = .001$, or cue $F(1,36) = .72, p = .60; \eta_p^2 = .008$. No interaction was found between language group and cue, $F(1,36) = .47, p = .50; \eta_p^2 = .01$. This indicates that, overall, English adults showed no statistical difference in their LDN responses compared to Mandarin adults.

Table 4

Overall syllabic bias in participants indexed by mean MMN amplitudes.

	Stress on Syllable 1	Stress on Syllable 2
Pitch	-1.49(.28)	-3.25(.40)
Duration	-1.8(.32)	-2.62(.28)
Intensity	-.96(.27)	-1.22(.24)

Table 5
English and Mandarin adults' mean MMN amplitude to each syllable/cue condition.

	English adults		Mandarin adults	
	Stress on Syllable 1	Stress on Syllable 2	Stress on Syllable 1	Stress on Syllable 2
Pitch	-1.33(.37)	-3.06(.49)	-1.65(.41)	-3.44(.62)
Duration	-1.6(.33)	-1.58(.32)	-0.31(.31)	-0.85(.35)
Intensity	-2.06(.51)	-3.16(.51)	-1.54(.41)	-2.09(.37)

4. Discussion

To investigate the effects of language-specific influence and the ITL on stress cue-weighting, this experiment examined English and Mandarin speakers' neural responses to pitch, duration and intensity changes at first and second syllable positions in a multi-feature paradigm. In terms of perceptual hierarchy, English adults showed the strongest neural encoding for intensity cues, followed by pitch and duration whereas Mandarin adults showed the strongest neural encoding for pitch, followed by intensity and duration, indexed by the MMN amplitudes. In terms of syllabic bias, results point to the language-specific influence of English in English adults but language-specific influence from both English and Mandarin in Mandarin adults. These effects are indexed by novelty effects of the MMN (Honbolygó & Csépe, 2013) and familiarity effects of LDN components (Yu et al., 2017).

4.1. Cross-linguistic comparison in cue-weighting

Our results revealed that overall, English and Mandarin adults had similar perceptual responses in our cue-weighting paradigm, including larger MMN amplitudes to the second syllable position for pitch cues ($P1 < P2$), and similar MMN amplitudes across the two syllable positions for duration ($D1 = D2$). The LDN amplitudes were also similar across the two language groups. However, cross-linguistic differences in the MMN and LDN amplitudes were also observed. For the MMN responses, results have demonstrated differing cue-weighting patterns across the language groups. While English adults were most sensitive to intensity cues and least sensitive to duration cues, Mandarin speakers were most sensitive to pitch cues and least to duration cues. In previous behavioural studies, it has been reported that Mandarin adults primarily use pitch cues in stress perception (Chrabaszcz et al., 2014; Ou, 2010; Vickie & Andruski, 2010; Wang, 2008). Mandarin adults' perceptual hierarchy found in this experiment thus aligns with cue-weighting patterns in behavioural studies. However, results for the English adults did not match cue-weighting studies in Chrabaszcz et al. (2014) where pitch was found to be the most important cue for English speakers, or Mattys (2000) where intensity was found to be the least important cue. This discrepancy of perceptual hierarchy from previous studies may be caused by differences in specific acoustic parameters of the cues used across studies (e.g., using a pitch deviant with a smaller pitch incremental change). Thus, future studies should take into account the acoustic parameters of the individual cues.

Moreover, English speakers showed a larger MMN response to duration changes than did Mandarin adults, pointing to enhanced sensitivity to stress signaled by duration cues in English adults compared to Mandarin adults. English adults also differed from Mandarin adults in their sensitivity specifically to D1 in that a significant MMN response to D1 was found in English adults only. This speaks to previous studies suggesting stress-timed languages such as English have a larger syllable durational variability compared to syllable-timed languages (Grabe & Low, 2002; Ramus, Nespore, & Mehler, 1999). It is thus possible that higher durational variability in a language leads to a better neural encoding of durational changes in native speakers of this language. For intensity cues, English adults showed a larger MMN amplitude to the second syllable position compared to the first syllable position ($I1 < I2$),

