The Efficacy of the Functional Load and L1 Background Principles' Predictions of the Development of Mutual Intelligibility during Interaction

George O'Neal and Lance Latham

This study assesses the predictions of the functional load principle and the L1 background principle as regards the development of pronunciation intelligibility during an interaction. The functional load principle predicts that high functional load phonemic contrasts will be more difficult to adjust into more intelligible variants during an interaction than low functional load phonemic contrasts. The L1 background principle, on the other hand, predicts that L2 phonemic contrasts that can be conflated with L1 phonemes will be more difficult to adjust into more intelligible variants during an interaction. Eleven Chinese-Japanese dyads participated in this experiment, which required them to complete four tests in which they had to adjust phonemic contrasts into more intelligible variants. Each test contained some combination of high and low functional load phonemic contrasts as well as problematic and non-problematic phonemic contrasts. This study hypothesized that the phonemic contrasts that were both high functional load phonemic contrasts and problematic for L2 English speakers from a Chinese or Japanese L1 background would be the most resistant to adjustment into more intelligible variants. The results offer tentative evidence for some of the hypotheses and suggest that interaction's positive effects on intelligibility operate differently among different kinds of phonemic contrasts.

Keywords: Functional Load, Intelligibility, Accommodation, L2 Pronunciation

1. Introduction

This study reexamines a seminal Second Language Acquisition (SLA) theory as it applies to Second Language (L2) English pronunciation development: the interaction hypothesis. Long's (1996) seminal interaction hypothesis states that adult SLA can be facilitated through conversational interaction between non-native English speakers (NNESs) and native English speakers (NESs) because such interactions provide opportunities to notice both incorrect linguistic forms and correct linguistic forms when interactants encounter communication breakdowns. During the collaboration to overcome such communication breakdowns, interactants first notice which linguistic forms are not successful, and then negotiate a more successful linguistic form. In a word, the interaction

hypothesis states that conversational interaction provides both negative and positive evidence as to the efficacy of linguistic forms, which helps learners develop second language abilities (Long, 1996; Gass & Mackey, 2015).

Although a host of SLA research supports Long's (1996) interaction hypothesis (see e.g., Gass & Varonis, 1994; Mackey, 1999; McDonough, 2006; Goo & Mackey, 2013; Saito & Akiyama, 2017), this does not mean that the interaction hypothesis does not warrant a critical reexamination. This is for two main reasons. First, the interaction hypothesis assumes that, or is operationalized in such a way that, interaction with NESs is a prerequisite for the development of language abilities (e.g., Gass & Varonis, 1994; Mackey, 1999; Baker & Trofimovich, 2005; Gass & Mackey, 2015; Saito & Akiyama, 2017). However, reasons abound to doubt that interaction with NESs is the only way to gain the benefits derived from the positive and negative evidence that is provided through interactions designed to overcome miscommunications. This is because NNESs can also provide positive and negative evidence as to the efficacy of linguistic forms, and this efficacy does not even need to be equated to NES linguistic forms either (see e.g., Holliday, 2006; Gilner, 2016; D'Angelo, 2017). Second, there is still a gap within interaction hypothesis research. Experiments have been used to demonstrate that the interaction hypothesis can explain the development of syntax (see e.g., Mackey, 1999), vocabulary (see e.g., Ellis & Sheen, 2006), and even the comprehensibility of pronunciation (see e.g., Saito & Akiyama, 2017), but no research to date has yet explored the interplay among the predictions of the functional load principle, the L1 background principle, the interaction hypothesis as regards the development of intelligible pronunciation during interactions between NNESs. As this study deals with all of these concepts, a detailed explanation of these ideas is warranted before proceeding any further.

2. Previous Literature

2.1. Intelligibility

Pronunciation is often conceptualized as a superordinate phenomenon that encompasses three subordinate phenomena: intelligibility, comprehensibility, and accentedness. Intelligibility is often defined as the extent to which an interlocutor understands the speaker's pronunciation. Comprehensibility, in contrast, is defined as the extent to which the pronunciation is subjectively considered to be easy to understand. Accentedness is defined as the extent to which the pronunciation deviates from a native speaker accent (Derwing & Munro, 2015; Levis, 2018).

