

# Integrating syntactic and prosodic boundary cues in second language processing: Evidence from L1-Chinese L2-English learners and L1 English speakers

Second Language Research

1–24

© The Author(s) 2026



Article reuse guidelines:

[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)

DOI: 10.1177/02676583261445731

[journals.sagepub.com/home/slr](https://journals.sagepub.com/home/slr)

Jinxin Ji<sup>1,2</sup>, Wencui Liu<sup>1</sup>, Jiaxin Li<sup>2</sup>,  
Xiaohu Yang<sup>1</sup> and Gang Peng<sup>2</sup>

## Abstract

It remains unclear whether second language prosody–syntax integration resembles native language mechanisms. This study examines how first language (L1) Chinese second language (L2) English learners integrate prosody and syntax compared with native English speakers, and explores L1 Chinese and L2 English processing within L2 learners. Using the Rapid Prosody Transcription method, 89 participants, including native English speakers and Chinese L2 learners, judged prosodic boundaries under manipulated acoustic and syntactic conditions. Results showed that Chinese L2 learners exhibited stronger syntactic constraints in English sentence judgments compared with native speakers. Furthermore, they relied more on syntactic information when processing L2 English than L1 Chinese. These results were examined in relation to the Shallow Structure Hypothesis (SSH), as heuristic processing appears to dominate in L2, leading learners to prioritize syntactic over acoustic cues for rapid and global interpretation. This study proposes that ‘shallow’ in the SSH indicates that L2 processing prioritizes higher-level information.

## Keywords

prosody–syntax integration, shallow structure hypothesis, rapid prosody transcription, speech perception, second language processing

<sup>1</sup>School of Foreign Languages, Tongji University, Shanghai, China

<sup>2</sup>The Hong Kong Polytechnic University, Hongkong SAR, China

## Corresponding author:

Gang Peng, Department of Language Science and Technology, The Hong Kong Polytechnic University, Hong Kong SAR, China

Email: [gang.peng@polyu.edu.hk](mailto:gang.peng@polyu.edu.hk)

## I. Introduction

Despite the growing focus on the interaction and mapping among linguistic modules (such as syntax, morphology, semantics, and prosody) in second language acquisition over the past two decades, the integration of prosody and syntax in second language (L2) learners remains relatively underexplored (e.g. Nickels and Steinhauer, 2018; White, 2011). Research has suggested that prosody–syntax integration in second language (L2) learners might be influenced by several factors, including language transfer (Xue and Yue, 2024), second language processing mechanisms (Ip and Cutler, 2018), and language proficiency (Nickels and Steinhauer, 2018). Previous studies have only identified potential factors; the underlying mechanisms remain to be explored in relation to current L2 theories. Moreover, research is needed to investigate whether L2 processing aligns with the patterns observed in first language (L1) processing. The present study aims to compare how L1 English speakers and L2 English learners process prosody–syntax integration, and to examine how processing mechanisms in L2 English learners’ L1 Chinese differ from those in L2 English.

### *1.1. Interaction between prosody and syntax among L1 listeners*

The interactions between the phonetic and syntactic levels have been extensively explored in the processing of L1. Linguistic prosody, composed of multiple acoustic cues, hierarchically organizes information into different sentence types and conveys both explicit and implicit meanings (Dankovičová et al., 2004; Peters, 2005). One of the key functions of prosody is speech segmentation. Relatively large boundaries, referring to prosodic boundaries beyond the word level, disambiguate sentences and convey information through three main prosodic cues: pause, final lengthening, and pitch change. The use of these cues for speech segmentation has been shown to be universal across different languages (Chinese: Ji et al., 2024; Yang et al., 2014; English: Carlson et al., 2001; Clifton et al., 2002; Morrill et al., 2014; Streeter, 1978). Pause refers to the silent interval that typically occurs between syntactic units (Peters, 2005; Scott, 1982). Final lengthening refers to the increased duration of the word preceding the prosodic boundary. A prosodic boundary is signaled by pitch changes, such as a pitch reset (suspension of downstep with a return to a higher pitch after the boundary) or a pitch rise (the rise of intonation before the boundary). For example, Chinese speakers use pitch reset to indicate a prosodic boundary (Yang et al., 2014), whereas German and English speakers tend to use pitch rise as a prosodic boundary cue (Petroni et al., 2017). These prosodic cues determine how words are grouped, often coinciding with syntactic boundaries (Ji et al., 2024). For instance, the phrase ‘mango salad and juice’ can convey different meanings based on these boundaries: [[mango salad] and juice] suggests two items, and [[mango] [salad] and [juice]] implies three.

Listeners not only rely on prosodic information to disambiguate syntactic structures, but also use syntactic and semantic information to anticipate prosodic structure. Bishop (2012) found that listeners’ judgments of prosodic prominence varied significantly across different question contexts and information structures, even when presented with

the same acoustic information. Van der Burght et al. (2021) investigated how pitch accents influenced both syntactic and semantic expectations in sentence processing. The results showed that syntactic expectations were strong enough to interfere with sentence comprehension when violated, indicating a primary reliance on syntax for guiding auditory sentence comprehension.

The annotation methods further revealed how syntax affects prosodic information. Rapid Prosody Transcription (RPT) is one of the latest methods used to capture listeners' judgments on prosodic features in auditory stimuli (Cole et al., 2017). To assess prosodic boundaries, listeners made binary judgments (e.g. 1 for boundary, 0 for no-boundary) before each word in the auditory materials, meaning any word could potentially be marked. With a group of annotators, these scores indicate the likelihood that a new annotator from the same population would perceive a prosodic boundary before each word. Cole et al. (2010) used the RPT method to investigate the relationship between syntactic and prosodic phrase structures in spontaneous speech using data from the Buckeye corpus and real-time prosodic transcription by untrained listeners. The results demonstrated that syntactic context was the most reliable predictor, suggesting that listeners' judgments were heavily influenced by their syntactic expectations. Building on Cole et al.'s (2010) method, subsequent studies (Kim et al., 2016; Pintér et al., 2014; You, 2012) have continued to explore this interaction between syntax and prosody. For instance, Buxó-Lugo and Watson (2016) investigated the effect of syntactic expectations on boundary perception using the RPT method. The sentences used in the experiment were unambiguous, such as 'Put the big bowl on the tray'. The strength of acoustic cues was systematically controlled at 9 steps and placed at both syntactically licensed locations (e.g. after *bowl*) and syntactically unlicensed locations (e.g. after *big*). The results showed that listeners identified more boundaries at the syntactically licensed location even with the same acoustic strength, indicating that both acoustic cues and syntactic expectations played a role in processing relatively large prosodic boundaries.

The mutual influence between higher-level syntactic anticipation and bottom-up acoustic cue processing is supported by parallel-process models (e.g. Brodbeck et al., 2023; Kim and Osterhout, 2005; Tabor and Tanenhaus, 1999). Although these models differ in specifics, they generally propose that multiple systems process speech input in parallel. Listeners receive phonemes and words as local units of representation, and this mechanism processes local information. In addition to this local processing, there are two other processing streams: top-down and bottom-up. In the bottom-up manner, auditory information represented by phonemes combines to form words, which then convey syntactic information. In the top-down manner, prior words provide context that sets expectations for possible word and phoneme candidates, and the following sentence context further refines these expectations. These three systems operate in parallel and influence each other during speech comprehension. The interaction between top-down and bottom-up processes has been widely investigated among L1 speakers (e.g. Donhauser and Baillet, 2019). Although such interactions have also been examined in L2 learners (e.g. Wiener and Liu, 2021), how these models align with other second language theories remains unclear.

## 1.2. Interactions of prosody and syntax in L2 learners

Compared with studies on L1 speakers, research on the integration of prosody and syntax among L2 learners is relatively underdeveloped. Several studies have shown that L2 learners are less precise than native speakers when using acoustic cues for sentence comprehension. For example, Vanlancker-Sidtis (2003) tested how L1 speakers and L2 learners at different proficiency levels use auditory information to distinguish between prosodic patterns signaling idiomatic and literal meanings in ambiguous sentences. The results showed that L1 speakers significantly outperformed L2 speakers across different proficiency levels. Ip and Cutler (2020) found that L1-Chinese L2-English speakers used prosodic cues to anticipate the position of accented words in their L1, and a similar effect was observed among L1 English speakers in their own language. However, the Chinese learners were unable to use this entrainment strategy in their L2 English.