whereas Mandarin adults showed similar amplitudes to intensity cues across the two syllable positions ($I1 = I2$). These met our hypotheses that English adults show trochaic bias in line with the language-specific influence of English and the ITL whereas Mandarin show no bias in line with the language-specific influence of Mandarin. In terms of the LDN results, English adults had a significant LDN only to the P1 and D1 positions whereas Mandarin adults had significant LDN responses to P1, P2 and D1. Based on previous findings that pitch cues are the cues that Mandarin speakers rely upon the most (Ou, 2010; Vickie & Andruski, 2010; Wang, 2008), the LDN results for pitch do not seem to represent difficulty in processing. Similarly, it is very unlikely that English adults would experience difficulty perceiving P1 and D1 cues, as the initial syllable position should be the most salient according to the metrical stress strategy (Cutler & Norris, 1988). Thus, we argue that the LDN results reflect a language-specific familiarity effect (Yu et al., 2017) and/or the establishment of more permanent internal representations of recurring regularities in the incoming auditory stream (Zachau et al., 2005). Both of these possibilities point to enhanced sensitivity at the syllable positions with significant LDN amplitudes. In this case, our results are in line with the language-specific effect, as Mandarin adults were attentive to pitch cues at both the first and second syllable positions whereas English adults were attentive to pitch change at the first syllable position only. In terms of perceptual hierarchy, however, the LDN results point to more important language-specific effects of pitch and duration cues compared to intensity in both language groups, which does not match the MMN results in terms of intensity and duration. More research is thus needed to understand the MMN and LDN responses of intensity and duration cue-weighting. Below, we discuss the MMN and LDN effects by syllabic bias for each cue within each language group.

4.2. MMN and LDN within the English group

For pitch cues, significant MMN responses were elicited for both syllable positions, with amplitudes larger in P2 compared to P1. In comparison, significant LDN responses were only seen for the first syllable position. As we interpret the MMN response as a novelty effect and our LDN responses as a language-specific familiarity effect (Yu et al., 2017) and/or the establishment of more permanent internal representations of recurring regularities in the incoming auditory stream (Zachau et al., 2005), results from both the MMN and LDN responses are in line with the trochaic bias/the ITL for pitch perception.

For duration cues, significant but similar MMN amplitudes were found across the syllable positions, whereas significant LDN responses were only present for stress at the first syllable position. While similar MMN amplitudes across the two syllable positions do not conform to trochaic/iambic bias previously found in the literature, the LDN results with significant responses at D1 represent a familiarity effect, thus point more towards a trochaic bias, in line only with the metrical segmentation strategy/language-specific influence. However, the MMN finding here is surprising, as Peter et al. (2012) found larger MMN amplitudes in long-short compared to short-long stress patterns in (Australian) English speakers, in both speech and music stimuli, consistent with the ITL predictions. Their study manipulated a disyllabic non-word (/dada/), and the longer syllable is also around 30% longer than the control syllable, similar to the current study. It is then safe to exclude the possibility that the results difference was due to cue salience. Nevertheless, Peter et al. (2012) differs from our study in that only a duration cue was investigated, and the authors used a design where a number of standards varying in length were used. It is thus possible that the ITL is more easily elicited in an MMN paradigm for duration cues when no other cues are present.

For intensity cues, significant MMN responses were found with a trochaic pattern but no such significant LDN responses were found in English adults. This perceptual pattern supports both the metrical segmentation strategy/language-specific influence and the ITL. As to why there is an absence of LDN responses for intensity but not for pitch and

duration cues, it could be that perceptual priority is given to pitch and durations cues, as intensity cues are relied upon the least for stress perception (Mattys, 2000).

Taken together, adult English speakers exhibited a trochaic perceptual bias for pitch (seen in the MMN and LDN results) and intensity cues (seen in the MMN results) and can perceive duration trochaically (such as seen in the LDN results) or wholly without trochaic/iambic bias (as seen in the MMN results). While replicating previous studies investigating English-learning/speaking populations which have found a trochaic bias both in line with the native stress perception (e.g., Cutler & Norris, 1988) and the ITL in behavioural and MMN paradigms (Bhatara et al., 2013; Crowhurst, 2016; Crowhurst & Olivares, 2014; Hay & Diehl, 2007; Peter et al., 2012), our findings seem to be more in line with the account of language-specific influence. Moreover, our results reveal a more sophisticated pattern in perceptual sensitivity to these cues gained from interpreting results in neural responses, i.e., the MMN and LDN responses.

