ASHA American Speech-Language-Hearing Association

JSLHR

Research Note

Phonetic Encoding Contributes to the Processing of Linguistic Prosody at the Word Level: Cross-Linguistic Evidence From Event-Related Potentials

Luodi Yu,[a,b] Jiajing Zeng,[a,b] Suiping Wang,[a,b] and Yang Zhang[c]

[a] Key Laboratory of Brain, Cognition and Education Sciences, Ministry of Education, South China Normal University, Guangzhou [b] School of Psychology, Center for Studies of Psychological Application, and Guangdong Key Laboratory of Mental Health and Cognitive Science, South China Normal University, Guangzhou [c] Department of Speech-Language-Hearing Sciences, University of Minnesota, Twin Cities, Minneapolis

ARTICLE INFO

Article History:
Received January 24, 2021
Revision received June 8, 2021
Accepted August 4, 2021

Editor-in-Chief: Bharath Chandrasekaran
Editor: Chao-Yang Lee

https://doi.org/10.1044/2021_JSLHR-21-00037

ABSTRACT

Purpose: This study aimed to examine whether abstract knowledge of word-level linguistic prosody is independent of or integrated with phonetic knowledge.
Method: Event-related potential (ERP) responses were measured from 18 adult listeners while they listened to native and nonnative word-level prosody in speech and in nonspeech. The prosodic phonology (speech) conditions included disyllabic pseudowords spoken in Chinese and in English matched for syllabic structure, duration, and intensity. The prosodic acoustic (nonspeech) conditions were hummed versions of the speech stimuli, which eliminated the phonetic content while preserving the acoustic prosodic features.
Results: We observed language-specific effects on the ERP that native stimuli elicited larger late negative response (LNR) amplitude than nonnative stimuli in the prosodic phonology conditions. However, no such effect was observed in the phoneme-free prosodic acoustic control conditions.
Conclusions: The results support the integration view that word-level linguistic prosody likely relies on the phonetic content where the acoustic cues embedded in. It remains to be examined whether the LNR may serve as a neural signature for language-specific processing of prosodic phonology beyond auditory processing of the critical acoustic cues at the suprasyllabic level.

Neurophysiological studies on phonological processing have mostly examined speech stimuli at the segmental or monosyllabic level. However, speech perception and comprehension are beyond individual sounds but rely on the structure of sound organization. Prosody guides the suprasegmental organization of speech sounds and conveys critical linguistic information. Human newborn infants show signs of prenatal learning of native prosody (DeCasper & Spence, 1986; Mampe et al., 2009), which serves important "bootstrapping" functions in early language acquisition (Gervain & Mehler, 2010). For example, prosodic cues that mark phrasal boundaries can be utilized by infants during the first year of life to facilitate word learning (Shukla et al., 2011). In mature listeners, it is well established that correct prosody facilitates speech recognition, whereas incorrect prosody results in interference (Kjelgaard & Speer, 1999). Therefore, it is important to elucidate how linguistic prosody as opposed to emotional prosody (Diamond & Zhang, 2016) and phonetic knowledge are represented in the brain.

The investigation of prosody processing typically addresses three levels of analysis: lexical analysis at word level, syntactic analysis at sentence level, and global analysis at discourse level (Cutler et al., 1997). As the characteristics and functions of prosody differ across these levels in language processing, their degree of independence from phonetic information may also differ. This study chose to have its focus on the word-level prosody. One prominent

Correspondence to Luodi Yu: yuluodi@m.scnu.edu.cn, and Suiping Wang: wangsuiping@m.scnu.edu.cn. ***Disclosure:*** *The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

debate in this area is whether prosodic knowledge is independent of or integrated with phonetic knowledge. Phonetic knowledge or segmental phonetic knowledge concerns the acoustic-articulatory features of individual speech sounds or segmental cues, whereas prosodic knowledge concerns the acoustic organization of suprasyllabic units through pitch, intensity, and duration variations. Answers to this question have been mixed in the literature. Using prosodic priming paradigm, studies have shown that prosodic cues alone are sufficient in producing priming effects on subsequent word recognition (C. K. Friedrich et al., 2004; Schild et al., 2014). For example, C. K. Friedrich et al. (2004) observed that pitch-matching prime words accelerated target word recognition and attenuated event-related potential (ERP) amplitude regardless of whether its phonemes matched with the target or not, indicating that the prosodic pitch cue and phonetic context operate independently in word processing. This observation was replicated in Schild et al. (2014) using more naturalistic stress manipulation. Other evidence came from developmental research. Wade-Woolley (2015) observed that children's prosodic awareness and phonemic awareness independently contributed to their word reading skills. Becker et al. (2018) used similar stimuli as in Schild et al. and reported divergent developmental paths for phonemic and prosodic priming in infants of 3–9 months. Taken together, these studies suggest that prosodic cues might influence language processing in terms of abstract prosodic representations that are independent of phonetic representations.

There is considerable counterevidence against the independent status of prosodic knowledge. For instance, a series of experiments in Slowiaczek et al. (2006) showed no prosodic priming effects on auditory word naming or lexical decision in English, but significant priming for target words with a metrical stress pattern of strong onset syllables. The presence of metrical stress effect along with the absence of prosodic priming effect lends support to the phoneme-dependent view of prosody representation. Indirect evidence also came from brain imaging data involving sentence processing. Ischebeck et al. (2008) compared cortical activation patterns for sentence-level prosody contained in natural speech and in hummed sentences, which served as a nonspeech acoustic control. They found a network of left-hemisphere brain regions in temporal, frontal, and parietal lobes for natural speech processing and did not observe any brain areas that were activated more strongly for hummed stimuli than for natural speech, which indicates that prosody processing is integrated with speech processing (Ischebeck et al., 2008). However, this finding may be confined to sentence-level prosody as sentential and word-level prosody may differ in their degree of independence from phonetic information. Nonetheless, a recent study of word processing failed to observe stress (nonmetrical) priming effect on lexical decision except for segment-matching fragment primes (Protopapas et al., 2016). Together, there are reasons to claim that prosodic cues are only meaningful for word processing when integrated with phonetic information.

The disagreement continues and some methodological factors may have contributed to the inconsistent findings. One issue is that most studies only included speech material without prosodic acoustic control. Lack of acoustic control may limit our ability to detect potential independent representation of prosodic characteristics. Although there are studies deliberately made word prosody orthogonal to phonetic content (e.g., Schild et al., 2014), a speech mode of processing was inevitable; therefore, it might undermine the proposed "phoneme-free" prosody representation. Despite that acoustic control such as hummed sounds was used for sentence-level processing in Ischebeck et al. (2008), the participants were told that all stimuli were spoken sentences and required to monitor a probe word amidst each hum trial, and therefore, the task-related lexical processing might have influenced the results of prosody activation (Ischebeck et al., 2008). In fact, most work used attentive tasks relying on priming paradigm and lexical decision. Although a priming-lexical decision task with real words is excellent in determining whether prosodic features constrain lexical access, it might be difficult to tease apart prosody processing from lexico-semantic processing and decision making.

