PASAA Journal
Volume 72, January – June 2026, 357–379

Beyond Proficiency: How Speech Attribute Alignment Shapes L2 Listening Comprehension

Simon Moxon[a]* and Nantana Sittirak[b]

[a] Walailak University, Nakhon Si Thammarat, Thailand
[b] Prince of Songkhla University, Trang, Thailand
**Corresponding author: simon.mo@mail.wu.ac.th*

Article information	
Abstract	This study employed a quantitative observational design to investigate whether L2 learners' speaking proficiency, temporal-prosodic speech attributes, and speaker-listener temporal-prosodic alignment are associated with performance on a L2 listening test. Speaker-listener temporal-prosodic alignment was operationalized as the degree of similarity between learners' produced speech and the temporal-prosodic attributes of the listening test input. This approach assumes that similarity in timing and pitch patterns may reduce perceptual processing demands during speech recognition. 91 Thai undergraduates submitted ten weekly audio journals using ALL-Talk before completing a listening test structured on the TOEIC listening format. Learner speech and listening test input were analyzed for temporal-prosodic attributes using PRAAT. The student speaking proficiency scores were obtained using Microsoft Azure via the ALL-Talk platform. These attributes were then compared against student performance in the listening test using three analytical approaches: 1) one-way ANOVA and bootstrapped Spearman correlations to examine the relationship between speaking proficiency and listening comprehension. 2) bootstrapped generalized linear models to assess the influence of specific temporal-prosodic attributes. 3) generalized linear mixed-effects models to investigate the role of speaker-listener temporal-prosodic alignment. While L2 speaking proficiency did not significantly predict comprehension performance, generalized linear modelling revealed that Mean Pitch, Syllable Count, and Pause Count were significant predictors. Contrary to speech alignment expectations, generalized linear mixed effects modelling showed that greater speaker-listener divergence in Articulation Rate significantly predicted higher listening test scores, while shorter learner average pause durations relative to the input were also associated with improved performance. These findings complicate similarity-based accounts of speaker-listener temporal prosody alignment and suggest that the speaking-listening relationship may be driven more by specific temporal and prosodic features than by overall proficiency measures or uniform prosodic similarity.
Keywords	L2 Listening Comprehension, Thai EFL Learners, Prosodic Alignment

E-ISSN: 2287-0024

APA citation:	Moxon, S., & Sittirak, N. (2026). Beyond Proficiency: How Speech Attribute Alignment Shapes L2 Listening Comprehension. *PASAA*, *72*, 357–379.

1. Introduction

Listening comprehension (LC) is widely regarded as a critical facet of second language (L2) acquisition (Liu, 2020). Yet it remains the most underexplored and conceptually underdeveloped of the four core skills (Reed & Liu, 2020; Rukthong, 2021; Vandergrift, 2007). LC is a complex, multi-layered process shaped by the interaction of acoustic, lexical, syntactic, and discourse-level information. A substantial body of research has demonstrated that prosodic and temporal cues, such as speech rate, pitch variation, accent, and pausing, play an important role in facilitating segmentation and interpretation by marking boundaries and signaling prominence (e.g., Blau, 1990; Chiu & Chen, 2023; Cutler et al., 1997; Derwing & Munro, 2001; Fujita, 2017; Gilakjani & Ahmadi, 2011; Hisagi et al., 2024).

To date, research has mostly studied these influences from the speaker's perspective and focused on how manipulation of temporal-prosodic speech attributes (TPSA), such as speech rate, can aid or hinder LC (Blau, 1990; Fujita, 2017; Griffiths, 1990, 1992; Hayati, 2010; Medina et al., 2020). In the present study, TPSA refers to measurable suprasegmental features of speech, including articulation rate, pause distribution, pitch height, and pitch range, which characterize temporal-prosodic patterns in both learner production and input speech. While prior work has demonstrated how input characteristics influence LC, comparatively little attention has been paid to the potential role of listener-specific TPSA or to the influence of speaker-listener TPSA alignment.

If production and perception rely on partially shared temporal and prosodic representations, then listeners' own speech attributes may shape how incoming input is segmented and interpreted. From this perspective, speaker-listener TPSA alignment, defined here as the degree of similarity or divergence between a listener's TPSA and those present in the listening input, offers a theoretically grounded means of examining how production-based representations interact with perceptual processes during LC. Investigating alignment is therefore significant not only for refining theoretical models of the production-perception interface, but also for pedagogy, as comprehension may be influenced by the relationship between learners' speech patterns and input characteristics.

Despite this theoretical and pedagogical relevance, listener-specific TPSA and speaker-listener TPSA alignment remain underexamined in L2 listening research. In response to these gaps, the present study investigated how L2 LC test performance is influenced by L2 learners' speaking proficiency, their L2 TPSA, and the degree of speaker-listener TPSA alignment. By examining these interrelated facets, the study aims to clarify the role of temporal-prosodic variation in L2 listening and contribute to a more integrated account of speaking and listening development.

2. Literature Review

The present review is grounded in production-perception accounts of L2 speech processing, which propose that speaking and listening rely on partially shared phonological representations (Bloomfield et al., 2010; Ernestus et al., 2017). Variation in temporal and pitch-related speech attributes may affect LC not only by altering the acoustic signal, but by shaping

how listeners map input onto shared phonological representations during comprehension (Cutler et al., 1997; Field, 2008).

Despite the central role of prosodic cues in supporting LC, empirical research has given comparatively limited attention to how these cues are processed in instructed L2 classroom contexts, particularly with respect to how listeners' own speech patterns may shape comprehension outcomes (Reed & Liu, 2020; Rukthong, 2021; Vandergrift, 2007). LC is shaped by interacting factors, including TPSA like speech rate, pitch, pausing, and fluency (Chiu & Chen, 2023; Hisagi et al., 2024). Temporal features, such as articulation rate and syllable prominence, have been shown to significantly affect speaking and listening proficiency, particularly for less proficient learners, who may struggle to parse and comprehend speech characterized by rapid articulation rates and reduced syllable prominence (Kallio et al., 2022). However, limited research has examined the extent to which L2 speech production and speaker-listener TPSA alignment influence L2 LC.

L2 speaking proficiency is typically measured using combined ratings of fluency, accuracy, and intelligibility. However, such global proficiency measures may obscure the specific phonetic attributes that shape speech production (Derwing & Munro, 2001; Thomson, 2015). From a production-perception perspective, speaking reflects the learner's internalized phonological representations and habitual TPSA, which may not align with the cues most salient during listening (Bloomfield et al., 2010; Lesnov et al., 2020). Consequently, while speaking and listening draw upon overlapping phonological resources (Vandergrift & Goh, 2012), overall speaking proficiency may not directly predict LC, particularly if their relationship depends on more specific prosodic features.

In this study, TPSA encompasses both temporal features (articulation rate, rhythm, pause distribution) and tonal features (pitch height and pitch range). These attributes function together to signal prominence, mark boundaries, and organize discourse during speech processing (Cutler et al., 1997; Field, 2008). Temporal cues support speech segmentation and processing speed (Blau, 1990; Field, 2008), whereas pitch-related cues contribute to prominence detection and discourse-level interpretation (Ladd, 2008; Vandergrift & Goh, 2012). Examining these features collectively as TPSA allows for a more precise investigation of how structured acoustic variation may influence LC. The following sections first examine the temporal and tonal dimensions of TPSA separately, then consider speaker-listener TPSA alignment.

2.1 Speech Rate

The effect of speech rate on LC has received substantial attention from the reviewed literature, although the findings regarding its influence, particularly on L2 LC, have failed to converge on a definitive answer. Broadly speaking, slower speech rates tend to support lower-proficiency learners by affording them more processing time for segmentation and lexical access (Chiu & Chen, 2023; Fujita, 2017). In contrast, excessively slow speech has been shown to disrupt the natural processing of rhythmic and prosodic cues that typically support parsing (Griffiths, 1990; Hayati, 2010). Medina et al. (2020) argue that faster speech rates primarily affect higher-proficiency learners, suggesting that the influence is only evident with more proficient learners because those with a lower-proficiency may already be constrained by baseline comprehension limitations, which may be unaffected by increases in speech rate. These findings suggest that speech rate alone is neither universally facilitative nor detrimental.

