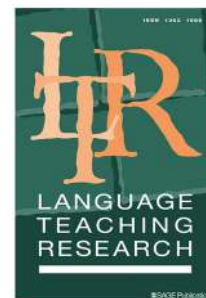


To be published in: *Language Teaching Research* (Sage publication)

## Which Segmental Features Matter for Successful L2 Comprehensibility? Revisiting and Generalizing the Pedagogical Value of the Functional Load Principle



**Yui Suzukida**<sup>1</sup>

Institute of Education, University College London, London

**Kazuya Saito**

Institute of Education, University College London, London

### Abstract

Building up on Derwing and Munro, the current study set out to reexamine and generalize the Functional Load (FL) principle (Brown, 1988) as a tool to identify a set of relatively crucial segmental features for successful understanding in L2 communication. In Experiment 1, 40 Japanese learners of English in English-as-Foreign-Language settings engaged in a semi-structured task (i.e., timed picture description). Their speech was assessed by native speaking raters for overall comprehensibility (ease of understanding); and then coded for the number of high vs. low FL segmental substitution errors according to the FL principle. The results showed that it was only high FL consonant substitutions (e.g., mispronunciation of /l/ as /r/ or /v/ as /b/) that negatively impacted on native listeners' comprehensibility judgements. In Experiment 2, 40 Japanese learners of English with a wide range of immersion experience in English-speaking countries participated. The results replicated the significant impact of high FL consonant substitutions as observed in Experiment 1. Taken together, this study suggests that the FL principle may greatly help teachers and students prioritize communicatively important segmental features, a crucial step towards improving L2 oral proficiency in an effective and efficient way.

**Keywords:** Second language pronunciation, segmentals, assessment, comprehensibility, task, functional load

---

<sup>1</sup> We are grateful to Professor Pavel Trofimovich and the *Language Teaching Research* reviewers for their insightful comments and advice on earlier versions of our paper. Parts of this study were presented at the 2017 EuroSLA conference (Reading, UK). The project was partially funded by the Grant-in-Aid for Scientific Research in Japan (No. 26770202) and the Birkbeck Additional Research Funds. **Corresponding author:** Yui Suzukida. Institute of Education, University College London, 20 Bedford Way, London, United Kingdom WC1H 0AL Email: [yui.suzukida.18@ucl.ac.uk](mailto:yui.suzukida.18@ucl.ac.uk)

### **Introduction**

With an emphasis on learners' communicative success in a second language (L2) as the ultimate goal of acquisition (Derwing and Munro, 2005; Field, 2005; Levis, 2005), scholars have extensively worked on identifying dimensions of speech directly relevant to listeners' understanding of accented speech (Kang, 2010; Kang, Rubin, and Pickering, 2010; Munro and Derwing, 2001; Saito, Trofimovich and Isaacs 2017; Trofimovich and Isaacs, 2012). While evidence from such research has advanced our understanding of two interrelated yet different constructs of speech (comprehensibility and accentedness), a growing amount of attention has been given to investigating which pronunciation features are relatively important for attaining successful comprehensibility regardless of accentedness (Levis, 2005). The relative importance of various prosodic features (e.g., word stress, sentence stress, intonation, rhythm) has been examined; however, segmentals are often treated as a single measurement (Isaacs and Trofimovich, 2012). Furthermore, all the relevant findings emerge from one single task condition (i.e., picture description task)—a methodological problem that may prevent researchers from investigating potential task effects on learners' segmental productions and narrowing down the list of essential segmentals to be taught. In what follows, we will first highlight Munro and Derwing's (2006) study, wherein the researchers took a very first step towards examining which segmental substitutions impaired comprehensibility via the Functional Load (FL) principle (Brown, 1988). Next, we will report general findings based on an empirical study which examined the role of high and low FL segmental errors in native listeners' comprehensibility of Japanese speakers of English under two experimental conditions (different task type, different sample type).

### **Background**

#### **Second Language Comprehensibility**

Despite ESL learners' desire to speak like native speakers (Derwing, 2003; Foote, Holtby, and Derwing, 2012; Kang 2010; Scales, Wennerstrom, Richard, and Wu, 2006; Timmis, 2002), second language acquisition (SLA) researchers have reported that hardly any L2 learners ever accomplish native-like control over the phonological dimensions of speech due to having retained foreign accents (Abrahamsson and Hyltenstam, 2009; Flege, Munro, and MacKay, 1995; Saito, 2015; Ortega, 2009). Moreover, researchers have found that foreign-accented speech can be highly intelligible (Murphy, 2014). In the language teaching context, English as a Second Language/English as a Foreign Language teachers have realized that native-like phonological control is not an essential goal for L2 learners as long as learners continue to progress towards relevant levels of intelligibility (Kanellou, 2009; Munro and Derwing, 2011). Accordingly, the general emphasis of the current agenda for L2 speech development and teaching has shifted towards the achievement of effective and efficient communication, with a particular focus on ease of understanding, rather than on the suppression of a foreign accent (Derwing and Munro, 2009). As Levis (2005) pointed out, a word comprehensibility is frequently used in L2 pronunciation research as the synonym of intelligibility (for more discussion of the term, see Isaacs, 2008), a rather clear account has been provided by Munro & Derwing (1995, 1999). According to their explanation, intelligibility and comprehensibility both indicate the listener's ability to understand L2 speech in the broader meaning, yet, in a narrower sense, they are different in the way they tap into listeners' processing during L2 speech judgements. Intelligibility relates to the amount of actual understanding, typically operationalized via orthographical transcriptions. Comprehensibility

indicates the ease of understanding through scalar ratings. In the current study, we focused on comprehensibility instead of intelligibility.

In previous research, the listeners' degree of understanding of L2 speech is often measured via raters' intuitive, scalar judgements of comprehensibility. These studies have shown that comprehensibility encompasses a broad range of features (for a comprehensive review, see Isaacs, Trofimovich, and Foote, 2018), as follows:

1. Temporal, such as pause, discussed by Kang, Rubin, and Pickering (2010), and speech rate investigated by Crowther, Trofimovich, Isaacs, and Saito (2015),
2. Prosodic, such as word stress studied by Field (2005) and Hahn (2004), tone choice by Pickering (2001) and Wennerstrom (2001), and pitch range and syllable length by Kang et al. (2010), Tajima, Port, and Dalby (1997), and Winters and O'Brien (2013),
3. Lexico-grammatical, such as grammatical accuracy as mentioned by Munro and Derwing (1995), Saito, Webb, Trofimovich, and Isaacs (2016), and Varonis and Gass (1982),
4. Lexical richness/appropriateness, as noted by Appel, Trofimovich, Saito, Isaacs, & Webb, 2019; Fayer and Krasinski (1987) and Saito et al. (2016, 2017), and
5. Segmental features, as identified by Derwing and Munro (1997), Isaacs and Trofimovich (2012), Munro and Derwing (1995), Saito et al. (2017) and Trofimovich and Isaacs (2012).