Several factors may influence the integration of prosody and syntax among L2 learners. Language proficiency plays a crucial role in shaping their integration strategies. Dekydtspotter et al. (2008) investigated how L1-English L2-French learners use prosodic and syntactic information to process ambiguous relative clause attachment sentences. The results revealed that college-level L2 learners at different proficiency levels used different processing strategies. Fourth-semester learners demonstrated an improved ability to revise their initial parse strategy based on prosodic cues compared with second-semester college-level learners. Language background, as another significant factor, might interact with proficiency to affect how L2 learners perceive prosody and syntax. Nickels and Steinhauer (2018) used event-related potentials (ERP) methods to investigate the online processing of garden-path sentences among L1 listeners and L2 learners. The results showed that although the Chinese participants had greater L2 immersion at a Canadian university compared with the German learners tested in Germany, they exhibited different perception patterns from both L1 speakers and the German L2 learners. This might be due to the prosodic structures of German and English being more similar to each other than to those of Chinese and English. Therefore, further studies focusing on Chinese learners of English are necessary to explore how prosodic and syntactic integration is influenced by differences in prosodic structures between their L1 and target L2 languages.

Language transfer is another key factor influencing L2 processing, as it reflects the influence of learners' L1 on their L2 processing. Atterer and Ladd (2004) demonstrated that L1 German speakers transferred the feature of prenuclear rising accents at the segmental level to their L2 English. In a recent study, Xue and Yue (2024) found that Chinese-English L2 learners relied solely on maximum F0 for focus marking in English speech production, whereas L1 English speakers used a combination of pitch cues. This difference could be attributed to L1 language transfer, given that Chinese has a narrower range of prosodic cues for indicating focus compared with English.

It should be noted that language transfer is not an all-or-nothing phenomenon (O'Brien et al., 2014), as some studies have reported only limited language transfer effects. For instance, O'Brien et al. (2014) examined how German-English and English-German L2 learners used prosodic cues to disambiguate prepositional phrase attachment in both German and English. The findings indicated that both groups used the same pitch cues

when speaking English and German, suggesting that L2 learners did not fully transfer prosodic cues from their L1. Furthermore, Ip and Cutler (2018) reported that L2 listeners did not automatically transfer L1 strategies to L2 but instead developed new strategies for processing syntactic ambiguities in L2.

Given that most previous studies used ambiguous sentences as materials, Zhang and Ding (2020) further proposed that awareness of ambiguity significantly influences how L2 learners use prosodic cues to resolve syntactic ambiguities. They found that without awareness of ambiguity, L2 learners' judgments were less accurate and they used a 'good-enough' heuristic strategy, leading to incorrect interpretations. However, explicit awareness of ambiguity significantly enhanced L2 learners' ability to integrate prosodic and syntactic information.

Although previous studies have revealed that the relationship between syntax and prosody among L2 learners differs from that of L1 speakers, the underlying mechanisms remain unclear. The present study aims to further investigate this relationship within the framework of L2 theories.

### *1.3. The shallow structure hypothesis*

A related theory concerning syntactic constraints is the Shallow Structure Hypothesis (SSH), which proposes that L2 learners are less sensitive to syntactic constraints than L1 speakers and tend to rely more on non-structural information, such as lexical, pragmatic, and real-world knowledge (Clahsen and Felser, 2006b; Papadopoulou, 2005; Papadopoulou and Clahsen, 2006). The SSH is grounded in a multiple-pathway model of language processing (Clahsen and Felser, 2018), which posits that morphologically complex words can be processed through more than one route: a heuristic pathway that relies on global interpretation, and an algorithmic pathway that constructs detailed grammatical representations. The relationship between these two types of processing and the parallel mechanisms of top-down and bottom-up models is one of correlation rather than direct mapping. Heuristic processing relies more on semantic content through top-down processing, whereas algorithmic processing relies primarily on bottom-up mechanisms, organizing unfolding input according to successive linguistic rules (Karimi and Ferreira, 2016).

The SSH raises two important aspects for further discussion. First, although Clahsen and Felser (2018) noted that the SSH has mainly focused on between-group comparisons between L1 and L2 speakers, research examining the influence of learners' L1 on L2 acquisition remains limited (Gerth et al., 2017; Marinis et al., 2005). To achieve a deeper understanding of the SSH, it is necessary to combine L1–L2 between-group comparisons with cross-linguistic investigations among L2 learners.

Second, it is important to consider how acoustic cues interact with syntactic processing in L2 listening. Both the original proposal of the SSH (Clahsen and Felser, 2006a) and the subsequent note by the authors (Clahsen and Felser, 2018) emphasized that L2 learners rely more on non-structural information. The non-structural information also includes acoustic information such as prosody (Clahsen and Felser, 2006b). However, later studies examining the SSH have focused on the interaction between semantic and syntactic cues (Felser et al., 2003; Song et al., 2020; Witzel et al., 2012). Few studies

have examined the SSH in the auditory modality. Fernandez et al. (2017) investigated the cognitive cost of speech perception in German L2 speakers of English, examining how participants used prosodic cues to process both complex and simpler syntactic constructions. The results showed that L2 learners and native speakers had similar grammaticality ratings across different sentence types. The increase in cognitive cost for complex sentences compared to simpler ones, as indicated by pupil dilation, did not differ significantly between groups.

Two unsolved problems remain in their study. First, the study examined sentences with different syntactic structures in the auditory modality but did not, as acknowledged by the authors, compare different acoustic conditions. As a result, the findings might only reflect listeners' performance under one acoustic condition, which could differ when acoustic cues such as intonation change. Secondly, the results were interpreted as inconsistent with the original proposal of the SSH, as L2 learners did not exhibit shallower processing of syntactic constraints. However, whether the underlying mechanism of the SSH could explain the results warrants further discussion. The SSH pattern may differ in speech perception tasks compared with reading tasks. In previous studies on sentence reading and lexical selection tasks, top-down information mainly involves heuristic processing and originates from semantic memory to syntactic structure. It should be noted that in speech perception tasks, syntactic licensing is generally considered a higher-level process relative to acoustic cues. Using syntactic information to predict acoustic patterns constitutes a top-down process, which is associated with heuristic processing (Brodbeck et al., 2023). Thus, in the relationship between syntactic and acoustic cues, the SSH's claim that heuristic processing dominates in L2 appears to contradict its original proposal that L2 learners prioritize non-structural over syntactic information.

## 2. The present study

Based on previous literature, we proposed two research questions.

**Research question 1:** How do syntactic cues and prosodic boundary cues integrate in L2 processing?

L2 processing in this study is examined from two perspectives: (1) cross-group comparison to investigate whether prosody–syntax integration differs between L1 English speakers and Chinese L2 English learners, and (2) cross-language comparison to examine how L2 learners integrate prosody and syntax in their L1 (Chinese) versus their L2 (English).

This study differs from previous research in terms of the materials used. First, we specifically investigated prosodic boundary judgment using acoustic continua. Previous studies on syntax and prosody integration have tended to investigate prosody as a whole (e.g. Clahsen and Felser, 2018). They examined prosodic features such as intonation or focus across entire sentences, rather than focusing on specific acoustic components or continua of acoustic features. The use of acoustic continua contributes to revealing the dynamic and gradual transitions between bottom-up and top-down mechanisms. Bottom-up mechanisms involved the perception of prosodic cues under

different syntactically licensed conditions, and top-down mechanisms used syntactic information to predict prosodic boundary cues. Second, most studies on L2 prosody–syntax integration have focused on how listeners perceive ambiguous sentences. However, the processing of such sentences may be influenced by factors such as limited cognitive resources, L2 parsing strategies (Ip and Cutler, 2018), and awareness of ambiguity (Zhang and Ding, 2020) due to their structural complexity. Therefore, this study used sentences with clear and simple syntactic structures, allowing for a more precise and controlled investigation of the mechanisms underlying prosody–syntax integration.