Not all aspects of pronunciation are equally studied, and the focus of research tends to differ according to the research paradigm. English as a Lingua Franca (ELF) pronunciation research tends

to focus solely on intelligibility (see e.g., Matsumoto, 2011; Zhang, 2015; Zoghbor, 2018; O'Neal, 2015, 2019; O'Neal & Matsumoto, 2019). SLA pronunciation research, however, tends to focus on comprehensibility (see e.g., Crowther, Trofimovich, Saito, & Isaacs, 2015; Saito & Akiyama, 2017; Levis, 2018). This study assesses the extent to which L2 English speakers understand one another's pronunciation and how this changes due to interaction, and thus the focus of this study is intelligibility. As such, this study does not attempt to ascertain the extent to which pronunciation is considered easy to understand, nor does this study attempt to determine how much pronunciation differs from a NES's accent.

2.2. The Interaction Hypothesis

Experimental research begins from theory, and there is one SLA theory that is particularly applicable to the predictions concerning intelligibility development and interaction among L2 English speakers. Long (1983) hypothesizes that adult SLA can be facilitated through conversational interaction with other NESs because such interactions provide many opportunities to notice so-called incorrect and correct linguistic forms when NES-NNES dyads encounter communication breakdowns. Long (1996) claims that during the collaboration to overcome such communication breakdowns, interactants first notice which linguistic forms are not successful, which yields negative feedback that an unsuccessful linguistic form is both ineffective and not correct, and then the interactants negotiate a more successful linguistic form, which yields positive feedback that the successful linguistic form is both effective and correct.

Long's (1983) interaction hypothesis remained untested until the Gass & Varonis (1994) study, which is the first SLA study to directly test the predictions of the interaction hypothesis. Gass & Varonis (1994) used data gathered from an information-gap task among sixteen NES-NNES dyads in order to assess the relationship between interaction and "communicative success", which was operationalized as placing an object on a board at the location described by one's partner. Half of the NES-NNES dyads were allowed to interact (i.e., ask for clarification, repeat themselves, etc.) and half of the dyads were not allowed to interact. The Gass & Varonis (1994) study is a significant study for two reasons: first, it created a way to experimentally control for the phenomenon of "interaction", which allows for it to be tested through the scientific method; second, it tested and validated the predictions of the interaction hypothesis in the sense that the NES-NNES dyads who were allowed to interact during the task were much more successful at the task than the NES-NNES dyads who were not allowed to interact.

Numerous follow up studies, many of which fully or partially replicated Gass & Varonis's (1994) experimental procedure, further tested and validated the predictions of the interaction

hypothesis. Polio & Gass (1998) replicated a portion of the Gass & Varonis (1994) experiment, which demonstrated the reliability of the original experiment and further validated the predictions of the interaction hypothesis. Both Mackey (1999) and Mackey & Philp (1998) experimentally investigated the relationship among interaction, recasts, and question syntax formation. Both studies yielded positive and significant results, and thus, Mackey & Philp (1998) concluded, that "there does appear to be evidence for a significant relationship between [interaction and] development, as measured by question production and exposure to recasts" (pp. 347-348, brackets added). Saito & Akiyama (2017) directly tested the relationship between interaction and the comprehensibility of pronunciation in a longitudinal study, and the experiment garnered positive and significant results for several linguistic attributes, not just comprehensibility. Thus, in the aggregate, these experiments all provide evidence to support the interaction hypothesis.

The specific forms of positive and negative evidence that the interaction hypothesis predicts for the development of intelligible pronunciation manifest in what is called a segmental repair sequence (Matsumoto, 2011; O'Neal, 2015, 2019). As the following example of a segmental repair sequence will demonstrate, negative evidence as to the unintelligibility of a pronunciation appears when one interactant reacts to a pronunciation with a signal of less than complete understanding. Positive evidence as to the intelligibility of a pronunciation, on the other hand, appears when one of the interactants re-articulates the problematic pronunciation in a new form, to which the other interlocutor orients to as intelligible. In the following example of a segmental repair sequence, Bai, a Chinese exchange student, and Marika, a Japanese undergraduate student, are talking about what they ate for lunch, and Bai mentions that she ate a pastry from the university's bakery (see Appendix 1 for transcription system details). Marika initially displays less than complete understanding of Bai's articulation of "blueberry", which operates as negative evidence. After Bai adjusts her pronunciation into a more intelligible variant, Marika displays that she understandings through the deployment of the "ah" discourse marker, which operates as positive evidence.

