Prosodic Boundaries in Spoken Sentences: A Window into Speech Comprehension

Authors: Fang Lan, Yuanyi Zheng, Jin Han, Li Xiaoqing, Yang Yufang, Ruiming Wang, Wang Ruiming

Date: 2020-10-26T00:00:00+00:00

Abstract

Prosodic boundary processing is closely related to speech comprehension and has gradually become a research focus in psychology and linguistics over the past decade. The prosodic system comprises several prosodic units ranging from small to large, with prosodic constituents at different levels exhibiting varying boundary strengths, which are manifested in different parameters across three acoustic cues: pitch, lengthening, and pause. During sentence listening comprehension, listeners employ acoustic cue weighting strategies to process the acoustic cues of prosodic boundaries. At the neural level, the brain exhibits independent and specific neural mechanisms for prosodic boundary processing. The ability to process prosodic boundaries develops with age after birth, gradually deteriorates in old age, and appears to be transferable to a second language. Future research should expand the scope of investigation into the acoustic manifestations of prosodic boundaries, further clarify the processing mechanisms of prosodic boundaries, further elucidate the relationship between prosodic boundary processing and syntactic processing, and further attend to the development of prosodic boundary processing abilities in L2 learners.

Full Text

Prosodic Boundaries in Speech: A Window to Spoken Language Comprehension

FANG Lan[1],[2]; ZHENG Yuanyi[3],[4]; JIN Han[2]; LI Xiaoqing[3],[4]; YANG Yufang[3],[4]; WANG Ruiming[1]*

(1 Key Laboratory of Brain, Cognition and Education Sciences, Ministry of Education, & Center for Studies of Psychological Application, School of Psychology, South China Normal University, Guangzhou 510631)
(2 School of Foreign Studies, Guangzhou University, Guangzhou 510030)

(3 Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing 100101)
(4 Department of Psychology, University of Chinese Academy of Sciences, Beijing 100149)

Abstract

Prosodic boundary processing is closely related to spoken language comprehension and has gradually become a focal research topic in psychology and linguistics over the past decade. The prosodic system comprises hierarchical units of varying sizes, with different prosodic constituents exhibiting different boundary strengths, which are manifested through distinct parameters across three acoustic cues: pitch, final lengthening, and pause. During auditory sentence comprehension, listeners employ an acoustic cue weighting strategy to process prosodic boundary cues. At the neural level, the brain demonstrates independent and specific neural mechanisms for prosodic boundary processing. The ability to process prosodic boundaries develops with age after birth, gradually declines in older adulthood, and appears to be transferable to a second language. Future research should expand the scope of acoustic manifestations examined, further clarify the processing mechanisms of prosodic boundaries, elucidate the relationship between prosodic boundary processing and syntactic processing, and pay greater attention to the development of prosodic boundary processing abilities in second language learners.

Keywords: prosodic boundary; spoken language comprehension; acoustic characteristics; neural oscillation; perceptual development

1 Introduction

Prosody is central to spoken language comprehension (Frazier et al., 2006; Holzgrefe et al., 2013). As a crucial component of spoken prosodic features, prosodic boundaries serve as an organizational framework that segments continuous speech into hierarchical chunks (Frazier et al., 2006). Both the placement and acoustic realization of these boundaries influence listeners' sentence comprehension. For instance, prosodic boundaries can facilitate sentence understanding, resolve syntactic ambiguities, help listeners decompose continuous speech signals, extract meaningful units, and thereby acquire language (Holzgrefe-Lang et al., 2016; Wagner & Watson, 2010). Prosodic boundary processing ability is also closely related to reading fluency (Holliman et al., 2017). Clearly, to understand prosody and language processing, prosodic boundaries represent an indispensable window. Researchers have conducted numerous studies on prosodic boundaries, particularly in the past decade, with technological advances yielding many new discoveries.

2 The Hierarchical Structure of Prosodic Boundaries

The Strict Layering Hypothesis posits that the prosodic system contains six hierarchical levels: syllable, foot, clitic, phonological phrase, intonational phrase, and utterance (Selkirk, 2005). In phonological structure, each prosodic constituent is composed of constituents from the next lower level, with each constituent possessing its own boundary. Constituents at the same level are distinguished through these boundaries. Thus, prosodic boundaries, together with other prosodic features such as stress, constitute the framework for lexical segmentation and syntactic parsing in spoken communication.

The hierarchical levels of prosodic boundaries can be distinguished by both speakers and listeners in oral communication, based on boundary strength corresponding to each level, which is represented by parameters of boundary acoustic cues (Cole, 2015; Krivokapic, 2012). These cues include: (1) pitch change, manifested through pitch rise, pitch fall, or pitch reset (fall followed by rise back to the original high level) to indicate the presence of a boundary tone (also called f0, with the former described from a perceptual perspective and the latter from an articulatory perspective); (2) duration, referring to the lengthening of the pre-boundary syllable; and (3) pause, the intermittent interruption between two prosodic units (Männel & Friederici, 2016).

Researchers have investigated the relationship between prosodic hierarchy and grammatical hierarchy, concluding that they are related but not in one-to-one correspondence. Some researchers argue that a mapping relationship exists between prosodic hierarchy and syntactic structure, with each prosodic level having a corresponding syntactic unit. Prosodic structure boundaries often align with syntactic structure boundaries, particularly in read speech. Based on this phenomenon, Truckenbrodt (1995) proposed the Generalized Wrap Theory, suggesting that intonational phrase boundaries tend not to break syntactic phrases, meaning prosodic boundaries align with syntactic nodes. However, Frazier et al. (2004) manipulated intonational phrase boundary positions based on grammatical nodes and sense units respectively, using naturalness judgment tasks, and found that syntactic nodes are not absolute determinants of prosodic boundary placement. Clearly, prosodic structure is far less complex than syntactic structure, making a simple one-to-one relationship between prosodic and syntactic boundaries impossible. Moreover, everyday experience shows that the relationship between prosodic and syntactic boundaries is even looser in natural conversation. What patterns underlie such linguistic phenomena remains to be further explored.

3 Processing Cues for Prosodic Boundaries

During speech perception, although multiple acoustic cues are presented simultaneously, listeners assign different weights to these cues based on their discriminative power. This theory is known as the Speech Perceptual Cue Weighting Theory (Qin et al., 2016). According to this theory, speech perception is multidimensional, with different categories and languages assigning different weights to acoustic cues, making it an important theoretical framework for studying the perception of boundary acoustic cues.

Native adult speakers' utilization of prosodic boundary acoustic cues is language-specific. For example, English native speakers rely more on pre-boundary segment lengthening than on pause when judging clause boundaries (Aasland & Baum, 2003), whereas for Mandarin native speakers, pause is the most important boundary cue (Yang et al., 2014). German native speakers use both pause categories and pre-boundary segment lengthening to judge boundaries, with less reliance on pitch (Petrone et al., 2017). Notably, most prosodic boundary perception studies have used the presence or absence of a particular cue to determine its importance, potentially overlooking acoustic thresholds. That is, within a certain range, a particular acoustic cue may be effective, but beyond that range, it loses its effectiveness. Petrone et al. (2017) were the first to use the concept of threshold to explore acoustic cue weighting, providing a new direction for future research: we can further deepen boundary cue perception research from categorical and continuous perspectives.

