

The Interaction Effect of Pronunciation and Lexicogrammar on Comprehensibility: A Case of Mandarin-Accented English

Yongzhi Miao, Heath Rose, and Sepideh Hosseini

Abstract

Scholars have argued that *comprehensibility* (i.e., ease of understanding), not nativelike performance, should be prioritized in second language learning, which inspired numerous studies to explore factors affecting comprehensibility. However, most of these studies did not consider potential interaction effects of these factors, resulting in a limited understanding of comprehensibility and less precise implications. This study investigates how pronunciation and lexicogrammar influences the comprehensibility of Mandarin-accented English. A total of 687 listeners were randomly allocated into six groups and rated (a) one baseline and (b) one of six experimental recordings for comprehensibility on a 9-point scale. The baseline recording, a 60 s spontaneous speech by an L1 English speaker with an American accent, was the same across groups. The six 75-s experimental recordings were the same in content but differed in (a) speakers' degree of foreign accent (American, moderate Mandarin, and heavy Mandarin) and (b) lexicogrammar (with errors vs. without errors). The study found that pronunciation and lexicogrammar interacted to influence comprehensibility. That is, whether pronunciation affected comprehensibility depended on speakers' lexicogrammar, and vice versa. The results have implications for theory-building to refine comprehensibility, as well as for pedagogy and testing priorities.

Keywords

Comprehensibility, speech perception, listening, speaking

This is an author-generated pdf of a published paper. Please see the journal for the definitive version:

Miao, Y., Rose, H., & Hosseini, S. (2023). The Interaction Effect of Pronunciation and Lexicogrammar on Comprehensibility: A Case of Mandarin-Accented English. *Language and Speech*, 00238309231156918.

<https://doi.org/10.1177/00238309231156918>

1 Introduction

The English language is now actively used by both first language (L1) and second language (L2) speakers. Accordingly, many scholars have argued that comprehensibility, more so than nativelike performance, is key to communicative success and should be the goal of L2 teaching, learning, and testing (Cook, 2016; Harding, 2017; Jenkins, 2015).

Comprehensibility in this study is defined as listeners' perception of how difficult an L2 speaker is to understand (Derwing & Munro, 1997, 2015).

Research within comprehensibility can be largely divided into two strands: (a) comprehensibility development from a pedagogical perspective and (b) the conceptualization of comprehensibility from a theoretical perspective. In the first strand, comprehensibility is used as a global measure of L2 speakers' proficiency, and relevant studies explore how speakers' comprehensibility can be developed (e.g., Derwing et al., 1998; Zhang & Yuan, 2020). The second strand explores factors affecting comprehensibility, seeking to conceptually refine the construct to provide implications for pedagogy and assessment (see Isaacs & Trofimovich, 2012; Munro & Derwing, 1995, 2020; K. Saito et al., 2019). The present study sits within the second strand of this body of research.

The concept of comprehensibility implies two-way communication between both speakers and listeners in a specific context. Previous studies have focused largely on *speaker variables* (e.g., pronunciation; see Isaacs & Trofimovich, 2012), *listener variables* (e.g., listeners' familiarity with accented English; see Kennedy & Trofimovich, 2008), and contexts of communication (see, for example, Crowther et al., 2018). The current study narrows its scope and focuses solely on speaker variables, specifically, pronunciation and lexicogrammar, in relation to comprehensibility. Pronunciation in this study is broadly defined as encompassing *segmental* (i.e., vowels and consonants) and *suprasegmental features* (e.g., word stress, prosody). Lexicogrammar is defined as lexical and grammatical features (i.e., accuracy and appropriateness).

In exploring factors affecting comprehensibility, the focus of previous studies included pronunciation, temporal factors, and other linguistic variables such as lexicogrammar. Specific variables of interest include segmental accuracy (see Derwing & Munro, 1997; Pérez-Ramón et al., 2022), speech rate (see Kang, 2010), word stress (see Isaacs & Trofimovich, 2012), grammatical accuracy, and lexical richness (see Trofimovich & Isaacs, 2012).

While this line of research has been highly productive, several studies do not consider potential *interaction effects* of speaker variables on comprehensibility (i.e., the influence of one speaker variable on comprehensibility may depend on another speaker variable). This is because most studies cited above are correlational in nature. When using multiple regression, for example, including all speaker variables to predict comprehensibility, the interaction between the predictors was overlooked (see Kline, 2015). Therefore, this field of research requires studies with different research designs (e.g., experimental) to address this lack. Any overlaps or discrepancies of findings would provide additional insights into the conceptualization of comprehensibility.

The present study, inspired by Varonis and Gass (1982), explores potential interaction effects of speaker variables on comprehensibility using Mandarin-accented English as an exemplar L2. Theoretically, this study seeks to enrich our understanding of comprehensibility

by investigating the interaction effects of speaker variables. Methodologically, this study adopts a different design (i.e., experimental), hoping to add knowledge to the conceptualization of comprehensibility. The concept of comprehensibility was developed in the 1990s (e.g., Munro & Derwing, 1995). While there are many studies exploring factors affecting comprehensibility (e.g., Isaacs & Trofimovich, 2012), we have yet to develop a systematic knowledge of the construct itself. A more thorough theoretical understanding of the concept would benefit comprehensibility-informed empirical studies for L2 pronunciation development.

2 Literature review

2.1 Factors affecting comprehensibility

Previous research has explored pronunciation and temporal factors in relation to comprehensibility, including segmentals (e.g., consonants), suprasegmentals (e.g., stress), and fluency. For instance, Caspers (2010) had 25 L1 Dutch speakers evaluate the comprehensibility of L1 French and L1 Mandarin speakers reading Dutch words. Results suggested that segmental errors more negatively influenced comprehensibility than word stress errors, and when both errors co-occur, comprehensibility was the most impeded.

Regarding temporal features, Munro and Derwing (2001) investigated 48 L1 English speakers' evaluation of read-aloud sentences from 48 L2 speakers with diverse L1s and found that speech rate was significantly correlated with comprehensibility. While Munro and Derwing's findings pertained to the sentence level, Kang (2010) corroborated and extended these findings to the discourse level in L2 English users' lectures. Kang had 51 U.S. undergraduate students evaluate 11 international teaching assistants' speech and found speech rate to be the best predictor of comprehensibility ratings among several prosodic features.