To a greater extent, its effects may depend on the pace of the input speech in relation to the processing capacity of the listener.

Interpreting the different theories and findings regarding the effects of speech rate on LC is complicated by methodological inconsistencies in the way in which speech rate itself is measured and defined (Blau, 1990; Chui & Chen, 2023; Zhao, 1997). In the reviewed literature, the authors found that speech rate was reported in a range of units, including words per minute (WPM), syllables per second (SPS), or phonemes per second (PPS). According to Griffiths (1990), each of these units capture partially distinct temporal properties. In addition, there is little consensus between studies regarding what constitutes fast or slow speech (Chui & Chen, 2023). Beyond these measurement issues, speech rate is typically treated as a speaker-driven variable, despite the emerging evidence suggesting that speaker-listener relationships may play an important role (Lesnov et al., 2020). This raises the question of whether the listeners' habitual L2 temporal speech patterns significantly influence their ability to process spoken input.

2.2 Pitch and Pitch Processing in Tonal-L1 Listeners

For tonal-language learners in particular, L2 pitch differences may prove to be a problematic aspect of LC, particularly when the listener's focus is on detecting boundary markings, prominence, and discourse-level meaning. In tonal languages, such as Thai, pitch conveys lexical meaning (Burnham et al., 2015; Gandour et al., 1994; Liu et al., 2022). Conversely, in English, it primarily functions at the suprasegmental level. As a result, learners of such tonal languages may approach English LC tasks with heightened pitch sensitivity (Lyu et al., 2024). This, in turn, may influence the way in which they process intonation and rhythm.

While the effects of pitch are well-documented in L2 speech production (e.g., Fang et al., 2024; Kallio et al., 2022; Thomson, 2015; Trofimovich & Baker, 2006), its influence on L2 LC remains undetermined. There is limited empirical evidence to determine whether tonal sensitivity facilitates or interferes with comprehension, particularly where speaker-listener TPSA alignment is concerned (Lesnov et al., 2020; Vandergrift & Goh, 2012).

2.3 Pauses

Pauses facilitate the segmentation of speech into manageable units, thereby aiding word recognition and syntactic parsing (Blau, 1990; Derwing, 1990; Zhao, 1997), which is particularly beneficial for lower-proficiency learners. However, research has shown that excessive or poorly timed pauses can disrupt the natural fluency of speech and break syntactic and semantic continuity, thereby hindering comprehension (Coulange et al., 2024; Griffiths, 1992; Thomas, 2015). Hisagi et al. (2024) suggest that advanced learners may have the ability to compensate for such disruptions by utilizing their prosodic skills.

Nonetheless, the relationship between listeners' own L2 pausing patterns and their ability to process pauses in input speech remains underexamined in the reviewed literature. Moreover, the effects of pauses are often examined in isolation without considering their interaction with other TPSA, such as pitch, speech rate, and fluency (Coulange et al., 2024; Fang et al., 2021). For example, Thai speakers listening to English may misinterpret mid-sentence pauses as boundary markers due to L1 prosodic expectations (Iwasaki & Ingkaphirom, 2005), potentially increasing processing demands under limited working memory conditions (Sweller, 1994). These findings suggest that comprehension appears to rely on the interaction

of multiple prosodic cues rather than isolated features (Bloomfield et al., 2010; Ernestus et al., 2017).

2.4 Speaker Input

Despite advances in understanding how temporal and pitch-related features influence L2 listening (Chiu & Chen, 2023; Cutler et al., 1997; Field, 2008; Fujita, 2017), comparatively little attention has been given to the dynamic relationship between the acoustic properties of input speech and the listener's own habitual speech patterns. Research in dialogue and phonetic convergence suggests that speakers and listeners tend to align at multiple linguistic levels, including prosody, through adaptive mechanisms that support comprehension and mutual understanding (Pickering & Garrod, 2004; Shockley et al., 2004). Extending this perspective to L2 listening, speaker-listener TPSA alignment refers to the degree of similarity or divergence between a listener's TPSA and those present in the input signal.

From a perceptual standpoint, when the listener's existing TPSA closely resembles incoming speech, segmentation may be facilitated and processing demands reduced, as listeners map acoustic cues onto their own TPSA (Field, 2008; Vandergrift & Goh, 2012). When they differ, processing may become more difficult or, in some contexts, boundary cues may become more noticeable (Bloomfield et al., 2010; Ernestus et al., 2017). Alignment, therefore, should not be assumed to be inherently beneficial. Rather, it represents a theoretically plausible mechanism through which production-based representations may interact with perception during LC. However, empirical investigations directly examining speaker-listener TPSA alignment in relation to L2 LC remain scarce.

Building on the production-perception framework and the empirical gaps identified above, the present study examines the production-perception interface in L2 processing by investigating how LC relates to speaking proficiency, specific TPSA, and speaker-listener TPSA alignment. Rather than assuming that overall proficiency alone determines listening success, the study distinguishes between composite speaking proficiency scores and specific acoustic features, conceptualizing alignment as a mechanism through which production-based representations may influence perceptual processing.

To address the identified conceptual and empirical gaps, the following research questions guide this study:

RQ1: Does L2 speaking proficiency correlate with listening comprehension?

RQ2: To what extent do learners' L2 temporal-prosodic speech attributes affect their listening comprehension?

RQ3: Does speaker-listener temporal-prosodic alignment influence L2 listening comprehension?

Collectively, these questions aim to clarify how proficiency, specific TPSA, and speaker-listener TPSA alignment shape listening outcomes, with implications for understanding the production-perception relationship in L2 speech processing and for integrated speaking-listening pedagogy.

3. Methodology

The present study employed a quantitative observational design to examine the relationship between learners' TPSA and L2 LC. To address the research questions, analyses were conducted at both the student level (RQ1-RQ2) and item level (RQ3), thereby enabling

examination of overall proficiency effects alongside speaker-listener TPSA alignment across repeated listening tasks.

3.1 Research Design

The design incorporated both student-level and item-level analyses to address the research questions. At the student level, analyses examined the relationship between speaking proficiency and LC (RQ1) and the contribution of learner-produced TPSA to listening outcomes (RQ2). At the item level, analyses tested whether speaker-listener TPSA alignment or divergence predicted comprehension accuracy across repeated listening tasks (RQ3). This two-level approach enabled direct comparison between global proficiency-based predictors and fine-grained speaker-listener TPSA alignment measures.

3.2 Participants

Participants were 91 Thai undergraduate students (aged 18-20; male = 23, female = 68) enrolled in an English program at a university in southern Thailand. All participants reported Thai as their L1, with English learned as a foreign language through formal instruction within an EFL context. No participants reported bilingual or multilingual L1 status. Participants were recruited through convenience sampling from intact classes in a compulsory English communication course that incorporated weekly spoken reflection tasks.

As part of regular course activities, students submitted ten weekly audio journals reflecting on their classroom learning via the ALL-Talk platform (Moxon, 2024). Participation in the study was voluntary, and students were invited to complete a listening test structured on the TOEIC listening format at the end of the ten-week period.

Ethical approval for this study was obtained from the university's Human Research Ethics Committee. Participants were informed of the study procedures, and verbal informed consent was obtained prior to data collection.

3.3 Instruments

Two primary instruments were used for data collection: weekly learner-produced audio journals and a listening test structured on the TOEIC listening format, developed in-house using publicly available item specifications.

The audio-journal format was selected to capture learner speech under relatively low-stress conditions, thereby providing a more stable estimate of habitual temporal and pitch behavior than tightly constrained reading tasks. The use of multiple weekly submissions enabled aggregation of learner speech across repeated recordings, thereby reducing the influence of task-specific performance effects and supporting more stable estimates of habitual temporal and pitch-related behavior. Learner recordings were transcribed to provide reference text for scripted pronunciation assessment, after which both the audio and transcript were submitted to Microsoft Azure Cognitive Services (Microsoft, n.d.-a) for evaluation via the ALL-Talk interface.