4.3. MMN and LDN within the Mandarin group

Similar to English-speaking adults, Mandarin adults showed significant MMN responses for pitch cues at both syllable positions, with larger amplitudes for the second syllable position compared to the first syllable position. Comparatively, significant but similar LDN amplitudes were found for the pitch cues (i.e., no difference between P1 and P2). The MMN responses could be interpreted as a trochaic bias from the effects of the ITL, whereas the LDN responses seem to point to a lack of syllabic bias in line with the language specific influence of Mandarin. Another possibility is that the trochaic bias from MMN responses comes from the language-specific influence of English, as the trochaic perception indexed by the MMN results was in line with the metrical segmentation strategy. This is not surprising, as for our sample of Mandarin-speaking adults, English was taught in school at least from year seven to year 12, which is typical in the Taiwan education system. Although L2 English adults were previously reported to have difficulty in English stress perception and production, there is also evidence that the experience in English does improve stress perception. For example, L2 learners' self-reported daily use of English was found to be a significant predictor of their successful perception of stress (Tremblay, 2009).

Secondly for duration perception, significant MMN responses were only found at the second syllable position, whereas significant LDN responses were only seen at the first syllable position. This indicates that there might be unreliable perception of D1, as previous behavioural studies also reported reliance of pitch but not duration cue in the perception of lexical stress (Ou, 2010; Vickie & Andruski, 2010; Wang, 2008). Nevertheless, Mandarin adults might still be able to establish a new memory trace after being exposed to stress signaled by a duration change at this syllable position. This result is similar to Zora et al. (2015) where late responses were reported to pseudowords, which might be the LDN responses. Furthermore, both MMN and LDN responses seem to point to a trochaic bias for duration cues, which was not predicted by our hypotheses of the language-specific influence of Mandarin or the ITL. Similar to our speculation of the trochaic bias found in the pitch cue-weighting in Mandarin adults, this trochaic bias for duration cues may indicate second language influence from English. As our study did not foresee this L2 effect, we are not able to ascertain the effects of L2 in cue-weighting. Future studies are thus needed to understand the effects of L2 in cue-weighting.

For intensity cues, significant but similar MMN amplitudes were found in Mandarin adults. This means Mandarin-speaking adults were able to perceive stress signaled by intensity alone, and they exhibited similar sensitivity across the two syllable positions. This was predicted by our hypothesis in line with the language-specific influence, as (Taiwan) Mandarin is a syllable-timed language with no obvious iambic/trochaic syllabic bias (Qin et al., 2017). Similar to their English counterparts, no LDN was found for intensity cues in Mandarin-speaking

adults, indicating that the intensity cue was less essential in some way compared to pitch and duration cues (which exhibited LDN responses) in stress perception for them, possibly because intensity is not a lexical cue in Mandarin.

Taken together, adult Mandarin speakers showed language-specific influence from both English and Mandarin. They can perceive pitch trochaically (seen in the smaller MMN amplitudes for P1), but were quite sensitive to intensity cues at both syllable positions (seen in the comparable MMN amplitudes for this cue). For duration cues, it could be that they have a trochaic bias (non-significant MMN at D1 but significant MMN at D2). However, they also have unreliable perception for duration (non-significant MMN but significant LDN at D1).

In summary, our results indicate that, when listening to English stimuli, both English and Mandarin speakers show a language-specific influence of their native language. These are indexed by trochaic biases from MMN and LDN results for pitch and intensity but LDN results for duration for English adults, and LDN results for pitch and MMN results for intensity for Mandarin adults. Nevertheless, the trochaic bias in pitch and intensity could be argued as part of the ITL effect. Moreover, Mandarin adults' trochaic bias in duration (and possibly pitch) may come from their L2 experience of English. These are indexed by MMN results for pitch and both MMN and LDN results for duration in Mandarin adults. As such, these findings add to the body of evidence detailing differences in cue-weighting in line with language-specific influence (Cutler & Norris, 1988; Jusczyk, Houston, & Newsome, 1999) and/or the ITL (Bion et al., 2011; Hay & Diehl, 2007; Nespor et al., 2008), and suggest that not only the MMN component, but also other ERP components such as the LDN are relevant measurements in auditory discrimination of stress. The present study has also added to the evidence that clarifies the functional significance of LDN and its relation to the processes indexed by MMN.