**Research question 2:** How do the results of the integration of prosody and syntax in L2 relate to the SSH?

This theory has primarily been examined concerning the relationship between semantics and syntax, and the discussion of how this model fits into speech perception is still lacking. As Clahsen and Felser (2018) claimed, the SSH requires further examination to evaluate how different types of constraints are weighed. The SSH proposals that the heuristic pathway is more dominant in L2 (Clahsen and Felser, 2018) and that L2 learners rely more on non-structural information in parsing (Clahsen and Felser, 2006b) may not be simultaneously feasible in L2 listening. If L2 processing relies more on syntax than acoustics compared to L1 processing, then it is more likely to be top-down and mainly driven by heuristic processing. If L2 processing relies more on acoustics than syntax, then it is more likely to be bottom-up and mainly driven by algorithmic mechanisms. How the SSH fits into speech perception requires further discussion.

### 3. Method

#### 3.1. Participants

A total of 89 participants were recruited online, including 30 L1 English speakers from the United States ( $M=24.60$  years,  $SD=3.73$ ), 29 Chinese L2 learners with lower English proficiency ( $M=22.40$  years,  $SD=3.84$ ), and 30 Chinese L2 learners with higher English proficiency ( $M=22.27$  years,  $SD=2.77$ ). None of the L2 learners had lived or studied in English-speaking countries. The mean age of acquisition (AOA) was 6.60 years ( $SD=1.10$ ) for higher proficiency learners, and 7.07 years ( $SD=1.14$ ) for lower proficiency L2 learners. The higher proficiency group had achieved an IELTS score of 7 or above within the past two years, and the lower proficiency group had achieved an IELTS score between 5 and 6 within the past two years. All participants were compensated for their participation. All procedures were conducted with written consent from participants and with the approval of the Ethical Review Board (PolyU Institutional Review Board; Protocol number: HSEARS20240416002; Title: Syntax and Prosody Perception among L1 and L2 Listeners).

**Table 1.** English materials.

Content	Syntactic condition
(1a) Put the <b>teal bead</b>   in the jar.	Licensed
(1b) Put the <b>teal</b>   <b>bead</b> in the jar.	Unlicensed
(2a) Put the <b>big bowl</b>   on the tray.	Licensed
(2b) Put the <b>big</b>   <b>bowl</b> on the tray.	Unlicensed
(3a) Put the <b>brown dog</b>   on the couch.	Licensed
(3b) Put the <b>brown</b>   <b>dog</b> on the couch.	Unlicensed
(4a) Put the <b>black book</b>   on the chair.	Licensed
(4b) Put the <b>black</b>   <b>book</b> on the chair.	Unlicensed

Notes. Boundary locations are indicated by '|'. Words that were acoustically manipulated are in bold.

### 3.2. Materials

An American English L1 speaker recorded the English materials, and a L1 Chinese speaker recorded the Mandarin materials. Following Buxó-Lugo and Watson (2016), the English materials included four sentences, each produced at two syntactic boundary positions: one at an unlicensed boundary (e.g. between key nouns and modifiers) and the other at a licensed boundary (see Table 1). For example, item (2a) in Table 1 was produced with a relatively large boundary at a syntactically licensed location, and item (2b) was produced with a boundary at an unlicensed location.

The Chinese materials followed the same phrase structure with four tone combinations to enhance generalizability, reflecting the variability of tone combinations in Chinese (see Table 2). These controlled positions were selected based on prior research. Previous studies have indicated that relatively large prosodic boundaries were more likely to occur at major syntactic boundaries, such as the boundary between an object phrase and a prepositional phrase, rather than at non-major syntactic boundaries (e.g. between a noun and a modifier: Watson and Gibson, 2004; Wang et al., 2004).

Following Buxó-Lugo and Watson (2016), boundary continua were separately created at syntactically licensed locations – e.g. Table 1, item (1a) – and unlicensed locations; e.g. Table 1, item (1b). The endpoints of each boundary continuum, representing the absence of a boundary and the presence of a prosodic boundary, were based on L1 speaker productions. For example, the continuum at the syntactically licensed locations between *bowl* and *on* was constructed using productions from two sentences at the same sentence positions. Item (2b) in Table 1 indicates no prosodic boundary between *on* and *bowl*, whereas item (2a) in Table 1 indicates a prosodic boundary between *on* and *bowl*.

In each continuum, the three main prosodic cues (pause, final lengthening, and pitch change) were simultaneously manipulated across nine steps for both the licensed and unlicensed conditions in each language. The endpoints of each continuum were Step 1 and Step 9, with Step 9 representing a prosodic boundary and Step 1 representing no boundary. Specifically, Step 1 included the pause, final lengthening, and pitch change corresponding to the absence of a boundary. Step 5 represented the midpoint of the continuum, with the pause, final lengthening, and pitch change adjusted to intermediate

**Table 2.** Chinese materials.

Chinese	English translation	Syntactic condition
(1a) 把 <b>橙色书包</b> 放在椅子上 bǎ chéng sè shū bāo fàng zài yǐ zi shàng	Put the orange backpack on the chair	Licensed
(1b) 把 <b>橙色</b> 书包放在椅子上 bǎ chéng sè shū bāo fàng zài yǐ zi shàng	Put the orange backpack on the chair	Unlicensed
(2a) 把 <b>蓝色球鞋</b> 放在架子上 bǎ lán sè qiú xié fàng zài jià zi shàng	Put the blue sneakers on the shelf	Licensed
(2b) 把 <b>蓝色</b> 球鞋放在架子上 bǎ lán sè qiú xié fàng zài jià zi shàng	Put the blue sneakers on the shelf	Unlicensed
(3a) 把 <b>红色雨伞</b> 放在抽屉里 bǎ hóng sè yǔ sǎn fàng zài chōu ti lǐ	Put the red umbrella in the drawer	Licensed
(3b) 把 <b>红色</b> 雨伞放在抽屉里 bǎ hóng sè yǔ sǎn fàng zài chōu ti lǐ	Put the red umbrella in the drawer	Unlicensed
(4a) 把 <b>黑色外套</b> 放在柜子里 bǎ hēi sè wài tào fàng zài guì zi lǐ	Put the black coat in the cabinet	Licensed
(4b) 把 <b>黑色</b> 外套放在柜子里 bǎ hēi sè wài tào fàng zài guì zi lǐ	Put the black coat in the cabinet	Unlicensed

Notes. Boundary locations are indicated by '|'. Words that were acoustically manipulated are in bold.

values between the endpoints. Step 9 included all three cues at the boundary position, representing the presence of a prosodic boundary.

These manipulations formed a continuous prosodic gradient for sentences in each language. First, we used Praat to manipulate the pause duration across 9 steps. Then, for each pause step, we employed MATLAB with the TANDEM-STRAIGHT toolbox (Kawahara et al., 2009), which allows simultaneous adjustment of pitch changes and final lengthening to match the corresponding pause step. The three acoustic features were manipulated as follows.

- **Pause:** Pauses were adjusted across nine steps, ranging from no prosodic boundary to a prosodic boundary, based on visual inspection in Praat. These steps were based on the measurements of a naturally produced boundary – e.g. production of the sentence (1a) in Table 1 – and non-boundary phrases – e.g. production of the sentence (1b) in Table 1 – forming the two endpoints of a boundary continuum. Equally spaced intermediate steps were created to ensure a smooth transition between these endpoints, simulating the gradual perception of boundary strength (see Figures S1 and S2 in supplemental material).
- **Final Lengthening:** Vowel durations in both Chinese and English sentences were manipulated across 9 steps, ranging from no boundary to boundary, using the TANDEM-STRAIGHT toolbox (Kawahara et al., 2009). This manipulation was equally spaced intermediate steps to reflect a gradual increase in boundary-likeness (see Figures S3 and S4 in supplemental material).
- **Pitch Change:** F0 contours were measured for words in different positions. In English, only the pre-boundary words were manipulated from boundary to no

boundary (see Figure S5 in supplemental material), as pitch change is signaled by a rise in the pre-boundary word in English: the ‘*bow*l’ in Table 1 (2a). In Chinese, both pre-boundary and post-boundary words were manipulated, as pitch changes are indicated by a pitch reset (see Figure S6 in supplemental material), which is a suspension of downstep and a return to a higher pitch following the boundary (Wang et al., 2004). The F0 changes for both Chinese and English were manipulated in 9 steps from no boundary to boundary using the TANDEM-STRAIGHT MATLAB toolbox.

A total of 144 trials (2 languages  $\times$  4 sentences  $\times$  2 syntactic license conditions  $\times$  9 prosodic boundary steps) were created. The average intensity of sentences was normalized to 70 dB. Example spectrograms of the stimuli are shown in Figures S7 to S10 in supplemental material.