Microsoft Azure offers two methods of speech evaluation. 1) A scripted (referenced text) method, which compares the audio against the reference text to evaluate the speech in terms of completeness, fluency, prosody, and pronunciation accuracy. 2) An unscripted method, which uses Automatic Speech Recognition (ASR) to determine what was most likely said, and evaluate the speech based on fluency and the pronunciation accuracy of the words

detected by the ASR models. The scripted method of evaluation was selected over the unscripted version because unscripted (ASR) evaluation can obscure pronunciation and fluency scores by overlooking omitted and inserted words and incorrectly recognizing mispronounced words (Cámara-Arenas et al., 2023).

LC was assessed using an in-house listening test structured on the TOEIC listening format. The instrument comprised 40 multiple-choice items distributed across three section types: question-response, picture description, and short conversations. Each item was scored dichotomously (correct = 1, incorrect = 0), producing both total LC scores and item-level correctness data for subsequent modelling.

Test materials were developed in-house using publicly available TOEIC-style specifications and example formats to approximate the interactional and temporal demands of each listening task. Audio prompts were recorded in naturally paced spoken English and designed to reflect short-form communicative exchanges that require rapid processing of lexical, syntactic, and prosodic information. The listening stimuli were retained as individual audio files and subsequently analyzed in PRAAT to derive temporal-prosodic measures for the speaker-listener TPSA alignment analyses.

The TOEIC-style format was selected because it provides an ecologically valid framework for examining LC under real-time listening conditions while maintaining consistency in task structure across participants. Although not administered as an official TOEIC assessment, the use of standardized task types enabled systematic investigation of how speaker-listener TPSA alignment relates to item-level listening performance (Buck, 2001).

To evaluate the internal consistency of the LC instrument, Cronbach's alpha was calculated across test items. The resulting reliability coefficient (α = .76) indicated acceptable internal consistency (Babyak, 2004), suggesting that the test items functioned cohesively as a measure of LC ability.

3.4 Data Collection

Data collection proceeded in four stages. First, participants submitted one audio journal per week for ten consecutive weeks via the ALL-Talk platform (Moxon, 2024). Each journal consisted of a short, spoken reflection on recent classroom learning and was recorded and uploaded through the web-based interface.

Second, each audio submission was transcribed to produce a corresponding reference text. The paired audio and transcript were processed through the ALL-Talk interface for automated pronunciation assessment, and resulting scores were stored in the database for later aggregation.

Third, following the ten-week journalling period, participants completed a LC test structured on the TOEIC listening format, comprising question-response, picture description, and short conversation sections. Listening items were presented in a controlled classroom setting, and participant responses were recorded as correctness scores (0/1) for each item. The audio files corresponding to each LC test question were subsequently analyzed in PRAAT to derive temporal-prosodic measures used in the speaker-listener TPSA alignment analyses.

Finally, audio files for both learner journals and listening-test questions were retained for subsequent TPSA analysis, and the resulting speaking, listening, and item-level datasets were compiled for statistical modelling.

3.5 Data Processing and Preparation

Audio data from student journals and listening test prompts were analyzed using PRAAT v6.4.04 (Boersma & Weenink, 2024) and the syllable-rate script by De Jong and Wempe (2009). Extracted temporal-prosodic features included Pause Count, Syllable Count, Speech Duration, Phonation Time, Articulation Rate, Speech Rate, Average Syllable Duration (ASD), Average Pause Duration (APD), and pitch measures (minimum, maximum, and mean). Definitions and units for all derived variables are provided in Table 1. These features were selected to capture two broad prosodic domains central to the research questions: (a) temporal organization (speech rate, pausing, syllable timing), and (b) pitch implementation (height and variability).

Table 1

Definitions and Units for PRAAT and ALL-Talk Derived Variables

Variable	Unit	Operationalization
Pause Count		Total silent pauses detected in the recording.
Syllable Count		Total syllables detected in the recording.
Speech Duration	s	Total duration including pauses.
Phonation Time	s	Speech duration excluding pauses.
Speech Rate	SPS	Syllable Count / Speech Duration.
Articulation Rate	SPS	Syllable Count / Phonation Time.
Average Syllable Duration (ASD)	s	Phonation Time / Syllable Count.
Average Pause Duration (APD)	s	(Speech Duration - Phonation Time) / Pause Count.
Min Pitch	Hz	The lowest frequency detected in the recording.
Max Pitch	Hz	The highest frequency detected in the recording.
Mean Pitch	Hz	The average frequency of the recording.
Pronunciation	%	Closeness to nativelike pronunciation of each phoneme.
Fluency	%	Smoothness of speech, including pace and pausing patterns.
Completeness	%	Ratio of pronounced words based on the reference text.
Overall Speech	%	Weighted composite score integrating pronunciation, fluency, prosody, and completeness.

Note: s = seconds, SPS = syllables per second, Hz = Hertz

In this study, learners' L2 TPSA refer specifically to the set of temporal and pitch-based production measures derived from the PRAAT analysis, including articulation rate, mean pitch, pitch range, pause count, APD, ASD, and syllable count.

In addition to PRAAT-derived measures, learner speaking proficiency was evaluated using the scripted pronunciation assessment functionality of Microsoft Azure Cognitive Services. Audio recordings and corresponding reference transcripts were submitted via the ALL-Talk interface, which returned phrase-, word-, syllable-, and phoneme-level scores for pronunciation accuracy, fluency, prosody, and completeness. These measures were aggregated across the ten-week submission period to obtain stable estimates of overall speaking proficiency.

Three participants who completed the audio-journal component did not attend the listening test and were therefore excluded from subsequent analyses, resulting in a final analytical sample of 88 students. Normality diagnostics were conducted only for the continuous student-level predictor variables used in the ANOVA and GLM analyses (e.g., articulation rate, pitch, APD, syllable count, and speaking proficiency measures), rather than for the binary item-level correctness outcome used in the GLMM. Inspection of skewness and kurtosis values indicated minimal deviation from normality (skewness = 0.29, *SE* = 0.26; kurtosis = 0.08, *SE* = 0.51). No missing or implausible values were detected.

To facilitate the comparison of listening test audio and listener variables, Mean-, Max-, and Min-Pitch variables were converted from Hertz to semitones using the formula 12 x $\log_2$(F0 / 100), enabling Mean Pitch (height) and Pitch Range (variability) to be treated as distinct constructs (Ladd, 2008). Pitch Range was then calculated as the difference between maximum and minimum semitone values. This transformation was intended to improve interpretability by expressing pitch values in perceptually meaningful units. Remaining continuous predictors were z-standardized to mitigate scaling effects and further reduce multicollinearity within the ordinary least squares (OLS) modelling framework.

For RQ3, absolute difference scores were calculated for each of the TPSA (|Student - Question|), thereby creating item-level variables that reflect the degree of speaker-listener alignment. In this study, speaker-listener TPSA alignment was operationalized as the distance between learner and input TPSA, enabling direct testing of whether smaller distances (greater similarity) correspond to higher accuracy, or whether the reverse pattern emerges.

Because absolute difference scores only capture the magnitude of speaker-listener TPSA differences, supplementary post hoc analyses were also conducted using signed difference scores (Student - Question). Unlike absolute scores, signed scores preserve the direction of difference, thereby enabling examination of whether learners performed differently when their TPSA were higher or lower than those of the listening input. These exploratory analyses were conducted to determine whether the effects observed in the alignment models reflected similarity alone, or directional divergence between speaker and listener TPSA.

Two additional datasets were created:

1. A student-level dataset containing mean speech scores and overall comprehension accuracy.

2. An item-level dataset containing the students' difference scores for each question, and correctness (0/1).

This two-level preparation aligned the analyses with the research questions: RQ1 and RQ2 focus on between-learner patterns (student level outcomes), while RQ3 requires item-level modelling that accounts for repeated observations per learner.