This study aimed to examine the potentially independent representation of prosody by investigating whether the prosodic acoustics alone can produce language-specific effects on the neural coding of word-level stimuli. Tonal languages such as Mandarin Chinese have significantly different prosodic characteristics from nontonal languages such as English. Specifically, English is a stress language in which syllable-by-syllable stress variations convey linguistic messages, whereas Chinese is a stress-flexible language in which syllable-by-syllable pitch variations are employed (Chrabaszcz et al., 2014). We created prosodic phonology contrasts using Chinese and English disyllabic pseudowords and prosodic acoustic contrasts using nonspeech hummed versions of the pseudowords. The hums preserved the characteristic prosody of the speech material but without its phonetic content. A passive (inattentive) listening procedure was adopted to minimize task-related lexical and cognitive processes. If abstract knowledge of linguistic prosody at the word level existed independently from phonetic knowledge, we would observe native versus nonnative differences in the prosodic acoustic conditions similar to those in the prosodic phonology conditions. If prosodic knowledge were an integral part of phonetic knowledge, the language-specific effect would be confined to the prosodic phonology conditions.

The stimuli were expected to elicit exogenous ERP components represented by the N1 and P2 as well as endogenous responses. The N1–P2 complex peak at approximately

100 and 200 ms after stimulus onset with a fronto-central topography and are called the obligatory auditory responses, which are known to be driven by acoustic features such as sound onset, intensity, and speech acoustics (Martin et al., 2008; Paiva et al., 2016). Unlike the N1–P2, late responses after 200 ms are considered endogenous neural activities related to higher order linguistic and cognitive processes. For example, N400s are generally larger for nonsense words (pseudowords) than for real words (Kutas & Federmeier, 2011). Late negative responses (LNRs) have been found to index processing of higher order phonological structures (M. Friedrich & Friederici, 2005; Rossi et al., 2011). In Rossi et al. (2011), phonotactically legal pseudowords elicited larger centro-parietal N400 than illegal pseudowords during passive listening. Moreover, cross-language studies have typically found language-specific response to native phonemes in later response windows (Dehaene-Lambertz et al., 2005; Näätänen et al., 1997; Sharma & Dorman, 2000; Wagner et al., 2012). Besides the averaged ERPs, time–frequency analysis of electroencephalographic (EEG) oscillatory dynamics can provide in-depth information about the neural coding of speech sounds and linguistic processes (Ding et al., 2016; Edwards et al., 2009). The intertrial phase coherence (ITPC) measures trial-by-trial oscillatory phase alignment across frequencies that ERP is incapable to capture. For sentence processing, research have established the role of theta frequency phase alignment for suprasyllabic processing of linguistic content and rhythm-based speech parsing (Doelling et al., 2014; Giraud & Poeppel, 2012; Peelle et al., 2013). For spoken word processing, theta ITPC was found enhanced for targets with incongruent emotional prosody from visual primes in the N400 and later response windows (Diamond & Zhang, 2016). A study of children with autism and speech impairment has found smaller theta ITPC post–200 ms for both pure tone and word processing (Yu et al., 2018). With the linguistic function of theta activities in mind, we additionally explored the potential role of theta ITPC in the nonemotional word prosody processing. We hypothesized that cross-linguistic effects in the late ERP windows for prosodic acoustics would indicate a presence of abstract prosodic knowledge; otherwise, the effects would only be observed in the obligatory N1–P2 window that reflects acoustic differences of the stimuli.

Method

Participants

The participants were 18 normal-hearing native Chinese-speaking young adults (10 females and eight males, age M = 21.6 years, SD = 1.7). All the participants reported having no speech or language disorders, medical history of neurological disorders, brain injury, or cognitive deficits. They all have received basic English-as-a-second-language education at school.

Stimuli and Procedure

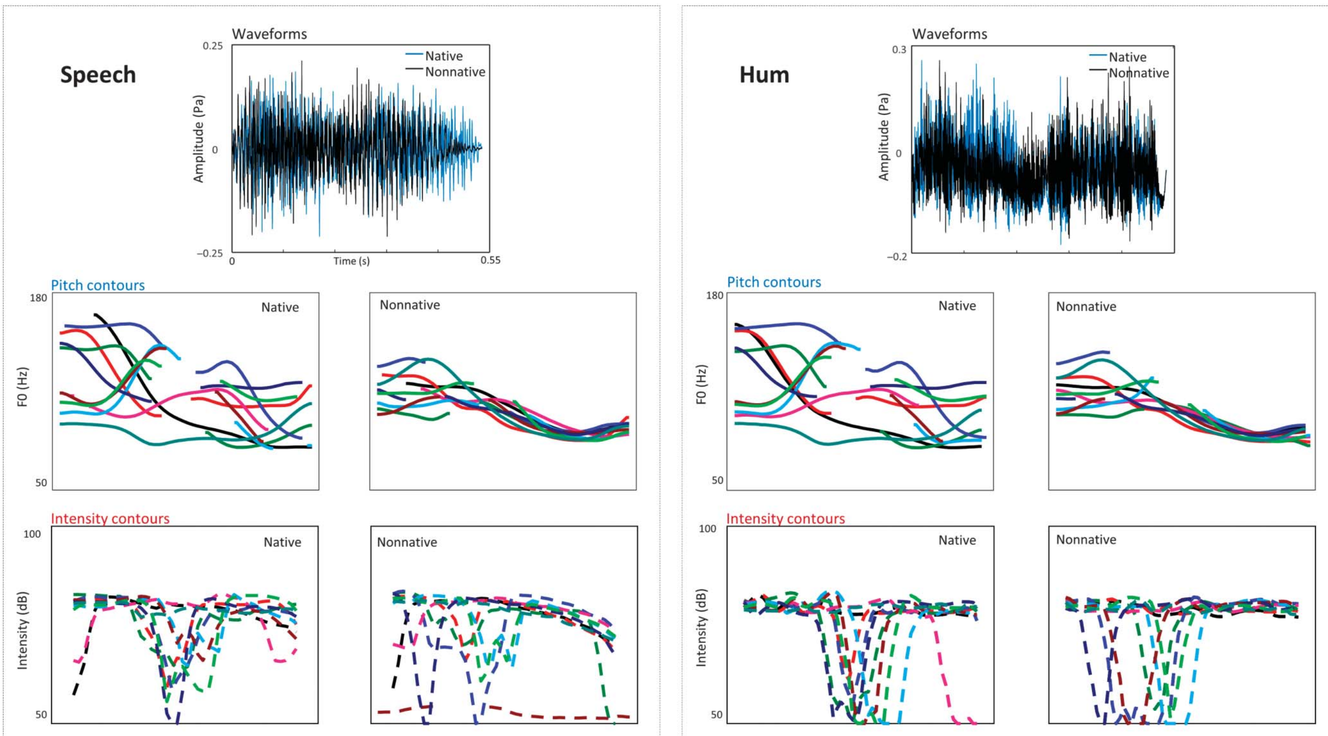
The experiment followed a 2 (type: speech vs. hum) × 2 (language: native vs. nonnative) design. The speech stimuli were 10 Chinese (native) nonsense disyllabic pseudowords and 10 English (nonnative) ones. Each pair of native and nonnative tokens matched on syllabic structure. For example, for a native token /mi3ling2/ in Pinyin, the paired English stimulus was "meeling." The utterances were first produced by a synthesized male voice using a commercial text-to-speech program and then edited to match overall intensity and duration. All tokens had a duration of 550 ms. The quality of the speech tokens was checked by native Chinese speakers and native English speakers and deemed high-quality utterances. The nonspeech stimuli were hummed versions of the speech stimuli. These hum stimuli were created using the "to sound (hum) function" in Praat (Boersma, 2002). To confirm that the hum stimuli can no longer be perceived as speech, five normal-hearing adults naïve to the stimuli rated the tokens using a 7-point Likert scale (1 = *definitely not speech*, 7 = *definitely speech*). The speech tokens received an average score of 6.6, and hum tokens received 2.1. Acoustic properties of the speech and hum stimuli were illustrated in Figure 1. The cross-linguistic differences of these properties were well characteristic of the language-specific realizations of Chinese versus English word prosody and largely consistent across stimulus types, therefore ensuring cross-linguistic comparison for both speech and hum.