There is also debate about whether prosodic boundary processing relies on local or global cues. One view (e.g., Marcus & Hindle, 1990) holds that prosodic boundary processing is based on cues immediately adjacent to the boundary, without requiring other prosodic information. Thus, prosodic boundary processing is local and context-independent. The aforementioned studies on boundary perception weighting are based on this view. Another view argues that prosodic boundary processing is holistic, with the strength of adjacent prosodic phrase boundaries influencing prosodic boundary perception. For example, in the spoken utterance "old men#1 and women #2 with very large houses," if the boundary strength at #1 is greater than at #2, the phrase means "old men and women with very large houses," with no information about whether the old men have large houses. If the boundary strength at #2 is greater than at #1, the meaning becomes "old men with very large houses and (old) women," where the women also have large houses (Clifton et al., 2002). In other words, listeners need to process prosodic boundary cues within a global framework.

Researchers have also explored the relationship between prosodic boundary processing and syntactic processing. Prosodic boundaries serve as an organizational framework for segmenting speech streams, while syntax constitutes the rules of language organization. Both have organizational functions and thus influence each other. On one hand, syntax has a predictive effect on intonational boundary perception. Cole et al. (2010) analyzed data from the Buckeye corpus and transcription results from 97 participants, finding that clause syntactic boundaries are the strongest predictor of boundary perception. Buxó-Lugo & Watson (2016) further discovered that expectations of syntactic phrase boundary positions also influence intonational boundary perception: participants' efficiency in perceiving prosodic boundaries at syntactically licensed positions was significantly higher than at unlicensed positions. For example, in the spoken sentences

(1) "Put the big#1 bowl on the tray" and (2) "Put the bowl that' s big#2 on the tray," prosodic boundaries were placed after "big." Boundary #1 is at a syntactically unlicensed position, while #2 is at a licensed position. Using acoustic software to synthesize prosodic boundaries of different strengths (by changing F0 contour, pause duration, and pre-boundary vowel lengthening ratio), the results showed that in prosodic boundary judgment tasks, regardless of boundary strength level, participants always reported prosodic boundaries more frequently at licensed than unlicensed positions. This indicates that listeners' detection of prosodic boundaries is not strictly driven by auditory factors but is influenced by syntactic factors. Clear syntactic structure has also been shown to facilitate prosodic boundary perception (Honbolygo et al., 2016).

On the other hand, prosodic boundaries influence the structural parsing of spoken sentences. Kjelgaard & Speer (1999) used garden-path sentences as materials and found that appropriate prosodic boundaries could overcome syntactic ambiguity. The authors argued that during sentence processing, prosodic boundaries act as indicators of sentence structure. This study suggests that prosodic boundaries are representations of syntactic structure at the phonetic level, allowing us to glimpse deep syntactic structure information through prosody. This effect was subsequently confirmed by ERP evidence in Steinhauer et al. (1999). The Boundary Deletion Hypothesis further strengthened the role of prosodic boundaries in syntactic parsing: adding a boundary where none should exist takes more time than missing a prosodic boundary (i.e., failing to pause when one should) (Pauker et al., 2011; Bögels et al., 2013). These studies suggest that in spoken language, prosodic structure may play a role in sentence structure parsing prior to syntax-driven parsing, rather than merely serving an auxiliary role in sentence comprehension. Later, Webman-Shafran & Fodor (2016) found that when processing short sentences with double prepositional phrase constructions such as "hid his anger about the divorce from Tami," if a prosodic boundary precedes the second prepositional phrase ("from Tami"), listeners tend to attach the second prepositional phrase to "hid" (high attachment). Without such a boundary, they treat it as a modifier of "the divorce" (low attachment). This result clearly challenges the Late Closure principle in linguistics. According to Late Closure, the second prepositional phrase in a double prepositional construction should be low-attached (i.e., modifying the nearest unit). However, due to the presence of a prosodic boundary, attachment of the second prepositional structure becomes ambiguous: when no prosodic boundary precedes the second prepositional structure, listeners adopt a low-attachment parsing strategy; when a prosodic phrase boundary exists, listeners tend to adopt a high-attachment strategy (Webman-Shafran & Fodor, 2016). This study further highlights the role of prosodic boundaries in sentence comprehension—they indeed determine sentence structure and meaning. Such results complicate our understanding of the role of prosodic boundaries in sentence processing but also provide new perspectives for sentence processing models.

The Module Approach and Interactive Approach represent two major theoretical schools concerning sentence processing mechanisms. The Module Approach posits that when processing sentences, listeners first establish a structural framework based on linguistic syntax, then assign words to different syntactic positions, and finally extract semantics to compose sentence meaning. The Interactive Approach argues that the application of syntactic and non-syntactic information (such as semantics, pragmatics, and prosody) is parallel. Existing models emphasize the contribution of semantics/pragmatics but discuss prosodic contributions very little (the two theories alternately become mainstream in relevant fields, with their pendulum-like history see Clifton & Duffy, 2001). However, as described above, prosodic boundaries participate in syntactic parsing through prosodic chunking during spoken sentence comprehension, influencing listeners' understanding of sentence meaning. Specifically, listeners first perceive acoustic features and then project them onto sentence structure to complete prosodic chunking, yielding parsing results. Yet existing module-based processing models overlook the role of spoken prosody. For spoken language processing, prosodic boundaries can also segment different syntactic constituents, thereby facilitating the assignment of lexical items to different positions and syntactic roles, so prosodic rules should be considered to update models. On the other hand, for the Interactive Approach, prosodic boundaries clearly constrain and guide linguistic syntactic parsing, with both participating in online sentence processing in parallel, supporting its core argument. However, how prosody interacts and integrates with other dimensional information deserves in-depth exploration. Overall, traditional sentence processing models have focused primarily on the contributions of linguistic syntax and lexical meaning to sentence processing, while discussions on the interactive relationship between prosodic boundaries and syntax will have important implications for expanding existing methodologies in spoken sentence processing (for specific models regarding module and interactive approaches see Clahsen & Felsh, 2017).

4 Neural Responses to Prosodic Boundary Processing

In the past decade, researchers have paid particular attention to the neural responses underlying prosodic boundary processing, exploring how prosodic boundaries are acquired by the human auditory system to facilitate speech perception and what the underlying cognitive neural mechanisms are.

4.1 CPS and Prosodic Boundary Processing

First, researchers raised a key question: Is prosodic boundary processing an independent mechanism or merely a "byproduct" of spoken language processing? In other words, are there specific cognitive neural mechanisms for prosodic boundary processing? To address this, researchers employed ERP (Event-related Potential) technology. Steinhauer et al. (1999) first reported the specific EEG component for prosodic boundary processing, CPS (Closure Positive Shift), in Nature's subsidiary journal Neuroscience. The study required 20 college student participants to judge whether prosody was appropriate during the experiment. Results revealed a specific EEG component at intonational boundaries with a latency of 400-500 ms, distributed over mid-posterior scalp regions. To 剥离 semantic and syntactic influences and verify the specificity of CPS, Pannekamp et al. (2005) conducted an ERP study examining participants' processing of four types of sentences with progressively decreasing information: (1) normal sentences, (2) jabberwocky sentences where content words were replaced with pseudowords (no semantic information), (3) pseudo sentences where both content and function words were replaced with pseudowords (no syntactic or semantic information), and (4) hummed sentences composed entirely of humming sounds (only prosodic information). The results showed that CPS components were observed at prosodic phrase boundaries in all sentence types, including hummed sentences. This indicates that CPS can be elicited even in the absence of syntactic and semantic information, suggesting that prosodic boundary processing exists independently of syntactic, semantic, and phonemic processing. Honbolygo et al. (2016) further verified the independent existence of prosodic boundary perception using both real and pseudo sentences with embedded clause structures (e.g., "The grandfather, who entered, was thirsty."). CPS has also been shown to have cross-linguistic universality, having been confirmed in English (e.g., Pauker et al., 2011; Steinhauer et al., 2010), Dutch (e.g., Bögels et al., 2010), Chinese (Li & Yang, 2009), and Korean (Hwang & Steinhauer, 2011).