While the above study focused solely on pronunciation and temporal factors in relation to comprehensibility, other studies have included such features as vocabulary, grammar, and discourse. For example, Derwing and Munro (1997) had 48 L2 English learners record a picture description task which was evaluated by 26 L1 English speakers for comprehensibility. Results revealed that listeners' ratings were correlated with speakers' prosody, speech rate, and also grammatical errors. A similar result is replicated in O'Brien (2014), which explored the perception of L2 German speech. O'Brien found that speech rate, lexicogrammar, and fluency were correlated with listeners' comprehensibility ratings. Corroborating the findings above, K. Saito et al. (2019) in a large-scale study found that both phonology and lexicogrammar contributed to comprehensibility in L2 Japanese.

These studies suggest that comprehensibility is associated with diverse speaker variables such as pronunciation and lexicogrammar, in line with numerous other studies in the literature (see Crowther et al., 2015; K. Saito et al., 2017; K. Saito & Shintani, 2016). However, most of these studies based their findings on speech data elicited from picture description tasks, which are criticized for simplicity, thus challenging the ecological validity of the findings (see Crowther et al., 2018; Robinson, 2005; Skehan, 2009). Nonetheless, it was found that regardless of task type, pronunciation and lexicogrammar could both contribute to comprehensibility (see Crowther et al., 2018; Suzuki & Kormos, 2020). To build on these previous studies, the present study used speech stimuli elicited from another task to improve the generalizability of findings across different communicative contexts and to diversify the use of tasks across studies into comprehensibility.

2.2 Interaction effect of speaker variables: empirical evidence

Comprehensibility is a complex construct which diverse speaker variables contribute to (e.g., pronunciation, lexicogrammar). However, most studies have treated these variables relatively independently, instead of considering potential interaction. Accordingly, our current understanding of speaker factors in comprehensibility may be incomplete. Below, we review a seminal study within comprehensibility that provides evidence for this interaction effect, which motivated the present study. To the best of our knowledge, there do not seem to be other studies that examine such interaction effect.

Varonis and Gass (1982) recorded L2 English speakers reading aloud ungrammatical and grammatical sentences. The recordings were evaluated for pronunciation and for comprehensibility by L1 English listeners. The study found that (a) listeners' evaluation of speakers' pronunciation was confounded by grammaticality and (b) pronunciation and grammaticality interacted to influence comprehensibility. Specifically, when pronunciation was rated as either extremely good or extremely bad, lexicogrammar did not influence comprehensibility to the same degree.

Overall, Varonis and Gass (1982) tentatively suggested that pronunciation and lexicogrammar could interact to influence comprehensibility, a finding that has not been explored in many contemporary studies. The present study seeks to further validate and extend this finding by addressing some of the methodological caveats. First, we included a larger listener sample to improve the generalizability of findings. Second, we used discourse as our stimuli, not decontextualized, isolated words or sentences, to improve the external validity of findings. Third, we present both descriptive and inferential statistics for the findings. Fourth, we adopt an experimental design, controlling for the independent variables (i.e., pronunciation and lexicogrammar) for a more straightforward observation of any interaction effect.

3 Methodology

3.1 Research questions

Overall, the research questions (RQs) of this study are as follows:

RQ1: What is the effect of pronunciation and lexicogrammar on comprehensibility?

RQ2: Does pronunciation and lexicogrammar interact to influence comprehensibility?

3.2 Research design

The present study included six experimental groups, to which our listener participants were randomly allocated. In each group, listeners were exposed to the same baseline recording and a different experimental recording for comprehensibility ratings. The baseline recording was an L1 English speakers' planned speech. In terms of the experimental recordings, we had three other speakers to record two scripts: (a) an L1 English speaker from the United States, (b) an L2 English speaker with a moderate Mandarin accent, and (c) an L2 English speaker with a heavy Mandarin accent. One script contained grammatical errors, and the other had these errors corrected. Table 1 provides an outline of the group assignment.

Table 1. Group Assignment.

Group	Recording	Condition
1	Baseline	North American Accent × Grammatical
	Experimental	North American Accent × Ungrammatical
2	Baseline	North American Accent × Grammatical
	Experimental	Moderate Mandarin Accent × Ungrammatical
3	Baseline	North American Accent × Grammatical
	Experimental	Heavy Mandarin Accent × Ungrammatical
4	Baseline	North American Accent × Grammatical
	Experimental	North American Accent × Grammatical
5	Baseline	North American Accent × Grammatical
	Experimental	Moderate Mandarin Accent × Grammatical
6	Baseline	North American Accent × Grammatical
	Experimental	Heavy Mandarin Accent × Grammatical

3.3 Participants

3.3.1 Speakers

Four speakers were recruited in this study (22 years old). All were female, to reduce possible gender effects on the results (see [Kang, 2010](#)). Speakers 1 and 2 were L1 English speakers. Both were raised in California, USA. Speakers 3 and 4 spoke L1 Mandarin and were raised in Guangdong, China. [Table 2](#) provides a summary of speaker demographics. Specifically, of the two L2 speakers, Speaker 4 was perceived to be more accented in English than Speaker 3, and Speaker 3, more accented than Speaker 2 (i.e., an L1 English speaker). This was supported both quantitatively and qualitatively as described below.

Table 2. Speaker Demographics.