Due to many of the temporal variables being mathematically interrelated, Articulation Rate (syllables per second during phonation) was selected as the primary measure of speech-rate. Speech Rate in WPM was avoided because longer, multisyllabic words could artificially depress WPM scores, which could reduce interpretability when comparing learners who may differ in lexical choice or morphological complexity. Likewise, SPS was retained over PPS to maintain cross-linguistic comparability (Pellegrino et al., 2011) and to reduce the influence of coarticulation and reduction effects that can alter phoneme-level counts without necessarily reflecting timing changes (De Jong & Wempe, 2009). Retaining APD and ASD enabled

analysis of pausing and syllable timing behavior while reducing redundancy from other temporal variables.

Pronunciation and Fluency were initially included in preliminary diagnostic models. However, both showed extreme multicollinearity with Overall Speech (Pronunciation: VIF = 27.41, Tolerance = .036; Fluency: VIF = 5.44, Tolerance = .184; Overall Speech: VIF = 41.73, Tolerance = .024). This was expected as Azure evaluates Overall Speech by sorting the scores for Pronunciation, Fluency, Prosody, and completeness from lowest to highest (S_1-S_4) before applying the weighted formula: Overall Speech = $0.4S_1 + 0.2S_2 + 0.2S_3 + 0.2S_4$ (Microsoft, n.d.-b). Moreover, Fluency was also found to be closely aligned with speech and articulation rates as well as pausing attributes. Therefore, Overall Speech was retained as the most reliable measure of speaking proficiency for RQ1, while specific temporal and pitch measures were retained to address RQ2 and RQ3.

3.6 Data Analysis

The original selection of temporal and pitch related variables was informed by Lesnov et al. (2020), who studied the relationship and interplay between the learner's TPSA and their L2 LC performance. However, analytical procedures and certain speech variables were adapted to address multicollinearity issues observed in the present dataset.

Prior to model specification, multicollinearity diagnostics were performed on the full set of temporal and pitch-related variables. Initial collinearity tests on raw speech measures revealed substantial redundancy among predictors. Max Pitch was automatically excluded by SPSS due to a tolerance value of .000, while Min Pitch and Pitch Range were strongly associated (Condition Index = 43.08; Variance Proportions = .73, .75), indicating that these variables were statistically difficult to separate and risked producing unstable estimates (Belsley et al., 1980; Stevens, 2012). These diagnostics informed subsequent predictor transformation and selection decisions prior to model estimation to mitigate instability arising from overlapping variance.

In addition, Spearman's rho and, where applicable, 5,000 gender-stratified BCa bootstrapped confidence intervals were employed to enhance estimate robustness and reduce reliance on strict distributional assumptions. Table 2 summarizes the outcome variable, predictor type, and statistical model specified for each research question.

Table 2

Analytical Models for Each Research Question

	Outcome Variable	Predictor Type	Predictors	Statistical Procedure
RQ1	LC	Percentile-based learner scores	Overall Speech – (Low, Med, High)	One-Way ANOVA / Spearman's correlation
RQ2	LC	Learner-produced TPSA	Articulation Rate Mean Pitch Pitch Range Pause Count APD Syllable Count Proficiency Level (Fixed Factor)*	GLM

RQ3	LC	Learner-input TPSA difference scores	Articulation Rate Mean Pitch Pitch Range Pause Count APD Proficiency Level (Fixed Factor)*	GLMM

Note. RQ = Research Question. LC = Listening Comprehension. APD = Average Pause Duration. TPSA = Temporal-Prosodic Speech Attributes. ASD = Average Syllable Duration. Low, Med, and High refer to the lower, intermediate, and upper percentile scores, respectively, * = Used in supplementary / exploratory comparisons. GLM (multicollinearity diagnostics conducted prior to model specification).

To examine the influence of the learners' overall L2 speaking proficiency on their performance in a LC test (RQ1), a one-way ANOVA was conducted based on 5,000 gender-stratified BCa bootstrapped samples. The LC scores were compared across three speaking proficiency level cut-off points (lower < 50th percentile; intermediate 50th - 75th percentiles; and upper > 75th percentile). To complement ANOVA analysis, Spearman rank-order correlations were calculated using 5,000 gender-stratified BCa bootstrap samples to assess the relationship between speaking proficiency and LC. Group differences and associations were interpreted using *F* statistics, correlation coefficients (ρ), and associated probability values.

For RQ2, preliminary OLS regression diagnostics indicated severe multicollinearity among the raw predictors (see Table 3). Condition indices exceeded 50, and multiple predictors showed strong cross-loadings across shared eigen-dimensions, indicating instability in coefficient estimation. Subsequent correlation analysis confirmed substantial shared variance, particularly between Pause Count and Syllable Count and between Pitch Range and Syllable Count (see Table 4).

Table 3

Multicollinearity Diagnostics for the Full ENTER Regression

				Variance Proportions					
Dimension	Eigenvalue	Condition Index	Constant	Articulation Rate	Mean Pitch	Pitch Range	Pause Count	APD	Syllable Count
1	6.712	1.000	.00	.00	.00	.00	.00	.00	.00
2	.177	6.166	.00	.00	.02	.00	.18	.01	.07
3	.049	11.705	.00	.00	.13	.00	.28	.08	.37
4	.039	13.055	.00	.01	.34	.00	.28	.07	.27
5	.014	22.032	.01	.12	.20	.12	.00	.60	.13
6	.007	30.820	.01	.30	.09	.77	.11	.04	.01
7	.002	55.090	.98	.57	.22	.11	.15	.20	.15

Table 4

Correlations Among Speech Predictors

Variable	Articulation Rate	Mean Pitch	Pitch Range	Pause Count	APD	Syllable Count
Articulation Rate	—	–.20	.13	–.04	.03	.30
Mean Pitch	–.20	—	.17	–.18	–.26	–.03
Pitch Range	.13	.17	—	.36	–.20	.44
Pause Count	–.04	–.18	.36	—	–.19	.66
APD	.03	–.26	–.20	–.19	—	–.21
Syllable Count	.30	–.03	.44	.66	–.21	—

Note. Bold values indicate correlations between predictor variables identified as contributing to multicollinearity in preliminary regression diagnostics.

Based on these initial diagnostic tests, a bootstrapped generalized linear model (GLM) was constructed to assess the contribution of learner-produced TPSA to LC outcomes at the student level. Z-standardized variables were used to reduce the effects of scaling and collinearity related distortions. Additionally, to assess whether these relationships were reliant on overall speaking proficiency, a GLM including Proficiency Level as a fixed factor was estimated. Predictor effects were evaluated using standardized regression coefficients (β) and corresponding significance levels within the GLM framework.

For RQ3, a binomial generalized linear mixed-effects model (GLMM) with a logit link was estimated, treating each student's 40 listening test responses as repeated measures. The model included a random intercept for Student ID to account for between-learner variability and allowed item-level variability through a diagonal covariance structure. Fixed-effect estimates from the GLMM were used to assess whether speaker-listener TPSA alignment predicted the likelihood of item-level listening accuracy.

4. Results

4.1 Research Question 1: Does L2 speaking proficiency correlate with listening comprehension?

As illustrated in Table 5, the ANOVA showed no statistical significance between the three proficiency levels. Furthermore, bootstrapped Bonferroni adjusted post hoc comparisons confirmed the absence of any meaningful differences between the three groups, with all confidence intervals spanning zero and, therefore, indicating a high degree of uncertainty around any potential difference.

Table 5

ANOVA and Post-Hoc Test Results for Listening Comprehension Accuracy by Speaking Proficiency

Proficiency Measure	Test	Statistic	*df*1	*df*2	*p*	η^2	95% CI (Post-Hoc)
Overall Speech	Levene's Test (Mean)	1.60	2	85	.207		
	ANOVA	0.48	2	85	.619	.011	
	Bonferroni (Max Δ)	1.36					[–1.73, 4.42]

Note. Bonferroni maximum pairwise difference (Max Δ) values are based on bootstrapped estimates (5,000 stratified samples by gender). *p* values are two-tailed.