To further explore the relationship between CPS and different hierarchical levels of prosodic boundaries, Li and Yang (2009) used Chinese sentences as materials and found that different hierarchical levels of prosodic boundaries could stably elicit CPS. Even when pause cues at boundaries were removed, CPS still occurred. Moreover, there was a systematic relationship between CPS latency and prosodic hierarchy: as prosodic level increased (prosodic phrase, intonational phrase, clause boundary, utterance boundary), CPS latency increased. This suggests that the human brain can not only independently perceive prosodic boundaries but also specifically represent different hierarchical levels of prosodic boundaries, thereby segmenting and integrating long speech streams into multi-level nested linguistic structures.

4.2 Neural Oscillatory Activity and Prosodic Boundary Processing

While CPS reflects time-locked and phase-locked neural responses related to prosodic boundaries, neural oscillatory activity can further reveal time-locked but non-phase-locked rhythmic neural activity in the brain, more intuitively reflecting the brain's responsive activity to prosodic boundaries in auditory speech signals. Researchers using EEG and MEG to record brain signals while participants listened to speech found that neural oscillatory activity in human auditory cortex showed strong synchronous rhythmic changes with low-frequency (<10 Hz) speech envelopes (e.g., Brodbeck et al., 2018; Ding et al., 2016; Ding et al., 2018). This frequency band corresponds to the rhythm of different linguistic levels segmented by prosodic boundaries: for example, delta band (0-3 Hz) modulation envelopes are related to sentence or phrase rhythm, while theta

(4-10 Hz) modulation envelopes are related to syllable rhythm (Poeppel, 2003). Researchers metaphorically describe this mechanism as the brain' s "entrainment" effect on speech signals, suggesting that neural oscillatory activity acts like a "neural recorder" that reconstructs the rhythmic changes of speech in the nervous system. Based on this mechanism, the auditory system can obtain prosodic boundary perception, segmenting the input long speech stream into small-scale speech units for processing. Entrainment has been shown to optimize speech comprehension, being stronger when listening to meaningful speech than meaningless speech, and low-frequency entrainment can predict speech comprehension performance (Gross et al., 2013; Park et al., 2015).

Although entrainment is closely related to prosodic boundary perception, the underlying processing mechanism remains unclear. One view holds that entrainment is purely caused by acoustic envelopes and has no necessary connection with linguistic information comprehension, as people can produce entrainment when listening to incomprehensible reversed speech (Howard & Poeppel, 2010). Another view supports a close relationship between entrainment and language comprehension, arguing that entrainment only occurs under intelligible speech conditions or is stronger under high-intelligibility conditions (e.g., Gross et al., 2013; Park et al., 2015). For example, Doelling et al. (2014) used linguistic materials (digit sequences presented at delta rhythm, around 3 Hz) and found that after removing rhythmic fluctuation cues related to temporal information, the brain' s entrainment activity in the delta band (2-4 Hz) significantly decreased, while adding energy fluctuation cues restored entrainment. Additionally, the strength of this entrainment activity positively correlated with listeners' perceptual ratings of speech sharpness and intelligibility. Regarding the mechanism of such entrainment in speech comprehension, Giraud and Poeppel (2012) proposed the Oscillation-based Hierarchical Binding Model, suggesting that envelope modulation signals at the left edge of speech units can cause phase reset of delta-theta band (1-8 Hz) neural oscillatory activity in the auditory system, which then adjusts the amplitude of gamma band (30-70 Hz) neural oscillatory activity, aligning neural excitatory fluctuations with acoustic envelope fluctuations and providing a "temporal framework" and "neural excitability preparation" for processing subsequent speech information. In this process, the auditory cortex achieves synchronous "tuning" to the speech signal envelope, producing entrainment and thereby segmenting and decoding speech stream information.

Bottom-up parsing of acoustic envelope signals is not the only cause of entrainment; certain top-down processes have also been shown to induce brain entrainment to speech. For example, Ding et al. (2016) investigated whether entrainment still exists when prosodic boundary acoustic cues are removed. They used fixed-rhythm four-character Chinese short sentences as materials (e.g., "sheep eat grass"), making syllable, lexical, and sentence rhythms 4 Hz, 2 Hz, and 1 Hz respectively, retaining only semantic and syntactic cues while removing acoustic cues like pauses at boundaries and pre-boundary syllable lengthening. Analysis based on neural oscillatory activity found significant entrainment at rhythms corresponding to different hierarchical speech structures (4, 2, 1 Hz). Ding and Jin (2019) further showed that entrainment is related to rule-based speech chunking patterns based on syntax or task, rather than being a byproduct of processing lexical-semantic connections. In cocktail party scenarios, listeners show better entrainment to attended speech streams (Brodbeck et al., 2018; Ding et al., 2018). Other evidence from acoustic cue perspectives shows that besides envelope modulation rhythm information, fine frequency information representing only "phonemic" features unrelated to prosodic boundaries can also lead to entrainment (Ding et al., 2014; Zoefel & VanRullen, 2015, 2016). These findings indicate that top-down speech, semantic, and syntactic analysis processes, as well as selective attention, can modulate neural activity for speech stream segmentation, leading to entrainment. These results suggest that entrainment across different frequency bands is more likely a "mixture" of the brain' s interactive process between bottom-up processing and top-down regulation for speech stream segmentation and reorganization, meaning that the brain's entrainment to speech envelopes also contains certain "active" perceptual processing components.

Does entrainment also reflect the top-down processing of prosodic boundaries themselves? Some existing evidence suggests that entrainment is indeed related to temporal expectations of prosodic boundaries. Researchers divided continuous speech streams into context and target phrases, finding that listeners could predict subsequent speech boundaries based on the segmentation patterns of prosodic boundaries in the context. For example, Dilley & Pitt (2010) used sentences where function word presence did not affect fluency (e.g., "Deena doesn' t have any leisure or time⋯"), slowing the rhythm of the context part ("Deena doesn' t have any lei⋯") to 1.9 times its original duration while keeping the target phrase rhythm unchanged ("-sure or time"), or keeping the context rhythm unchanged while speeding up the target word rhythm to 0.6 times its original duration. Influenced by context rhythm, participants were more likely to perceive no boundary between the suffix "-sure" and the function word "or," and less likely to report hearing "or." When both context and target word rhythms were changed, the probability of participants reporting "or" was no different from normal speech rate (similar results in Maslowski et al., 2019). Kösem and Wassenhove (2016) replicated this at the neural level using MEG recording methods, finding that when context rhythm remained unchanged but target word rhythm suddenly changed, the entrainment effect triggered by the original context boundary rhythm persisted for some time and influenced the speech perception outcome of the target word. These results suggest that entrainment is not only related to bottom-up processing of speech acoustic envelopes but may also reflect the influence of top-down, active temporal expectations on prosodic boundary perception.