	Gender	Age	L1	Accentedness
Speaker 1	Female	22	English	American English
Speaker 2	Female	22	English	American English
Speaker 3	Female	22	Mandarin	Moderate Mandarin-accented English
Speaker 4	Female	22	Mandarin	Heavy Mandarin-accented English

As part of quantitative assessments, after Speakers 2, 3, and 4 made the experimental recordings, nine L1 Mandarin teachers who taught L2 English (88.89% female, $M_{\text{age}} = 33.56$, $SD = 5.92$, $Mdn = 36.00$) listened to a short clip of the three speakers' recordings (20 s each) and rated their pronunciation on a 9-point scale via an online questionnaire (1 = *very bad*, 9 = *very good*; see [Varonis & Gass, 1982](#), for a similar methodological practice). The teachers had extensive experience in teaching English ($M_{\text{year}} = 10.44$, $SD = 4.93$, $Mdn = 10.00$). Experienced teachers with a shared L1 Mandarin were chosen because they were more aware of Mandarin–English phonological transfer and thus could provide more consistent speech evaluation (see [Suzukida & Saito, 2019](#)). The nine experienced teachers expressed great confidence in their pronunciation ratings on a 9-point scale (1 = *not confident*, 9 = *very confident*; $M_{\text{confidence}} = 8.44$, $SD = 0.73$, $Mdn = 9.00$). The experienced teachers' ratings of the speakers were rank ordered. All listeners unanimously assigned higher pronunciation ratings to Speaker 2 compared with Speaker 3, and to Speaker 3 compared with Speaker 4. The only one exception was that one teacher ranked Speaker 2 as equal to Speaker 3 in terms of pronunciation. Overall, this analysis supported that Speakers 2, 3, and 4 had three distinguishable degrees of accent.

Moreover, Speaker 3's and 4's recordings of the grammatical script were coded based on six phonological features: errors in segmentals, syllable structure (phoneme insertion/deletion), stress, vowel reduction, and pitch contour (modified from [Isaacs & Trofimovich, 2012](#)). The coding was initially completed independently by two experienced coders with an advanced degree in applied linguistics and training in phonology. After the initial coding, they held a reconciliation meeting to address any discrepancies. Acknowledging that different varieties have different segmental and suprasegmental features, the coders were instructed to mark any feature that did not conform to L1 English varieties (e.g., Northern American varieties, British varieties). [Table 3](#) illustrates the independent coding for both coders and their final coding after the reconciliation meeting.

Table 3. Phonological Coding of Speakers 3's and 4's Grammatical Recording.

	Speaker 3			Speaker 4		
	Final	Coder 1	Coder 2	Final	Coder 1	Coder 2
Segmental	34	14	24	42	27	26
Syllable structure	4	3	3	30	18	16
Word stress	0	0	0	1	1	0
Vowel reduction	1	2	1	2	7	0
Pitch contour	0	0	0	3	0	3

It is important to note that the final coding is not a simple mathematical addition of two coders' coding. For segmental features (Speakers 3 and 4), the final coding was significantly greater than the initial coding for each coder. This was because while coder 1 attended to more consonant features, coder 2 paid more attention to vowels. In the reconciliation meeting, they went through the recording line by line, listening and relistening, with special attention to the discrepancies. Then, they agreed with the other coder regarding many segmental features that they previously overlooked. For syllable structure features (Speaker 4), Coders 1 and 2 each had some features that they were not sure of. These features were not included in the initial coding. At the reconciliation meeting, the two coders listened to the recording line by line and decided to include many of these features that they were not sure about earlier, hence the increased final coding. For vowel reduction features (Speaker 4), Coder 1 was more strict during the initial coding. Many of the decisions were negotiated with Coder 2 at the reconciliation meeting. This resulted in a decline in the final coding.

According to the results of the coding, Speaker 4 had more phonological errors in all five coded categories, in which the most prominent difference was syllable structure error. Overall, this coding suggested that Speaker 4 had a less nativelike pronunciation than Speaker 3. This, coupled with the numerical findings above, suggested that Speakers 2, 3, and 4 had different degrees of pronunciation.

3.3.2 Listeners

Initially, our target listener population included L1 and L2 English users at any proficiency level, although this differed from our final inclusion criteria (see [Supplementary Material 7](#) for details). In total, 687 listeners were included for data analysis. Listeners across the six groups were roughly balanced on general demographics. Each group had 109 to 118 listeners with around 50% to 60% L1 English listeners, and 60% to 70% of them were multilinguals. Each had 10% to 20% experts (defined as having both L2 teaching experience and linguistic training such as an advanced degree in TESOL, i.e., teaching English as a second language). All groups' reported familiarity with Mandarin-accented English as 5.0 to 5.5 on a 9-point

scale. All groups' mean age ranged from approximately 25 to 30 years, and all groups contained around 65% to 70% female listeners.

Moreover, the multilingual L1 English listeners in our sample spoke very diverse additional languages, including 5 to 9 language families and 13 to 18 languages per group, of which Indo-European languages were the most common (around 70%–85%). Similarly, L2 English listeners had very diverse L1s, including 5 to 7 language families and 14 to 22 languages per group, of which Indo-European (around 35%–65%) and Sino-Tibetan languages (around 25%–50% [mostly Mandarin]) were the most common. [Table 4](#) presents demographic information about the sample in this study.

Table 4. Listener Demographic Information.

	M_{age} (SD)	Gender	L1 status	Salient L1s (for L2 listeners)
Group 1 ($n = 118$)	29.22 (12.74)	63.56% female 36.44% male 0% other	62.71% L1 English	Sino-Tibetan (50.00%) Indo-European (34.10%)
Group 2 ($n = 115$)	32.27 (14.71)	65.23% female 33.03% male 1.74% other	60.00% L1 English	Sino-Tibetan (47.83%) Indo-European (36.96%)
Group 3 ($n = 115$)	28.60 (11.94)	69.57% female 28.69% male 1.74% other	53.04% L1 English	Indo-European (42.59%) Sino-Tibetan (38.89%)
Group 4 ($n = 114$)	26.02 (9.23)	69.30% female 30.7% male 0% other	56.14% L1 English	Indo-European (66.00%) Sino-Tibetan (24.00%)
Group 5 ($n = 116$)	26.91 (9.97)	70.69% female 29.31% male 0% other	50.00% L1 English	Indo-European (58.62%) Sino-Tibetan (34.48%)
Group 6 ($n = 109$)	26.50 (8.99)	72.48% female 27.52% male 0% other	54.13% L1 English	Indo-European (43.10%) Sino-Tibetan (32.76%)

3.4 Questionnaire and online task

An online questionnaire was used for data collection. It was advertised via email within the research team's immediate networks and via social media (e.g., Facebook, Instagram, Twitter, LinkedIn). As an incentive, all listeners could enter a prize draw for £100. Only one response was permitted from each device. The majority of the listeners were able to complete the questionnaire within approximately 6 min.