According to previous L2 speech studies with native listeners, there is some research evidence suggesting that prosodic accuracy should be prioritized (Derwing, Munro, and Weibe, 1998; Derwing and Rossiter, 2003; Hahn, 2004; Warren, Elgort, and Crabbe, 2009). However, such evidence does not mean that the accurate use of segmentals is irrelevant to L2 learner's communicative success. Instead, a number of L2 speech studies with various conditions (such as speakers' first language [L1] backgrounds and proficiency levels) have suggested that making certain segmental errors cannot be ignored for native listeners' successful understanding of L2 speech (Isaacs and Trofimovich, 2012). For example, segmental errors are associated with the comprehensibility of learners with certain L1 backgrounds, such as Francophone, Japanese and Chinese. In their investigation of mixed L1 speakers, Crowther, Trofimovich, Saito, and Isaacs (2015) demonstrated that the negative impact of segmental substitutions on comprehensibility rated by native listeners differed depending on the speakers' L1s. According to their results, involving Chinese, Hindi and Farsi learners of English, only Chinese speakers' segmental errors were associated with a lower comprehensibility rating. Their finding illustrated a unique influence of speakers' L1-specific patterns on the weight of segmental accuracy concerning comprehensibility.

### **Segmental Issues in L2 speech**

In the context of the research on attaining global intelligibility by L2 speakers from various L1 backgrounds, Jenkins (2000) established the framework of the Lingua Franca Core (LFC) and suggested the importance of learning core segmentals as a key component for producing intelligible speech. LFC is based on conversational and information gap task data she collected from non-native speakers of English. For example, among all the English segmentals, learners are suggested to approximate most consonants (with few exceptions, such as /θ, ð/) and certain vowel articulations

(e.g., /æ/ in *sad* and *sat*: for more detail please see Jenkins, 2000). LFC has led to a quantity of empirical research that has investigated the elements of mutual intelligibility among non-native speakers of English (Deterding and Kirkpatrick, 2006; Osimk, 2009; Pickering, 2009; Pitzl, 2005). As the results of such research are quite mixed (some evidence supporting LFC fully and other evidence not), further research is required to determine the contribution of segmental accuracy to global intelligibility (for a review, see Luchini and Kennedy, 2013).

Compared to the number of studies that focused on the interactions among non-native speakers of English (Dauer, 2005; Jenkins, 1998, 2000, 2007; Smith and Nelson, 2006; Walker, 2010), native listener-based research, which assumes the interaction between non-native and native speakers of English, lacks empirical evidence to prove the importance of prioritizing certain segmentals for attaining higher comprehensibility. It can be argued that the method of segmental analysis in the aforementioned comprehensibility research usually specifies any deviations from the native baseline or identifies substitutions for other sounds (for example, *think* pronounced as *tink*; see Trofimovich and Isaacs, 2012). Therefore, “a more nuanced approach (Isaacs and Trofimovich, 2012, p. 498)” is called for to identify the segmentals that hinder native listeners’ understanding more than others.

### **Functional Load Principle**

Notably, Munro and Derwing's (2006) study aimed to identify crucial segmentals by adopting the Functional Load (FL) principle (Brown, 1988; Catford, 1987). FL is a list of segmental contrasts that are ranked based on their communicative value. These contrasts were developed from minimal pairs in frequently used words, the degree of neutralization among regional English dialects, and the segmental position within a word. Using this concept, they further divided the segmental contrasts into high and low FL categories by simply dividing the ranked numbers in two—10 to 6 as high FL, and 5 to 1 as low FL according to Brown's (1988) ranking, and 100 % to 51 % as high FL, and 50 % to 1 % as low FL based on Catford's (1978) ranking. Using speech samples of Cantonese speakers elicited via a sentence-reading task, they identified that substitutions of high FL consonants (for example, *Need* is pronounced as *Leed*) led to significantly lower comprehensibility ratings by native listeners than did low FL consonant substitutions (such as  $\theta$  pronounced f, as in *mouTH* is pronounced as *mouF*).

In essence, Munro and Derwing's (2006) study confirmed that there is variation in the acquisitional values of individual sounds (high and low FL distinctions of segmentals), suggesting that teachers should prioritize high FL segmentals in foreign language classrooms in which the time to teach pronunciation is often limited (Derwing and Munro, 2005; Munro, Derwing, and Thomson, 2015). However, it is noteworthy that virtually no other studies have examined the varying importance of segmentals on listeners’ comprehensibility nor have they developed an efficient and effective pronunciation syllabus by adapting FL (for a study of FL in the oral assessment context, see Kang and Moran, 2014). One of the goals of the current study was, therefore, to provide further evidence to support the high/low FL approach for L2 pronunciation research.

### **Role of Tasks**

While research on L2 pronunciation and speech comprehensibility has usually employed some form of picture description task (Derwing and Munro, 2009; Hopp and Schmid, 2013; Isaacs and Trofimovich, 2012; Munro and Mann, 2005; Saito et

al., 2015) to elicit extemporaneous speech, other aspects of L2 speech research such as fluency research (Foster and Skehan, 1996; Skehan and Foster, 1999) have demonstrated that the quality of the speech produced is attributed to the nature of a task. The impact of task type on L2 speech performance has been discussed actively in the area of fluency along with accuracy and complexity (e.g., Bygate, 1996; Ejzenberg, 1992; Skehan and Foster, 1999). For instance, Derwing, Rossiter, Munro, and Thomson (2004) provided an account of task differences based on the degree of freedom speakers had in terms of lexical choices, structures, and content during task completion. Derwing et al. reported that, while a personal information exchange task and a collaborative decision-making task were found to be associated with higher fluency ratings, a picture description task received significantly lower fluency ratings. Thus, Derwing et al. suggested that the picture description task may be the most difficult for their participants to complete due to the speakers' limited control over the content and lexical choices about the content they needed to describe, as they were compelled to adhere to the topic shown in the pictures. On the other hand, a personal information exchange task and a collaborative decision-making task may be relatively easy for the speakers because the task formats allow for a great deal of topical, structural and lexical freedom. In those two tasks, thus, the participants could scaffold their speech easily by accessing the familiar topics they had conceptualized previously, either in their L2 or their L1.

In the context of L2 pronunciation research, however, impact of the task type has received less attention. More recently, Crowther, Trofimovich, Isaacs et al. (2015) demonstrated how different types of task affected the linguistic correlates of comprehensibility using two tasks with different complexity (The International English Language Testing System [IELTS] and Test of English as a Foreign Language [TOEFL] tasks<sup>1</sup>). While some variations among the speakers' L1s could be observed (Chinese, Hindi/Urdu, Farsi and Roman), the overall result showed that comprehensibility rated by native listeners was linked to pronunciation and fluency measures in the task with lower cognitive demands (IELTS task), whereas it was associated with a wider range of measures ranging from lexicogrammatical aspects to pronunciation variables in the task with high cognitive demands (TOEFL task).