Based on this, researchers have attempted to isolate frequency bands, time courses, and brain network characteristics related to the top-down processing mechanism of prosodic boundaries from this "mixture." Park et al. (2015) found that delta and theta band neural oscillatory activity might be related to top-down processing of prosodic boundaries. They had participants listen to a 7-

minute story while recording MEG signals. Using causal temporal correlation analysis and reversed speech listening as a baseline condition, they used the predictive power of neural signals in other brain regions on auditory cortex signal phases under certain time delay conditions as a neural index of top-down regulation (TDI, top-down index). Analysis locked to prosodic boundaries (edges) revealed that regulatory effects of different brain regions on left auditory cortex peaked before or after the boundary. The researchers speculated this might be related to the roles these brain regions play in temporal expectation. The left inferior frontal gyrus might regulate left auditory cortex activity through delta band oscillatory activity, generating temporal expectations before actual acoustic signal input, while the left premotor area and right fronto-temporal junction might regulate real-time temporal expectation matching processes through theta band oscillatory activity based on acoustic signals already input after the boundary, updating the temporal expectation model. Unfortunately, this study did not manipulate acoustic cues of prosodic boundaries, prosodic hierarchy, or other top-down factors, and thus could not provide direct evidence for temporal expectations of prosodic boundaries. Kayser et al. (2015) used story speech as material, manipulating the variability (jitter) of "pauses" between speech segments with jitter rates of 0%, 30%, 60%, and 90% while maintaining overall rhythm and comprehensibility. Analysis locked to anomalous "pauses" and the onset of subsequent words found that before the boundary (-0.16 to -0.06 s), alpha band (8-12 Hz) power decreased in left prefrontal brain regions, and after the boundary, phase consistency between delta band (0.5-2 Hz) neural oscillatory activity and speech envelope decreased in fronto-temporal brain regions, meaning entrainment to bottom-up input information weakened, with a significant positive correlation between the two. Entrainment activity and ERP effects in other frequency bands showed no significant differences compared to normal conditions. These results suggest that before acoustic signal input at speech boundaries, left prefrontal brain regions may have already generated temporal expectations of prosodic boundaries (indicated by decreased alpha band power) and transmitted expectation signals to auditory perception brain regions, thereby regulating online processing of prosodic boundaries in real time. Delta band neural oscillatory activity may be the product of the brain' s temporal expectations for prosodic boundaries based on speech rhythm, relatively independent of bottom-up acoustic signal processing. The above findings reveal that the perception-motor system from prefrontal to temporal brain regions may participate in the top-down temporal expectation process of prosodic boundaries, regulating online segmentation of speech streams through prefrontal alpha band, fronto-temporal delta band, or theta band neural oscillatory activity.

Researchers have rarely elaborated on the top-down expectation mechanism of prosodic boundaries from a theoretical perspective. However, the above findings show that listeners' bottom-up perceptual processing processes and results are influenced by top-down expectation processing, which may reflect a "synthesis-by-analysis" processing process. The Analysis-by-Synthesis Hypothesis (Halle & Stevens, 1962) posits that listeners, based on analysis of already obtained (acoustic or linguistic) context information, "synthesize" hypotheses about upcoming speech signal patterns. After actual signal input, listeners "analyze" based on internally modeled "expected" signals and external input signals to obtain actual perception results of long speech streams. When the two match, perceptual processing is facilitated. This "synthesis-by-analysis" approach has been proven to improve spoken language processing efficiency in semantic and syntactic processing. Does prosodic boundary processing also follow this synthesis-by-analysis process? How do the interactive processes of top-down and bottom-up processing mechanisms for prosodic boundaries work, and are there spatiotemporal characteristics that separate them? These questions remain to be studied.

4.3 The Relationship Between CPS and Neural Oscillations

Both CPS and low-frequency brain neural oscillatory activity are important neural indices of prosodic boundary perception. Recent studies further suggest there may be a certain correspondence between them, as both are more sensitive to prosodic boundaries above the phrase level. Some researchers have found that CPS effects elicited by prosodic boundary processing exhibit a "length constraint effect," meaning CPS effects only occur at boundaries of specific speech structure lengths. Li and Yang (2009) used Chinese sentences and found no significant CPS was elicited at prosodic word boundaries. Similarly, Holzgrefe et al. (2013) used bracketed lists ("Pat or Jay and Lee") as experimental materials and found that if the structure was parsed as (Pat) (or (Jay and Lee))—early closure structure—no CPS was elicited after "Pat," but if parsed as (Pat or Jay) and Lee, CPS was elicited after "Jay." The authors interpreted this as CPS reflecting the integration of prosodic boundary information during sentence comprehension, meaning that whether a boundary is used for sentence comprehension is the key to eliciting CPS. However, Pannekamp et al. (2005) confirmed that CPS exists independently of linguistic information, conflicting with this integration explanation. Therefore, the integration explanation seems insufficient. Another possibility is that only boundaries of certain speech structure lengths can elicit CPS, such as prosodic phrase level and above (e.g., Li & Yang, 2009). This explanation is consistent with results from implicit prosody studies by Hwang & Steinhauer (2011). In brain neural oscillatory activity, prosodic phrase level and above corresponds to the delta band (0-3 Hz), whose entrainment effect has been shown to play an important role in both bottom-up tracking (e.g., Doelling et al., 2014; Brodbeck et al., 2018) and top-down expectation (e.g., Kayser et al., 2015; Park et al., 2015) of prosodic boundaries.

Given that both CPS and delta band neural oscillatory activity are related to boundary perception at phrase level and above, what is their underlying cognitive significance? Considering the rhythmic characteristics of natural language and the brain' s language processing features, they may actually reflect the brain' s preference for chunking speech streams into phrase-sized units. Corpus analysis shows that intonational phrases (i.e., one intonational prosodic unit) in human language predominantly consist of 4-8 syllables (Zhou & Liu, 2017), which corresponds exactly to delta band rhythm. According to neurocognitive research findings, the brain has two optimal integration time windows for external information: a short window of 150-300 ms (right-lateralized) and a long window of 3 s, corresponding to theta band (4-8 Hz) and delta band (0-3 Hz) neural oscillatory activity (Yang, 2016). These frequency bands, especially delta band, have been shown to be related to temporal expectations of prosodic boundaries. A fixed-pace reading study found that people also produce CPS at phrase boundaries during reading, with CPS occurring cyclically at 2-3 s intervals (Roll et al., 2012; Schremm et al., 2015), consistent with delta band rhythm. These findings suggest that based on the chunking rhythm patterns in daily spoken language, the brain may prefer to segment and integrate speech streams using certain speech structure lengths as chunks (e.g., phrases), within which syllables and words are further segmented, or different chunks are connected to form higher-level sentence and paragraph structure perception. The close relationship between CPS, delta band neural oscillatory activity, and prosodic boundary processing above phrase level may reflect this chunking preference of the brain (Meyer et al., 2017). The relationship among these three could be further verified in future studies by manipulating the chunking patterns of language materials themselves (using languages with different rhythmic characteristics) or chunking patterns obtainable through implicit learning (inducing the brain to segment language according to certain patterns) and examining their relationship with brain CPS responses and delta band oscillatory activity. Additionally, whether this chunking preference is universal in language comprehension and how important it is for language understanding are also worth investigating.

5 Development of Prosodic Boundary Processing Ability

Prosodic boundaries are crucial for speech comprehension and represent linguistic cues used earlier than syntactic structure parsing in spoken language processing. However, the perception of prosodic boundaries is not innate but rather a result of postnatal learning that declines with age and is also influenced by other factors.