Each token was presented 10 times, providing 100 trials for each stimulus condition. The interstimulus interval was randomized between 1000 and 1200 ms. Each condition was presented with 10 short blocks, resulting in 40 blocks in total. The speech and hum stimuli were presented separately with two long blocks between which the participants received short breaks. The stimuli were delivered through ER-1 ear inserts at 70 dB SPL. The participant was instructed to sit still and watch a muted movie of own choice while ignoring the sounds.

EEG Recording and Data Analysis

Continuous EEG was recorded using a 32-channel BrainAmps DC amplifier system at a 1000-Hz sampling rate (Brain Products). The left mastoid and AFz were used as the reference and ground, respectively. Electrode impedance was kept below 10 kΩ. Data analysis was performed using EEGLAB and ERPLAB (Delorme & Makeig, 2004; Lopez-Calderon & Luck, 2014). The data were re-referenced to linked mastoids and bandpass filtered

Figure 1. (Top) Sound waveform averaged across tokens. (Middle) Pitch contours of all native and nonnative tokens. (Bottom) Intensity contours of all native and nonnative tokens.



at 0.1 Hz–40 Hz. Epochs of 1200 ms including 200-ms baseline were extracted. Trials with instantaneous values exceeding ± 80 μV were rejected. ERP peak detection windows were determined based on the grand mean ERP and global field power (GFP) waveforms (see Figures 2 and 3). N1 and P2 peaks were searched within the poststimulus windows of 70–140 ms and 140–240 ms, respectively. Amplitude of the LNR or N400-like response was computed as the mean amplitude from 300 to 800 ms. Frontal–central midline electrodes Fz, FCz, and Cz where the ERP response maximized were chosen for statistical analysis. Trial-by-trial time–frequency analysis of ITPC was performed following the procedures described in Yu et al. (2018). The maximum theta ITPC values within the designated time windows of the N1–P2 and mean ITPC in the LNR of 300–800 ms were obtained for each participant in each condition for statistical analysis.

Linear mixed-effects (LME) regression was performed for each outcome measure. Language (native vs. nonnative), type (speech vs. hum), region (Fz, FCz, and Cz), and possible interactions were entered as fixed effects; participant was entered as a random effect. To examine the relationships between ERP amplitude and ITPC within the corresponding windows in each group, LME model with theta ITPC as predictor variable was fit for each ERP amplitude measure. In each LME model, language and type variables were first included as blocking variables; then, theta ITPC was entered as a fixed effect, and subject as a random effect. Two-tailed significance level was used for all analyses. Standard error bars on all bar plots were corrected for within-subject design (Cousineau, 2005).

Results

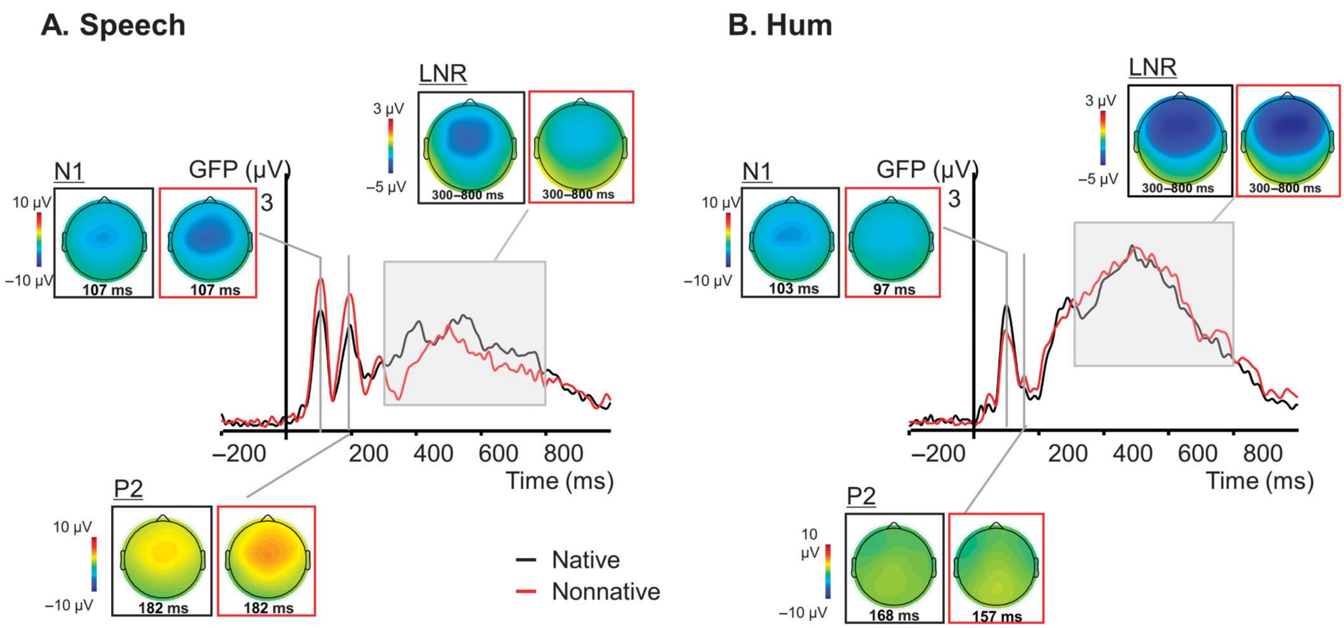
ERP Measures

N1 and P2 Responses

For the latency measure, LME regression revealed significant main effects of type that the hum stimuli elicited earlier N1 and P2 than the speech stimuli (N1, $F(1, 187) = 19.19$, $p < .001$; P2, $F(1, 187) = 202.77$, $p < .001$; see Table 1). No language, region, or interaction effect was observed for the latency measures ($ps > .368$).