As shown in Table 6, correlations were non-significant for the full sample as well as within each proficiency band. Effect sizes ranged from weak to moderate, but none approached thresholds typically associated with practical significance (Plonsky & Oswald, 2014). The null findings were consistent across parametric and non-parametric analysis and across proficiency strata.

Table 6

Spearman's ρ Correlations Between L2 Speaking Proficiency and Listening Test Score (With Bootstrapped Estimates)

					Spearman's Correlation Bootstrap		
Predictor	*Group*	*N*	*ρ*	*p*	Bias	*SE*	95% CI
Overall Speech	All	88	0.10	.356	–0.002	0.109	[–0.12, 0.31]
	Lower	22	0.38	.081	–0.014	0.211	[–0.06, 0.73]
	Intermediate	44	0.05	.731	0.002	0.154	[–0.26, 0.35]
	Upper	22	–0.14	.522	0.009	0.212	[–0.50, 0.27]

Note. Bootstrapped estimates based on 5,000 samples stratified by gender. *p* values are two-tailed.

4.2 Research Question 2: To what extent do learners' L2 temporal-prosodic speech attributes affect their listening comprehension?

The GLM significantly predicted comprehension outcomes, $F(6, 81) = 2.40$, $p = .035$, accounting for 15.1% of the variance ($R^2 = .151$; Adjusted $R^2 = .088$). As can be seen in Table 7, three predictors (Syllable Count, Pause Count, and Mean Pitch) were found to be statistically significant, whereas Articulation Rate, Pitch Range, and APD were non-significant.

Table 7

Bootstrapped GLM Predicting Listening Comprehension (Overall Model)

Predictor	*B*	*p*	Partial η²	BCa 95% CI
Articulation Rate	−0.39	.419	.006	[−1.29, 0.46]
Mean Pitch	−0.30	.032*	.064	[−0.59, −0.02]
Pitch Range	0.16	.450	.007	[−0.29, 0.57]
Pause Count	−1.28	.032*	.039	[−2.50, −0.03]
APD	0.75	.149	.026	[−0.30, 1.77]
Syllable Count	1.55	.013*	.050	[0.25, 2.80]

Note. Model fit: $F(6, 81) = 2.40$, $p = .035$; $R^2 = .151$, Adjusted $R^2 = .088$. Confidence intervals based on 5,000 BCa bootstrapped samples stratified by gender. *p* values are two-tailed. * Statistically significant at $p < .05$.

To assess whether these relationships were reliant on overall speaking proficiency, a supplementary GLM including Proficiency Level as a fixed factor was estimated. This model yielded a comparable pattern of results: Proficiency Level itself did not significantly predict comprehension, $F(2, 79) = 1.05$, $p = .355$, nor did it moderate the effects of the TPSA. Importantly, the inclusion of Proficiency Level did not meaningfully alter the direction or magnitude of the significant predictors. Given the absence of moderation and the principle of model parsimony, the initial GLM was retained as the primary analysis for RQ2.

4.3 Research Question 3: Does speaker-listener temporal-prosodic alignment influence L2 listening comprehension?

Model fit indices for the GLMM indicated an acceptable balance between explanatory power and complexity (-2 log likelihood = 15259.89; AICc = 15340.84; BIC = 15586.47), consistent with recommended criteria (Burnham & Anderson, 2002). The overall fixed-effects model was statistically significant, $F(5, 3514) = 2.47$, $p = .031$. As shown in Table 8 and

illustrated in Figure 1, difference scores for Articulation Rate and APD emerged as statistically significant predictors of listening accuracy, whereas difference scores for Mean Pitch, Pitch Range, and Pause Count were not (all $p > .34$).

Table 8

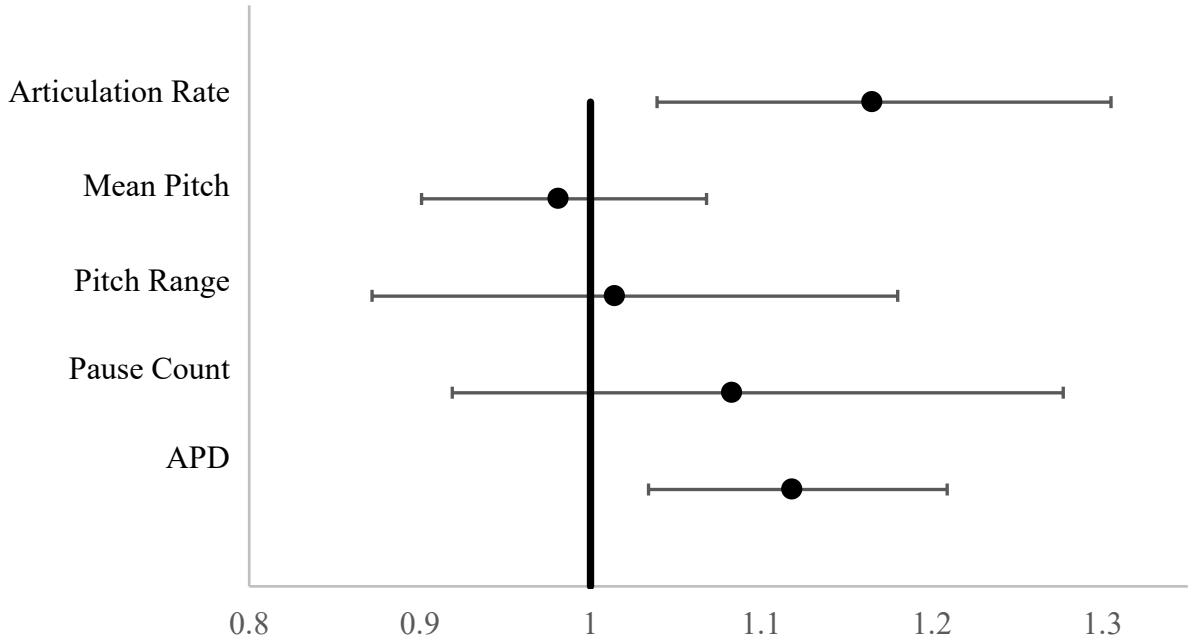
Generalized Linear Mixed-Effects Model Predicting Listening Accuracy

Predictor	*F*	*df1*	*df2*	*B*	*SE*	*t*	*p*	95% CI	Exp (*B*)	95% CI for Exp(*B*)
Diff Articulation Rate	7.771	1	3514	.111	.0400	2.788	.005*	[0.03, 0.19]	1.118	[1.03, 1.21]
Diff Mean Pitch	0.904	1	3514	.080	.0840	.951	.342	[−0.09, 0.25]	1.083	[0.92, 1.28]
Diff Pitch Range	0.033	1	3514	.014	.0773	.182	.856	[−0.14, 0.17]	1.014	[0.87, 1.18]
Diff Pause Count	0.194	1	3514	−.019	.0432	−0.441	.659	[−0.10, 0.07]	0.981	[0.90, 1.07]
Diff APD	6.870	1	3514	.152	.0582	2.621	.009*	[0.04, 0.27]	1.165	[1.04, 1.31]

Note. Model fit: $F(5, 3514) = 2.47$, $p = .031$. *B* = logit coefficient. CI = confidence interval. Exp(*B*) = odds ratio. Predictors reflect absolute difference scores calculated from z-standardized speaker and listener values. Model estimated with a binomial distribution and logit link. * Statistically significant at $p < .05$.

Figure 1

Forest Plot of Odds Ratios (95% CI) for Speech Attribute Difference Scores Predicting Listening Accuracy



A post hoc directional model revealed that the APD effect was contingent upon the direction of mismatch, remaining a significant predictor of LC, $F(1, 3514) = 5.386$, $p = .020$. Inspection of the fixed-effects coefficients further indicated that the directional effect of APD was negative ($\beta = -.130$, $p = .020$), such that increasing learner APD relative to the input was associated with a reduction in the odds of a correct response (OR = 0.878, 95% CI [0.787, 0.980]). In contrast, the effect for Articulation Rate did not retain significance when directional differences were considered ($p = .334$), suggesting that its influence on performance was symmetric with respect to whether learner speech exceeded or fell below the temporal characteristics of the listening input.