5.1 Development of Prosodic Boundary Processing in Native Speakers

Newborn infants can already extract syntactic information through prosodic boundaries. Although infants' brain processing of prosodic boundaries is still at a low level, they can segment speech streams by perceiving prosodic boundaries (Hawthorn & Gerken, 2014). However, children of different ages utilize acoustic cues differently. Four-month-old English infants rely on all three acoustic cues when discriminating clause boundaries (Seidl & Cristia, 2008), while six-month-old English infants mainly rely on pitch (Seidl, 2007). In another study with 3- and 6-year-old children, researchers found that 3-year-olds prioritized pause cues over length cues, while only by age 6 could children judge prosodic boundaries using just length and pitch without relying on pause (Holzgrefe-Lang et al., 2016; Männel et al., 2013). In other words, for children over 6 years old, the combination of pitch and length alone can achieve prosodic boundary perception and elicit CPS. Clearly, as age increases, children' s reliance on boundary acoustic cues changes. Ommena et al. (2020) compared German- and French-learning infants (6 and 8 months) in their perception of French bracketed lists and found that since German prosodic boundaries rely more on pre-boundary lengthening and pitch than French boundaries, 8-month-old German-learning infants developed sensitivity to pre-boundary lengthening and pitch, while French infants did not show this phenomenon. This result shows that the developmental change in acoustic cue weighting is influenced by the native language.

The above phenomenon is called developmental weight shift, which occurs in both phonetic and prosodic acquisition (Seidl, 2007). Seidl and Cristia (2008) found that children must undergo a developmental weight shift process to acquire the same prosodic boundary acoustic cue utilization strategy as adults, with the shift path being from holistic to analytic. Existing studies suggest that language proficiency level influences the development of children' s prosodic boundary perception ability, but it should be noted that cognitive development also occurs concurrently with language development in children. Whether cognitive development also plays a role in the development of prosodic boundary perception ability remains to be further studied.

On the other hand, Steinhauer et al. (2010) found that although prosodic boundary cues also guide syntactic analysis in older adults, helping them resolve syntactic ambiguity, compared to young adults, older adults' processing patterns are basically consistent with young people' s, but they are less sensitive to conflicts between syntax and prosodic boundaries. This result seems to indicate that contrary to infants' rapidly developing prosodic boundary perception ability with age, older adults' ability to use prosodic boundary cues declines with age.

In summary, prosodic boundary perception develops with age, but after reaching a certain age, it may decline with age. This seems to imply that prosodic boundary perception ability is related to cognitive development.

5.2 Development of Prosodic Boundary Processing in Second Language Learners

Prosodic boundaries provide key information about sentence structure and meaning and are important prosodic cues relied upon by native speakers in speech processing (Nickels et al., 2013). Do second language learners process target language prosodic boundaries like native speakers? The academic community holds different views on this question.

Nickels et al. (2013) used English native speakers and German-English (L2) learners as participants, with pause as the manipulation method for prosodic boundaries (intonational boundaries) and naturalness judgment as the experimental task to examine sensitivity to prosodic boundaries in listeners from different language backgrounds. Spoken sentences like (1) “When a bear is approaching the people#1 the dogs come running.” and (2) “When a bear is approaching #2 the people#3 come running.” were used, with four conditions: a. no intonational boundary in the sentence; b. boundary at #1; c. boundary at #2; d. boundaries at both #2 and #3. Results showed no significant difference between L2 learners and native speakers in naturalness judgment scores, with EEG components only slightly delayed in time compared to native speakers but showing the same overall pattern. In their further study (2018), participant groups were expanded to include native speakers and two learner groups with different language distances: German-English (L2) learners and Chinese-English (L2) learners, with other experimental components unchanged. Results showed that Chinese learners also exhibited behavioral and EEG results consistent with the trend shown by native speakers and German learners, but acceptance in condition d was significantly higher than for German and English participants, and the EEG amplitude reflecting the erroneous boundary (at #3) did not increase with target language proficiency like German learners, seeming to indicate a more bottom-up processing approach rather than the top-down predictive processing shown by the other two groups. The authors interpreted Chinese learners’ performance as possibly related to the verbs used or perception of pre-boundary prosodic contours. These two experiments seem to imply that the application of prosodic boundaries may involve not only the perception of boundary acoustic cues but also the integration of multiple dimensions and even multiple modalities of cues, and this process is influenced by native language experience. Future research should adopt a global perspective on prosodic boundary perception.

However, other researchers hold different views, arguing that L2 learners cannot use prosodic boundaries like native speakers. Pennington and Ellis (2000) selected Cantonese-English (L2) bilingual participants for a discrimination task where they judged whether the English sentence they heard matched a previously heard sentence. The listening task had two phases. In the first phase, participants listened to 24 target sentences. In the second phase, they listened to 48 test sentences and judged which ones had appeared in the first phase. Test sentences were divided into three categories: the first category had different words but the same syntax, the second category were sentences selected from target sentences, and the third category had the same words but different prosodic cues. Results showed participants performed far above chance level for the first and second categories but far below chance for the third category. Even when reminded to pay attention to prosodic differences (second experiment), results did not improve. The authors interpreted this as classroom foreign language learners, especially those in education environments with a recitation culture, being accustomed to learning prosodic features bundled with sentence meaning without specific training in perceiving phonetic cues. Additionally, learners might not understand the projection relationship between prosodic cues (including boundaries) and syntax/meaning. The authors did not further investigate whether learners' failure was due to inability to perceive boundary acoustic cues, inability to make correct projections, or excessive cognitive resource consumption by the task that left learners with no attention to spare for prosodic cues. Schmidt et al. (2019) derived their view from memory experiments. They used English native speakers and Greek-English (L2) learners as participants, with digit memory (similar to digit span memory measurement) as the task, comparing memory performance for digits in native speakers and foreign learners under conditions with and without prosodic boundaries. Results showed that both groups performed better on native digit memory with prosodic boundaries than without, but foreign digit memory performance was unaffected by the presence of prosodic boundaries. The authors argued that prosodic boundaries are deep structures, and therefore foreign language learners cannot apply them like native speakers. This conclusion is inconsistent with the traditional view that prosodic boundaries are surface structures of language. Undoubtedly, this view provides a new direction for understanding prosodic boundaries. These four studies leave multiple questions worth further exploration in the field of L2 prosodic boundary perception development.

Training can promote prosodic boundary perception ability. Yang (2016) found in classroom teaching that shadowing target language expressions or imitating natural sound patterns can help learners reproduce the connection between speakers' prosodic boundaries and syntactic comprehension in their minds. Current results have not revealed the mechanism by which training affects prosodic boundary perception. We do not know what factors affect boundary perception or which specific cues in boundary perception are affected. Given foreign learners' common complaints about listening difficulties and native speakers speaking too fast, these may also be caused by deficient prosodic boundary perception. The acquisition mechanism and training of prosodic boundary perception deserve more in-depth research.

6 Future Directions for Prosodic Boundary Research

As a physical manifestation of prosody, prosodic boundaries have more cognitive functions than stress and intonation, serving as an important window for observing spoken language processing and acquisition, and thus have been used as observation and manipulation objects in numerous studies. Through comprehensive summary and analysis of prosodic boundary research in psychology and linguistics, it is evident that many issues require further investigation.