Transcript 1: A Segmental Repair Sequence

1	Bai:	it tastes like [ˈblɔːbɔːɻiː].
2		(0.7)
3	Marika:	like [ˈblɔ]?
4	Bai:	li- like [ˈblɔːbɔːɻiː].
5		(0.8)
6	Marika:	wha- what's that?
7		(0.4)

8	Bai:	↑oh
9		(.)
10		[ˈblɯ.bɜː.ɻiː].
11	Marika:	↑ah. [blɯ.bɜː.ɻiː]. {ahhahahahaha}
12	Bai:	{yes. hahahaha}haha.

In lines 3 and 6, Marika orients to Bai's articulation of "blueberry" as unintelligible, which both manifests negative evidence as to the efficacy of the [ˈblɜːbɜː.ɻiː] pronunciation and launches the segmental repair sequence. After Bai seems to realize why her pronunciation is unintelligible in line 8, she adjusts her articulation to [ˈblɯ.bɜː.ɻiː], modifying the vowel in the first syllable from an unrounded rhotacized central vowel to an unrounded high back vowel. In line 11, Marika reacts to this new variant of the pronunciation as intelligible, which shows that Marika and Bai were able to co-construct a phonemic contrast between an unintelligible unit and an intelligible word (O'Neal, 2020). This example demonstrates the efficacy of the predictions of the interaction hypothesis among L2 English speakers: in lines 3 and 6, Marika provides Bai negative evidence, and in response, Bai adjusted her pronunciation in line 10; in line 11, Marika provides positive evidence.

A host of research on the relationship between interaction and the development of intelligibility has demonstrated that the interaction hypothesis adequately predicts that interaction will change unintelligible pronunciations into intelligible pronunciations (Matsumoto, 2011; O'Neal, 2015, 2019). But the interaction hypothesis is not without its own blind spots. One problem with the interaction hypothesis is that its predictions do not change depending on the phonemic contrasts present in the segmental repair sequence. That is, the interaction hypothesis does not predict that interacting to change a /i/ phoneme into a /ɪ/ phoneme would be any more difficult than interacting to change a /k/ phoneme into a /g/ phoneme. However, it seems reasonable to assume that some phonemic contrast adjustments would be more difficult to achieve than others. But for theory-based predictions as to which phonemic adjustments would be better facilitated through interaction and which would be resistant to change even within an interaction, we turn to two principles, both of which make predictions as to which phonemic changes would be conducive to adjustment during an interaction.

2.3. The Functional Load Principle

The Functional Load (FL) principle is the idea that the greater frequency of a phonemic contrast within the vocabulary of a language (i.e., the greater the number of minimal pairs in a language that one pair of phonemes creates) equates to the greater importance of that phonemic contrast to

intelligibility (Brown, 1988; Catford, 1987; Gilner & Morales, 2008, 2010; Munro & Derwing, 2006; Sewell, 2017; Suzukida & Saito, 2019). According to the FL principle, because the phonemic contrast between /i/ and /ɪ/ distinguishes a large number of vocabulary in English (i.e., creates a large number of high frequency minimal pairs such as *sleep/slip* and s*eat/sit*), the phonemic contrast between /i/ and /ɪ/ is more important to intelligibility than the phonemic contrast between /ʊ/ and /u/, which does not distinguish a lot of vocabulary in English (i.e., creates only a few low frequency minimal pairs such as *look/Luke* and *should/shoed*). Thus, according to the FL principle, because the /i/ and /ɪ/ phonemic contrast does more work than the /u/ and /ʊ/ phonemic contrast to keep vocabulary separate, the /i/ and /ɪ/ phonemic contrast has a high FL and the /u/ and /ʊ/ phonemic contrast has a low FL. Furthermore, the FL principle states that this distinction between high and low FL phonemic contrasts correlates to how important they are to the maintenance of intelligibility. In other words, according to the FL principle, substituting either the /i/ or the /ɪ/ phoneme with the other would harm intelligibility far more than substituting either the /u/ or the /ʊ/ phoneme with the other because high FL phonemic contrasts are more important to intelligibility than low FL phonemic contrasts.