### 3.3. Procedure

All recordings were uploaded to Gorilla. The procedure followed Buxó-Lugo and Watson (2016) by using the RPT method. The instructions explained that speakers often group utterances into chunks, and that these chunks are often divided by what we call boundaries. They were informed that words before boundaries have a ‘different’ sound compared with those that do not. The instructions were chosen to ensure listeners would not consciously search for specific cues, such as pauses, to identify boundaries (see Appendix A).

For each question, participants saw a media player icon with the instructions: ‘There is a boundary after: . . .’ The sentence was displayed below the instructions. Participants were asked to check the boxes under the words they believed were followed by a boundary. They could replay the recordings and mark as many words as they wanted. We analyzed the perceived boundary responses at the controlled boundary position after the controlled words for each recording.

L1 English speakers were required to complete the section for the judgment of English sentences, while Chinese L2 participants had an additional section for the judgment of Chinese sentences. The order of the sections was randomized for L2 participants, and the recordings were randomized within each section.

### 3.4. Analysis

We obtained binary boundary judgments for each word in the presented sentences. Previous studies have suggested that the results of the RPT method experiments can be analyzed at specific positions (e.g. Buxó-Lugo and Watson, 2016; Turnbull et al., 2017). Therefore, we analyzed responses in the syntactically licensed condition – e.g. boundary judgment for the syntactically licensed position after *bow*l in sentence Table 1(2a) – and unlicensed condition – e.g. boundary judgment for the syntactically unlicensed position after *big* in sentence Table 1 (2b) – under different prosodic boundary steps. The first analysis included the responses for 89 listeners across three groups of listeners (L1 English speakers and L2 learners with higher English proficiency and L2 learners with

lower English proficiency) to 72 recordings of English sentences, totaling 6,408 data points. The second analysis included responses from 59 listeners (L2 learners with higher and lower English proficiency) on 144 recordings of English and Chinese sentences, totaling 8,496 data points.

We constructed mixed-effects logistic regression models using the ‘lme4’ package (Bates et al., 2015) in R Version 4.4.1 (R Core Team, 2024). The model construction followed previous studies on speech perception that involved multiple factors, as overall models were built first, followed by separate models based on specific groups or conditions (e.g. Durvasula and Kahng, 2016; Kwon, 2017; Zhang and Peng, 2021). In this study, two overall models incorporating all possible factors were initially constructed; the first analysis compared how different groups (L1 English speakers and L2 English learners) process English, and the second analysis compared languages (L1 Chinese and L2 English) within the group of L2 English listeners. Subsequently, separate models were then built for different groups and conditions to investigate the interactions observed in the overall models.

The first overall model for cross-group comparison was a mixed-effects logistic regression incorporating all relevant factors. The dependent variable was the listener’s response in the English section (0=no boundary judgment and 1=boundary judgment). The model included main effects for syntactic condition (syntactically licensed and unlicensed locations), group (Chinese listeners with higher English proficiency, Chinese listeners with lower English proficiency, and L1 English listeners), and prosodic boundary step (9 boundary steps), and their possible interactions. The random effects structure selected was the most complex one justified by the study design (Barr et al., 2013), including the intercepts of subject and item. The items consisted of the four sentence types under two conditions, totaling 8 items.

The second overall model for cross-language comparison was a mixed-effects logistic regression incorporating all relevant factors, built for two groups of English L2 learners. The dependent variable was the listeners’ response (0=no boundary judgment and 1=boundary judgment) in both the English and Chinese sections. The model included main effects for syntactic condition (syntactically licensed and unlicensed locations), group (Chinese listeners with higher English proficiency, Chinese listeners with lower English proficiency), language (Chinese L1 and English L2) and prosodic boundary step (9 boundary steps), and their possible interactions. The random effects included the intercepts of subject and item.

## 4. Results

### 4.1. Cross-group comparison of English L1 listeners and Chinese L2 listeners

The first analysis examines whether prosody–syntax integration differs between L1 English speakers and Chinese L2 English learners. Figure 1 shows the boundary response percentage for L1 English speakers and L2 English learners across different boundary steps under syntactically licensed and unlicensed locations. The results for the first overall model are shown in Table 3. There was a significant three-way interaction of Syntactic

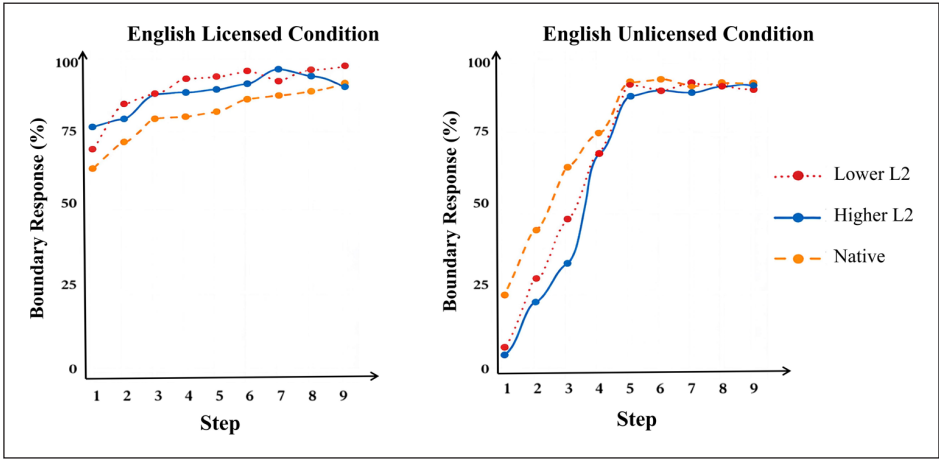


Figure 1. Mean percentage of boundary responses for English across three groups.

Table 3. Results of the overall model for cross-group comparison in English sentence processing.

Term	$\chi^2$	df	p
Intercept	9.59	1	.002**
Con	39.08	1	<.001***
Step	47.27	1	<.001***
Group	3.10	2	.21
Con: Step	29.82	1	<.001***
Con: Group	21.67	2	<.001***
Step: Group	5.06	2	.08
Con: Step: Group	6.04	2	.048*

Notes. Con = Syntactic Condition; Step = Prosodic Boundary Step; Reference Levels: Group = L1 English Speakers; Condition = Licensed; Step = 1. \* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

Condition  $\times$  Group  $\times$  Step. These results suggested that L1 English listeners had different prosodic and syntactic integration patterns compared with Chinese L2 English listeners, highlighting the need for further analysis to explore these integration patterns in detail.

To compare how syntactic information influenced listeners’ prosodic boundary perception, three mixed-effects logistic regression models were fitted for each of the three groups. The dependent variable was the listener’s response. In each group model, the fixed effects included syntactic condition (syntactically licensed and unlicensed conditions) and prosodic boundary step (9 boundary steps). The random effects included the intercepts for both subject and item. The results are presented in Table 4. In each of the three models, there were significant main effects of syntactic condition and prosodic boundary step and a significant interaction of prosodic boundary step  $\times$  syntactic

**Table 4.** Model output for each group in English sentence processing.

Term	Group								
	Native			Higher L2			Lower L2		
	$\chi^2$	df	p	$\chi^2$	df	p	$\chi^2$	df	p
Intercept	8.32	1	.003**	16.66	1	<.001***	23.38	1	<.001***
Step	55.74	8	<.001***	56.11	8	<.001***	28.92	8	<.001***
Con	28.74	1	<.001***	55.27	1	<.001***	64.61	1	<.001***
Step: Con	44.91	8	<.001***	32.78	8	<.001***	74.95	8	<.001***

Notes. Con = Syntactic Licensed Condition; Step = Prosodic Boundary Step; Native = L1 English Speakers; Higher L2 = L2 Learners with Higher English Proficiency; Lower L2 = L2 Learners with Lower English Proficiency; Reference Levels for Each Model: Condition = Licensed; Step = 1. \*\* $p < .01$ . \*\*\* $p < .001$ .

condition. Bonferroni post hoc tests indicated that the effect of prosodic boundary step was significant in both unlicensed and licensed conditions (all  $ps < .001$ ). This indicated that acoustic cues significantly influenced the boundary judgments under different syntactic conditions. In addition, all three groups of listeners made more prosodic boundary judgments under the syntactically licensed locations compared with the syntactically unlicensed locations when controlling for prosodic boundary step (all  $ps < .05$ ). These results suggested that with the same acoustic cue strength, the judgments of all three groups of listeners were significantly influenced by syntactic information. Overall, these findings implied that both syntactic expectations and acoustic cues influenced prosodic boundary processing.