For the N1 amplitude measure, there was a Language × Type interaction, $F(1, 187) = 41.06$, $p < .001$ (see Figure 3). Post hoc analysis indicated that nonnative stimuli elicited larger N1 than the native stimuli did in the speech condition, $F(1, 89) = 37.77$, $p < .001$, but the opposite in the hum condition, $F(1, 89) = 26.32$, $p < .001$. No other effect was found for the N1 amplitude ($ps > .104$). For the P2 amplitude, the analysis revealed significant

Figure 2. GFP waveforms in the (A) speech condition and (B) hum condition. The vertical lines mark the N1–P2, and the shaded areas represent the LNR intervals. GFP = global field power; LNR = late negative response.



region effect, $F(1, 187) = 3.31$, $p < .05$, with maximum response at the FCz electrode. There was also a Language × Type interaction, $F(1, 187) = 14.33$, $p < .001$, that nonnative stimuli produced larger P2 than native stimuli in the speech conditions only, $F(1, 187) = 14.33$, $p < .001$, but not in the hum conditions, $F(1, 187) = 14.33$, $p < .001$.

LNR

The analysis showed significant Language × Type interaction, $F(1, 187) = 19.99$, $p < .001$ (see Table 2 and Figure 3). That is, native speech elicited larger LNR than the nonnative speech did, $F(1, 89) = 30.22$, $p < .001$, whereas no difference was found between the native and nonnative hums, $F(1, 89) = 2.19$, $p = .142$. Hum elicited overall greater LNR amplitude than speech (native, $F(1, 89) = 64.12$, $p < .001$; nonnative, $F(1, 89) = 238.2$, $p < .001$). No other interaction was found for this measure ($ps > .246$).

ITPC and Its Relationship With ERP Amplitude

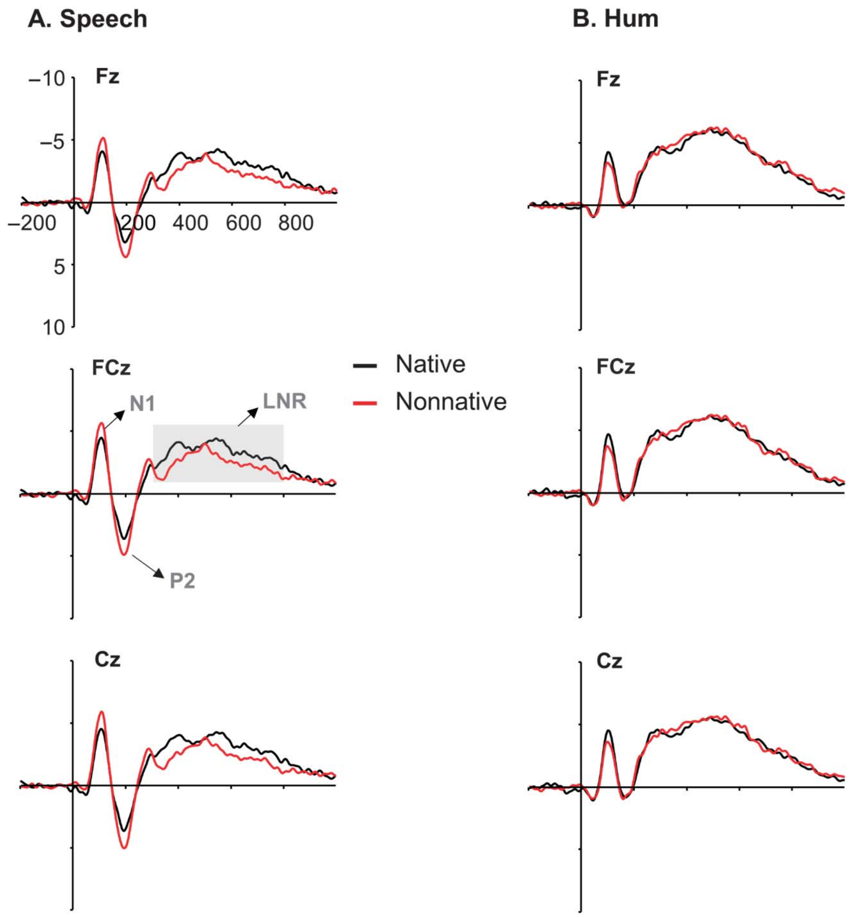
The analysis selected the FCz electrode as a representation of region of interest revealed by ERP results and scalp maps. In the N1 and P2 windows, speech sounds produced greater theta ITPC than the hum stimuli (N1, $F(1, 51) = 69.88$, $p < .001$; P2, $F(1, 51) = 75.65$, $p < .001$). No significant language effect was observed for either window (N1, $F(1, 51) = 0.19$, $p = .669$; P2, $F(1, 51) = 3.17$, $p = .081$). There was significant Language × Type interaction on theta ITPC in both N1 window, $F(1, 51) = 11.90$, $p < .01$, and P2 window, $F(1, 51) = 20.88$, $p < .001$. The nonnative sound elicited greater theta ITPC than the native sound did in the speech condition (N1, $F(1, 17) = 14.41$, $p < .01$; P2, $F(1, 17) = 37.93$, $p < .001$), but the ITPC difference patterns were reversed in the hum condition (N1, $F(1, 51) = 9.65$, $p < .01$; P2, $F(1, 17) = 7.62$, $p < .05$; see Table 1 and Figure 4). The LME regression model also demonstrated the predictive role of theta ITPC in N1, $F(1, 51) = 71.72$, $p < .001$, $\beta = -19.40$, and P2, $F(1, 51) = 8.64$, $p < .01$, $\beta = 10.64$, across language and stimulus type conditions. The β value indicated that greater theta synchrony was associated with greater N1 and P2 amplitude.

Similar to the LNR findings, hum produced greater theta ITPC than the speech sounds, $F(1, 51) = 5.45$, $p < .05$. Neither language effect, $F(1, 51) = 2.0$, $p = .163$, nor Language × Type interaction, $F(1, 51) = 2.35$, $p = .131$, was significant. The LME models did not reveal any significant relationship between theta ITPC and LNR amplitude, $F(1, 51) = 0.65$, $p = .423$ (see Figure 5).