In the task, listeners listened to a “test recording” to adjust the volume of their equipment. Then, they listened to the baseline and the experimental recording. After each recording, they rated the speaker's comprehensibility on a 9-point scale (1 = *very difficult to understand*, 9 = *very easy to understand*). All listeners rated the same baseline recording before experimental recordings.

Listeners also provided their background information. Their age and gender (i.e., male, female, other) were collected, as too was their L1 status (i.e., whether they spoke L1 English). Listeners' accent familiarity was measured by a self-evaluation of familiarity with Mandarin-accented English on a 9-point scale (1 = *not familiar at all*, 9 = *very familiar*). Two questions explored listeners' linguistics-related experience, including whether they had language teaching experience or graduated/were graduating from a linguistic-related degree. In line with [K. Saito et al. \(2017\)](#), we defined experts as those who fulfilled both conditions.

Three questions probed L1 English listeners' multilingual background, include

1. whether they were *simultaneous bilinguals* (i.e., having ≥ 2 L1s)
2. whether they studied abroad using an L2, and
3. self-evaluation of their most proficient L2 from on a 9-point scale.

Multilingual L1 English listeners in this study were defined as either being simultaneous bilinguals, having studied abroad using an L2, or self-evaluating their most proficient L2 as ≥ 4 from 1 to 9. Comparatively, monolingual L1 English listeners were defined as having no L2 or self-evaluating their most proficient L2 as ≤ 3 from 1 to 9.

Additional questionnaire items explored L2 English listeners' language background, including whether they had taken English proficiency tests. Those who had taken tests were asked to provide the name, score, and year of the test. Those without test scores were asked to evaluate their English proficiency on a 9-point scale.

3.5 Materials (recordings)

Seven recordings were prepared for this study (i.e., one baseline and six experimental recordings). All were normalized for peak amplitude at -3 dB using Audacity, an audio editing software; thus, they were similarly loud to the listeners. All were transformed into mp3 files at 11025 Hertz and 32 bit, ensuring similar recording quality. All were under 300 KB, minimizing the likelihood of any buffering issues.

The baseline recording was a planned speech from an L1 English speaker, in which she described her morning routine (e.g., having breakfast). This topic was chosen because it was a non-academic, day-to-day topic, catering to a wide audience. The baseline recording was 60 s in length.

Six experimental recordings were prepared for this study. We had Speakers 2, 3, and 4 record themselves reading two experimental scripts. To make the scripts, we had a structured interview with six intermediate English learners (L1 = Mandarin), in which participants talked about a party experience based on some prompt questions (e.g., who attended the party). We extracted their lexicogrammar mistakes and formulated an ungrammatical script based on the mistakes and the themes they mentioned. Then, we amended the mistakes, leading to a grammatical script (both 192 words long).

All experimental recordings were 75 s in length. Being the most accented, Speaker 4 was asked to record the two scripts first and ensure that the recordings were equal in length (measured in seconds). Based on her rate of speech, Speakers 2 and 3 were asked to record at a similar pace as did Speaker 4. They were asked to make sure the length of the recording was exactly that of Speaker 4's (measured in seconds), to control for fluency. To help control their rate of speech, they listened to Speaker 4's recording before re-recording (see [Kang et al., 2019](#)).

4 Results

To explore how pronunciation and lexicogrammar influenced comprehensibility, two between-group comparisons were performed. The first model tested to what extent listeners' ratings of the baseline recording differed across groups. Few group differences would indicate listeners' similar interpretation of the scales across groups. Because listeners' ratings of the baseline recording deviated significantly from normality (i.e., there was an expected ceiling effect due to the native accent of the speaker), a Kruskal–Wallis test was used to address this question.

The second model tested the interaction effect of pronunciation and lexicogrammar in listeners' ratings of the experimental recordings. The data met the assumptions of parametric test. Therefore, a 2×3 factorial analysis of variance (ANOVA) was performed to assess the interaction effect of pronunciation and lexicogrammar in comprehensibility. Specifically, the independent variables in the model were lexicogrammar and pronunciation; the dependent variable was the differences between listeners' comprehensibility ratings of the baseline recording and of the experimental recording. The choice of the dependent variable was to control for listeners' rating leniency as observed in the baseline recording on the findings.

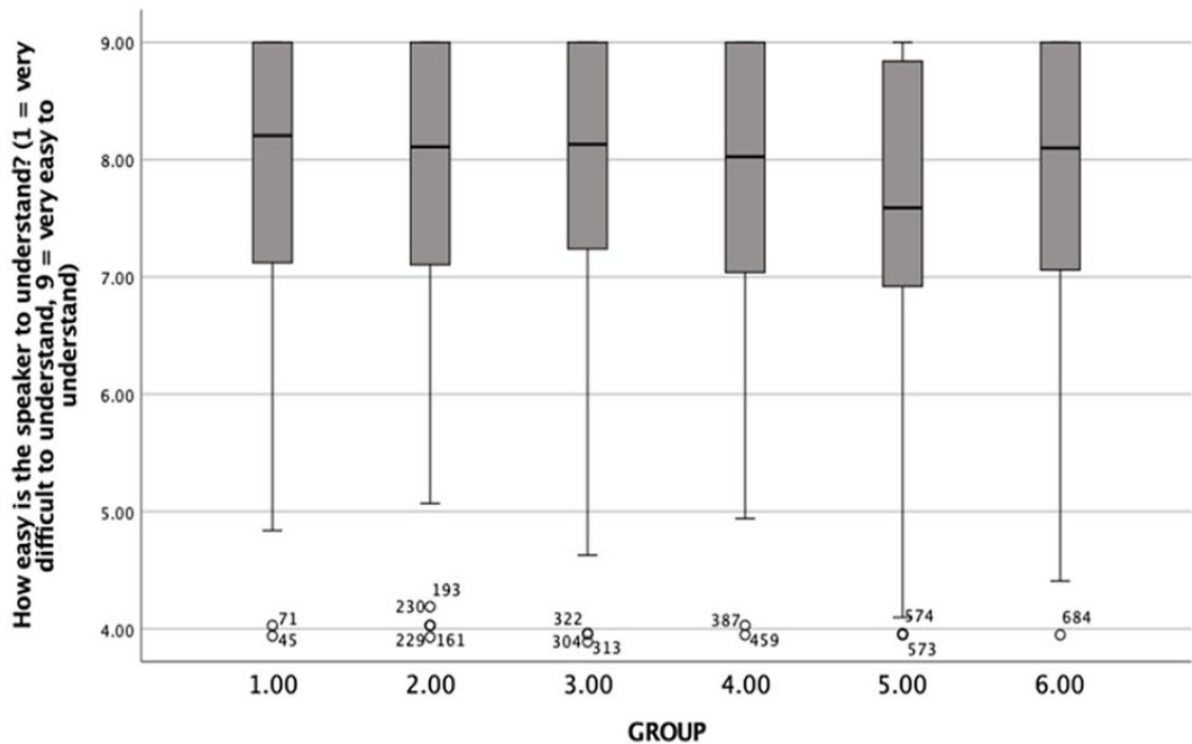
4.1 Listener ratings of the baseline recording