To further investigate the group differences in syntactic and prosodic integration strategies, two mixed-effects logistic regression models were fitted separately for the syntactically licensed and unlicensed conditions. The dependent variable was listeners' response under each condition. In each model, the fixed effects included group (Chinese listeners with higher English proficiency, Chinese listeners with lower English proficiency, and L1 English listeners), prosodic boundary step (9 boundary steps), and their possible interactions. The random effects included the intercepts for both subject and item. The results are shown in Table 5. Under the syntactically licensed condition, there was a significant main effect of prosodic boundary step, indicating that prosodic cues had a significant effect on listeners. However, there was no significant main effect of group and no significant interaction of prosodic boundary step  $\times$  group, indicating that the three groups had similar integration strategies for syntactic and acoustic cues. Under the syntactically unlicensed condition, there were significant main effects of prosodic boundary step and group. There was no significant interaction of prosodic boundary step  $\times$  group. To explore the significant main effect of group, Bonferroni post-hoc tests were conducted. The results revealed no significant difference between L2 listeners with higher proficiency and those with lower proficiency ( $\beta = 0.20$ ,  $SE = 0.30$ ,  $z = 0.67$ ,  $p = 1.00$ ), indicating that listeners' proficiency levels did not influence their processing strategy. However, L2 listeners with lower proficiency ( $\beta = -1.01$ ,  $SE = 0.33$ ,  $z = -3.12$ ,  $p = .006$ ) and higher proficiency ( $\beta = -0.81$ ,  $SE = 0.32$ ,  $z = -2.51$ ,  $p = .036$ ) differed from L1 English

**Table 5.** Model output for the two syntactic conditions.

Term	Condition					
	Licensed			Unlicensed		
	$\chi^2$	<i>df</i>	<i>p</i>	$\chi^2$	<i>df</i>	<i>p</i>
Intercept	12.61	1	<.001***	7.76	1	.005**
Step	33.99	8	<.001***	115.39	8	<.001***
Group	3.45	2	.18	9.20	2	.010*
Step: Group	12.12	16	.74	9.79	16	.90

Notes. Condition = Syntactic Condition; Step = Prosodic Boundary Step; Licensed = Syntactic Licensed Condition; Unlicensed = Syntactic Unlicensed Condition; Reference Levels for Each Model: Group = L1 English Speakers; Step = 1. \* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

listeners. This suggested that L1 listeners made more prosodic boundary judgments under the syntactic unlicensed condition, and they were less influenced by syntactic information compared with the two groups of L2 learners.

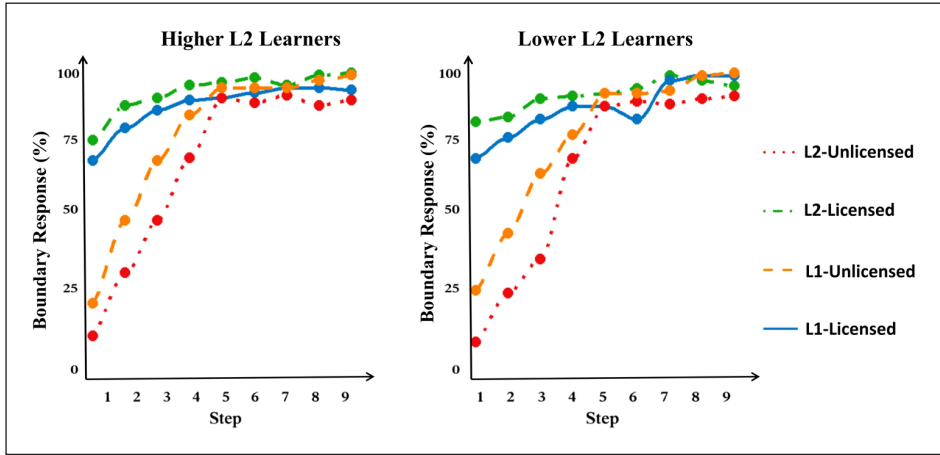
#### 4.2. Cross-language comparison of Chinese L1 and English L2

The second analysis examines how L2 English learners integrate prosodic and syntactic cues in their L1 Chinese and L2 English. Figure 2 presents the boundary response percentages for L2 learners with higher and lower English proficiency for processing L1 Chinese and L2 English across different boundary steps.

The results for the second overall model are shown in Table 6. A significant four-way interaction of syntactic condition  $\times$  prosodic boundary step  $\times$  group  $\times$  language was found, indicating that L2 listeners might have different integration patterns in L1 and L2. Further models were needed to investigate this four-way interaction.

To further investigate this four-way interaction, two models were built for each L2 group to examine how L1 Chinese listeners integrate prosodic and syntactic information in their L1 and L2. The dependent variable was the response from each group of listeners. Each model included main effects for syntactic condition (licensed and unlicensed locations), prosodic boundary step (9 boundary steps), language (L1 Chinese and L2 English), and all possible interactions. The random effects included the intercepts of subject and item. The results are shown in Table 7.

For L2 listeners with higher English proficiency, there was a significant three-way interaction of prosodic boundary step  $\times$  syntactic condition  $\times$  language (see Table 7). To further investigate this three-way interaction, two separate models were built for the licensed condition and unlicensed condition for the responses of L2 listeners with higher English proficiency. The results of these two models are presented in Table 8. Under the licensed condition, there was a significant main effect of prosodic boundary step, indicating that L2 listeners with higher English proficiency identified more prosodic boundaries when stronger acoustic cues were present. However, there was no



**Figure 2.** Mean percentage of boundary responses in first language (L1) Chinese and second language (L2) English.

**Table 6.** Overall model output for Chinese and English processing in L2 learners.

Item	$\chi^2$	df	p
(Intercept)	14.07	1	<.001***
Con	54.84	1	<.001***
Step	41.80	1	<.001***
Group	1.07	1	.30
Lang	0.15	1	.70
Con: Step	52.07	1	<.001***
Con: Group	0.37	1	.54
Step: Group	0.99	1	.32
Con: Lang	0.92	1	.34
Step: Lang	3.40	1	.07
Group: Lang	2.43	1	.11
Con: Step: Group	0.80	1	.37
Con: Step: Lang	7.16	1	.009**
Con: Group: Lang	3.82	1	.05
Step: Group: Lang	4.86	1	.03*
Con: Step: Group: Lang	5.50	1	.019*

Notes. Con=Syntactic Licensed Condition; Step=Prosodic Boundary Step; Lang=Language; Reference Levels: Group=Higher-level L2 Learners; Condition=Licensed; Step=1; Language=Chinese. \* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

significant effect of language, nor a significant interaction of language  $\times$  prosodic boundary step, suggesting that L2 listeners with higher English proficiency used similar processing strategies in both L1 and L2 built for the syntactically licensed conditions.

**Table 7.** Model output for Chinese and English processing in each Chinese learner group.

Term	Group					
	Higher L2 Learners			Lower L2 Learners		
	$\chi^2$	<i>df</i>	<i>p</i>	$\chi^2$	<i>df</i>	<i>p</i>
Intercept	15.75	1	<.001***	5.31	1	.02**
Step	41.11	1	<.001***	58.07	1	<.001***
Lang	0.24	1	.63	3.63	1	.06
Con	60.94	1	<.001***	35.29	1	<.001***
Step: Lang	3.65	1	.06	1.73	1	.19
Step: Con	52.51	1	<.001***	35.80	1	<.001***
Lang: Con	1.07	1	.30	7.62	1	.006**
Step: Lang: Con	7.80	1	.005**	0.42	1	.52

Notes. Licensed = Syntactic Licensed Condition; Unlicensed = Syntactic Unlicensed Condition; Step = Prosodic Boundary Step; Con = Syntactic Condition; Lang = Language; Reference Levels for Each Model: Condition = Licensed; Step = 1; Language = Chinese. \*\*  $p < .01$ . \*\*\*  $p < .001$ .