4.4. Language-specific influence vs. the ITL

From the above, results from both the English-speaking adults and Mandarin-speaking adults were largely in line with the language-specific influence account. As Mandarin-speaking adults also had some English knowledge from school, their language-specific influence seems to come from both English and Mandarin. Recall that a confound exists in differentiating domain-specific and domain-general accounts of stress cue-weighting, in that language-specific influence for stress-timed language speakers is based off trochaic grouping (at least for pitch and intensity cues) and thus can be argued as just a part of the ITL. Therefore, part of our results are also in line with the ITL. In this vein, the ITL as a domain-general mechanism may have been modulated across languages if language-specific properties do not support the maintenance of the ITL-bias (Boll-Avetisyan, Bhatara, Unger, Nazzi, & Höhle, 2020). Effects of the ITL may be present from birth (Abboub, Boll-Avetisyan, Bhatara, Höhle, & Nazzi, 2016). With the acquisition of the native language, sensitivity to the ITL would decrease when acquiring languages with features non-conforming to the ITL, while this initial ITL-bias would be maintained when acquiring languages with features conforming to the ITL. Future studies should look at younger Mandarin monolingual participants (e.g., children) to possibly exclude the effects of a second language. It is also useful to compare stress cue-weighting before and after the time of perceptual attunement in Mandarin infants, at which time trochaic bias develops in stress-timed language-learning peers.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This research was funded by Western Sydney University MARCS Candidature Funds and the Australian Linguistic Society Research Grant 2019 awarded to the first author. The second author was further supported by a European Union Horizon 2020 research and innovation programme under Marie Skłodowska-Curie grant agreement 798658, hosted by the University of Oslo Center for Multilingualism across the Lifespan, financed by the Research Council of Norway through Center of Excellence funding grant No. 223265.