A Kruskal–Wallis H test showed that there was a statistically significant difference in listeners' comprehensibility ratings of the baseline recordings across the six groups, $\chi^2(5) = 13.3, p = .0204, \eta^2 = .01$, 95% confidence interval (CI) [.00, .04]. However, p values are sensitive to the sample size. Specifically, with a large sample size (as in the case of the present dataset), even negligible differences would yield a significant p value. By observing the effect size and the CIs around it, it is clear that the differences across the six groups regarding the ratings of the baseline recording were marginal. This was corroborated by examining (a) the descriptive statistics presented in [Table 5](#) and (b) the boxplot presented in [Figure 1](#). Specifically, the means (and SD s), and medians were very similar across groups (see [Table 5](#)). Moreover, visual inspection of the boxplot also confirmed that the differences across groups appeared to be negligible. Taken together, different groups had relatively homogeneous understanding of the Likert-type scale used to measure comprehensibility.

Table 5. Descriptive Statistics of Listeners' Ratings of the Baseline Recording by Group.

Groups (Grammar \times Accent) ^a	<i>M</i> (<i>SD</i>)	<i>Mdn</i>
1 (UNGRAM \times LI)	7.90 (1.34)	8.21
2 (UNGRAM \times Moderate)	7.80 (1.31)	8.11
3 (UNGRAM \times Heavy)	7.87 (1.32)	8.13
4 (GRAM \times LI)	7.54 (1.44)	8.03
5 (GRAM \times Moderate)	7.44 (1.44)	7.59
6 (GRAM \times Heavy)	7.71 (1.43)	8.10

^aUNGRAM=the experimental recording contained grammatical errors; GRAM=the experimental recording did not contain grammatical errors; LI = the experimental recording was recorded by the LI English speaker; Moderate=the experimental recording was recorded by the moderately accented speaker; Heavy=the experimental recording was recorded by the heavily accented speaker.

**Figure 1.** Listeners' comprehensibility ratings of the baseline recording by group.

4.2 Listener ratings of the experimental recordings

A 2×3 factorial ANOVA was performed to investigate the effect of pronunciation and lexicogrammar on the differences between listeners' comprehensibility ratings of the baseline recording and of the experimental recording. The F test confirmed that the overall model was significant, $F(5, 681) = 57.81, p < .001$. The interaction effect of lexicogrammar and pronunciation on the ratings was also statistically significant, $F(2, 681) = 3.10, p = .030$. Taken together, this model yielded an R^2 value of 29.80% (adjusted $R^2 = 29.30\%$), revealing that approximately 30% of the variance in the differences between listeners' comprehensibility ratings of both recordings could be accounted for by speakers' pronunciation, lexicogrammar, and the interaction between them.

As a significant interaction effect was observed, pairwise comparisons were performed using Tukey's HSD (honestly significant difference) test. Descriptive statistics are presented

in Table 6. Several main findings are summarized below, which point to a general finding suggesting the interaction effect of pronunciation and lexicogrammar on comprehensibility:

Table 6. Listeners' Comprehensibility Ratings of the Experimental Recordings and the Differences Between Listeners' Two Ratings by Group.

Groups (Grammar \times Accent) ^a	M (SD) ^b	M _{diff} (SD) ^c
1 (UNGRAM \times L1)	7.41 (1.44)	0.50 (1.71)
2 (UNGRAM \times Moderate)	7.10 (1.49)	0.71 (1.51)
3 (UNGRAM \times Heavy)	5.28 (1.81)	2.63 (2.04)
4 (GRAM \times L1)	8.10 (1.00)	-0.44 (1.51)
5 (GRAM \times Moderate)	7.03 (1.47)	0.39 (1.71)
6 (GRAM \times Heavy)	5.23 (1.76)	2.52 (1.91)

^aUNGRAM = the experimental recording contained grammatical errors; GRAM = the experimental recording did not contain grammatical errors; L1 = the experimental recording was recorded by the L1 English speaker; Moderate = the experimental recording was recorded by the moderately accented speaker; Heavy = the experimental recording was recorded by the heavily accented speaker. ^bM (SD) = the mean and standard deviation values for listeners' comprehensibility ratings of the experimental recordings. ^cM_{diff} (SD) = the mean and standard deviation values for listeners' comprehensibility ratings of the baseline recording minus the experimental recordings.

1. lexicogrammar affected comprehensibility in the L1 English accent condition;
2. lexicogrammar did not affect comprehensibility in the moderate accent condition and in the heavy accent condition;
3. pronunciation affected comprehensibility in the most ungrammatical and grammatical conditions;
4. pronunciation did not affect comprehensibility in the ungrammatical condition between L1 English accent and moderate Mandarin accent.

5 Discussion

The purpose of our study was to explore how pronunciation and lexicogrammar influenced the comprehensibility of Mandarin-accented English. We found that both pronunciation and lexicogrammar could affect comprehensibility. Moreover, pronunciation and lexicogrammar interacted to influence comprehensibility. Below, we discuss these findings in relation to previous studies.