The study provided evidence that task type affected the linguistic variables to which native listeners paid attention when judging comprehensibility. While pronunciation (including segmental accuracy) and fluency are crucial for L2 comprehensibility regardless of task complexity, the importance of other features of speech such as lexis, grammar, and discourse tend to increase as the complexity of the task increases. Although such evidence suggests that refined segmental accuracy plays a crucial role in L2 learners' communicative success, it remains unclear which segmental errors are consequential to L2 comprehensibility than others, and whether such categorizing is resistant to different conditions. Therefore, the second goal of the current study is to address the generalizability of the crucial segmental features for L2 comprehensibility by conducting a separate experiment with different task type and sample type.

### **Motivation for the Current Study**

Despite the general lack of further scrutiny of segmental correlations with raters' L2 comprehensibility judgments (Trofimovich and Isaacs, 2012), few studies have investigated how both vowel and consonant substitutions affect listeners' perception of comprehensibility, and how such results vary according to different task conditions. Furthermore, despite researchers' interest in the use of FL to identify

segmental impacts on communication breakdowns (Zielinski, 2008), virtually no studies have been conducted except for Munro and Derwing's (2006) study with a sentence-leading task (cf. Kang and Moran, 2014 for an FL application in an oral assessment context). Considering the task effect on L2 comprehensibility (Crowther, Trofimovich, Isaacs et al., 2015; Derwing et al., 2004), further examinations of the segmental influences on L2 comprehensibility are required to demonstrate how such findings could be generalizable across different task conditions, and to advance the current agenda of effective and efficient segmental instruction. Therefore, the current study investigated this topic (i.e., which segmental features are crucial for L2 comprehensibility judgements) by conducting two separate experiments with adopting Functional Load principle.

We first report the results of our original experiment, where we tested the predictive power of the Functional Load principle in the context of a picture description task (Experiment 1). As this type of task has been often used to elicit a sufficient length of spontaneous speech from L2 speakers of various proficiency levels in the research on L2 speech development (for example, Saito and Hanzawa, 2016 for inexperienced L2 learners; Saito, Trofimovich, & Isaacs, 2016 for intermediate to advanced L2 learners). To reexamine and replicate the results obtained in Experiment 1, the second experiment (Experiment 2) is conducted with different participants under different task conditions. Experiment 2 highlighted a relatively unstructured task—an IELTS long-turn interview task (henceforth, IELTS task) to elicit a longer stretch of more extemporaneous speech, as this task format allowed the speakers much more structural and conceptual freedom than would the picture description task (Foster and Skehan, 1996). The speakers recruited for Experiment 2 also differ from Experiment 1 in their length of residence (LOR) (LOR = 0 in Experiment 1 vs. LOR = 0–24 in Experiment 2), indicating that variety in speech proficiency level of the two groups was assumed to be different (cf. Flege, Takagi and Mann, 1996; Munro and Derwing, 2008; Trofimovich and Baker, 2006).

Extending Munro and Derwing's (2006) research framework, (a) timed picture description for Experiment 1 and (b) IELTS task for Experiment 2 were selected as these tasks were sufficiently different in topical, structural and lexical freedom for comparing the results. According to previous studies, comprehensibility ratings were negatively affected by segmental errors when L2 speech was elicited from the timed picture description task and the IELTS task (Crowther, Trofimovich, and Saito et al., 2015 and Saito et al., 2017).

The participants' L2 speech in Experiment 1 was first assessed by four native speakers of English concerning comprehensibility (for a similar decision concerning number of raters, see Zielinski, 2006; and then analyzed acoustically by three Japanese coders who are highly proficient in English to identify high- and low-FL vowel and consonant substitutions. The same procedure was taken for the L2 speech sample in Experiment 2. Subsequently, the results of Experiment 1 were discussed together with the results of Experiment 2 to draw a general finding concerning segmental correlates of L2 comprehensibility.

## Experiment 1

### Speakers

A total of 40 Japanese learners of English participated in the current study in 2016 summer. They were freshman students at universities in Tokyo, Japan ( $M_{age}=18.42$ ;  $Range = 18-19$ ). Their TOEIC<sup>2</sup> scores widely varied from 450 and 910 (equivalent of B1–C2 in The Common European Framework of Reference for

Languages [CEFR]), indicating that their L2 proficiency levels could be considered intermediate to advanced. At the time of the experiment, all the participants reported that they had no prior study abroad or living abroad experience, suggesting that their L2 learning had been taken place in English as a foreign language setting (cf. Muñoz, 2008).

### **Speaking Task: Timed Picture Description**

Timed picture description tasks are often used as a way to elicit learners' spontaneous speech production in L2 oral development and speech intelligibility studies to examine their state of speech proficiency (e.g., Saito et al., 2016). The task is regarded as a simple information transmission that requires relatively little cognitive requirements—understanding an image and construct a response by using the provided keywords (for example, speakers describe an image of a cloud in the blue sky with keywords blue sky, road, and cloud). The task does not involve conceptualizing of their experience, interpreting causal relationships, and integrating multiple perspectives for forming their opinions.

The speech samples used in the experiment were elicited via a timed picture description task of three prompt pictures. They depicted a table left out in a driveway in heavy rain (keywords: rain, table, and driveway), three men playing rock music with one singing a song and the other two playing guitars (keywords: three guys, guitar, and rock music), and a long stretch of road under a cloudy blue sky (keywords: blue sky, road, and cloud). A careful selection of the keywords was made to include challenging segmentals and syllables features for Japanese speakers. Prior to the three target pictures, four practice pictures were used to control for speakers' lack of familiarity with the task. In addition, speakers were given 5 seconds of planning time before describing each picture to minimize the conscious speech monitoring (see Ellis, 2005).

### **Global Analysis**

**Material Preparation.** For the speech rating, the first 30 seconds of the entire speech was excised from each of the 40 speakers' performance (hence, approximately three minutes each) for preparation of the ratings in order to (a) be in line with previous speech judgement studies (e.g., Derwing and Munro, 1997; Iwashita, Brown, McNamara, and O'Hagan, 2008), and to avoid raters' fatigue, which affects the accuracy of the rating result due to its time-consuming nature. In addition, since this study's focus is on pronunciation, not lexicogrammar or fluency, it was decided that 30 seconds is long enough to identify pronunciation errors on the basis of the empirical evidence shown by Munro, Derwing, and Burgess (2010), which revealed that native speakers could detect accents even with a single word.

**Raters.** To assess the comprehensibility of the 40 speech samples of Japanese learners of English, four English native judges ( $M_{\text{age}} = 31.8$  years; range = 29-37) were recruited in London. According to the results of background questionnaire, they are ESL/EFL teachers with teaching experience ( $M = 5.8$  years; Range = 3-8.5) and extensive phonological knowledge. Their familiarity with Japanese accented English was quite high on a 6-point scale ( $M = 5.5$  years; Range = 4-6).