References

- Abboub, N., Boll-Avetisyan, N., Bhatara, A., Höhle, B., & Nazzi, T. (2016). An exploration of rhythmic grouping of speech sequences by French- and German-learning infants. *Frontiers in Human Neuroscience, 10*, 292.
- Abercrombie, D. (1964). *A phonetician's view of verse structure*.
- Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh: Edinburgh University Press.
- Adams, C., & Munro, R. R. (1978). 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*(3), 125–156. <https://doi.org/10.1159/000259926>
- Bhatara, A., Boll-Avetisyan, N., Unger, A., Nazzi, T., & Höhle, B. (2013). Native language affects rhythmic grouping of speech. *The Journal of the Acoustical Society of America, 134*(5), 3828. <https://doi.org/10.1121/1.4823848>
- Bion, R. A., Benavides-Varela, S., & Nespor, M. (2011). Acoustic markers of prominence influence infants' and adults' segmentation of speech sequences. *Language and Speech, 54*(1), 123–140.
- Boersma, P., & Weenink, D. (2013). *Praat: Doing phonetics by computer [Computer program]*. Version 5.3.56. Retrieved September 15, 2013.
- Boll-Avetisyan, N., Bhatara, A., Unger, A., Nazzi, T., & Höhle, B. (2020). Rhythmic grouping biases in simultaneous bilinguals. *Bilingualism: Language and Cognition, 23*(5), 1070–1081. <https://doi.org/10.1017/S1366728920000140>
- Bolton, T. L. (1894). Rhythm. *The American Journal of Psychology, 6*(2), 145–238.
- Bull, D., Eilers, R. E., & Oller, D. K. (1984). Infants' discrimination of intensity variation in multisyllabic stimuli. *The Journal of the Acoustical Society of America, 76*(1), 13–17. <https://doi.org/10.1121/1.391110>
- Chen, A., Peter, V., Wijnen, F., Schnack, H., & Burnham, D. (2018). Are lexical tones musical? Native language's influence on neural response to pitch in different domains. *Brain and Language, 180*, 31–41.
- Cheour, M., Korpiälhti, P., Martynova, O., & Lang, A.-H. (2001). Mismatch negativity and late discriminative negativity in investigating speech perception and learning in children and infants. *Audiology and Neurotology, 6*(1), 2–11.
- 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, 57*(4), 1468–1479. https://doi.org/10.1044/2014_JSLHR-L13-0279
- Chung, W.-L., & Bidelman, G. M. (2016). Cortical encoding and neurophysiological tracking of intensity and pitch cues signaling English stress patterns in native and nonnative speakers. *Brain and Language, 155*, 49–57.
- Cook, P., Rouse, A., Wilson, M., & Reichmuth, C. (2013). A California sea lion (*Zalophus californianus*) can keep the beat: Motor entrainment to rhythmic auditory stimuli in a non vocal mimic. *Journal of Comparative Psychology, 127*(4), 412.
- Courchesne, E. (1977). Event-related brain potentials: Comparison between children and adults. *Science, 197*(4303), 589–592. <https://doi.org/10.1126/science.877575>
- Courchesne, E. (1978). Neurophysiological correlates of cognitive development: Changes in long-latency event-related potentials from childhood to adulthood. *Electroencephalography and Clinical Neurophysiology, 45*(4), 468–482.
- Crowhurst, M. (2016). Iambic-trochaic law effects among native speakers of Spanish and English. *Laboratory Phonology, 7*(1).
- Crowhurst, M. J. (2020). The iambic/trochaic law: Nature or nurture? *Language and Linguistics Compass, 14*(1), e12360. <https://doi.org/10.1111/ln3.12360>
- Crowhurst, M. J., & Olivares, A. T. (2014). Beyond the Iambic-Trochaic Law: The joint influence of duration and intensity on the perception of rhythmic speech. *Phonology, 31*(1), 51–94.
- Cutler, A., & Carter, D. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language, 2*, 133–142.
- Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 14*(1), 113.
- Ellis, R. J., & Jones, M. R. (2009). The role of accent salience and joint accent structure in meter perception. *Journal of Experimental Psychology: Human Perception and Performance, 35*(1), 264.
- Escera, C., Alho, K., Schröger, E., & Winkler, I. W. (2000). Involuntary Attention and Distractibility as Evaluated with Event-Related Brain Potentials. *Audiology and Neurotology, 5*(3–4), 151–166. <https://doi.org/10.1159/000013877>
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191.
- Fitch, W. T. (2012). The biology and evolution of rhythm: Unraveling a paradox. *Language and Music as Cognitive Systems, 73–95*.
- Fitzgerald, K., & Todd, J. (2020). Making sense of mismatch negativity. *Frontiers in psychiatry, 11*, 468.
- Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech, 1*(2), 126–152.
- Gómez, J., Yc, K., Ayala, Y. A., Dotov, D., Prado, L., & Merchant, H. (2018). Predictive rhythmic tapping to isochronous and tempo changing metronomes in the nonhuman primate. *Annals of the New York Academy of Sciences, 1423*, 396–414.