**Table 8.** Model output for different conditions for higher proficiency L2 learners.

Term	Condition					
	Licensed			Unlicensed		
	$\chi^2$	<i>df</i>	<i>p</i>	$\chi^2$	<i>df</i>	<i>p</i>
Intercept	9.74	1	.002**	48.00	1	<.001***
Step	25.86	1	<.001***	164.32	1	<.001***
Language	0.33	1	.56	1.29	1	.25
Step: Language	3.45	1	.06	4.06	1	.04*

Notes. Step = Prosodic Boundary Step; Licensed = Syntactic Licensed Condition; Unlicensed = Syntactic Unlicensed Condition; Reference Levels for Each Model: Step = 1; Language = Chinese. \*  $p < .05$ . \*\*  $p < .01$ . \*\*\*  $p < .001$ .

Under the syntactically unlicensed condition, there were significant main effects of prosodic boundary step and language. There was a significant interaction effect of prosodic boundary step  $\times$  language. Bonferroni post-hoc tests on the interaction effect revealed that listeners made more prosodic judgments in Chinese L1 than English L2 ( $\beta = 0.98$ ,  $SE = 0.22$ ,  $z = 4.48$ ,  $p < .001$ ) in the syntactic unlicensed condition.

For L2 listeners with lower English proficiency, significant main effects of step and syntactic condition were found. Two-way interactions were found between prosodic boundary step  $\times$  syntactic condition and language  $\times$  syntactic condition (see Table 7). To investigate the interaction between step and syntactic condition, Bonferroni post-hoc tests revealed a significant difference between the syntactically licensed and unlicensed conditions when the prosodic boundary step was controlled ( $\beta = 0.93$ ,  $SE = 0.24$ ,  $z = 3.96$ ,

$p < .001$ ). This indicated a significant effect of syntactic cues when presented with the same prosodic cue strength. To investigate the interaction between language and syntactic condition, Bonferroni post-hoc tests were conducted. The results showed no significant difference between Chinese L1 and English L2 under the syntactically licensed condition ( $\beta = -0.34$ ,  $SE = 0.34$ ,  $z = 1.02$ ,  $p = .31$ ). However, under the syntactically unlicensed condition, listeners made more prosodic boundary judgments in L1 Chinese than L2 English ( $\beta = 0.91$ ,  $SE = 0.33$ ,  $z = 2.75$ ,  $p = .006$ ). These results indicated that Chinese L2 listeners with lower English proficiency had similar processing strategies between L1 Chinese and L2 English under syntactic licensed condition, but had different processing strategies between L1 Chinese and L2 English in the unlicensed condition.

## 5. Discussion

This study examined the prosodic boundary and syntactic license integration in L2 processing. We investigated how L1-Chinese L2 -English learners integrated syntactic and prosodic boundary cues by comparing the processing of English in L1 speakers and L2 learners, and the processing of L1 Chinese and L2 English among L2 English learners. The results demonstrated that syntactic expectations influenced prosodic boundary judgments. L2 English learners were more influenced by syntactic information than L1 English speakers. These results were discussed in relation to the SSH by suggesting that ‘shallow’ processing reflects the prioritization of higher-level information, as learners rely more on syntactic cues when acoustic information is less informative.

To address research question 1 (how syntactic and prosodic boundary cues are integrated in L2 processing), both cross-group and cross-language comparisons were conducted. The cross-group analysis examined prosody–syntax integration between L1 English speakers and L2 English learners. Both groups showed a similar dynamic pattern in integrating syntactic and prosodic boundary cues. When acoustic information was weak or absent, listeners primarily relied on syntactic expectations. However, when prosodic cues were strong (i.e. involving a long pause, increased final lengthening, and sharp pitch changes, corresponding to steps 5–9 in our experiment), both L1 and L2 listeners identified boundaries even in syntactically unlicensed conditions, indicating that prosodic information alone can trigger boundary perception. These results replicate previous findings (Buxó-Lugo and Watson, 2016; Cole et al., 2010), demonstrating that L1 listeners’ prosodic judgments are influenced by syntactic cues. Importantly, this effect was also observed in L2 listeners, consistent with studies showing that L2 listeners can form prosodic expectations based on syntactic information in a manner similar to L1 listeners (e.g. Fernandez et al., 2017; Ganga et al., 2024).

Although the general pattern was similar, there were notable differences in syntactically unlicensed conditions. In the licensed conditions, both English L1 speakers and L2 English learners exhibited similar boundary judgments, as syntactic information alone provided sufficient cues. All groups of listeners showed over 70% boundary judgments even in the absence of prosodic boundary cues. In unlicensed conditions, where syntactic and acoustic information conflicted, L2 English learners relied more on syntax and made

fewer boundary judgments based on acoustic cues. This may be due to L2 listeners relying more on syntactic information compared to acoustic cues, reflecting a top-down processing manner. These differences could also be influenced by factors such as L1 transfer or L2 processing strategies that differ from those used in L1. To better understand the prosody–syntax integration mechanism in L2 processing, we conducted a comparison between L1 Chinese and L2 English processing among L2 English learners.

We investigated how L1 Chinese speakers integrate prosody and syntax in their L1 Chinese and L2 English. The results indicated that L2 English learners employed similar strategies in both L1 and L2 under syntactically licensed conditions, suggesting that syntactic expectations influenced listeners' judgments in both languages. However, in the syntactically unlicensed condition, L2 English learners relied more on syntactic information in their L2 English than in their L1 Chinese. This demonstrated a different processing strategy in L2 when prosody and syntax conflicted. These findings implied that the increased syntactic constraints in L2 processing compared with L1 were due to the L2 processing strategy rather than L1 transfer, consistent with research showing differences between L1 and L2 processing strategies (O'Brien et al., 2014). In addition, previous studies on Chinese syntax have indicated that Chinese is an analytic language with a more flexible syntactic structure, which tends to be less constrained by syntactic information compared with English (e.g. Li and Thompson 1989). Therefore, it was unlikely that the greater reliance on syntactic cues among L2 learners was transferred from their L1 Chinese.

Another important finding of our study was that no significant differences were observed among listeners of different proficiency levels. The first analysis compared the groups in their perception of English phrases, showing that both groups of Chinese listeners made prosodic boundary judgments when acoustic cues were salient, regardless of syntactic condition. While L2 listeners were more influenced by syntactic information under unlicensed conditions compared with L1 English listeners, both higher and lower proficiency L2 learners demonstrated similar constraints in both the licensed and unlicensed conditions. The second analysis compared English and Chinese phrase perception among L2 English listeners of different proficiency levels, revealing that both groups were more constrained by syntax in L2 than in L1 in the unlicensed conditions. The results suggested that the reliance on syntactic cues in L2 remained stable and did not change with increased English proficiency.

We examined how the results relate to the SSH to answer research question 2. Our results compared L1 speakers with L2 learners, and L2 learners' L1 with their L2, indicating that L2 processing was more constrained by syntactic information than L1 processing was. This finding seems to contrast with the initial proposal of the SSH (Clahsen and Felser, 2006a), which claims that L2 learners' sentence comprehension is shallower and less detailed than that of L1 speakers. However, our results support the SSH's underlying mechanism, suggesting that heuristic processing dominates L2 processing (Clahsen and Felser, 2018). Our findings indicate that this dominance is maintained among L2 learners with different levels of English proficiency. The greater reliance on syntactic cues in L2 processing is likely due to the limited role of acoustic features in achieving precise sentence comprehension. In contrast, grammatical information provides an overall structural representation of the sentence, allowing for a rapid and global interpretation

of meaning. This reflects a ‘fast and frugal’ heuristic processing strategy that reduces cognitive effort in L2 processing, even for learners who have attained a higher proficiency level (Karimi and Ferreira, 2016).

Building on this finding, we reconsider the SSH to clarify what ‘shallow’ processing represents. We argue that ‘shallow’ processing reflects the prioritization of higher-level information. According to parallel processing models, semantic information for overall understanding represents a higher level than syntactic information, and syntactic information represents a higher level than acoustic cues. In L2 comprehension, the shallower processing observed in L2 learners depended on the type of available information, with learners relying more on semantic and pragmatic cues than on syntactic cues, while giving precedence to syntactic cues when acoustic cues were less informative.