**Procedures.** The rating sessions were conducted individually in a quiet room using the rating software *Praat* (Boersma and Weeink, 2017). Each rater participated in rating sessions on different day and time. With the software, raters judged speeches by clicking a number in a box-shaped 9-point Likert scales (1 as hard to understand, 9 as easy to understand) depending on their perceived comprehensibility level.

All the rating results were automatically recorded in the software. The raters listened to the audio through a pair of earphones connected to the researcher's laptop computer. For each rating, the raters were asked to listen to the whole audio (30 seconds) before they made the judgements. The software played the speech only once in a randomized order, and the next speech was played right after the raters made the judgement for the previous speech.

Prior to the actual rating sessions, a language background questionnaire and sufficient instruction for the rating were given to the raters. First, the researcher asked the raters to complete the language background questionnaire to obtain the information about (a) the languages they speak, and (b) their degree of the familiarity with Japanese and Japanese accented English. Second, a training session was directed by the researcher to ensure that the raters adequately understood the concept of comprehensibility. In the training phase, the concept of comprehensibility (how effortless it is to understand L2 speech) was introduced to the raters with a 9-point scale (*1 = hard to understand, 9 = easy to understand*) based on the definition used in the previous research on comprehensibility (e.g., Trofimovich and Isaacs, 2012). They were asked to make intuitive judgements and use the scales flexibly. To confirm the raters' accurate understanding of the scale, they completed a practice rating session with the rating software and discussed the results with the researcher before the actual rating.

In order to examine the consistency of the four native raters' judgments, inter-rater reliability was calculated. Cronbach's  $\alpha$  of the raters was .91 ( $p < .001$ ). The Cronbach's  $\alpha$  value indicates that three raters were internally consistent and reacted to the speech samples in a very similar way. Therefore, we proceeded to calculate mean comprehensibility scores of each speaker by averaging the scores rated by the four raters.

### Segmental Analysis

Building on the existing fine-grained coding approach in L2 speech proficiency studies (e.g., Saito, Suzukida, and Sun, 2018), the speech samples were objectively analyzed in terms of high- vs. low-FL segmental substitutions.

**Coders.** A total of three Japanese natives who were highly proficient in L2 English conducted the segmental analyses. They were considered "experienced," as they had a great deal of EFL teaching experience ( $M = 4.3$  years of teaching experience), completed MA degree in TESOL or Applied Linguistics, and have knowledge of English phonological and phonetics. As conceptualized and demonstrated in Riney, Takada, and Ota (2000), experienced coders, who share the same L1 background as the speakers, are assumed to be better capable of detecting a subtle impact of the L1 phonological transfer on speakers' interlanguage (for similar decisions, see Saito, Suzukida, and Sun, 2018).

**Procedure.** Firstly, the researcher who is a native speaker of Japanese and familiar with Japanese-accented English completed the coding. Any substitutions of English segmentals with Japanese sounds were identified and counted as segmental substitutions. For instance, salient substitutions observed in the speech samples include the use of Japanese liquid /r/ for English /l/ in cloud or challenge, and Japanese /b/ for English /v/ in driveway.

Subsequently, in order to (a) ensure the correctness of such coding by the researcher, and (b) pursue a fine-grained objective analysis of the speech, the result of the first coding was re-coded by two Japanese coders. Prior to the re-coding procedure, the two additional coders were sufficiently instructed for the accurate



detection of the substitutions (i.e., any substitution of English segmentals with Japanese sounds). The results of the two additional codings did not show a major difference from the first coding result. The three coders' Cronbach's  $\alpha$  was .91 ( $p < .001$ ), suggesting that the three coders' judgements were internally highly consistent. Table 1 illustrates all the substitutions (e.g., /v/  $\rightarrow$  /b/) identified in the speech samples and actual examples of those substitutions (e.g., driveway).

Table 1  
*Type of Segmental Substitutions Found in Experiment 1*

Segmentals	FL category	Substitutions	Examples
Vowel	High FL	/ei/ $\rightarrow$ /e/	<b>table</b>
	Low FL	/ɑ:/ $\rightarrow$ /ɒ/	<b>guitar</b>
		/aʊ/ $\rightarrow$ /ɔ:/	<b>cloud</b>
Consonant	High FL	/r/ $\rightarrow$ /l/	<b>rock</b>
		/l/ $\rightarrow$ /r/	<b>cloud</b>
		/v/ $\rightarrow$ /b/	<b>drive</b>
	Low FL	/s/ $\rightarrow$ /ʃ/	<b>sing</b>
		/z/ $\rightarrow$ /dʒ/	<b>music</b>
		/ð/ $\rightarrow$ /z/	<b>the</b>
		/θ/ $\rightarrow$ /s/	<b>three</b>

*Note.* Bold characters in Examples indicate the locations of the substitution.

After ensuring the internal consistency of the coders, the total number of segmental substitutions was calculated for each sample. Subsequently, in order to examine the gravity of high FL and low FL substitutions, the total number of substitutions was divided into four subcategories based on the FL lists by Brown (1988) and Munro and Derwing's (2006) distinction of high and low FL (i.e., 1-5 are low FL, 6-10 are high FL)—high FL and low FL vowel substitutions and high FL and low FL consonant substitutions. For instance, the use of Japanese /r/ for English /l/ was categorized as a consonant substitution, then further subcategorized as a high FL consonant substitution as /r/-/l/ substitution is ranked 10<sup>th</sup> in Brown's (1988) FL list and 10<sup>th</sup> is regarded as high FL in Munro and Derwing's (2006) distinction.

Overall, seven variables (i.e., overall segmental substitution, high FL vowel substitutions, low FL vowel substitutions, high FL consonant substitutions, and low FL consonant substitutions) were calculated. To calculate the ratio of each type of segmental substitutions, the total numbers of the seven variables were then divided by the total number of segments articulated for each sample (e.g., Kang and Moran, 2014; Saito et al., 2018; Trofimovich and Isaacs, 2012). Table 2 shows the results of categorization and calculation of the ratio for seven segmental measures.

## Result

In order to identify crucial segmental features for L2 comprehensibility, a set of correlation analyses was computed in Experiment 1. Seven segmental measures and comprehensibility scores of 40 speech samples were submitted to a Spearman's correlation coefficient (summarized in Table 3). According to the results, significant negative correlations with comprehensibility judgment were found in overall segmental substitutions ( $r = -.557, p < .001$ ), high FL substitutions ( $r = -.561, p < .001$ ), and high FL consonant substitutions ( $r = -.560, p < .001$ ). However, correlation

values of high and low FL vowel substitutions and low FL consonant substitutions did not reach statistical significance ( $p > .05$ ).