6.1 Acoustic Manifestations of Prosodic Boundaries Require Further Investigation

Research results on acoustic signal processing of prosodic boundaries are abundant. Based on native speaker processing, researchers have conducted in-depth investigations into prosodic boundary hierarchical levels, acoustic cue representation of boundary strength, and perceptual weight distribution of prosodic boundaries (including single cues and cue combinations). These findings provide a solid empirical foundation for designing hearing assistive devices (such as cochlear implants), speech recognition, and artificial language synthesis. However, existing studies have not examined the influence of syntax, and given the close relationship between prosodic boundaries and language processing, this aspect needs strengthening. Additionally, most such research has been conducted in laboratories using read speech materials, reducing ecological validity. Future research could expand the scope by using oral language materials of different stylistic genres, such as storytelling or casual conversations between friends, and examining listeners' use of prosodic boundary acoustic cue characteristics in different background environments (e.g., with/without noise). Finally, future descriptions of acoustic cue perceptual characteristics could be further refined, not merely examining the effect of single cue presence/absence on boundary perception, but viewing the influence of acoustic cues on perception from a continuous perspective (e.g., threshold).

6.2 The Processing Mechanism of Prosodic Boundaries Needs Further Clarification

Previous researchers have primarily used CPS and low-frequency "entrainment" effects (i.e., neural oscillatory activity) to elaborate the processing mechanisms of prosodic boundaries, confirming that prosodic boundaries can be processed independently of other linguistic information and are closely related to linguistic structure perception and semantic or syntactic comprehension. However, many unresolved issues remain regarding the cognitive neural mechanisms of prosodic boundaries. First, many questions about the spatiotemporal characteristics of prosodic boundary cognitive neural mechanisms require further investigation. For example, previous studies have preliminarily found that entrainment can reflect both bottom-up processing of prosodic boundaries in auditory brain regions and top-down regulation of temporal-occipital regions by higher frontal brain regions based on expectations of prosodic boundaries, but systematic discussion is lacking on the specific time courses and brain network characteristics of each process when these two processes interact. Additionally, unlike other linguistic cues, acoustic prosodic boundaries do not have a one-to-one correspondence with semantic or syntactic boundaries. It remains unclear which process has higher relative weight in influencing prosodic boundary perception when expected prosody conflicts with bottom-up acoustic prosody. Furthermore, recent studies suggest that the relationship among CPS, delta band entrainment, and the brain' s preference for chunking language in phrase-sized units requires further verification. Expanding research to L2 learners and special populations will also help understand the relationship between brain processing characteristics of prosodic boundaries and language characteristics/development. Second, the cognitive importance of entrainment to prosodic boundaries remains unclear. In quiet environments, prosody is considered an auxiliary rather than necessary cue for spoken language comprehension. Even when prosodic pitch fluctuation information (Xu et al., 2013) or pause cues (e.g., Li & Yang, 2009; Kayser et al., 2015) are removed, listeners can still understand spoken sentences normally, while in noisy environments, prosodic structure perception ability correlates with spoken language perception performance (Slater & Kraus, 2016). Is top-down temporal expectation of prosodic boundaries a universal mechanism in prosodic boundary processing (i.e., not dependent on contextual conditions), or does it only play a role under difficult conditions (e.g., noise masking, degraded speech quality)? These questions await future research. Finally, much evidence shows that prosodic boundary processing affects higher-level language processing (e.g., semantic, syntactic processing). At the individual level, the ability to process linguistic prosodic temporal structure is related to speech perception and reading comprehension abilities (e.g., Cason et al., 2015; Cumming et al., 2015; Flaugnacco et al., 2015; Hidalgo et al., 2017; Schön & Tillmann, 2015; Slater & Kraus, 2016), suggesting that the specific mechanisms by which different individuals promote higher-level language functions through prosodic boundary processing may differ. However, current research rarely discusses this directly. Investigating this issue will help understand the potential relationship between prosodic boundary processing and other higher-level language cognitive functions and provide possible intervention solutions for populations with insufficient language ability development (e.g., children with dyslexia) or degradation (e.g., elderly people, patients with language disorders).

6.3 The Relationship Between Prosodic Boundary Processing and Syntactic Processing Needs Further Clarification

Existing research has found that prosodic boundaries affect sentence processing, syntax can predict prosodic boundaries, and the two constrain each other. These views are supported by behavioral experiments and neuroscience evidence. However, these research results are based on the assumption that semantics, syntax, and prosody are hierarchical (Ischebeck et al., 2008), with research frameworks focusing on the status and relationship of prosodic information (and other non-syntactic information, such as lexical, pragmatic, etc.) and syntactic information in sentence processing—that is, whether listeners first use syntactic information to decompose sentences or simultaneously use multiple information sources to parse sentences, with prosodic boundary cues in an auxiliary position. However, existing results show that the relationship between prosodic boundaries and syntax is not a simple one-to-one relationship but is much more complex than we know. Prosodic boundaries may not be merely a surface structure that participates in sentence comprehension through interaction with syntax, but may essentially be an embodied manifestation of language rules (Giraud & Poeppel, 2012; Kreiner & Eviatar, 2014), serving as an interface connecting abstract rules, the physical world, and human experience. The size of prosodic boundaries (represented by acoustic features) shows the closeness of relationships between sentence constituents, reproducing the hierarchical organizational structure of speech to reflect how language information is arranged in the time domain, facilitating the brain to capture abstract language organizational rules by simulating its rhythm through oscillations. For example, research on prosodic boundary length has found that the prosodic boundary-specific EEG component CPS cannot be elicited merely by acoustic cues but is a response to segmenting and integrating speech streams using certain-length chunks (e.g., phrases). Similarly, Ding et al. (2016) cross-linguistic research found that the strength of EEG components generated at sub-intonational prosodic boundaries was significantly greater than that generated at positions within prosodic units (see earlier text). Future research could break away from traditional structural linguistics frameworks and, from an embodied perspective, use EEG and MEG techniques to manipulate prosodic boundary perception using languages with different rhythmic patterns, filtering, and speech reversal, examining the relationship between neural oscillatory characteristics of prosodic boundary perception and syntactic parsing to deeply reveal the nature and function of prosodic boundaries and further clarify the relationship between prosodic boundary processing and syntactic processing.

6.4 Development of Prosodic Boundary Processing Ability Requires Further Attention

The ability to process acoustic cues of prosodic boundaries is closely related to language comprehension, but we know little about the development of prosodic boundary processing ability in L2 learners. Unfortunately, literature on L2 learners' processing of prosodic boundaries is extremely scarce. We are unclear about the specific aspects of processing differences between L2 learners and native speakers, such as whether they involve perception differences of specific acoustic cues. We also do not know the factors causing these differences. Are they caused by L2 learners adopting different cue weighting patterns from the target language, or by L2 learners adopting different processing mechanisms, such as different processing directions from native speakers? Although existing research suggests that L2 learners with large background language differences may be more inclined to use bottom-up cues (Nickels & Steinhauer, 2018), no study has specifically answered this question. Additionally, some have proposed that CPS is actually a manifestation of integration results (Nickels, 2013). Do L2 learners also have the same integration ability as native speakers? From a neurological perspective, do L2 learners and native speakers show different characteristics in time courses and brain networks for processing prosodic boundaries? These questions are all worth further exploration.