The FL principle does not specifically predict which phonemic contrasts would be more difficult to adjust during a segmental repair sequence, but it does clearly predict the consequences of not distinguishing between high FL and low FL phonemic contrasts during an interaction. The FL principle predicts that not differentiating between high FL phonemic contrasts would necessitate segmental repair sequences to reestablish mutual intelligibility. For example, if a Chinese-Japanese dyad does not differentiate between a high FL phonemic contrast like /ɹ/ and /l/, then intelligibility is likely to falter during an interaction and necessitate a segmental repair sequence. In contrast, the FL principle also predicts that not differentiating between low FL phonemic contrasts would not necessitate segmental repair at all because intelligibility is unlikely to falter due to substitutions of low FL phonemic contrasts. Thus, if a Chinese-Japanese dyad does not differentiate between a low FL phonemic contrast like /u/ and /ʊ/, then intelligibility is not likely to falter at all. Accordingly, the FL principle predicts that failing to differentiate high FL phonemic contrasts would necessitate more segmental repair sequences while failing to differentiate low FL phonemic contrasts would not. The consequence of these predictions is that the FL principle predicts that it would be more difficult to adjust high FL phonemic contrasts into more intelligible forms (i.e., requires more segmental repair sequences) while it would be relatively easy to adjust low FL phonemic contrasts into more intelligible forms.

2.4. The L1 Background Principle

The FL principle is not the only principle that predicts that some L2 English phonemic contrasts

will be more difficult to adjust into more intelligible forms during a segmental repair sequence. The L1 Background principle also does this, and it is the idea that certain L2 English phonemic contrasts will be difficult to differentiate because the phonemic categories from the L2 are conflated with similar L1 phonemic categories (Flege, 1987; Baker, Trofimovich, Flege, Mack, & Halter, 2008; Bradlow & Bent, 2008). This study includes Chinese-Japanese dyads, and thus the L1 Background principle has very specific predictions as to which phonemic contrasts will be difficult to distinguish among these dyads. For example, the L1 Background principle predicts that the /ɹ/ and /l/ phonemic contrast will be problematic for L2 English speakers from a L1 Japanese background because these speakers might conflate the L2 English /ɹ/ and /l/ phonemic categories with the /ɾ/ phonemic category from L1 Japanese. Similarly, the L1 Background principle predicts that the /eɪ/ and /ɛ/ phonemic contrast will be problematic for L2 English speakers from a L1 Chinese background because these speakers might conflate the L2 English /eɪ/ and /ɛ/ phonemic categories with the /eɪ/ phonemic category from L1 Chinese.

What needs to be emphasized for this study is that the FL and L1 Background principles make different predictions as regards which phonemic contrasts will be the most difficult to adjust during a segmental repair sequence. As regards L2 English speakers from a L1 Japanese background, both the FL and the L1 Background principles predict that the /ɹ/ and /l/ phonemic contrast will be difficult to adjust into more intelligible forms during a segmental repair sequence, although the reasoning is very different. According to the FL principle, the /ɹ/ and /l/ phonemic contrast is important because this phonemic contrast distinguishes a lot of vocabulary, and thus during an interaction the inability to differentiate it would lower intelligibility and necessitate segmental repair sequences to restore intelligibility. The L1 Background principle, on the other hand, states that the /ɹ/ and /l/ phonemic contrast is difficult for L2 English speakers from a Japanese L1 background to differentiate because it conflates L2 phonemic categories with L1 phonemic categories. However, when it comes to the /θ/ and /s/ phonemic contrast, the FL principle predicts not differentiating these two phonemes will not harm intelligibility because not distinguishing them would not lower intelligibility, but the L1 Background principle predicts that it will be very difficult for L2 English speakers from either a Chinese or a Japanese L1 background to differentiate these two English phonemes because they are conflated with the /s/ phoneme in each language, and thus segmental repair would not be very effective.

This study assesses the predictions of the FL and L1 background principles as regards the development of mutually intelligible pronunciation among interacting Chinese-Japanese dyads of L2 English speakers. As such, this study attempts to answer the following research questions: 1) Does the FL principle predict which L2 English phonemes will be more difficult to adjust into more

intelligible variants during an interaction among Chinese-Japanese dyads of L2 English speakers? 2) Does the L1 background principle predict which L2 English phonemes will be more difficult to adjust into more intelligible variants during an interaction among Chinese-Japanese dyads of L2 English speakers?