Table 2

*Descriptive Statistics of Segmental Errors (ratio) in Experiment 1*

Segmental measures	<i>M</i>	<i>SD</i>	<i>Range</i>
Segmentals (overall)	.03	.031	0–.12
Segmentals (High FL)	.024	.025	0–.1
Segmentals (Low FL)	.003	.006	0–.02
Segmentals (vowel high FL)	.001	.004	0–.02
Segmentals (vowel low FL)	.003	.005	0–.02
Segmentals (consonant high FL)	.023	.024	0–.09
Segmentals (consonant low FL)	.003	.005	0–.02

Table 3

*Results of Correlations between Comprehensibility and Seven Segmental Measures at Experiment 1*

Segmental measures	Comprehensibility	
	<i>r</i>	<i>p</i>
Segmentals (overall)	-.557*	<.001
Segmentals (High FL)	-.561*	<.001
Segmentals (Low FL)	-.314	.049
Segmentals (vowel high FL)	-.151	.353
Segmentals (vowel low FL)	-.228	.156
Segmentals (consonant high FL)	-.560*	<.001
Segmentals (consonant low FL)	-.250	.119

Note. \*indicates  $p < .005$  (Bonferroni corrected)

### Experiment 2

Although Experiment 1 aimed to identify the segmental correlates of L2 comprehensibility, the findings here could have been specific to the particular task—i.e., timed pictured description. Considering the task effect on L2 speech production as observed in previous literature (e.g., Derwing et al., 2004), we need to reexamine the segmental correlates under different task conditions. This follow-up experiment (henceforth Experiment 2) was conducted with different task and another group of speakers in 2017 spring.

### Speakers

In order to cover wider variety of L2 speech proficiency level, 40 Japanese learners with varied L2 immersion experience (measured via LOR in English-speaking countries) were participated. The participants were comprised of college-level students in Japan (without any experience overseas) and Japanese learners of English who resided in Calgary, Canada at the time of the experiment (LOR:  $M = 4.6$  years,  $Range = 0.1$ -24 years). According to the findings from the previous studies in L2 speech development (e.g., Flege, Takagi and Mann, 1996; Munro and Derwing, 2008; Trofimovich and Baker, 2006), L2 learners can greatly improve their pronunciation accuracy and fluency as a function of increased LOR. This suggests that the participants in Experiment 2 could be considered to have more varied,

advanced L2 speech proficiency than those in Experiment 1 (i.e., EFL students without any experience abroad).

### **Speaking task: IELTS Interview**

Compared to timed picture description tasks, IELTS tasks require a greater number of cognitive activities to complete the task (e.g., favorite place to visit). These include interpreting the prompt, recalling their previous personal experience, organizing the ideas in a sequenced manner, and articulating their answers.

The IELTS task used in the experiment followed a condition applied to the actual IELTS speaking test. The test is conducted in an interview style with an examiner. In the long-term part of the IELTS speaking test (i.e., the second part of the test), the examiner provides a test-taker with a prompt sheet and asks him/her to speak after one minute of preparation time. Following this procedure, a prompt "Describe the hardest and toughest challenge in your life" with few supporting questions was developed and used for the elicitation of the speech samples, and the speakers were given one-minute preparation time and talked for three minutes.

### **Global Analysis**

Speech samples were prepared (30 seconds each) and rated (comprehensibility rating on a 9-point scale) in the same way as the speech samples described in Experiment 1.

**Raters.** Four raters participated in Experiment 1 were recruited (see Experiment 1 for the rater profile). In order to examine the consistency of the four native raters' judgments, inter-rater reliability was calculated (Cronbach's  $\alpha = .87$ ,  $p < .001$ ). Since the raters demonstrated consistent rating pattern, mean comprehensibility scores of each speaker were calculated.

### **Segmental Analysis**

Coding procedure and coders of segmental analysis were the same as Experiment 1. The three coders' Cronbach's  $\alpha$  was .94 ( $p < .001$ ), suggesting that three coders' judgements were internally highly consistent. Table 4 illustrates all the substitution types (e.g., /ð/ → /z/) and examples identified in the speech samples (e.g., **that**). As it is for Experiment 1, seven variables (i.e., overall segmental substitution, high FL substitutions, low FL substitutions, high FL vowel substitutions, low FL vowel substitutions, high FL consonant substitutions, and low FL consonant substitutions) were prepared (see Table 5 for a summary of error ratios of five segmental measures in Experiment 2).

### **Result**

A set of Spearman's correlations was run for examining segmental correlates of comprehensibility (summarized in Table 6). Similar to the results observed in Experiment 1 with the timed picture description task, the variables that showed statistical significance in the IELTS task included overall segmental substitutions ( $r = -.798$ ,  $p < .001^*$ ), high FL substitutions ( $r = -.685$ ,  $p < .001^*$ ), and high FL consonant substitutions ( $r = -.617$ ,  $p < .001^*$ ) but high and low FL vowel substitutions and low FL consonant substitutions.

### **Discussion and Conclusion**

With using two different tasks and groups of speakers (Experiment 1 with timed picture description task and EFL Japanese learners of English vs. Experiment 2 with IELTS task and Japanese learners of English with varied immersion experience),

Table 4

*Type of Segmental Substitutions Found in Experiment 2*

Segmentals	FL category	Substitutions	Example
Vowel	High FL	/ei/ → /e/	<b>m</b> ajor
		/i/ → /e/	pre- <b>t</b> est
		/ɒ/ → /ʌ/	was
Consonant	Low FL	/ɑ:/ → /ɒ/	<b>h</b> ard
	High FL	/r/ → /l/	gr <b>a</b> de
		/l/ → /r/	<b>l</b> ife
		/v/ → /b/	l <b>i</b> ving
	Low FL	/s/ → /ʃ/	un <b>i</b> versity
		/z/ → /dʒ/	<b>b</b> usy
		/ð/ → /z/	<b>t</b> hat
		/θ/ → /s/	<b>t</b> hird

*Note.* Bold characters in Examples indicate the locations of the substitution.

Table 5

*Descriptive Statistics of Segmental Errors (ratio) in Experiment 2*

Segmental measures	<i>M</i>	<i>SD</i>	<i>Range</i>
Segmentals (overall)	.017	.015	0–.06
Segmentals (High FL)	.014	.011	0–.04
Segmentals (Low FL)	.004	.008	0–.03
Segmentals (vowel high FL)	.001	.003	0–.01
Segmentals (vowel low FL)	.002	.004	0–.02
Segmentals (consonant high FL)	.012	.011	0–.04
Segmentals (consonant low FL)	.002	.005	0–.02

Table 6

*Results of Correlations between Comprehensibility and Seven Segmental Measures in Experiment 2*

Segmental measures	Comprehensibility	
	<i>r</i>	<i>p</i>
Segmentals (overall)	-.798	<.001*
Segmentals (High FL)	-.685	<.001*
Segmentals (Low FL)	-.234	.146
Segmentals (vowel high FL)	-.380	.016
Segmentals (vowel low FL)	-.190	.241
Segmentals (consonant high FL)	-.617	<.001*
Segmentals (consonant low FL)	-.186	.250

*Note.* \*indicates  $p < .005$  (Bonferroni corrected)

the current study examined how both high- and low-functional load segmental errors differentially impact raters' L2 comprehensibility judgements.