These results indicate that only temporal attributes (Articulation Rate and APD) demonstrated statistically significant speaker-listener TPSA alignment effects, whereas pitch-related attributes (Mean Pitch and Pitch Range) and Pause Count did not significantly predict listening accuracy.

5. Discussion

The present study examined the relationship between L2 learners' LC and their overall speaking proficiency, their TPSA, and speaker-listener TPSA alignment. The findings from the three research questions contribute to existing knowledge on the production-perception interface and provide a more detailed account of speaker-listener TPSA alignment in L2 listening.

5.1 Speaking Proficiency and Listening Comprehension (RQ1)

The results of this study showed no significant relationship between speaking proficiency and LC. These findings align with prior research suggesting that speaking and listening rely on overlapping but differentially weighted cognitive resources (Vandergrift & Goh, 2012). Speaking proficiency primarily measures articulatory control, phonological accuracy, and speaking fluency, while LC focuses more on perceptual parsing, lexical access, and real-time integration of acoustic and linguistic cues (Derwin & Munro, 2001; Vandergrift & Goh, 2012).

The present findings suggest that global speaking proficiency measures may obscure more specific production-perception relationships. This viewpoint aligns with the findings of Lesnov et al. (2020). Furthermore, in listening assessments structured on the TOEIC format, performance may reflect the combined influence of multiple factors, including lexical knowledge, familiarity with task demands, and the ability to process spoken input under time constraints (Rukthong, 2021; Wang & Treffers-Daller, 2017). As a result, overall speaking proficiency may offer limited advantage unless it aligns with perceptual mechanisms directly relevant to the listening task.

From a production-perception perspective, shared representations do not imply identical task performance, as global proficiency measures may aggregate multiple components that do not directly map onto perceptual mechanisms (Bloomfield et al., 2010; Derwing & Munro, 2001). The present findings therefore suggest that only specific TPSA, rather than composite proficiency scores, align with the processing demands engaged during LC (Ernestus et al., 2017; Field, 2008). Taken together, these results indicate that L2 listening comprehension is not primarily determined by general articulatory proficiency, but by the efficiency with which learners map acoustic cues onto internal phonological representations during real-time processing.

5.2 Learners' Temporal-Prosodic Speech Attributes as Predictors of Comprehension (RQ2)

Pitch and temporal fluency have been previously highlighted as key predictors of LC (Cutler et al., 1997; Kallio et al., 2022). However, only specific production features were significant contributors in the present study. The analysis revealed that, while Overall Speech was unrelated to comprehension, specific TPSA were associated with listening performance under multivariate modelling conditions. Higher Mean Pitch and greater Pause Count were

associated with lower comprehension accuracy, while higher Syllable Count showed a positive association. These findings suggest that individual suprasegmental features may index underlying processing strategies engaged during listening. Rather than exerting a direct causal influence on LC, learners' TPSA may reflect how attentional resources are allocated and how acoustic cues are parsed, weighted, and mapped onto internal phonological representations during real-time comprehension, processes that are assumed to operate across both speech production and perception (Cutler et al., 1997; Field, 2008; Vandergrift, 2007). In this sense, temporal-prosodic production patterns may function as behavioral indicators of segmentation efficiency and phonological organization, which are central to successful listening comprehension.

The negative association between Mean Pitch and LC may tentatively reflect differences in how Thai learners process pitch variation during L2 speech perception. As a tonal language, Thai relies heavily on pitch height to distinguish lexical meaning, whereas in English, pitch variation primarily signals prominence and discourse structure. It is therefore possible that some learners may attend differently to pitch-related cues during listening, which could influence segmentation and interpretive processing. This interpretation is broadly consistent with accounts of L2 speech perception suggesting that L1 prosodic experience may influence sensitivity to L2 prominence and boundary cues (Cutler et al., 1997; Ernestus et al., 2017). However, as the present study did not directly assess perceptual sensitivity to pitch, this explanation should be regarded as speculative and requiring further empirical investigation (Burnham et al., 2015; Gandour et al., 1994; Liu et al., 2022).

Likewise, given that a high rate of pauses in spoken language often reflects heightened cognitive load and reduced automaticity (Kormos, 2006; Tavakoli & Skehan, 2005), such effects could generalize to LC through shared attentional and working memory constraints (Hulstijn, 2011; Kahng, 2014). If production-related pausing indexes reduced processing automaticity, similar temporal constraints may operate during perception, limiting the efficiency with which listeners parse rapidly unfolding input (Field, 2008; Hulstijn, 2011). However, caution is warranted when attributing causal status to individual TPSA, as these features likely reflect processing tendencies shaped by task demands and characteristics of the input rather than functioning as independent drivers of LC performance (Field, 2008; Vandergrift & Goh, 2012).

In comparison with Lesnov et al. (2020), the present findings reveal points of divergence that warrant consideration. Lesnov et al. reported that wider pitch range and higher speech rate were statistically associated with improved LC, particularly amongst the lower-proficiency learners. However, the present study found no effect of Pitch Range, and no significant interactions with Proficiency Level were observed. In contrast, the results of the current study found that temporal divergence between speaker and listener, specifically in Articulation Rate and APD, were statistically significant predictors of LC performance.

These differences may reflect variation in how specific TPSA relate to processing demands during comprehension. Production features such as pausing and syllable timing may reflect how learners allocate attention or segment speech, processes that are also engaged during listening (Field, 2008; Vandergrift, 2007). From this perspective, certain TPSA may not directly influence LC performance, but instead reflect how efficiently learners manage overlapping processing demands in production and perception (Bloomfield et al., 2010; Hulstijn, 2011). Overall, the findings suggest that listening comprehension is more closely

associated with temporal organization and cue management than with broad indices of speaking proficiency.

5.3 Temporal-Prosodic Alignment Between Listener and Input (RQ3)

From the perspective of speaker-listener TPSA alignment, L2 listening theories suggest that comprehension stems from the interplay between input speech, linguistic structure, and listener-specific processing habits (Bloomfield et al., 2010; Ernestus et al., 2017). Within this framework, speaker-listener TPSA alignment may operate in an attribute-specific and context-dependent manner rather than as a uniformly facilitative mechanism. Consistent with this possibility, the present study did not find a general benefit of speaker-listener TPSA similarity. Absolute difference-score analyses showed that pitch- and pause- related alignment measures were not significantly associated with LC, and proficiency did not significantly moderate alignment effects within the present sample. However, greater absolute divergence in Articulation Rate was associated with higher comprehension scores.

Supplementary post hoc analyses using signed difference scores further indicated that lower learner APD relative to the listening input was associated with improved performance. Unlike the absolute difference score analyses, which captured only the magnitude of speaker-listener divergence, the signed analyses preserved the direction of the difference and were therefore interpreted more cautiously as exploratory findings. Together, these results suggest that the relationship between speaker-listener TPSA alignment and LC may differ according to both the specific temporal attribute examined and the direction of the speaker-listener difference.

Rather than contradicting alignment-based perspectives, the findings suggest that speaker-listener TPSA relationships may operate differently across temporal dimensions. While some forms of similarity may facilitate processing, the present results indicate that divergence in specific temporal attributes may also relate to listening performance under certain conditions. This interpretation aligns with accounts emphasizing the role of variability, segmentation, and adaptive perceptual processing in LC (Cutler et al., 1997; Ernestus et al., 2017). The findings therefore suggest that speaker-listener TPSA alignment may be more attribute-specific and direction-sensitive than similarity-based accounts alone would predict.

5.4 The Role of Proficiency as a Moderating Variable

Although proficiency level was included as a fixed factor in the multivariate models, no significant moderation effects were observed. This finding should be interpreted cautiously. First, the proficiency measure was derived from aggregated automated ratings of pronunciation, fluency, and temporal features, which may not have captured finer-grained perceptual abilities relevant to speaker-listener TPSA alignment. Second, the relatively homogeneous proficiency range of the participant group may have attenuated potential moderation effects, as restricted variance can reduce the likelihood of detecting interaction terms in multivariate modelling (Babyak, 2004). The absence of moderation should not be interpreted as evidence that proficiency is irrelevant to prosodic processing. Instead, it may indicate that global proficiency measures do not sufficiently differentiate learners with respect to the perceptual mechanisms underlying speaker-listener TPSA alignment.