Listening comprehension difficulty has always been a major problem in L2 teaching. Foreign language learners often complain that native speakers speak too fast, possibly because L2 learners cannot effectively use prosodic boundaries to segment speech streams. If prosodic boundary perception ability can be trained, there will be a new solution to L2 listening difficulties. However, current research on the development process of L2 learners' prosodic boundary processing ability is almost non-existent. Does L2 learners' boundary perception ability develop holistically to locally like native speakers, or does it become fossilized due to native language transfer? Or does it gradually develop subskills that conform to target language processing, eventually approaching native speaker levels? If so, what is the developmental path of prosodic interlanguage, and is it influenced by native language experience to have specific developmental pathways? Can pedagogical intervention improve prosodic boundary processing ability? For example, can training on boundary acoustic cue perception enhance bottom-up processing ability, or can rhythm training increase top-down processing ability? To answer these questions, future research must focus on L2 learners' online processing mechanisms of prosodic boundary acoustic cues, acquisition processes, and the effectiveness of different training methods.

Additionally, attention should be paid to the potential impact of individual differences in acoustic cue perception to make teaching materials and instructional design more targeted.

References

Yang, Y. (2016). *Psycholinguistics.* Beijing: Science Press.

Zhou, Y., & Liu, F. (2017). Comparative analysis of Chinese prosodic structure and syntactic structure. In Chinese Information Processing Society of China, Speech Information Special Committee (Ed.), *Proceedings of the 14th National Conference on Man-Machine Speech Communication (NCMMSC'2017)* (pp. 205-210). Chinese Information Processing Society of China Speech Information Special Committee, Tsinghua National Laboratory for Information Science and Technology.

Aasland, W., & Baum, S. (2003). Temporal parameters as cues to phrasal boundaries: A comparison of processing by left-hemisphere-damaged and right-hemisphere-damaged individuals. *Brain and Language, 87*(3), 385-399.

Bögels, S., Schriefers, H., Vonk, W., Chwilla, D. J., & Kerkhofs, R. (2010). The interplay between prosody and syntax in sentence processing: The case of subject-and object-control verbs. *Journal of Cognitive Neuroscience, 22*(5), 1036-1053.

Bögels, S., Schriefers, H., Vonk, W., Chwilla, D. J., & Kerkhofs, R. (2013). Processing consequences of superfluous and missing prosodic breaks in auditory sentence comprehension. *Neuropsychologia, 51*(13), 2715-2728.

Brodbeck, C., Hong, L. E., & Simon, J. Z. (2018). Rapid transformation from auditory to linguistic representations of continuous speech. *Current Biology, 28*(24), 3976-3983.

Buxó-Lugo, A., & Watson, D. G. (2016). Evidence for the influence of syntax on prosodic parsing. *Journal of Memory and Language, 90*, 1-13.

Clahsen, H., & Felser, C. (2018). Some notes on the shallow structure hypothesis. *Studies in Second Language Acquisition, 40*(3), 693-706.

Cason, N., Astésano, C., & Schön, D. (2015). Bridging music and speech rhythm: Rhythmic priming and audio-motor training affect speech perception. *Acta Psychologica, 155*, 43-50.

Clifton, C., & Duffy, S. (2001). Sentence and text comprehension: Roles of linguistic structure. *Annual Review of Psychology, 52*, 167-196.

Clifton, C., Carlson, K., & Frazier, L. (2002). Informative prosodic boundaries. *Language and Speech, 45*(2), 87-114.

Cole, J., Mo, Y., & Baek, S. (2010). The role of syntactic structure in guiding prosody perception with ordinary listeners and everyday speech. *Language and Cognitive Processes, 25*(7), 1141-1177.

Cole, J. (2015). Prosody in context: A review. *Language, Cognition and Neuroscience, 30*(1), 1-31.

Cumming, R., Wilson, A., Leong, V., Colling, L. J., & Goswami, U. (2015). Awareness of rhythm patterns in speech and music in children with specific language impairments. *Frontiers in Human Neuroscience, 9*, 1-21.

Dilley, L. C., & Pitt, M. A. (2010). Altering context speech rate can cause words to appear or disappear. *Psychological Science, 21*(11), 1664-1670.

Ding, N., Chatterjee, M., & Simon, J. Z. (2014). Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure. *Neuroimage, 88*, 41-46.

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience, 19*, 158-164.

Ding, N., Pan, X., Luo, C., Su, N., Zhang, W., & Zhang, J. (2018). Attention is required for knowledge-based sequential grouping: Insights from the integration of syllables into words. *Journal of Neuroscience, 38*(5), 1178-1187.

Ding, N., & Jin, P. (2019). Low-frequency neural activity reflects rule-based chunking during speech listening. *bioRxiv*, Article 742585v1. https://www.biorxiv.org/content/10.1101/742585v1.full.pdf

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*, 761-768.

Flaugnacco, E., Lopez, L., Terribili, C., Montico, M., Zoia, S., & Schön, D. (2015). Music training increases phonological awareness and reading skills in developmental dyslexia: A randomized control trial. *PLOS ONE, 10*(9), Article e0138715. https://doi.org/10.1371/journal.pone.0138715

Frazier, L., Clifton, C., & Carlson, K. (2004). Don’ t break, or do: Prosodic boundary preferences. *Lingua, 114*(1), 3-27.

Frazier, L., Carlson, K., & Clifton, C. (2006). Prosodic phrasing is central to language comprehension. *Trends in Cognitive Sciences, 10*(6), 244-249.

Giraud, A. L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience, 15*(4), 511-517.

Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., & Garrod, S. (2013). Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLOS Biology, 11*(12), Article e1001752. https://doi.org/10.1371/journal.pbio.1001752

Halle, M., & Stevens, K. (1962). Speech recognition: A model and a program for research. *IRE Transactions on Information Theory, 8*(2), 155-159.

Hawthorn, K., & Gerken, L. (2014). From pauses to clauses: Prosody facilitates learning of syntactic constituency. *Cognition, 133*(2), 420-428.

Hidalgo, C., Falk, S., & Schön, D. (2017). Speak on time! Effects of a musical rhythmic training on children with hearing loss. *Hearing Research, 351*, 11-18.

Holliman, A. J., Gutiérrez Palma, N., Critten, S., Wood, C., Cunnane, H., & Pillinger, C. (2017). Examining the independent contribution of prosodic sensitivity to word reading and spelling in early readers. *Reading and Writing, 30*, 509-521.

Holzgrefe, J., Wellmann, C., Petrone, C., Truckenbrodt, H., Höhle, B., & Wartenburger, I. (2013). Brain response to prosodic boundary cues depends on boundary position. *Frontiers in Psychology Language Sciences, 4*, 421.

Holzgrefe-Lang, J., Wellmann, C., Petrone, C., & Räling, R. (2016). How pitch change and final lengthening cue boundary perception in German: Converging evidence from ERPs and prosodic judgments. *Language, Cognition and Neuroscience, 31*(7), 1-17.

Honbolygo, F., Török, Á., Bánréti, Z., Hunyadi, L., & Csépe, V. (2016). ERP correlates of prosody and syntax interaction in case of embedded sentences. *Journal of Neurolinguistics, 37*, 22-33.

Howard, M. F., & Poeppel, D. (2010). Discrimination of speech stimuli based on neuronal response phase patterns depends on acoustics but not comprehension. *Journal of Neurophysiology, 104*(5), 2500-2511.

Hwang, H., & Steinhauer, K. (2011). Phrase length matters: The interplay between implicit prosody and syntax in Korean "garden path"sentences. *Journal of Cognitive Neuroscience, 23*(11), 3555-3575.

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.