To this end, Brown's (1988) FL principle and Munro and Derwing's (2006) approach of high/low FL distinctions were adapted to provide a full-fledged picture of the impact of different types of segmental substitutions on L2 comprehensibility. The correlation analyses between seven segmental phonetics variables and the comprehensibility rating in Experiment 1 showed statistical significance in overall segmentals, high FL segmentals (consonants, vowels combined), and high FL consonants, and the results of Experiment 2 demonstrated the same pattern: while overall segmentals, high FL segmentals, and high FL consonants showed statistical significance, none of the vowels and low FL consonants showed any correlations. Overall, the importance of high FL segmentals (and especially, high FL consonants) was highlighted in Experiment 1, and these findings, which may be potentially biased due to its use of single task and specific participants, were confirmed by the results obtained in Experiment 2.

First and foremost, our results (from Experiment 1 and 2) showed the relative weights of high FL errors over low FL errors in terms of their detrimental effects on L2 comprehensibility judgements. More specifically, our results revealed that high FL consonants substitutions lowered the comprehensibility more than that of high FL vowel substitution. The findings here (high FL > low FL; consonants > vowels) are in line with the result of Munro and Derwing (2006). In their study, it was found that sentences with high FL consonant substitutions received significantly lower comprehensibility rating than those with low FL consonant substitutions. Furthermore, it is important to note here that previous L2 pronunciation studies have focused more on consonants than vowels in pronunciation instruction (e.g., Munro et al., 2015). In contrast, other L2 pronunciation research has offered evidence that L2 vowel errors are less likely to hinder L2 communication especially between L2 users (Jenkins, 2000). Our results can be taken as indicative of this very same lack of effect of vowel errors on comprehensibility when native listeners are involved in the speech judgement task.

Second, to seek generalizability in segmental correlates of L2 comprehensibility, the current study conducted two separate experiments (Experiment 1 and 2). Since the results obtained from single speaking task can be subject to the effect of task type (e.g., Munro et al., 2004), each experiment employed different tasks. Following Derwing et al. (2004) and Foster and Skehan (1996), a selection of the two tasks—timed picture description, IELTS Interview—was made based on the structural, topical, and lexical freedom they offer to the speakers—relatively little freedom in the timed picture description task, and relatively more freedom in the IELTS task. Interestingly, unlike fluency studies that found the differences in the results in picture description task from that of the monologue and conversation tasks (Derwing et al., 2004; Foster and Skehan, 1996), any differences in the correlation results were not found—high FL consonant substitutions hampered comprehensibility significantly than the others regardless of task conditions.

The null effect of task type in the current study is worthy of attention. In the timed picture description task, the participants had to use a total of nine key words to describe three different pictures. As such, the participants were induced to make certain segmental errors (including a number of high FL errors) in Experiment 1. For example, the frequency of mispronunciations of /z/ in the timed picture description was maybe due to a fixed keyword *music* that need to be used when the speakers described the given picture, and a number of speakers mispronounced as *mugic*. Thus, the strong correlations between FL and L2 comprehensibility in Experiment 1 were not surprising. However, the FL-comprehensibility link was successfully replicated in

the IELTS task (Experiment 2), where the participants were encouraged to provide extemporaneous, free speech. The results of Experiment 1 and 2 provide strong evidence that the ratio of segmental errors with high-level FL hinders L2 comprehensibility in various social settings. Therefore, overall, the same results obtained from two experiments (different task and different samples) may serve as positive evidence of the importance of accurate articulation of high FL segmentals for comprehensibility irrespective of other mediating variables (e.g., task type, topic, sample type). Based on these findings in the current study, a pedagogical implication can be made for Japanese learners of English. In order for their English speech to be comprehensible, it is crucial to work on accurate production of high FL consonants (e.g., /v/ → /b/ in *drive*). Subsequently, they should move on to other segmental such as high FL vowels (e.g., /i/ → /e/ in *pre-test*) and low FL segmental to attain higher comprehensibility.

Lastly, it is important to provide a range of directions that future researchers should take in order to further expand the findings of the current study. First, the raters in this study were limited to native speakers of English. Considering the fact that there are more English users who have different L1 backgrounds than native speakers (Trudgill and Hannah, 2017; for English as lingua franca see Seidlhofer, 2011) and listener factor has received considerable attention in regard to L2 oral interaction (e.g., Kang, 2012; Munro, 2010; Zielinski, 2008), it is crucial to take into account the perceptions of native speakers and L2 users from various L1 backgrounds. However, L2 listeners could be characterized as a multifaced phenomenon, as their backgrounds may differ to a great degree in terms of L1-L2 distance, experience, familiarity, attitude, and cognitive/sociopsychological profiles. Indeed, it needs be noted here that existing research so far has yielded mixed results regarding non-native listeners. For example, while some studies demonstrated that L1 listeners rate differently from L2 listeners when speakers' L1s are closer to or the same as that of L1 listeners (e.g., Foote and Trofimovich, 2018), other studies have shown that L1 and L2 listeners arrive at similar rating results, and shared L1 backgrounds would have little to do with rating behavior (e.g., Gallardo-del-Puerto, García-Lecumberri, and Gómez-Lacabex, 2015 with trained L2 listeners; Major, 2007; Munro, Derwing, and Morton, 2006; MacKay, Flege, and Imai, 2006). The discrepancy here indicates a necessity of generalization of the findings in order to forward L2 speech research to examine potential listener-oriented effect on segmental correlates of L2 comprehensibility. We strongly call for future studies which will carefully examine the complex relationship between different types of segmental features, different groups of listeners and L2 comprehensibility judgements while elaborating a controlled design (cf. Saito, Tran, Suzukida, Sun, Magne, & Ilkan, in press). Secondly, although the current study employed two different tasks to elicit L2 speech, contribution to an ongoing discussion of task effect on L2 production (e.g., Crowther et al., 2015; Derwing et al., 2004) was out of the scope for the current study. Therefore, the current study acknowledges that investigations of task effect are a future direction of L2 pronunciation proficiency. Such future studies should control for a range of task variables which are assumed to influence the accuracy, complexity and fluency aspects of L2 speech performance (see Foster and Skehan, 1996). Finally, the current study only focused on Japanese speakers as the participants. Considering the pedagogical practicality for the variety of learner background, the study should have included different L1 speakers. Thus, further study should include speakers from other L1 backgrounds to make pedagogical implication more general.

**Note**

1. IELTS stands for The International English Language Testing System. It's speaking section consists of three parts: part 1 is an introduction and a short interview from an examiner, part 2 is individual long turn, and part 3 is a two-way discussion with the examiner. TOEFL iBT stands for Test of English as a Foreign Language Internet-Based test. There are six parts in the test and broadly categorized into two: independent speaking task (part 1, 2) and integrated speaking task (part 3, 4, 5, 6). Candidates are asked to speak to a microphone provided and their answers are recorded for evaluation. The first two parts ask candidates to express their opinions on given topics, and the rest of four tasks ask them to talk based on listening and reading passages.
2. TOEIC stands for The Test of English for International Communication.