6. Limitations and Future Research

The present study has several limitations that must be acknowledged. A restricted set of TPSA was examined, which excluded rhythm, stress, and full intonation contours. These features are important as they play established roles in speech perception (Trofimovich & Baker, 2006). In addition, the speaking proficiency scores used in this study were derived from automated assessments of pronunciation, fluency, and temporal speech features, which primarily reflect learners' speech delivery rather than their ability to organize and communicate meaning. The absence of a relationship between speaking proficiency and LC may reflect limited overlap between the perceptual-motor skills captured by this measure and the higher-level semantic and discourse processing required for successful LC performance. In speaking assessments where content development and communicative effectiveness form part of the construct, a relationship with LC may be more likely to emerge (Vandergrift & Goh, 2012).

The GLMM also did not incorporate item-level linguistic difficulty, such as lexical, syntactic, or semantic complexity. It is therefore possible that the observed benefit of temporal divergence reflects adaptive temporal adjustment to linguistically demanding input rather than a standalone speaker-listener TPSA alignment effect. Furthermore, although the sample size was adequate for mixed-effects modelling, the relatively homogeneous proficiency range of participants may have reduced the ability to detect associations or interaction effects involving overall speaking proficiency. As Babyak (2004) cautions, restricted variance may attenuate associations, which could partly explain the non-significant finding in the present study.

Future research should incorporate item-level linguistic difficulty (e.g., lexical, and syntactic complexity) to allow for a more fine-grained examination of how TPSA divergence interacts with input characteristics. Integrating these measures would extend the present findings and help clarify how such differences operate in relation to task demands and linguistic complexity in shaping listening outcomes. Examining the direction of speaker-listener TPSA misalignment across learners from different L1 backgrounds may also provide further insight into the conditions under which prosodic divergence supports comprehension.

7. Conclusion

The present study extends prior work by examining how L2 speech production relates to L2 LC. Overall speaking proficiency did not significantly predict listening performance or moderate the effects of TPSA benefits, although some production-based measures were linked to comprehension in certain models. Most notably, speaker-listener TPSA alignment did not uniformly benefit listening accuracy. Greater temporal divergence in Articulation Rate was associated with improved listening test performance, while shorter learner APD relative to the input was linked to higher accuracy. The results underscore the importance of modelling learner and input TPSA together and suggest that speaker-listener TPSA alignment effects may differ according to the specific temporal attribute examined.

These findings also carry several theoretical, methodological, and pedagogical implications. First the results suggest that speaker-listener TPSA alignment may not operate as a uniformly facilitative mechanism in L2 listening. Instead, the effects observed in the present study varied according to the temporal attribute involved and the direction of the difference between speaker and listener speech patterns. This supports perspectives that emphasize adaptive perceptual processing and the role of temporal variation in segmentation and cue salience during L2 listening.

Methodologically, the study highlights the importance of modelling learner and input speech characteristics simultaneously, rather than treating prosodic input features as fixed properties of the listening signal. Incorporating both item-level and listener-specific TPSA may provide a more detailed account of L2 listening performance, particularly in task-based assessment contexts.

From a pedagogical perspective, the findings suggest that temporal sensitivity may operate independently of global speaking proficiency. However, whether targeted training in temporal aspects of speech perception and production can improve L2 listening performance remains uncertain and represents an avenue for future research.

8. About the Authors

Dr. Simon Moxon is an EFL lecturer at Walailak University, Nakhon Si Thammarat, Thailand. He holds a first-class honors degree in Applied Computing from Staffordshire University, UK, as well as a master's degree and PhD in Teaching and Technology from Assumption University, Thailand. His research interests include pronunciation and CALL.

Dr. Nantana Sittirak is an EFL lecturer at Prince of Songkhla University, Trang, Thailand. She holds a PhD in English Language Studies from Thammasat University, Thailand. Her research interests include English language teaching (ELT), translation, and intercultural communication.

9. Acknowledgement

9.1 Declaration of Interest Statement

The authors of this research did not receive any form of grant or funding. All expenses were met by the authors.

One of the authors is employed by the institution where the data were collected. The authors affirm that this affiliation had no role in study design, data collection, analysis, interpretation of the findings, or the decision to submit the manuscript for publication.

The authors declare no conflicts of interest regarding the publication of this article.

10. Declaration of AI Use

The authors declare that no AI tools were used in preparation of the manuscript.

The authors take full responsibility for the content.

11. References

Babyak, M. A. (2004). What you see may not be what you get: A brief introduction to overfitting in regression-type models. *Psychosomatic Medicine, 66*(3), 411–421. https://doi.org/10.1097/00006842-200405000-00021

Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). *Regression diagnostics: Identifying influential data and sources of collinearity.* Wiley. https://doi.org/10.1002/0471725153

Blau, E. K. (1990). The effect of syntax, speed, and pauses on listening comprehension. *TESOL Quarterly, 24*(4), 746–753. https://doi.org/10.2307/3587129

Bloomfield, A., Wayland, S. C., Rhoades, E., Blodgett, A., Linck, J., & Ross, S. (2010). *What makes listening difficult? Factors affecting second language listening comprehension (Technical Report No. TTO 81434 E.3.1)*. University of Maryland, Center for Advanced Study of Language. https://apps.dtic.mil/sti/tr/pdf/ADA550176.pdf

Boersma, P., & Weenink, D. (2024). *PRAAT: Doing phonetics by computer (Version 6.4.04)* [Computer program]. http://www.praat.org/

Buck, G. (2001). *Assessing Listening.* Cambridge University Press. https://doi.org/10.1017/CBO9780511732959

Burnham, D., Kasisopa, B., Reid, A., Luksaneeyanawin, S., Lacerda, F., Attina, V., Rattanasone, N. X., Schwarz, I., & Webster, D. (2015). Universality and language-specific experience in the perception of lexical tone and pitch. *Applied Psycholinguistics, 36*(6), 1459–1491. https://doi.org/10.1017/S0142716414000496

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). Springer. https://doi.org/10.1007/b97636

Cámara-Arenas, E., Tejedor-García, C., Tomas-Vázquez, C. J., & Escudero-Mancebo, D. (2023). Automatic pronunciation assessment vs. automatic speech recognition: A study of conflicting conditions for L2-English. *Language Learning & Technology, 27*(1), 1–19. https://doi.org/10.64152/10125/73512

Chiu, C.-W., & Chen, T.-P. (2023). Speech rate and young EFL learners' listening comprehension. *English Language Teaching*, *16*(7), 74–80. https://doi.org/10.5539/elt.v16n7p74

Coulange, S., Kato, T., Rossato, S., & Masperi, M. (2024). Enhancing language learners' comprehensibility through automated analysis of pause positions and syllable prominence. *Languages*, *9*(3), 78–93. https://doi.org/10.3390/languages9030078

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

De Jong, N. H., & Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods*, *41*(2), 385–390. https://doi.org/10.3758/brm.41.2.385

Derwing, T. M. (1990). Speech rate is no simple matter. *Studies in Second Language Acquisition, 12*(3), 303–313. https://doi.org/10.1017/S0272263100009189

Derwing, T. M., & Munro, M. J. (2001). What speaking rates do non-native listeners prefer? *Applied Linguistics, 22*(3), 324–337. https://doi.org/10.1093/applin/22.3.324

Ernestus, M., Kouwenhoven, H., & van Mulken, M. (2017). The direct and indirect effects of the phonotactic constraints in the listener's native language on the comprehension of reduced and unreduced word pronunciation variants in a foreign language. *Journal of Phonetics*, *62*, 50–64. https://doi.org/10.1016/j.wocn.2017.02.003

Fang, L., Liu, W., Wu, R., Schwieter, J. W., & Wang, R. (2024). The role of prosodic sensitivity and executive functions in L2 reading: The moderated mediation effect. *Bilingualism: Language and Cognition,* 1–12. https://doi.org/10.1017/S1366728924000129