5.1 Factors affecting comprehensibility

The present study found that listeners across different groups demonstrated homogeneous ratings of the baseline recording in terms of its comprehensibility. This meant that they had similar interpretations of the 9-point comprehensibility scale as used in the present study. Moreover, this finding echoed extant literature where scholars have found that comprehensibility scales have been reliably used across listeners with different backgrounds, ranging from linguistic experts to laypersons and from L1 English to L2 English listeners (see K. Saito, 2021).

The study also found that pronunciation and lexicogrammar, in many cases, affected comprehensibility. This supports many previous studies that have found a range of pronunciation features to be correlated with comprehensibility, including segmental errors (Caspers, 2010), word stress (Field, 2005), prominence (sentence stress), rhythm (Isaacs & Trofimovich, 2012), and pitch range (Kang, 2010). This also echoes findings where research found that grammatical (Isaacs & Trofimovich, 2012) and lexical features (K. Saito et al., 2017) affected comprehensibility.

Regarding stimuli type, studies within comprehensibility used decontextualized words (e.g., Caspers, 2010), picture description tasks (e.g., Derwing & Munro, 1997; O'Brien, 2014; K. Saito et al., 2019), and more complex task types (e.g., Crowther et al., 2018; Suzuki & Kormos, 2020). They found that, among other factors, comprehensibility could be affected by pronunciation and lexicogrammar. Our study used a read-aloud story which replicated the findings observed in the above-cited studies. Taken together, we argue that both pronunciation and lexicogrammar have the potential to affect comprehensibility across a range of communication contexts.

5.2 Interaction effect of pronunciation and lexicogrammar on comprehensibility

Our study also found that pronunciation and lexicogrammar interacted to affect comprehensibility. That is, whether pronunciation affected comprehensibility depended on lexicogrammar, and vice versa, which extended the knowledge gleaned from the above studies.

Specifically, the present study found that lexicogrammar affected comprehensibility only in the L1 English accent condition; lexicogrammar in other accent conditions did not affect comprehensibility. To interpret this finding from a psycholinguistic point of view, it is possible that the absence of foreign accents increased listeners' speech processing fluency (see Dragojevic, 2020; Munro & Derwing, 1995, 2020) allowing them to listen beyond "accent" to detect any lexicogrammatical features. Comparatively, listeners in the moderate and heavy accent conditions attended first to pronunciation features and processed lexicogrammatical features secondarily. The assumption underlying this interpretation is that listeners perceive L2 speech first impressionistically at surface level focusing on accent or pronunciation before a more analytic, finely grained approach to lexicogrammar (see Ruivivar & Collins, 2019, for the effect of pronunciation on the perception of grammaticality).

To explain why lexicogrammar did not affect comprehensibility in neither moderate nor heavy accent conditions, it is possible that the lexicogrammar errors presented in the scripts were not necessarily sufficient to decrease comprehensibility. The scripts were structurally clear (see [Supplementary Material 5](#)). They had an introduction explaining the event, time, and place. Each theme or topic (e.g., food, music, conversation) were expanded logically and supported with examples. Moreover, the ungrammatical script only included minor lexicogrammar errors (e.g., morphological errors). It is possible that they did not substantially impede communication. Overall, the enhanced coherence might have made the script very comprehensible, despite the relatively low-functioning lexicogrammar errors (see Isaacs & Trofimovich, 2012, for the relationship between discourse coherence and comprehensibility). Nonetheless, this speculation needs empirical support. Future studies could (a) manipulate lexicogrammar errors based on their severeness in comprehensibility or (b) use naturally occurring speech samples that were not as structured as the ones used in the present study.

The present study also found that pronunciation affected comprehensibility in most ungrammatical and grammatical conditions. This could mean that speakers' pronunciation

indeed influences comprehensibility, corroborating with other studies (e.g., Caspers, 2010). From a sociolinguistic perspective, this could also lend support to listeners' general negative biases in evaluating L2 speech. That is, the lower ratings observed in this study could also be explained by listeners' negatively biased attitudes toward L2 speech (see Ghanem & Kang, 2021; Lev-Ari & Keysar, 2010; Lindemann et al., 2014). Future research could tease apart L2 accents from listener attitudes to explore their (perhaps interactional) effect on comprehensibility.

The study also found that pronunciation did not affect comprehensibility in the ungrammatical condition between L1 English accent and moderate Mandarin accent. To explain this finding, it is possible that listeners did not expect there to be lexicogrammar errors when hearing L1 English accents. The presence of such errors might have violated listeners' expectations, resulting in a lower rating (see Burgoon, 1993, 2015, for a detailed discussion of expectancy violations theory). This explanation again points to the complex interactional effect of pronunciation and lexicogrammar on comprehensibility.

Varonis and Gass (1982) found when speakers' pronunciation was either "good" or "bad," lexicogrammar did not affect comprehensibility, but when their pronunciation was moderate, lexicogrammar influenced comprehensibility considerably (inferential statistics unreported). In comparison, our study found that lexicogrammar affected comprehensibility when speakers' pronunciation was nativelike, but it did not when their pronunciation was moderately or heavily accented.

To explain the discrepancies of our findings, our study has a considerably larger sample size than Varonis and Gass (1982), thus minimizing the likelihood of random variations more susceptible to small samples. Second, the stimuli used in Varonis and Gass were short, decontextualized sentences while we used a coherently constructed story. Listeners in our study could infer the meaning of any lexicogrammar mistakes based on the context, while in the Varonis and Gass, this would be much more difficult. In other words, our study contextualized lexicogrammar mistakes, which might have assisted in comprehensibility. Despite these differences, our study supports Varonis and Gass's conclusions that overall, pronunciation and lexicogrammar interacted to affect comprehensibility. This finding suggests that the construct of comprehensibility is complex and dynamic. Therefore, it is difficult to pinpoint what factors most affect comprehensibility without considering the potential interaction among these factors. The fact that speakers' linguistic factors do not contribute to comprehensibility in a linear fashion is important in offering implications for language learning, teaching, and testing, which we detail in the conclusion section.

5.3 Limitations and future directions

Overall, the findings of the study should be interpreted with caution, taking into account the methodological limitations of the study. First, although this study used read-aloud speech stimuli, future research may want to build upon the findings and use naturally occurring interactional data. This would further improve the ecological validity of the findings.