### References

- Abrahamsson N and Hyltenstam K (2009) Age of acquisition and nativelikeness in a second language: Listener perception vs. linguistic scrutiny. *Language Learning* 59: 249-306.
- Appel R, Trofimovich P, Saito K, Isaacs T and Webb S (2019). Lexical Aspects of Comprehensibility and Nativeness from the Perspective of Native-Speaking English Raters. *International Journal of Applied Linguistics* 170: 24-52.
- Crowther D, Trofimovich P, Isaacs T and Saito K (2015) Does a Speaking Task Affect Second Language Comprehensibility? *The Modern Language Journal* 99: 80-95.
- Crowther D, Trofimovich P, Saito K and Isaacs T (2015) Second Language Comprehensibility Revisited: Investigating the Effects of Learner Background. *TESOL Quarterly* 49: 814-837.
- Boersma D and Weenink P (2017) *Praat: doing Phonetics by Computer*. Available at: [www.fon.hum.uva.nl/praat/](http://www.fon.hum.uva.nl/praat/) (accessed 25 June 2017).
- Brown A (1988) Functional Load and the teaching of pronunciation. *TESOL Quarterly* 22: 593-606.
- Bygate M (1996) Effects of task repetition: Appraising the developing language of learners. In: Willis J and Willis D (eds), *Challenge and Change in Language Teaching*. Oxford, England: Heinemann, pp. 136-146.
- Catford J C (1987) Phonetics and the teaching of pronunciation: A systemic description of English phonology. In: Morley J (ed), *Current Perspectives on Pronunciation: Practices Anchored in Theory*. Washington, DC: TESOL, pp. 87-100.
- Dauer R (2005) The Lingua Franca Core: A new model for pronunciation instruction? *TESOL Quarterly* 39: 543-550.
- Derwing T M (2003) What do ESL students say about their accents? *Canadian Modern Language Review* 59: 547-567.
- Derwing T M and Munro M J (1997) Accent, intelligibility, and comprehensibility. *Studies in Second Language Acquisition* 19: 1-16.
- Derwing T M and Munro M J (2005) Second language accent and pronunciation teaching: a research-based approach. *TESOL Quarterly* 39: 379-397.
- Derwing T M and Munro M J (2009) Putting accent in its place: Rethinking obstacles to communication. *Language Teaching* 42: 476-490.
- Derwing T M, Munro M J and Wiebe G (1998) Evidence in favor of a broad framework for pronunciation instruction. *Language Learning* 48: 393-410.
- Derwing T M and Rossiter M J (2003) The effects of pronunciation instruction on the accuracy, fluency and complexity of L2 accented speech. *Applied Language Learning* 13: 1-18.
- Derwing T M, Rossiter M J, Munro M J and Thomson R I (2004) Second language fluency: Judgments on different tasks. *Language Learning* 54: 655-679.
- Deterding D and Kirkpatrick A (2006) Emerging South-East Asian Englishes and intelligibility. *World Englishes* 3: 391-409.
- Ejzenberg R (1992) *Understanding nonnative oral fluency: The role of task structure and discourse variability*. Ann Arbor, MI: University Microfilms International.
- Ellis R (2005) Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies In Second Language Acquisition* 27: 141-172.
- Fayer J M and Krasinski E (1987) Native and nonnative judgments of intelligibility and irritation. *Language Learning* 37: 313-326.



- Field J (2005) Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly* 39: 399-423.
- Flege J E, Munro M J and MacKay I R A (1995) Effects of age of second-language learning on the production of English consonants. *Speech Communication* 16: 1-26.
- Flege J E, Takagi N and Mann V (1996) Lexical familiarity and English - language experience affect Japanese adults' perception of /ɪ/and/ɪ/. *The Journal of the Acoustical Society of America* 99: 1161-1173.
- Foote J A, Holtby A K and Derwing T M (2012). Survey of the teaching of pronunciation in adult ESL programs in Canada, 2010. *TESL Canada Journal* 29: 1-22.
- Foote J A and Trofimovich P (2018) Is It Because of My Language Background? A Study of Language Background Influence on Comprehensibility Judgments. *Canadian Modern Language Review* 74: 253-278.
- Foster P and Skehan P (1996) The influence of planning and task type on second language performance. *Studies in Second Language Acquisition* 18: 299-323.
- Gallardo del Puerto F, García Lecumberri M L and Gómez Lacabex E (2015) The assessment of foreign accent and its communicative effects by naïve native judges vs. experienced non-native judges. *International Journal of Applied Linguistics* 25: 202-224.
- Hahn L D (2004) Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly* 38: 201-223.
- Hopp H and Schmid M (2013) Perceived foreign accent in first language attrition and second language acquisition: The impact of age of acquisition and bilingualism. *Applied Psycholinguistics* 34: 361-394.
- Isaacs T (2008) Towards defining a valid assessment criterion of pronunciation proficiency in non-native English-speaking graduate students. *Canadian Modern Language Review* 64: 555-580.
- Isaacs T and Trofimovich P (2012) Deconstructing comprehensibility: Identifying the linguistic influences on listeners' L2 comprehensibility ratings. *Studies in Second Language Acquisition* 34: 475-505.
- Isaacs T, Trofimovich P and Foote J A (2018) Developing a user-oriented second language comprehensibility scale for English-medium universities. *Language Testing* 35: 193-216.
- Iwashita N, Brown A, McNamara T and O'Hagan S (2008) Assessed levels of second language speaking proficiency: How distinct? *Applied Linguistics* 29: 24-49.
- Jenkins J (1998) Which pronunciation norms and models for English as an International Language?. *ELT Journal* 52: 119-126.
- Jenkins J (2000) *The phonology of English as an international language*. Oxford, UK: Oxford University.
- Jenkins J (2007) *English as a Lingua Franca: Attitude and Identity*. Oxford, UK: Oxford University Press.
- Kanellow V (2009) The practice of pronunciation teaching in current ELT manuals and handbook: a review. *Speak Out: IATEFL Pronunciation Special Interest Group Newsletter* 40: 1-5.
- Kang O (2010) ESL learners' attitudes toward pronunciation instruction and varieties of English. In: Levis J and LeVelle K (eds) *Proceedings of the 1st Pronunciation in Second Language Learning and Teaching Conference*, Iowa State University, Sept. 2009. pp. 105-118, Ames, IA: Iowa State University