Fang, L., Xie, Y., Yu, K., Wang, R., & Schwieter, J. W. (2021). An examination of prosody and second language sentence processing through pause insertion. *International Journal of Bilingualism, 25*(5), 1473–1485. https://doi.org/10.1177/13670069211018753

Field, J. (2008). Listening in the language classroom. Cambridge University Press. https://doi.org/10.1017/CBO9780511575945

Fujita, R. (2017). Effects of speech rate and background noise on EFL learners listening comprehension of different types of materials. *The Journal of Asia TEFL, 14*(4), 638–653. https://doi.org/10.18823/asiatefl.2017.14.4.4.638

Gandour, J., Potisuk, S., & Dechongkit, S. (1994). Tonal coarticulation in Thai. *Journal of Phonetics, 22*(4), 477–492. https://doi.org/10.1016/S0095-4470(19)30296-7

Gilakjani, A. P., & Ahmadi, M. R. (2011). A study of factors affecting EFL learners' English listening comprehension and the strategies for improvement. *Journal of Language Teaching and Research, 2*(5), 977–988. https://doi.org/10.4304/jltr.2.5.977-988

Griffiths, R. (1990). Speech rate and NNS comprehension: A preliminary study in time-benefit analysis. *Language Learning, 40*(3), 311–336. https://doi.org/10.1111/j.1467-1770.1990.tb00666.x

Griffiths, R. (1992). Speech rate and listening comprehension: Further evidence of the relationship. *TESOL Quarterly, 26*(2), 385–390. https://doi.org/10.2307/3587015

Hayati, A. (2010). The effect of speech rate on listening comprehension of EFL learners. *Creative Education, 1*(2), 107–114. https://doi.org/10.4236/ce.2010.12016

Hisagi, M., Higby, E., Zandona, M., Acosta, A. P., Kent, J., & Tajima, K. (2024). Impact of speech rate on perception of vowel and consonant duration by bilinguals and monolinguals. *JASA Express Letters*, *4*(5), Article 055201. https://doi.org/10.1121/10.0025862

Hulstijn, J. (2011). Language proficiency in native and nonnative speakers. *Studies in Second Language Acquisition, 33*(4), 499–521. https://doi.org/10.1080/15434303.2011.565844

Iwasaki, S., & Ingkaphirom, P. (2005). *A reference grammar of Thai.* Cambridge University Press.

Kahng, J. (2014). Exploring utterance and cognitive fluency of L1 and L2 English speakers: Temporal measures and stimulated recall. *Language Learning, 64*(4), 809–854. https://doi.org/10.1111/lang.12084

Kallio, H., Suni, A., & Šimko, J. (2022). Fluency-related temporal features and syllable prominence as prosodic proficiency predictors for learners of English with different language backgrounds. *Language and Speech, 65*(3), 571–597. https://doi.org/10.1177/00238309211040175

Kormos, J. (2006). *Speech production and second language fluency*. Routledge.

Ladd, D. R. (2008). *Intonational phonology* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511808814

Lesnov, R., Wolhein Nava, S., & Bogorevich, V. (2020). Can successful L2 pronunciation facilitate listening comprehension? The role of speech rate and pitch range. In O. Kang, S. Staples, K. Yaw, & K. Hirschi (Eds.), *Proceedings of the 11th Pronunciation in Second Language Learning and Teaching Conference* (pp. 250–260). Iowa State University. https://www.iastatedigitalpress.com/pslit/article/id/15429/

Liu, L., Lai, R., Singh, L., Kalashnikova, M., Wong, P. C. M., Kasisopa, B., Chen, A., Onsuwan, C., & Burnham, D. (2022). The tone atlas of perceptual discriminability and perceptual distance: Four tone languages and five language groups. *Brain and Language,* 229, Article 105106. https://doi.org/10.1016/j.bandl.2022.105106

Liu, Y. (2020). Effects of metacognitive strategy training on Chinese listening comprehension. *Languages, 5*(2), 21. https://doi.org/10.3390/languages5020021

Lyu, S., Põldver, N., Kask, L., Wang, L., & Kreegipuu, K. (2024). Native language background affects the perception of duration and pitch. *Brain and Language, 256*, Article 105460. https://doi.org/10.1016/j.bandl.2024.105460

Medina, A., Socarrás, G., & Krishnamurti, S. (2020). L2 Spanish listening comprehension: The role of speech rate, utterance length, and L2 oral proficiency. *The Modern Language Journal, 104*(2), 439–456. https://doi.org/10.1111/modl.12639

Microsoft. (n.d.-a). *Transparency note and use cases for pronunciation assessment - Azure AI services*. Microsoft Learn. Retrieved September 26, 2025, from https://learn.microsoft.com/en-us/legal/cognitive-services/speech-service/pronunciation-assessment/transparency-note-pronunciation-assessment

Microsoft. (n.d.-b). *Use pronunciation assessment.* Microsoft Learn. Retrieved September 26, 2025, from https://learn.microsoft.com/en-us/azure/ai-services/speech-service/how-to-pronunciation-assessment?pivots=programming-language-javascript#pronunciation-score-calculation

Moxon, S. (2024). ALL-Talk: Enhancing EFL pronunciation with Microsoft Azure speech services. *ABAC Journal, 44*(4), 139–161. https://doi.org/10.59865/abacj.2024.58

Pellegrino, F., Coupe, C., & Marsico, E. (2011). A cross-language perspective on speech information rate. *Language, 87*(3), 539–558. https://doi.org/10.1353/lan.2011.0057

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*(2), 169–226. https://doi.org/10.1017/S0140525X04000056

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning, 64*(4), 878–912. https://doi.org/10.1111/lang.12079

Reed, M., & Liu, D. (2020). Technology-enhanced L2 listening: Triangulating perception, production, and metalinguistic awareness. In O. Kang, S. Staples, K. Yaw, & K. Hirschi (Eds.), *Proceedings of the 11th Pronunciation in Second Language Learning and Teaching Conference, Northern Arizona University* (pp. 173–185). Iowa State University. https://www.iastatedigitalpress.com/pslit/article/15422/galley/13503/view/

Rukthong, A. (2021). Complex interplay of cognitive and strategic processing in EFL listening: Implications for teaching, *rEFlections, 28*(3), 313–332. https://doi.org/10.61508/refl.v28i3.254570

Shockley, K., Sabadini, L., & Fowler, C. A. (2004). Imitation in shadowing words. *Perception & Psychophysics, 66*(3), 422–429. https://doi.org/10.3758/BF03194890

Stevens, J. P. (2012). *Applied Multivariate Statistics for the Social Sciences (5th ed.).* Routledge. https://doi.org/10.4324/9780203843130

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*(4), 295–312. https://doi.org/10.1016/0959-4752(94)90003-5

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure and performance testing. In R. Ellis (Ed.), *Planning and Task Performance in a Second Language* (pp. 239–273). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.11.15tav

Thomson, R. I. (2015). Fluency. In M. Reed & J. M. Levis (Eds.), *The handbook of English pronunciation* (pp. 209–226). John Wiley & Sons. https://doi.org/10.1002/9781118346952

Trofimovich, P., & Baker, W. (2006). Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition, 28*(1), 1–30. https://doi.org/10.1017/S0272263106060013

Vandergrift, L. (2007). Recent developments in second and foreign language listening comprehension research. *Language Teaching, 40*(3), 191–210. https://doi.org/10.1017/S0261444807004338

Vandergrift, L., & Goh, C. C. M. (2012). *Teaching and learning second language listening: Metacognition in action*. Routledge.

Wang, Y., & Treffers-Daller, J. (2017). Explaining listening comprehension among L2 learners of English: The contribution of general language proficiency, vocabulary knowledge and metacognitive awareness. *System*, *65*, 139–150. https://doi.org/10.1016/j.system.2016.12.013

Zhao, Y. (1997). The effects of listeners' control of speech rate on second language comprehension. *Applied Linguistics, 18*(1), 49–68. https://doi.org/10.1093/applin/18.1.49