- Goyet, L., de Schonen, S., & Nazzi, T. (2010). Words and syllables in fluent speech segmentation by French-learning infants: An ERP study. *Brain Research, 1332*, 75–89.
- Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In *Papers in laboratory phonology* (vol. 7 (1982)), pp. 515–546.
- Hasegawa, A., Okanoya, K., Hasegawa, T., & Seki, Y. (2011). Rhythmic synchronization tapping to an audio-visual metronome in budgerigars. *Scientific Reports, 1*(1), 120. <https://doi.org/10.1038/srep00120>
- Hay, J. S. F., & Diehl, R. L. (2007). Perception of rhythmic grouping: Testing the iambic/trochaic law. *Perception & Psychophysics, 69*(1), 113–122. <https://doi.org/10.3758/BF03194458>
- Hayes, B. (1995). *Metrical stress theory: Principles and case studies*. University of Chicago Press.
- Honbolyó, F., & Csépe, V. (2013). Saliency or template? ERP evidence for long-term representation of word stress. *International Journal of Psychophysiology, 87*(2), 165–172.
- Honbolyó, F., Kolozsvári, O., & Csépe, V. (2017). Processing of word stress related acoustic information: A multi-feature MMN study. *International Journal of Psychophysiology, 118*, 9–17.
- Houston, D., Santelmann, L., & Jusczyk, P. (2004). English-learning infants' segmentation of trisyllabic words from fluent speech. *Language and Cognitive Processes, 19*(1), 97–136.
- Iversen, J. R., Patel, A. D., & Ohgushi, K. (2008). Perception of rhythmic grouping depends on auditory experience. *The Journal of the Acoustical Society of America, 124*(4), 2263–2271.
- Jeon, H., & Arvaniti, A. (2016). *Rhythmic grouping in English*. Testing the iambic-trochaic law: Greek and Korean.
- Jacobsen, T., & Schröger, E. (2003). Measuring duration mismatch negativity. *Clinical Neurophysiology, 114*(6), 1133–1143. [https://doi.org/10.1016/S1388-2457\(03\)00043-9](https://doi.org/10.1016/S1388-2457(03)00043-9)
- Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development, 64*(3), 675–687.
- Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The Beginnings of Word Segmentation in English-Learning Infants. *Cognitive Psychology, 39*(3), 159–207. <https://doi.org/10.1006/cogp.1999.0716>
- Korpiälhti, P., Krause, C. M., Holopainen, I., & Lang, A. H. (2001). Early and Late Mismatch Negativity Elicited by Words and Speech-Like Stimuli in Children. *Brain and Language, 76*(3), 332–339. <https://doi.org/10.1006/brln.2000.2426>
- Kusumoto, K., & Moreton, E. (1997). *Native language determines the parsing of nonlinguistic rhythmic stimuli*. Acoustical Society of America [PhD Thesis].
- Ladefoged, P., & Johnson, K. (2014). *A course in phonetics*. Cengage learning.
- Lavin, R. S. (2002). Conceptual Foundations of a Prosodic Model for Mandarin. *Journal of Cognitive Science, 3*(1), 65–83.
- Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research, 50*(4), 940–967.
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods, 164*(1), 177–190. <https://doi.org/10.1016/j.jneumeth.2007.03.024>
- Martynova, O., Kirjavainen, J., & Cheour, M. (2003). Mismatch negativity and late discriminative negativity in sleeping human newborns. *Neuroscience Letters, 340*(2), 75–78. [https://doi.org/10.1016/S0304-3940\(02\)01401-5](https://doi.org/10.1016/S0304-3940(02)01401-5)
- Mattys, S. L. (2000). The perception of primary and secondary stress in English. *Perception & Psychophysics, 62*(2), 253–265. <https://doi.org/10.3758/BF03205547>
- Mok, P. (2009). On the syllable-timing of Cantonese and Beijing Mandarin. *Chinese Journal of Phonetics, 2*, 148–154.
- Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology, 118*(12), 2544–2590. <https://doi.org/10.1016/j.clinph.2007.04.026>
- Näätänen, R., Pakarinen, S., Rinne, T., & Takegata, R. (2004). The mismatch negativity (MMN): Towards the optimal paradigm. *Clinical Neurophysiology, 115*(1), 140–144. <https://doi.org/10.1016/j.clinph.2003.04.001>
- Nespor, M., Shukla, M., van de Vijver, R., Avesani, C., Schraudolf, H., & Donati, C. (2008). Different phrasal prominence realizations in VO and OV languages. *Lingue e Linguaggio, 7*(2), 139–168.
- 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*.
- Ou, S. (2010). Taiwanese EFL Learners. *Concentric: Studies. Linguistics, 36*(1), 1–23.
- Paavilainen, P. (2013). The mismatch-negativity (MMN) component of the auditory event-related potential to violations of abstract regularities: A review. *International Journal of Psychophysiology, 88*(2), 109–123. <https://doi.org/10.1016/j.ijpsycho.2013.03.015>
- Pakarinen, S., Lovio, R., Huotilainen, M., Alku, P., Näätänen, R., & Kujala, T. (2009). Fast multi-feature paradigm for recording several mismatch negativities (MMNs) to phonetic and acoustic changes in speech sounds. *Biological Psychology, 82*(3), 219–226.