- Kang O (2012) Impact of rater characteristics and prosodic features of speaker accentedness on ratings of international teaching assistants' oral performance. *Language Assessment Quarterly* 9: 249-269.
- Kang O and Moran M (2014) Functional Loads of Pronunciation Features in Nonnative Speakers' Oral Assessment. *TESOL Quarterly* 48: 176-187.
- Kang O, Rubin D O N, and Pickering L (2010). Suprasegmental measures of accentedness and judgments of language learner proficiency in oral English. *The Modern Language Journal* 94: 554-566.
- Levis J M (2005) Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly* 39: 369-378.
- Luchini P L and Kennedy S (2013) Exploring sources of phonological unintelligibility in spontaneous speech. *International Journal of English and Literature* 4: 79-88.
- Mackay I R, Flege J E and Imai S (2006) Evaluating the effects of chronological age and sentence duration on degree of perceived foreign accent. *Applied Psycholinguistics* 27: 157-183.
- Major R C (2007) Identifying a foreign accent in an unfamiliar language. *Studies in Second Language Acquisition* 29: 539-556.
- Muñoz C (2008) Symmetries and asymmetries of age effects in naturalistic and instructed L2 learning. *Applied Linguistics* 29: 578-596.
- Munro M J (2010) Intelligibility: Buzzword or buzzworthy?. In: Levis J and LeVelle K (eds) Proceedings of the 2nd Pronunciation in Second Language Learning and Teaching Conference, Sept. 2010, Ames, IA: Iowa State University, pp.7-16.
- Munro M J and Derwing T M (1995) Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech* 38: 289-306.
- Munro M J and Derwing T M (1999) Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning* 49: 285-310.
- Munro M J and 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: 451-468.
- Munro M J and Derwing T M (2006) The functional load principle in ESL pronunciation instruction: An exploratory study. *System* 34: 520-531.
- Munro M J and Derwing T M (2008) Segmental acquisition in adult ESL learners: A longitudinal study of vowel production. *Language Learning* 58: 479-502.
- Munro M J and Derwing T M (2011) The foundations of accent and intelligibility in pronunciation research. *Language Teaching* 44: 316-327.
- Munro M J, Derwing T M and Burgess C S (2010) Detection of nonnative speaker status from content-masked speech. *Speech Communication* 52: 626-637.
- Munro M J, Derwing T M and Morton SL (2006) The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition* 28: 111-131.
- Munro M J, Derwing T M and Thomson R I (2015) Setting segmental priorities for English learners: Evidence from a longitudinal study. *International Review of Applied Linguistics in Language Teaching* 53: 39-60.
- Munro M J and Mann V A (2005) Age of immersion as a predictor of foreign accent. *Applied Psycholinguistics* 26: 311-341.
- Murphy J M (2014) Intelligible, comprehensible, non-native models in ESL/EFL pronunciation teaching. *System* 42: 258-269.

- Ortega L (2009) *Understanding second language acquisition*. London: Hodder Arnold.
- Osimk R (2009) Decoding sounds: An experimental approach to intelligibility in ELF. *Vienna English Working Papers* 18: 64-89.
- Pickering L (2001) The role of tone choice in improving ITA communication in the classroom. *TESOL Quarterly* 35: 233-255.
- Pickering L (2009) Intonation as a pragmatic resource in ELF interaction. *Intercult. Pragmat* 6: 235-255.
- Pitzl M L (2005) Non-understanding in English as a lingua franca: Examples from a business context. *Vienna English Working Papers* 14: 50-71.
- Riney T J, Takada M and Ota M (2000) Segmentals and global foreign accent: The Japanese flap in EFL. *TESOL Quarterly* 34: 711-737.
- Saito K (2015) The Role of Age of Acquisition in Late Second Language Oral Proficiency Attainment. *Studies in Second Language Acquisition* 37: 713-743.
- Saito K and Hanzawa K (2016) Developing Second Language Oral Ability in Foreign Language Classrooms: The Role of The Length and Focus of Instruction and Individual Differences. *Applied Psycholinguistics* 37: 813-840.
- Saito K, Suzukida Y and Sun H (2018) Aptitude Effects on Second Language Pronunciation Proficiency Development in Classroom Settings: A Longitudinal Study. *Studies in Second Language Acquisition*. Advance online publication. doi.org/10.1017/S0272263117000432
- Saito K, Tran M, Suzukida Y, Sun H, Magne V and Ilkan M (in press) 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*.
- Saito K, Trofimovich P and Isaacs T (2015) Using Listener Judgments to Investigate Linguistic Influences on L2 Comprehensibility and Accentedness: A Validation and Generalization Study. *Applied Linguistics* 38: 439-462.
- Saito K, Trofimovich P and Isaacs T (2016) Second Language Speech Production: Investigating Linguistic Correlates of Comprehensibility and Accentedness for Learners at Different Ability Levels. *Applied Psycholinguistics* 37: 217-240.
- Saito K, Webb S, Trofimovich P and Isaacs T (2016) Lexical profiles of comprehensible second language speech: The role of appropriateness, fluency, variation, sophistication, abstractness, and sense relations. *Studies in Second Language Acquisition* 38: 677-701.
- Scales J, Wennerstrom A, Richard D and Wu S H (2006) Language learners' perceptions of accent. *TESOL Quarterly* 40: 715-738.
- Seidlhofer B (2011) *Understanding English as a Lingua Franca*. Oxford: Oxford University Press.
- Skehan P and Foster P (1999) The influence of task structure and processing conditions on narrative retellings. *Language Learning* 49: 93-120.
- Smith L E and Nelson C L (2006) World Englishes and issues of intelligibility. In: Kachru B, Kachru Y and Nelson C (eds) *The Handbook of World Englishes*. Malden, MA: Blackwell, pp. 428-445.
- Tajima K, Port R and Dalby J (1997) Effects of temporal correction on intelligibility of foreign-accented English. *Journal of Phonetics* 25: 1-24.
- Timmis I (2002) Native - speaker norms and International English: a classroom view. *ELT Journal* 56: 240-249.

- Trudgill P and Hannah J (2017) *International English: A guide to varieties of English around the world*. London: Routledge.
- Trofimovich P and 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: 1-30.
- Trofimovich P and Isaacs T (2012) Disentangling accent from comprehensibility. *Bilingualism: Language and Cognition* 15: 905-916.
- Varonis E M and Gass S (1982) The comprehensibility of non-native speech. *Studies in Second Language Acquisition* 4: 114-136.
- Walker R (2010) *Teaching the pronunciation of English as a lingua franca*. Oxford: Oxford University.
- Warren P, Elgort I and Crabbe D (2009) Comprehensibility and prosody ratings for pronunciation software development. *Language Learning and Technology* 13: 87-102.
- Wennerstrom A (2001) Intonation and evaluation in oral narratives. *Journal of Pragmatics* 33: 1183-1206.
- Winters S and O'Brien M G (2013) Perceived accentedness and intelligibility: The relative contributions of F0 and duration. *Speech Communication* 55: 486-507.
- Zielinski B W (2008) The listener: No longer the silent partner in reduced intelligibility. *System* 36: 69-84.