Discussion

This study aimed to address whether representation of word prosody is independent of the phonetic representation by examining whether language-specific neural responses can be elicited by prosodic acoustics without any phonetic content. The results demonstrated that the LNR was sensitive to native versus nonnative prosodic phonology

Figure 3. Event-related potential waveforms for the native and nonnative sounds. The vertical lines mark the N1–P2, and the shaded area represents the LNR that differed in native versus nonnative speech but not in native versus nonnative hum. LNR = late negative response.



(speech) but not prosodic acoustics (hum). Additionally, "speechness" (speech vs. hum) modulated both the N1–P2 and LNR, but in the opposite directions.

We observed robust interaction between language and speechness on LNR that language-specific effects were only present for speech but not for hum. It could indicate sustained influence of prosodic phonology over the course of automatic auditory processing. However, it is likely that phonetic variations also contributed to the speech-specific LNR difference between the native and nonnative conditions, given that the phonetic/segmental details of the Chinese and English speech were phonologically not possible to be identical. A rule-out of this possibility would require comparison between native and nonnative speech with identical prosody. If such phonetic contrast no longer produced cross-linguistic effect, it would indicate LNR as a correlate for prosodic phonology processing; if equivalent LNR effects can be produced, it would speak against the existence of prosodic knowledge in the phonological space.

Table 1. Latency, amplitude, and theta ITPC measures of N1 and P2 responses at FCz.

		N1			P2		
Type	**Language**	**Amplitude**	**Latency**	**Theta ITPC**	**Amplitude**	**Latency**	**Theta ITPC**
Speech	Native	–5.03 (2.58)	109 (12)	0.32 (0.07)	3.87 (3.15)	196 (11)	0.32 (0.07)
	Nonnative	–6.13 (2.73)	108 (7)	0.36 (0.07)	5.31 (3.59)	196 (11)	0.38 (0.06)
Hum	Native	–5.18 (2.98)	105 (10)	0.26 (0.10)	1.10 (3.10)	172 (18)	0.27 (0.08)
	Nonnative	–4.21 (2.53)	104 (11)	0.23 (0.10)	1.21 (2.86)	166 (19)	0.25 (0.08)

Note. Standard deviation in parentheses. ITPC = intertrial phase coherence.

Table 2. Mean amplitude and theta ITPC measures of the LNR at FCz.

Type	Language	LNR (300–800 ms)	
		Amplitude	ITPC
Speech	Native	−3.40 (1.93)	−0.03 (0.01)
	Nonnative	−2.52 (1.86)	−0.02 (0.01)
Hum	Native	−4.73 (1.80)	−0.03 (0.02)
	Nonnative	−4.90 (1.93)	−0.03 (0.02)

Note. Standard deviation in parentheses. ITPC = intertrial phase coherence; LNR = late negative response.

A post–200-ms ERP negativity could also represent auditory N400 or N400-like activity. Given that Chinese is a monosyllabic language with a large number of monosyllabic homophones (Duanmu, 2007), when hearing the Chinese pseudowords, the listeners may try to find the possible meaning and elicit a greater auditory N400. For example, it has been shown that Chinese disyllabic pseudowords elicited greater N400 activity than real words in Chinese listeners, indicating unsuccessful lexical access for pronounceable potential words (C. Zhang et al., 2015). The nonnative speech, however, might be quickly rejected as legitimate words in the context of native language phonology therefore elicited little N400 activities. Previous reports that the N400 effect can be elicited in the absence of attention to stimuli (Deacon & Shelley-Tremblay, 2000; Rossi et al., 2011) also lends some support for this interpretation.

Nonetheless, the absence of language-specific effects of prosodic acoustics in the hum conditions clearly opposes the independence view of prosodic representation. That is, prosodic cues might not operate independently but likely rely on the phonetic content where the prosodic cues embedded in. This observation disagrees with previous observations using sentences that hum can produce the same ERP and behavioral effects related to prosodic boundary processing as natural speech did (Ischebeck et al., 2008; Pannekamp et al., 2005). The different result patterns might reflect the inherently different degrees of dependence/independence from phonetic content between sentence prosody and word prosody. Additionally, linguistic processing demand and processing mode might also play a part. When speech mode of processing is required as in the previous studies, prosodic acoustics in hums would be coded within the language networks and linguistic demand increases; therefore, it might produce the speech-like effects. In contrast, when speech mode is restricted as

Figure 4. Intertrial phase coherence at 0.5–40 Hz as a function of time.