Kayser, S. J., Ince, R. A. A., Gross, J., & Kayser, C. (2015). Irregular speech rate dissociates auditory cortical entrainment, evoked responses, and frontal alpha. *Journal of Neuroscience, 35*(44), 14691-14701.

Kjelgaard, M. M., & Speer, S. R. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic ambiguity. *Journal of Memory and Language, 40*(2), 153-194.

Kösem, A., & Wassenhove, V. (2016). Distinct contributions of low- and high-frequency neural oscillations to speech comprehension. *Language, Cognition and Neuroscience, 32*, 536-544.

Kreiner, H., & Eviatar, Z. (2014). The missing link in the embodiment of syntax: Prosody. *Brain and Language, 137*, 91-102.

Krivokapic, J., & Byrd, D. (2012). Prosodic boundary strength: An articulatory and perceptual study. *Journal of Phonetics, 40*(3), 430-442.

Li, W. J., & Yang, Y. (2009). Perception of prosodic hierarchical boundaries in Mandarin Chinese sentences. *Neuroscience, 158*(4), 1416-1425.

Männel, C., Schipke, C. S., & Friederici, A. D. (2013). The role of pause as a prosodic boundary marker: Language ERP studies in German 3- and 6-year-olds. *Developmental Cognitive Neuroscience, 5*, 86-94.

Männel, C., & Friederici, A. D. (2016). Neural correlates of prosodic boundary perception in German preschoolers: If pause is present, pitch can go. *Brain Research, 1632*, 27-33.

Marcus, M., & Hindle, D. (1990). Description theory and intonation boundaries. In G. Altman (Ed.), *Cognitive Models of Speech Processing: Psycholinguistics and Computational Perspectives* (pp. 483-512). Cambridge, MA, USA: MIT Press.

Maslowski, M., Meyer, A. S., & Bosker, H. R. (2019). Listeners normalize speech for contextual speech rate even without an explicit recognition task. *The Journal of the Acoustical Society of America, 146*(1), 179-188.

Meyer, L., Henry, M. J., Gaston, P., Schmuck, N., & Friederici, A. D. (2017). Linguistic bias modulates interpretation of speech via neural delta-band oscillations. *Cerebral Cortex, 27*(9), 4293-4302.

Nickels, S., Opitz, B., & Steinhauer, K. (2013). ERPs show that classroom-instructed late second language learners rely on the same prosodic cues in syntactic parsing as native speakers. *Neuroscience Letters, 557*, 107-111.

Nickels, S., & Steinhauer, K. (2018). Prosody–syntax integration in a second language: Contrasting event-related potentials from German and Chinese learners of English using linear mixed effect models. *Second Language Research, 34*(1), 9-37.

Ommena, S. van., Boll-Avetisyanb, N., Larrazaa, S., Wellmannb, C., Bijeljac-Babica, R., Höhleb, B., & Nazzi, T. (2020). Language-specific prosodic acquisition: A comparison of phrase boundary perception by French- and German-learning infants. *Journal of Memory and Language, 112*, 104-108.

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.

Park, H., Ince, R. A., Schyns, P. G., Thut, G., & Gross, J. (2015). Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. *Current Biology, 25*(12), 1649-1653.

Pauker, E., Itzhak, I., Baum, S. R., & Steinhauer, K. (2011). Effects of cooperating and conflicting prosody in spoken English garden path sentences: ERP evidence for the boundary deletion hypothesis. *Journal of Cognitive Neuroscience, 23*(10), 2731-2751.

Pennington, M., & Ellis, N. C. (2000). Cantonese speakers' memory for English sentences with prosodic cues. *The Modern Language Journal, 84*(3), 372-389.

Petrone, C., Truckenbrodt, H., Wellmann, C., Holzgrefe-Lang, J., Wartenbruger, I., & Hohle, B. (2017). Prosodic boundary cues in German: Evidence from the production and perception of bracketed lists. *Journal of Phonetics, 61*, 71-92.

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as 'asymmetric sampling in time' . *Speech Communication, 41*(1), 245-255.

Qin, Z., Chine, Y., & Tremalay, A. (2016). Processing of word-level stress by Mandarin-speaking second language learners of English. *Applied Psycholinguistics, 38*(3), 1-30.

Roll, M., Lindgren, M., Alter, K., & Horne, M. (2012). Time-driven effects on parsing during reading. *Brain and Language, 121*(3), 267-272.

Schmidt, E., Pérez, A., Cilibrasi, L., & Tsimpli, I. (2019). Prosody facilitates memory recall in L1 but not L2. *Studies in Second Language Acquisition, 42*(1), 223-238.

Schön, D., & Tillmann, B. (2015). Short- and long-term rhythmic interventions: Perspectives for language rehabilitation. *Annals of the New York Academy of Sciences, 1337*, 32-39.

Schremm, A., Horne, M., & Roll, M. (2015). Brain responses to syntax constrained by time-driven implicit prosodic phrases. *Journal of Neurolinguistics, 35*, 68-84.

Seidl, A. (2007). Infants' use and weighting of prosodic cues in clause segmentation. *Journal of Memory & Language, 57*(1), 24-48.

Seidl, A., & Cristia, A. (2008). Developmental changes in the weighting of prosodic cues. *Developmental Science, 11*(4), 596-606.

Selkirk, E. (2005). Comments on intonational phrasing in English. In S. Frota, M. Vigario, & M. J. Freitas (Eds.), *Prosodies* (pp. 11-58). Berlin, Germany: Mouton de Gruyter.

Slater, J., & Kraus, N. (2016). The role of rhythm in perceiving speech in noise: A comparison of percussionists, vocalists and non-musicians. *Cognitive Processing, 17*(1), 79-87.

Steinhauer, K., Alter, K., & Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience, 2*(2), 191-196.

Steinhauer, K., Shani, H. A., Pauker, E., Itzhak, I., & Baum, S. R. (2010). Prosody-syntax interactions in aging: Event-related potentials reveal dissociations between on-line and off-line measures. *Neuroscience Letters, 472*(2), 133-138.

Truckenbrodt, H. (1995). *Phonological phrases: Their relation to syntax, focus and prominence* (Unpublished doctoral dissertation). Massachusetts Institute of Technology.

Wagner, M., & Watson, D. G. (2010). Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes, 25*(7-9), 905-945.

Webman-Shafran, R., & Fodor, J. D. (2016). Phrase length and prosody in online ambiguity resolution. *Journal of Psycholinguist Research, 45*(3), 447-474.

Xu, G., Zhang, L., Shu, H., Wang, X., & Li, P. (2013). Access to lexical meaning in pitch-flattened Chinese sentences: An fMRI study. *Neuropsychologia, 51*(3), 550-556.

Yang, L. C. (2016, March). Optimizing pronunciation and prosody teaching in second language learning. Paper presented at the meeting of ISAPh 2016 International Symposium on Applied Phonetics, Nagoya, Japan.

Yang, X., Shen, X., Li, W., & Yang, Y. (2014). How listeners weight acoustic cues to intonational phrase boundaries. *PLOS ONE, 9*(7), Article e102166. http://doi:10.1371/journal.pone.0102166

Zoefel, B., & VanRullen, R. (2015). Selective perceptual phase entrainment to speech rhythm in the absence of spectral energy fluctuations. *Journal of Neuroscience, 35*(5), 1954-1964.

Zoefel, B., & VanRullen, R. (2016). EEG oscillations entrain their phase to high-level features of speech sound. *Neuroimage, 124*, 16-23.

Note: Figure translations are in progress. See original paper for figures.