Second, this study only included four speakers. Although the findings provided insights into the interaction effect of pronunciation and lexicogrammar on comprehensibility, the findings were susceptible to the effect of speaker idiosyncrasies such as their voice quality. Future studies could include more speakers to minimize speaker individual differences on the findings to extend the generalizability of this study.

Specifically, the present study did not choose to use the "matched guise method" as used in numerous previous speech perception studies (see Lambert et al., 1960). This method has merits because it could minimize the unwanted influence of speakers' voice qualities on

speech ratings. We did not use this method because one of the variables of interest in the present study is accent (heavy, moderate, and native). If we were to use matched guise technique, we would have had a speaker with native accents mimic accented speakers to create the recordings. This process is limited because it poses validity issues as to whether the forced accent accurately represents the true target accent. Furthermore, to mimic an accent variety, a descriptive theoretical framework detailing the phonological features of a given variety is needed. While there are studies describing features in Chinese-accented English, they are limited in number and need more empirical support if they are to form a blueprint of speech production. Also, the inclusion of different guises implied the inclusion of many distractor recordings. This would have lengthened the duration of the questionnaire, resulting in fewer data points. This would have been a concern because there were six experimental groups, and an adequate number of listener participants was needed within each group for more robust statistical analysis.

6 Implications and conclusion

Our study has theoretical implications by way of its suggestions of an interaction effect of pronunciation and lexicogrammar in relation to comprehensibility. This suggests that comprehensibility may be more complex than previously assumed and points to a need for more research to build a thorough theoretical understanding of the construct before concrete implications can be made. Based on the interaction effect of pronunciation and lexicogrammar on comprehensibility, our study seeks to offer some preliminary implications for language learning and testing.

Scholars have argued for a need to focus on comprehensibility, not nativelike performance, in the learning and teaching of an L2 ([Levis, 2005, 2020](#)). However, the conceptualization of comprehensibility, or what factors caused comprehensibility, have not been clearly articulated in the literature. Specifically, many precursor studies have sought to explore factors affecting comprehensibility using a correlational design (e.g., [Isaacs & Trofimovich, 2012](#)), thus overlooking the potential interaction effect of different variables on comprehensibility. This has limited the theoretical understanding of the construct of comprehensibility, and in turn has limited pedagogical implications. To elaborate, while scholars argued for the emphasis on comprehensibility in L2 teaching and assessment, it remains unclear to classroom and testing practitioners what needs to be changed (see, for example, [Gordon & Darcy, 2016](#); [Y. Saito & Saito, 2017](#), for different realizations of comprehensibility-oriented speaking instruction). Accordingly, our study argues that while researchers implement different comprehensibility-based instructional studies, attention should be equally placed to addressing theoretically driven issues on the conceptualization of comprehensibility.

In line with this proposal, our study found that comprehensibility is dynamic, and to improve comprehensibility, language instructors and learners should focus on the differential needs based on learners' existing linguistic repertoire, instead of focusing solely on arguments such as whether segmental or suprasegmental instruction was more effective (see, for example, [Derwing et al., 1998](#)). For example, they could consider students' competencies in pronunciation and lexicogrammar to decide which area to prioritize to improve comprehensibility. In a similar vein, scholars within language testing have argued for a focus on comprehensibility in speaking assessment. The present study provides a more finely grained conceptualization of comprehensibility, upon which future studies could build in an interest to further develop speaking rubrics.

It is evident that the theoretical understanding of the construct of comprehensibility is still limited in the literature. This is reflected in the admittedly non-concrete implications

discussed above. We therefore argue for more studies into the conceptualizations of comprehensibility across a range of contexts, which use different research methods and designs to triangulate new findings with existing knowledge. This would advance our understanding of comprehensibility as a construct and allow us to provide more concrete implications for language learning, teaching, and testing.

Acknowledgments

This paper is based on the first author's master's dissertation. We would like to thank editors and reviewers for their support throughout the editorial process. We would also like to thank Jiying Zou for her expertise in statistics at the initial stages of this project.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Yongzhi Miao <https://orcid.org/0000-0002-1467-5688>

Supplemental material

Supplemental material for this article is available online.

References

- Burgoon J. K. (1993). Interpersonal expectations, expectancy violations, and emotional communication. *Journal of Language and Social Psychology*, 12(1–2), 30–48. [Crossref](#)
- Burgoon J. K. (2015). Expectancy violations theory. In Berger C. R., Roloff M. E., Wilson S. R., Dillard J. P., Caughlin J., Solomon D. (Eds.), *The international encyclopedia of interpersonal communication* (pp. 1–9). Wiley. [Crossref](#)
- Caspers J. (2010). The influence of erroneous stress position and segmental errors on intelligibility, comprehensibility and foreign accent in Dutch as a second language. *Linguistics in the Netherlands*, 27(1), 17–29. [Crossref](#)
- Cook V. (2016). *Second language learning and language teaching*. Routledge. [Crossref](#)
- Crowther D., Trofimovich P., Saito K., Isaacs T. (2015). Second language comprehensibility revisited: Investigating the effects of learner background. *TESOL Quarterly*, 49(4), 814–837. [Crossref](#)
- Crowther D., Trofimovich P., Saito K., Isaacs T. (2018). Linguistic dimensions of L2 accentedness and comprehensibility vary across speaking tasks. *Studies in Second Language Acquisition*, 40(2), 443–457. [Crossref](#)
- Derwing T. M., Munro M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition*, 19(1), 1–16. [Crossref](#)
- Derwing T. M., Munro M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. John Benjamins Publishing. [Crossref](#)
- Derwing T. M., Munro M. J., Wiebe G. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language Learning*, 48(3), 393–410. [Crossref](#)
- Dragojevic M. (2020). Extending the fluency principle: Factors that increase listeners' processing fluency positively bias their language attitudes. *Communication Monograph*, 87(2), 158–178. [Crossref](#)
- Field J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly*, 39(3), 399–423. [Crossref](#)
- Ghanem R., Kang O. (2021). ESL students' reverse linguistic stereotyping of English teachers. *ELT Journal*, 75(3), 330–340. [Crossref](#)