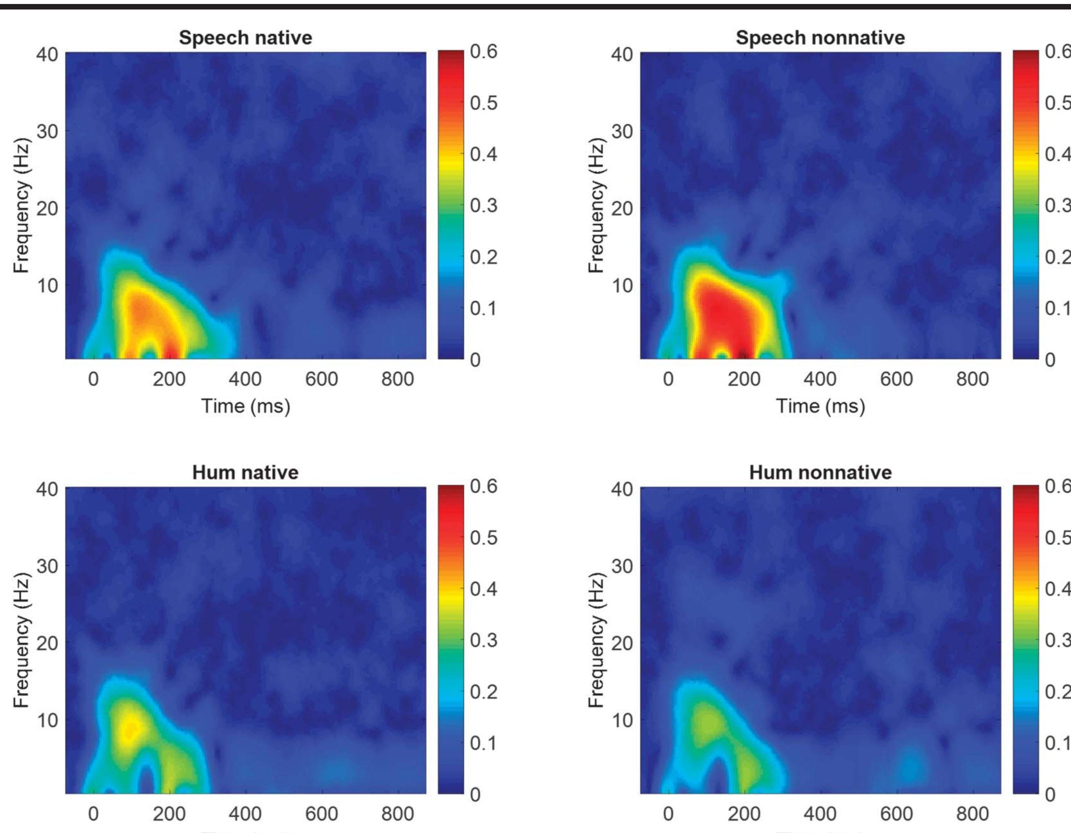
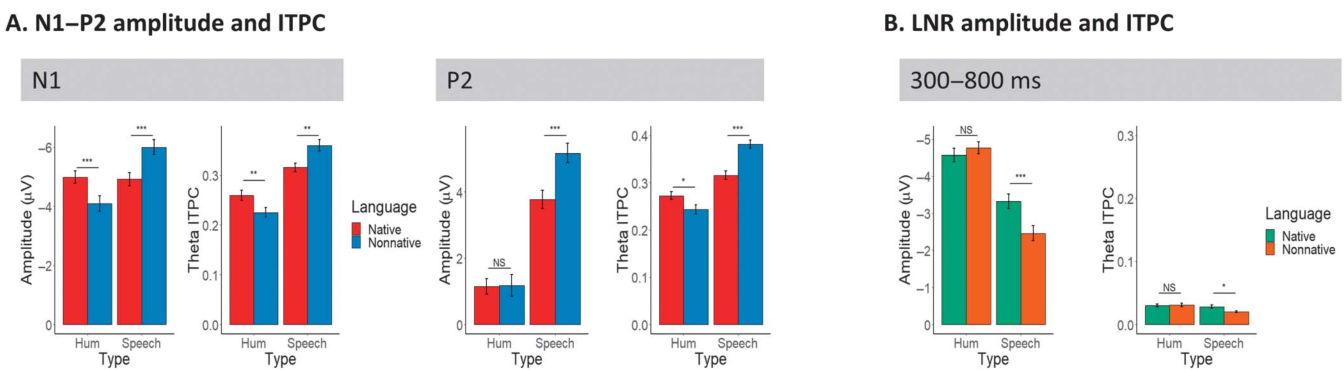


Figure 5. Bar graphs of event-related potential amplitude plotted in darker colors and the corresponding ITPC plotted in lighter colors. ITPC = intertrial phase coherence; LNR = late negative response.



in this study, prosody processing would be largely acoustic in nature. Another important relevant question is the time course of integration of phonemic and prosodic information for word processing. The current result pattern appears to suggest an early language-nonspecific processing stage for prosodic acoustics in the N1–P2 and a later phonological processing stage for integrated phonemic–prosodic object in the LNR. Although the current data cannot speak much about the potential substages within LNR, this rough timeline fits well with recent ERP reports on phonemic and prosodic processing. Using Chinese monosyllabic words, K. Zhang et al. (2021) found that linguistic context effects occurred earlier for vowels within 190–350 ms than for lexical tones within 220–500 ms, suggesting partially overlapped processes followed by integration after 200 ms.

Unlike previous work that showed left-hemisphere dominance for linguistic prosody processing when contrasted with nonlinguistic tasks (Arciuli & Slowiaczek, 2007; Gandour et al., 2004; Kreitewolf et al., 2014), we did not observe any hemisphere asymmetry in the ERPs to speech. Our result corresponds with studies using a passive listening procedure that does not require attending to speech content (Y. Zhang et al., 2005), and those using nonlinguistic prosody (Diamond & Zhang, 2016) and nonsense word stimuli (M. Friedrich & Friederici, 2005). The absence of leftward dominance was not uncommon, given that phonetic processing is thought to be bilateral in adults (Hickok & Poeppel, 2007).

Another major finding was that speech elicited larger N1 and P2 amplitude than hum, whereas the opposite was true for the LNR. The N1 component is sensitive to acoustic saliency (Näätänen & Picton, 1987; Pereira et al., 2014) and audibility (Martin et al., 1999). The result could indicate that speech is inherently more salient than hum to the human auditory system. The diminished P2 in response to hum could reflect the lack of acoustic richness compared to speech given that P2 has been found associated with acoustic complexity (Shahin et al., 2007). Nonetheless, the fact that P2 was virtually “shut down” by the hum stimuli let us lean toward the sound classification account that P2 might reflect the auditory system’s placement of sound categories (García-Larrea et al., 1992). Unlike speech, the hum sounds were unnaturalistic; therefore, they were unclassifiable by the auditory system.

Unlike the N1–P2, LNR amplitude was increased by hum as compared with the speech stimuli. Late ERP responses are generally driven by endogenous processes as oppose to exogeneous stimulus-driven processes, which can index higher order sensory integration and abstraction (Čeponienė et al., 2008). From a neural efficiency viewpoint, larger LNR might reflect widespread neuronal activation in the process of integrating unfamiliar auditory content. However, greater response to hum appears inconsistent with Ischebeck et al.’s (2008) observation that no cortical area activated more strongly for hum than for natural speech. As outlined in the introduction, different task demands could be responsible for the discrepancy. Their experiment employed a lexico-semantic task in which participants treated the hum material as sentences and judge whether target words had been contained in it (Ischebeck et al., 2008); therefore, the neural activation for hum might have been confined within the language regions as a result. Comparatively, the current task-free procedure had no such constrain on processing mode.

We additionally explored the trial-by-trial synchrony of oscillatory activities in relation to the auditory and phonological processing. We showed that greater theta ITPC was associated with larger N1 and P2 amplitude, which agrees with previous findings that oscillatory phase coherence contributes to the generation of sensory evoked potentials (Klimesch et al., 2007). The results also indicate that trial-by-trial synchrony became much weaker afterward. As the neural process moved up to higher levels in the auditory hierarchy, it became less phase-locked across trials with more sustained and variable time course. Moreover, there was no relationship or shared pattern of language effects between LNR and the associated theta ITPC. In fact, we observed trivial language effect in the ITPC. Therefore, at least in our case, theta synchrony may only have limited role for automatic processing of word prosody.