- Partanen, E., Vainio, M., Kujala, T., & Huotilainen, M. (2011). Linguistic multifeature MMN paradigm for extensive recording of auditory discrimination profiles. *Psychophysiology*, *48*(10), 1372–1380.
- Patel, A. D., Iversen, J. R., Bregman, M. R., & Schulz, I. (2009). Experimental evidence for synchronization to a musical beat in a nonhuman animal. *Current Biology*, *19*(10), 827–830.
- Peña, M., Mehler, J., & Nespors, M. (2011). The role of audiovisual processing in early conceptual development. *Psychological Science*, *22*(11), 1419–1421.
- Peter, V., McArthur, G., & Thompson, W. F. (2010). Effect of deviance direction and calculation method on duration and frequency mismatch negativity (MMN). *Neuroscience Letters*, *482*(1), 71–75.
- Peter, V., McArthur, G., & Thompson, W. F. (2012). Discrimination of stress in speech and music: A mismatch negativity (MMN) study. *Psychophysiology*, *49*(12), 1590–1600.
- Pike, K. L. (1945). *The Intonation of American English*.
- Qin, Z., Chien, Y.-F., & Tremblay, A. (2017). Processing of word-level stress by Mandarin-speaking second language learners of English. *Applied Psycholinguistics*, *38*(3), 541–570. <https://doi.org/10.1017/S0142716416000321>
- Ramus, F., Nespors, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, *73*(3), 265–292.
- Schachner, A., Brady, T. F., Pepperberg, I. M., & Hauser, M. D. (2009). Spontaneous Motor Entrainment to Music in Multiple Vocal Mimicking Species. *Current Biology*, *19*(10), 831–836. <https://doi.org/10.1016/j.cub.2009.03.061>
- Shestakova, A., Huotilainen, M., Čeponien, R., & Cheour, M. (2003). Event-related potentials associated with second language learning in children. *Clinical Neurophysiology*, *114*(8), 1507–1512.
- Shtyrov, Y., & Pulvermüller, F. (2002). Neurophysiological evidence of memory traces for words in the human brain. *Neuroreport*, *13*(4), 521–525.
- Sorokin, A., Alku, P., & Kujala, T. (2010). Change and novelty detection in speech and non-speech sound streams. *Brain Research*, *1327*, 77–90. <https://doi.org/10.1016/j.brainres.2010.02.052>
- Tong, X., Lee, S. M. K., Lee, M. M. L., & Burnham, D. (2015). A Tale of Two Features: Perception of Cantonese Lexical Tone and English Lexical Stress in Cantonese-English Bilinguals. *PLOS ONE*, *10*(11), e0142896. <https://doi.org/10.1371/journal.pone.0142896>
- Tremblay, A. (2009). Phonetic variability and the variable perception of L2 word stress by French Canadian listeners. *International Journal of Bilingualism*, *13*(1), 35–62. <https://doi.org/10.1177/1367006909103528>
- Vickie, Y. Y., & Andruski, J. E. (2010). A cross-language study of perception of lexical stress in English. *Journal of Psycholinguistic Research*, *39*(4), 323–344.
- Wang, Q. (2008). *Perception of English stress by Mandarin Chinese learners of English: An acoustic study* [Thesis]. <https://dspace.library.uvic.ca/handle/1828/1282>.
- Woodrow, H. (1909). *A Quantitative Study of Rhythm: The Effect of Variations in Intensity. Rate and Duration.*: Science Press.
- Woodrow, H. (1951). *Time perception*. In S. S. Stevens (Ed.), *Handbook of experimental psychology* (pp. 1224–1236). Wiley.
- Yu, Y. H., Shafer, V. L., & Sussman, E. S. (2017). Neurophysiological and behavioral responses of mandarin lexical tone processing. *Frontiers in Neuroscience*, *11*, 95.
- Zachau, S., Rinker, T., Körner, B., Kohls, G., Maas, V., Hennighausen, K., & Schecker, M. (2005). Extracting rules: Early and late mismatch negativity to tone patterns. *NeuroReport*, *16*(18), 2015–2019.
- Zora, H., Schwarz, I.-C., & Heldner, M. (2015). Neural correlates of lexical stress: Mismatch negativity reflects fundamental frequency and intensity. *Neuroreport*, *26*(13), 791–796. <https://doi.org/10.1097/WNR.0000000000000426>