- Gordon J., Darcy I. (2016). The development of comprehensible speech in L2 learners: A classroom study on the effects of short-term pronunciation instruction. *Journal of Second Language Pronunciation*, 2(1), 56–92. [Crossref](#)
- Harding L. (2017). What do raters need in a pronunciation scale? The users' view. In Isaacs T., Trofimovich P. (Eds.), *Second language pronunciation assessment: Interdisciplinary perspectives* (pp. 12–34). Multilingual Matters. [Crossref](#)
- Isaacs T., Trofimovich P. (2012). Deconstructing comprehensibility: Identifying the linguistic influences on listeners' L2 comprehensibility ratings. *Studies in Second Language Acquisition*, 34(3), 475–505. [Crossref](#)
- Jenkins J. (2015). *Global Englishes: A resource book for students* (3rd ed.). Routledge.
- Kang O. (2010). Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System*, 38(2), 301–315. <https://10.1016/j.system.2010.01.005>
- Kang O., Thomson R. I., Moran M. (2019). The effects of international accents and shared L1 on listening comprehension tests. *TESOL Quarterly*, 53(1), 56–81. [Crossref](#)
- Kennedy S., Trofimovich P. (2008). Intelligibility, comprehensibility, and accentedness of L2 speech: The role of listener experience and semantic context. *Canadian Modern Language Review*, 64(3), 459–489. [Crossref](#)
- Kline R. B. (2015). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.
- Lambert W. E., Hodgson R. C., Gardner R. C., Fillenbaum S. (1960). Evaluational reactions to spoken languages. *Journal of Abnormal and Social Psychology*, 60(1), 44–51. [Crossref](#)
- Lev-Ari S., Keysar B. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology*, 46(6), 1093–1096. [Crossref](#)
- Levis J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly*, 39(3), 369–377. [Crossref](#)
- Levis J. M. (2020). Revisiting the intelligibility and nativeness principles. *Journal of Second Language Pronunciation*, 6(3), 310–328. [Crossref](#)
- Lindemann S., Litzenberg J., Subtirelu N. (2014). Problematizing the dependence on L1 norms in pronunciation teaching: Attitudes toward second-language accents. In Levis J., Moyer A. (Eds.), *Social dynamics in second language accent* (pp. 171–194). De Gruyter Mouton. [Crossref](#)
- Munro M. J., Derwing T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 45(1), 73–97. [Crossref](#)
- Munro M. J., Derwing T. M. (2001). Modeling perceptions of the accentedness and comprehensibility of L2 speech: The role of speaking rate. *Studies in Second Language Acquisition*, 23(4), 451–468. [Crossref](#)
- Munro M. J., Derwing T. M. (2020). Foreign accent, comprehensibility and intelligibility, redux. *Journal of Second Language Pronunciation*, 6(3), 283–309. [Crossref](#)
- O'Brien M. (2014). L2 learners' assessments of accentedness, fluency, and comprehensibility of native and nonnative German speech. *Language Learning*, 64(4), 715–748. [Crossref](#)
- Pérez-Ramón R., García Lecumberri M. L., Cooke M. (2022). Foreign accent strength and intelligibility at the segmental level. *Speech Communication*, 137, 70–76. [Crossref](#)
- Robinson P. (2005). Cognitive complexity and task sequencing: Studies in a componential framework for second language task design. *International Review of Applied Linguistics in Language Teaching*, 43(1), 1–32. [Crossref](#)
- Ruivivar J., Collins L. (2019). Nonnative accent and the perceived grammaticality of spoken grammar forms. *Journal of Second Language Pronunciation*, 5(2), 269–293. [Crossref](#)
- Saito K. (2021). What characterizes comprehensible and native-like pronunciation among English-as-a-second-language speakers? Meta-analyses of phonological, rater, and instructional factors. *TESOL Quarterly*, 55(3), 866–900. [Crossref](#)

- Saito K., Shintani N. (2016). Do native speakers of North American and Singapore English differentially perceive comprehensibility in second language speech? *TESOL Quarterly*, 50(2), 421–446. [Crossref](#)
- Saito K., Tran M., Suzukida Y., Sun H., Magne V., Ilkan M. (2019). How do second language listeners perceive the comprehensibility of foreign-accented speech? Roles of first language profiles, second language proficiency, age, experience, familiarity and metacognition. *Studies in Second Language Acquisition*, 41(5), 1133–1149. [Crossref](#)
- Saito K., Trofimovich P., Isaacs T., Webb S. (2017). Re-examining phonological and lexical correlates of second language comprehensibility: The role of rater experience. In Isaacs T., Trofimovich P. (Eds.), *Second language pronunciation assessment: Interdisciplinary perspectives* (pp. 141–156). Multilingual Matters. [Crossref](#)
- Saito Y., Saito K. (2017). Differential effects of instruction on the development of second language comprehensibility, word stress, rhythm, and intonation: The case of inexperienced Japanese EFL learners. *Language Teaching Research*, 21(5), 589–608. [Crossref](#)
- Skehan P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency and lexis. *Applied Linguistics*, 30(4), 510–532. [Crossref](#)
- Suzuki S., Kormos J. (2020). Linguistic dimension of comprehensibility and perceived fluency: An investigation of complexity, accuracy, and fluency in second language argumentative speech. *Studies in Second Language Acquisition*, 42(1), 143–167. [Crossref](#)
- Suzukida Y., Saito K. (2019). Which segmental features matter for successful L2 comprehensibility? Revisiting and generalizing the pedagogical value of the functional load principle. *Language Teaching Research*, 25(3), 1–20. [Crossref](#)
- Trofimovich P., Isaacs T. (2012). Disentangling accent from comprehensibility. *Bilingualism: Language and Cognition*, 15(4), 905–916. [Crossref](#)
- Varonis E. M., Gass S. (1982). The comprehensibility of non-native speech. *Studies in Second Language Acquisition*, 4(2), 114–136. [Crossref](#)
- Zhang R., Yuan Z. (2020). Examining the effects of explicit pronunciation instruction on the development of L2 pronunciation. *Studies in Second Language Acquisition*, 42(4), 905–918. [Crossref](#)