Our study has several limitations. First, the cross-language design did not have a non-Chinese listener group. The LNR increment for the native relative to the nonnative pseudowords could reflect Chinese listeners' syllable-to-syllable lexico-semantic processing other than phonological processes, since each syllable of the pseudowords is associated with individual Chinese characters with possible semantic interpretations. While this possibility exists, semantic access was unlikely during the passive listening given that no leftward lateralization in the ERP responses was observed at any point. Second, this study did not have a prosody control condition to tease apart to what extent phonetic variations have contributed to the language effect on LNR in the speech condition. Third, we did not have behavioral measures of abstract prosodic knowledge. The functional significance of the LNR effects in relation to behavior needs further examination. Nonetheless, language-specific LNR to word-level linguistic prosody may serve as a neural signature of language development of children and potentially predict later language skills (M. Friedrich et al., 2009). So far, little developmental research has examined the mechanisms of suprasyllabic prosodic processing in Chinese children. The current adult study provides a preliminary reference for future work in this area.

Conclusions

The results in the present ERP study suggest that acoustic cues of word-level prosody might not operate independently but likely rely on the phonetic content where the prosodic cues embedded in. Further cross-linguistic and developmental studies are needed to verify whether the LNR can serve as a neural signature of language-specific processing of prosodic phonology.

Acknowledgments

Data in this research note were part of Luodi Yu's doctoral dissertation at the University of Minnesota, Twin Cities. This work was supported by the National Natural Science Foundation of China (31728009 and 31900775), the University of Minnesota College of Liberal Art Brain Imaging Research Project award (Yang Zhang), and the U of M Grand Challenges Research Grant (Yang Zhang). The authors would like to thank Guangzhou Rehabilitation and Research Center for Children with Autism (Guangzhou Cana School) for providing research equipment and our participants, as well as the staff and students who assisted with data collection.

References

Arciuli, J., & Slowiaczek, L. M. (2007). The where and when of linguistic word-level prosody. *Neuropsychologia, 45*(11), 2638–2642. https://doi.org/10.1016/j.neuropsychologia.2007.03.010

Becker, A., Schild, U., & Friedrich, C. K. (2018). Tracking independence and merging of prosodic and phonemic processing across infancy. *Developmental Science, 21*(2), e12525. https://doi.org/10.1111/desc.12525

Boersma, P. (2002). Praat, a system for doing phonetics by computer. *Glot International, 5*(9/10), 341–345.

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-L-13-0279

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology, 1*(1), 42–45. https://doi.org/10.20982/tqmp.01.1.p042

Cutler, A., Dahan, D., & Van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech, 40*(2), 141–201. https://doi.org/10.1177/002383099704000203

Čeponienė, R., Torki, M., Alku, P., Koyama, A., & Townsend, J. (2008). Event-related potentials reflect spectral differences in speech and non-speech stimuli in children and adults. *Clinical Neurophysiology, 119*(7), 1560–1577. https://doi.org/10.1016/j.clinph.2008.03.005

Deacon, D., & Shelley-Tremblay, J. (2000). How automatically is meaning accessed: A review of the effects of attention on semantic processing. *Frontiers in Bioscience, 5,* e82–e94. https://doi.org/10.2741/A569

DeCasper, A. J., & Spence, M. J. (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behavior & Development, 9*(2), 133–150. https://doi.org/10.1016/0163-6383(86)90025-1

Dehaene-Lambertz, G., Pallier, C., Serniclaes, W., Sprenger-Charolles, L., Jobert, A., & Dehaene, S. (2005). Neural correlates of switching from auditory to speech perception. *NeuroImage, 24*(1), 21–33. https://doi.org/10.1016/j.neuroimage.2004.09.039

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. https://doi.org/10.1016/j.jneumeth.2003.10.009

Diamond, E., & Zhang, Y. (2016). Cortical processing of phonetic and emotional information in speech: A cross-modal priming study. *Neuropsychologia, 82,* 110–122. https://doi.org/10.1016/j.neuropsychologia.2016.01.019

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience, 19*(1), 158–164. https://doi.org/10.1038/nn.4186

Doelling, K. B., Arnal, L. H., Ghitza, O., & Poeppel, D. (2014). Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. *NeuroImage, 85*(Pt. 2), 761–768. https://doi.org/10.1016/j.neuroimage.2013.06.035

Duanmu, S. (2007). *The phonology of standard Chinese*. OUP Oxford.

Edwards, E., Soltani, M., Kim, W., Dalal, S. S., Nagarajan, S. S., Berger, M. S., & Knight, R. T. (2009). Comparison of time–frequency responses and the event-related potential to auditory speech stimuli in human cortex. *Journal of Neurophysiology, 102*(1), 377–386. https://doi.org/10.1152/jn.90954.2008

Friedrich, C. K., Kotz, S. A., Friederici, A. D., & Alter, K. (2004). Pitch modulates lexical identification in spoken word recognition: ERP and behavioral evidence. *Cognitive Brain Research, 20*(2), 300–308. https://doi.org/10.1016/j.cogbrainres.2004.03.007

Friedrich, M., & Friederici, A. D. (2005). Phonotactic knowledge and lexical-semantic processing in one-year-olds: Brain responses to words and nonsense words in picture contexts. *Journal of Cognitive Neuroscience, 17*(11), 1785–1802. https://doi.org/10.1162/089892905774589172

Friedrich, M., Herold, B., & Friederici, A. D. (2009). ERP correlates of processing native and non-native language word stress in infants with different language outcomes. *Cortex, 45*(5), 662–676. https://doi.org/10.1016/j.cortex.2008.06.014

Gandour, J., Tong, Y., Wong, D., Talavage, T., Dzemidzic, M., Xu, Y., Li, X., & Lowe, M. (2004). Hemispheric roles in the perception of speech prosody. *NeuroImage, 23*(1), 344–357. https://doi.org/10.1016/j.neuroimage.2004.06.004

García-Larrea, L., Lukaszewicz, A.-C., & Mauguiére, F. (1992). Revisiting the oddball paradigm. Non-target vs neutral stimuli and the evaluation of ERP attentional effects. *Neuropsychologia, 30*(8), 723–741. https://doi.org/10.1016/0028-3932(92)90042-K

Gervain, J., & Mehler, J. (2010). Speech perception and language acquisition in the first year of life. *Annual Review of Psychology, 61,* 191–218. https://doi.org/10.1146/annurev.psych.093008.100408

Giraud, A. L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience, 15*(4), 511–517. https://doi.org/10.1038/nn.3063

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience, 8*(5), 393–402. https://doi.org/10.1038/nrn2113

Ischebeck, A. K., Friederici, A. D., & Alter, K. (2008). Processing prosodic boundaries in natural and hummed speech: An FMRI study. *Cerebral Cortex, 18*(3), 541–552. https://doi.org/10.1093/cercor/bhm083

Kjelgaard, M. M., & Speer, S. R. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language, 40*(2), 153–194. https://doi.org/https://doi.org/10.1006/jmla.1998.2620

Klimesch, W., Sauseng, P., Hanslmayr, S., Gruber, W., & Freunberger, R. (2007). Event-related phase reorganization may explain evoked neural dynamics. *Neuroscience and Biobehavioral Reviews, 31*(7), 1003–1016. https://doi.org/10.1016/j.neubiorev.2007.03.005