We need to note that this proposal might be influenced by sentence type, task type, and L2 learner proficiency. Clahsen and Felser (2018) acknowledged that their original hypothesis was too broad and required refinements for different grammatical structures. In this study, we used sentences with clear structures. It is possible that for ambiguous sentences, where interpretation depends on acoustic cues, L2 listeners must rely more on these cues to achieve an accurate understanding of the meaning. Task type might also influence the results. In the RPT, listeners judged auditory stimuli containing both acoustic and syntactic information, and then made prosodic judgments with the words in the sentence presented on screen. If the task required overall sentence interpretation rather than prosodic judgments, the results might be different. Further studies are needed to investigate whether L2 learners prioritize acoustic cues for integrating acoustic and syntactic information. It is also necessary to examine whether the effects we observed apply to different sentence types and task conditions, or only to sentences with clear structures using the RPT method. English proficiency might also influence the processing patterns. In this study, we focused on L2 English learners with an IELTS score above 5 who had not studied or lived in English-speaking countries. Future research is needed to examine whether English learners with lower proficiency, or those who have studied or lived in English-speaking countries, exhibit the same processing mechanisms.

This study has methodological implications. The advantage of RPT lies in its ability to efficiently and straightforwardly collect data on prosody perception. This method has proven feasible in both online and offline experiments across different participant groups, capturing listeners’ holistic perception of prosodic form and function. While several studies have applied this method to examine how L2 learners annotate prosody (Pintér et al., 2014; You, 2012), its feasibility in investigating the integration of syntax and prosody among L2 learners remains underexplored. Buxó-Lugo and Watson (2016) used this method to reveal how L1 listeners process prosodic cues alongside syntactic information. In this study, we demonstrated the feasibility of this method in exploring L2 syntax–prosody integration. Another methodological advantage of our study is the use of prosodic continua to reveal how listeners’ boundary judgments change with increasing acoustic cue salience under different syntactic licensing conditions.

## 6. Conclusions

We investigated how syntactic cues and prosodic boundary cues integrate in L2 processing and their relation to the SSH. The findings showed that both L1 and L2 listeners

made more boundary judgments in a syntactically licensed condition when acoustic information was controlled. However, L2 learners were more influenced by syntactic information than L1 speakers under syntactically unlicensed conditions. In addition, L2 learners demonstrated a stronger reliance on syntactic information in their L2 compared with their L1, indicating a limited influence of language transfer and highlighting a distinct L2 processing strategy when prosodic cues conflict with syntax. Both higher and lower proficiency L2 groups exhibited similar patterns in English processing, suggesting that proficiency might not significantly affect their reliance on syntactic cues in L2. These results seem to contrast with the original proposal of the SSH, but in fact align with its underlying mechanism, in that L2 processing relies more on a heuristic route than L1 processing does, leading learners to rely more on syntactic information than acoustic cues, enabling rapid and global interpretation. We reconsider the SSH by suggesting that ‘shallow’ processing prioritizes higher-level information, with syntactic cues taking precedence over acoustic cues in sentences with clear structure.

### ORCID iDs

Jinxin Ji  <https://orcid.org/0000-0002-0824-2472>

Gang Peng  <https://orcid.org/0000-0002-1465-1301>

### Ethical considerations

All procedures were conducted with written consent from participants and with the approval of the Ethical Review Board (The Hong Kong Polytechnic University Institutional Review Board; Protocol number: HSEARS20240416002; Title: Syntax and Prosody Perception among L1 and L2 Listeners).

### Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

### Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Supplemental material

Supplemental material for this article is available online.

### References

- Atterer M and Ladd DR (2004) On the phonetics and phonology of ‘segmental anchoring’ of F0: Evidence from German. *Journal of Phonetics* 32: 177–197. [https://doi.org/10.1016/S0095-4470\(03\)00039-1](https://doi.org/10.1016/S0095-4470(03)00039-1)
- Bates D, Mächler M, Bolker B et al. (2015) Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67: 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Barr DJ, Levy R, Scheepers C et al. (2013) Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68: 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>

- Bishop J (2012) Information structural expectations in the perception of prosodic prominence. In: Elordieta G and Prieto P (eds) *Prosody and Meaning*. Mouton de Gruyter, pp. 237–269. <https://doi.org/10.1515/9783110261790.239>
- Brodbeck C, Bhattasali S, Heredia AA et al. (2023) Parallel processing in speech perception with local and global representations of linguistic context. *eLife* 11: Article e72056. <https://doi.org/10.7554/eLife.72056>
- Buxó-Lugo A and Watson DG (2016) Evidence for the influence of syntax on prosodic parsing. *Journal of Memory and Language* 90: 1–13. <https://doi.org/10.1016/j.jml.2016.03.001>
- Carlson K, Clifton CJ, and Frazier L (2001) Prosodic boundaries in adjunct attachment. *Journal of Memory and Language* 45: 58–81. <https://doi.org/10.1006/jmla.2000.2762>
- Clahsen H and Felser C (2006a) How native-like is non-native language processing? *Trends in Cognitive Sciences* 10: 564–570. <https://doi.org/10.1016/j.tics.2006.10.002>
- Clahsen H and Felser C (2006b) Grammatical processing in language learners. *Applied Psycholinguistics* 27: 3–42. <https://doi.org/10.1017/S0142716406060024>
- Clahsen H and Felser C (2018) Some notes on the shallow structure hypothesis. *Studies in Second Language Acquisition* 40: 693–706. <https://doi.org/10.1017/S0272263117000250>
- Cole J, Mahrt T, and Roy J (2017) Crowd-sourcing prosodic annotation. *Computer Speech and Language* 45: 300–325. <https://doi.org/10.1016/j.csl.2017.02.008>
- Cole J, Mo Y, and Baek S (2010) The role of syntactic structure in guiding prosody perception with ordinary listeners and everyday speech. *Language and Cognitive Processes* 25: 1141–1177. <https://doi.org/10.1080/01690960903525507>
- Clifton CJ, Carlson K, and Frazier L (2002) Informative prosodic boundaries. *Language and Speech* 45: 87–114. <https://doi.org/10.1177/00238309020450020101>
- Dankovičová J, Pigott K, Wells B et al. (2004) Temporal markers of prosodic boundaries in children’s speech production. *Journal of the International Phonetic Association* 1: 17–36. <https://doi.org/10.1017/S0025100304001525>
- Dekydspotter L, Donaldson B, Edmonds AC et al. (2008) Syntactic and prosodic computations in the resolution of relative clause attachment ambiguity by English–French learners. *Studies in Second Language Acquisition* 4: 453–480. <https://doi.org/10.1017/S0272263108080728>
- Donhauser PW and Baillet S (2019) Two distinct neural timescales for predictive speech processing. *Neuron* 105(2): 385–393. <https://doi.org/10.1016/j.neuron.2019.10.019>
- Durvasula K and Kahng J (2016) The role of phrasal phonology in speech perception: What perceptual epenthesis shows us. *Journal of Phonetics* 54: 15–34. <https://doi.org/10.1016/j.wocn.2015.08.002>
- Felser C, Roberts L, Marinis T et al. (2003) The processing of ambiguous sentences by first and second language learners of English. *Applied Psycholinguistics* 24(3): 453–489. <https://doi.org/10.1017/S0142716403000237>
- Fernandez L, Höhle B, Brock J et al. (2017) Investigating auditory processing of syntactic gaps with L2 speakers using pupillometry. *Second Language Research* 34(2): 201–227. <https://doi.org/10.1177/0267658317722386>
- Ganga R, Ge H, Struiksma ME et al. (2024) Prosodic processing in sentences with ‘only’ in L1 and L2 English. *Studies in Second Language Acquisition* 46: 478–503. <https://doi.org/10.1017/S0272263124000019>
- Gerth S, Otto C, Nam Y et al. (2017) Strength of garden-path effects in native and non-native speakers’ processing of subject–object ambiguities. *International Journal of Bilingualism* 21: 125–144. <https://doi.org/10.1177/1367006915604401>
- Ip MHK and Cutler A (2018) Asymmetric efficiency of juncture perception in L1 and L2. In: Klessa K, Bachan J, Wagner A, Karpiński M, and Śledziński D (eds) *Proceedings of Speech*

- Prosody 2018*. International Speech Communication Association (ISCA), pp. 289–296. <https://doi.org/10.21437/SpeechProsody.2018-59>
- Ip MHK and Cutler A (2020) Universals of listening: Equivalent prosodic entrainment in tone and non-tone languages. *Cognition* 202: Article e104311. <https://doi.org/10.1016/j.cognition.2020.104311>
- Ji J, Zhao X, Li Y et al. (2024) Age effects on prosodic boundary perception. *Psychology and Aging* 39: 262–274. <https://doi.org/10.1037/pag0000811>
- Karimi H and Ferreira F (2016) Good-enough linguistic representations and online cognitive equilibrium in language processing. *The Quarterly Journal of Experimental Psychology* 69: 1013–1040. <https://doi.org/10.1080/17470218.2015.1053951>
- Kawahara H, Takahashi T, Morise M et al. (2009) Development of exploratory research tools based on TANDEM-STRAIGHT. In: *Proceedings of Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*. APSIPA, pp. 111–120. <http://hdl.handle.net/2115/39651>
- Kim A and Osterhout L (2005) The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language* 52: 205–225. <https://doi.org/10.1016/j.jml.2004.10.002>
- Kim J, Ramakrishna A, Lee S, and Narayanan SS (2016) Relations between prominence and articulatory-prosodic cues in emotional speech. In: Barnes J, Brugos A, Shattuck-Hufnagel S, and Veilleux N (eds) *Proceedings of Speech Prosody 2016*. International Speech Communication Association (ISCA), pp. 893–896. <https://doi.org/10.21437/SpeechProsody.2016-183>
- Kwon H (2017) Language experience, speech perception and loanword adaptation: Variable adaptation of English word-final plosives into Korean. *Journal of Phonetics* 60: 1–19. <https://doi.org/10.1016/j.wocn.2016.10.001>
- Li CN and Thompson SA (1989) *Mandarin Chinese: A functional reference grammar*. London and New York: University of California Press. <https://doi.org/10.1525/9780520352858>
- Marinis T, Roberts L, Felser C et al. (2005) Gaps in second language sentence processing. *Studies in Second Language Acquisition* 27: 53–78. <https://doi.org/10.1017/S0272263105050035>
- Morrill TH, Dilley LC, and McAuley JD (2014) Prosodic patterning in distal speech context: Effects of list intonation and F0 downtrend on perception of proximal prosodic structure. *Journal of Phonetics* 46: 68–85. <https://doi.org/10.1016/j.wocn.2014.06.001>
- Nickels S and 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: 9–37. <https://doi.org/10.1177/0267658316649998>
- O'Brien MG, Jackson CN, and Gardner CE (2014) Cross-linguistic differences in prosodic cues to syntactic disambiguation in German and English. *Applied Psycholinguistics* 35: 27–70. <https://doi.org/10.1017/S0142716412000252>
- Papadopoulou D (2005) Reading-time studies of second language ambiguity resolution. *Second Language Research* 21: 98–120. <https://doi.org/10.1191/0267658305sr254oa>
- Papadopoulou D and Clahsen H (2006) The shallow structure hypothesis of second language sentence processing: What is restricted and why? *Applied Psycholinguistics* 27: 49–62. <https://doi.org/10.1017/S014271640606005X>
- Peters B (2005) Weiterführende Untersuchungen zu prosodischen Grenzen in deutscher Spontansprache [Further investigations of prosodic boundaries in German spontaneous speech]. In: Kohler KJ, Kleber F, and Peters B (eds) *AIPUK. IPDS, Prosodic Structures in German Spontaneous Speech*. IPDS, pp. 203–345.
- Petrone C, Truckenbrodt H, Wellmann C et al. (2017) Prosodic boundary cues in German: Evidence from the production and perception of bracketed lists. *Journal of Phonetics* 61: 71–92. <https://doi.org/10.1016/j.wocn.2017.01.002>

- Pintér G, Mizuguchi S, and Tateishi K (2014) Perception of prosodic prominence and boundaries by L1 and L2 speakers of English. In: *Proceeding of Interspeech 2014*. ISCA, pp. 544–547. <https://doi.org/10.21437/Interspeech.2014-133>
- R Core Team (2024). *R: A language and environment for statistical computing* [software]. R Foundation for Statistical Computing. Available at: <https://www.R-project.org> (accessed May 2026).
- Scott DR (1982) Duration as a cue to the perception of a phrase boundary. *The Journal of the Acoustical Society of America* 71: 996–1007. <https://doi.org/10.1121/1.387581>
- Song H, Do Y, and Kim J (2020) Second language users exhibit shallow morphological processing. *Studies in Second Language Acquisition* 42: 923–945. <https://doi.org/10.1017/S0272263120000170>
- Streeter LA (1978) Acoustic determinants of phrase boundary perception. *The Journal of the Acoustical Society of America* 64: 1582–1592. <https://doi.org/10.1121/1.382142>
- Tabor W and Tanenhaus MK (1999) Dynamical models of sentence processing. *Cognitive Science* 23: 491–515. [https://doi.org/10.1016/S0364-0213\(99\)00013-0](https://doi.org/10.1016/S0364-0213(99)00013-0)
- Turnbull R, Royer AJ, Ito K et al. (2017) Prominence perception is dependent on phonology, semantics, and awareness of discourse. *Language, Cognition and Neuroscience* 32: 1017–1033. <https://doi.org/10.1080/23273798.2017.1279341>
- Van der Burght CL, Friederici AD, Goucha T et al. (2021) Pitch accents create dissociable syntactic and semantic expectations during sentence processing. *Cognition* 212: Article e104702. <https://doi.org/10.1016/j.cognition.2021.104702>
- Vanlancker-Sidtis D (2003) Auditory recognition of idioms by native and nonnative speakers of English: It takes one to know one. *Applied Psycholinguistics* 24: 45–57. <https://doi.org/10.1017/S0142716403000031>
- Wang B, Yang Y, and Lu S (2004) Hanyu yunlu cengji jieyou bianjie de shengxue fenxi [Acoustic analysis on prosodic hierarchical boundary of Chinese]. *Shengxue xuebao* 1: 29–36. <https://doi.org/10.15949/j.cnki.0371-0025.2004.01.006>
- Watson DG and Gibson E (2004) The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes* 19: 713–755. <https://doi.org/10.1080/01690960444000070>
- White L (2011) Second language acquisition at the interfaces. *Lingua* 121: 577–590. <https://doi.org/10.1016/j.lingua.2010.05.005>
- Wiener S and Liu J (2021) Effects of perceptual abilities and lexical knowledge on the phonetic categorization of second language speech. *JASA express letters* 1: Article e045202. <https://doi.org/10.1121/10.0004259>
- Witzel N, Witzel J, and Nicol J (2012) Deeper than shallow: Evidence for structure-based parsing biases in second-language sentence processing. *Applied Psycholinguistics* 33: 419–456. <https://doi.org/10.1017/S0142716411000427>
- Xue L and Yue M (2024) Discrepancy in prosodic disambiguation strategies between Chinese EFL learners and native English speakers. *International Review of Applied Linguistics in Language Teaching* 62: 1065–1092. <https://doi.org/10.1515/iral-2022-0004>
- Yang X, Shen X, Li W et al. (2014) How listeners weight acoustic cues to intonational phrase boundaries. *PLoS One* 9: Article e102166. <https://doi.org/10.1371/journal.pone.0102166>
- You H-J (2012) Determining prominence and prosodic boundaries in Korean by non-expert rapid prosody transcription. In: Ma Q, Ding H, and Hirst D (eds) *Proceedings of Speech Prosody 2012*. Tongji University Press, pp. 318–321. <https://doi.org/10.21437/SpeechProsody.2012-81>
- Zhang Y and Ding H (2020) The effect of ambiguity awareness on second language learners' prosodic disambiguation. *Frontiers in Psychology* 11: Article e573520. <https://doi.org/10.3389/fpsyg.2020.573520>

Zhang K and Peng G (2021) The time course of normalizing speech variability in vowels. *Brain and Language* 222: Article e105028. <https://doi.org/10.1016/j.bandl.2021.105028>

## Appendix A

### *Instructions*

When we speak, we naturally group our utterances into smaller segments, often divided by what we call ‘boundaries’. When words are right before boundaries, they tend to sound different to when they are not. For example, listen to the following sentence:

(Recording of a non-manipulated production with a prosodic boundary at a syntactic licensed condition: ‘Put the green frog | in the box.’)

In this sentence, there is a boundary after the word ‘frog’.

Now, listen to another example:

(Recording of a non-manipulated production with a prosodic boundary at a syntactic unlicensed condition: ‘Put the frog that is green | in the box.’)

In this sentence, there is a boundary after the word ‘green’.