Kreitewolf, J., Friederici, A. D., & von Kriegstein, K. (2014). Hemispheric lateralization of linguistic prosody recognition in comparison to speech and speaker recognition. *NeuroImage, 102*(Pt. 2), 332–344. https://doi.org/10.1016/j.neuroimage.2014.07.038

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology, 62,* 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

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. https://doi.org/10.3389/fnhum.2014.00213

Mampe, B., Friederici, A. D., Christophe, A., & Wermke, K. (2009). Newborns' cry melody is shaped by their native language. *Current Biology, 19*(23), 1994–1997. https://doi.org/10.1016/j.cub.2009.09.064

Martin, B. A., Kurtzberg, D., & Stapells, D. R. (1999). The effects of decreased audibility produced by high-pass noise masking on n1 and the mismatch negativity to speech sounds /ba/ and /da/. *Journal of Speech, Language, and Hearing Research, 42*(2), 271–286. https://doi.org/10.1044/jslhr.4202.271

Martin, B. A., Tremblay, K. L., & Korczak, P. (2008). Speech evoked potentials: From the laboratory to the clinic. *Ear and Hearing, 29*(3), 285–313. https://doi.org/10.1097/AUD.0b013e3181662c0e

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R. J., Luuk, A., Allik, J., Sinkkonen, J., & Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature, 385,* 432–434. https://doi.org/10.1038/385432a0

Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology, 24*(4), 375–425. https://doi.org/10.1111/j.1469-8986.1987.tb00311.x

Paiva, T. O., Almeida, P. R., Ferreira-Santos, F., Vieira, J. B., Silveira, C., Chaves, P. L., Barbosa, F., & Marques-Teixeira, J. (2016). Similar sound intensity dependence of the N1 and P2 components of the auditory ERP: Averaged and single trial evidence. *Clinical Neurophysiology, 127*(1), 499–508. https://doi.org/10.1016/j.clinph.2015.06.016

Pannekamp, A., Toepel, U., Alter, K., Hahne, A., & Friederici, A. D. (2005). Prosody-driven sentence processing: An event-related brain potential study. *Journal of Cognitive Neuroscience, 17*(3), 407–421. https://doi.org/10.1162/0898929053279450

Peelle, J. E., Gross, J., & Davis, M. H. (2013). Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cerebral Cortex, 23*(6), 1378–1387. https://doi.org/10.1093/cercor/bhs118

Pereira, D. R., Cardoso, S., Ferreira-Santos, F., Fernandes, C., Cunha-Reis, C., Paiva, T. O., Almeida, P. R., Silveira, C., Barbosa, F., & Marques-Teixeira, J. (2014). Effects of inter-stimulus interval (ISI) duration on the N1 and P2 components of the auditory event-related potential. *International Journal of Psychophysiology, 94*(3), 311–318. https://doi.org/10.1016/j.ijpsycho.2014.09.012

Protopapas, A., Panagaki, E., Andrikopoulou, A., Gutiérrez Palma, N., & Arvaniti, A. (2016). Priming stress patterns in word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 42*(11), 1739–1760. https://doi.org/10.1037/xhp0000259

Rossi, S., Jürgenson, I. B., Hanulíková, A., Telkemeyer, S., Wartenburger, I., & Obrig, H. (2011). Implicit processing of phonotactic cues: Evidence from electrophysiological and vascular responses. *Journal of Cognitive Neuroscience, 23*(7), 1752–1764. https://doi.org/10.1162/jocn.2010.21547

Schild, U., Becker, A. B. C., & Friedrich, C. K. (2014). Phoneme-free prosodic representations are involved in pre-lexical and lexical neurobiological mechanisms underlying spoken word processing. *Brain and Language, 136,* 31–43. https://doi.org/10.1016/j.bandl.2014.07.006

Shahin, A. J., Roberts, L. E., Miller, L. M., McDonald, K. L., & Alain, C. (2007). Sensitivity of EEG and MEG to the N1 and P2 auditory evoked responses modulated by spectral complexity of sounds. *Brain Topography, 20*(2), 55–61. https://doi.org/10.1007/s10548-007-0031-4

Sharma, A., & Dorman, M. F. (2000). Neurophysiologic correlates of cross-language phonetic perception. *The Journal of the Acoustical Society of America, 107*(5), 2697–2703. https://doi.org/10.1121/1.428655

Shukla, M., White, K. S., & Aslin, R. N. (2011). Prosody guides the rapid mapping of auditory word forms onto visual objects in 6-mo-old infants. *Proceedings of the National Academy of Sciences of the United States of America, 108*(15), 6038–6043. https://doi.org/10.1073/pnas.1017617108

Slowiaczek, L. M., Soltano, E. G., & Bernstein, H. L. (2006). Lexical and metrical stress in word recognition: Lexical or pre-lexical influences? *Journal of Psycholinguistic Research, 35*(6), 491–512. https://doi.org/10.1007/s10936-006-9026-7

Wade-Woolley, L. (2015). Prosodic and phonemic awareness in children's reading of long and short words. *Reading and Writing, 29*(3), 371–382. https://doi.org/10.1007/s11145-015-9600-1

Wagner, M., Shafer, V. L., Martin, B., & Steinschneider, M. (2012). The phonotactic influence on the perception of a consonant cluster /pt/ by native English and native Polish listeners: A behavioral and event-related potential (ERP) study. *Brain and Language, 123*(1), 30–41. https://doi.org/10.1016/j.bandl.2012.06.002

Yu, L., Wang, S., Huang, D., Wu, X., & Zhang, Y. (2018). Role of inter-trial phase coherence in atypical auditory evoked potentials to speech and nonspeech stimuli in children with autism. *Clinical Neurophysiology, 129*(7), 1374–1382. https://doi.org/10.1016/j.clinph.2018.04.599

Zhang, C., Xia, Q., & Peng, G. (2015). Mandarin third tone sandhi requires more effortful phonological encoding in speech production: Evidence from an ERP study. *Journal of Neurolinguistics, 33,* 149–162. https://doi.org/10.1016/j.jneuroling.2014.07.002

Zhang, K., Sjerps, M. J., & Peng, G. (2021). Integral perception, but separate processing: The perceptual normalization of lexical tones and vowels. *Neuropsychologia, 156,* 107839. https://doi.org/10.1016/j.neuropsychologia.2021.107839

Zhang, Y., Kuhl, P. K., Imada, T., Kotani, M., & Tohkura, Y. (2005). Effects of language experience: Neural commitment to language-specific auditory patterns. *NeuroImage, 26*(3), 703–720. https://doi.org/10.1016/j.neuroimage.2005.02.040