In order to assess these research questions, the following three hypotheses were formulated:

H_1: During a segmental repair sequence, it will be easiest to increase the mutual intelligibility of phonemic contrasts that are both low FL phonemic contrasts and non-problematic for L2 English speakers from a Chinese or Japanese L1 background.

H_2: During a segmental repair sequence, it will be next easiest to increase the mutual intelligibility of phonemic contrasts that are either 1) high functional load contrasts but non-problematic for L2 English speakers from a Chinese or Japanese L1 background or 2) that are low functional load contrasts but problematic for L2 English speakers from a Chinese or Japanese L1 background.

H_3: During a segmental repair sequence, it will be most difficult to increase the mutual intelligibility of phonemic contrasts that are both high functional load contrasts and problematic for L2 English speakers from a Chinese or Japanese L1 background.

3. Methodology

3.1. Participants

The participants in this study (N = 22) ranged in age from 29 to 19 with a Mean age of 21.5 years (SD = 2.99 years). They had begun English study at an average of 11 years of age (SD = 6.45 years). The participants were recruited through fliers that were posted across the campus of a large public Japanese university. The participants are either Japanese undergraduate students (N = 11) or Chinese exchange or graduate students (N = 11). A majority of the participants were female (female N = 18; male N = 4). All participants reported at least one measure of English proficiency, all of which were converted to equivalent TOEIC scores. The participants equivalent TOEIC scores ranged from 870 points to 400 points with a Mean of 631.36 points (SD = 108.93 points), and thus it can be said that the participants in this experiment in general have an intermediate level of proficiency in English. All participants reported that they did not know their partners before the experiment.

3.2. Experimental Design

The design of this experiment follows Gass & Varonis (1994), but the task is extensively modified so as to make the relationship between interaction and mutual intelligibility observable. The experiment utilized minimal pairs rather than pictures so that participants had to create intelligible phonemic contrasts to successfully complete the experimental tests. The specific minimal pairs that were used were taken from publicly available lists of minimal pairs (see Nilsen & Nilsen, 2010).

Because this experiment attempts to ascertain the extent to which interaction will make high/low functional load phonemic contrasts as well as problematic/non-problematic phonemic contrasts more mutually intelligible during a segmental repair sequence, we designed four experimental tests. In order to differentiate between high and low FL phonemes, following Munro & Derwing (2006), we decided to divide high and low FL contrasts at the 50% mark in Catford's (1987) rankings. Phonemic contrasts above 50% are considered to be high FL phonemic contrasts, and contrasts below 50% are considered to be low FL phonemic contrasts. In order to differentiate between problematic and non-problematic phonemic contrasts, we adopted Nilsen & Nilsen's (2010) categories of problematic and non-problematic phonemes for L2 English speakers from either a Chinese and Japanese L1 background.

The set of phonemic contrasts that dyads have to intelligibly convey to each other in order to successfully complete each of the four tests are different. Test 1 assesses how well Chinese-Japanese dyads can create mutually intelligible English phonemic contrasts during an interaction that are both high functional load contrasts and problematic for L2 English speakers from a Chinese or Japanese L1 background (see Appendices 2 and 3 for the worksheets used during Test 1). In order to succeed on Test 1, Chinese-Japanese dyads have to create mutually intelligible phonemic contrasts between minimal pairs that include /ɹ/ and /l/, /n/ and /l/, /ɑ/ and /ɔ/, /ɛ/ and /eɪ/, /s/ and /ʃ/, /f/ and /h/, /d/ and /l/, /æ/ and /ɑ/, /i/ and /ɪ/, /æ/ and /ʌ/, as well as /ɛ/ and /æ/. Both the FL principle and L1 background principle predict that it will be difficult for a Chinese-Japanese dyad to adjust these phonemic contrasts into more intelligible variants during a segmental repair sequence, and thus we expect that the intelligibility score on this test will be relatively low on test 1.

Test 2, on the other hand, assesses how well Japanese-Chinese dyads can create mutually intelligible phonemic contrasts during an interaction that are high FL phonemic contrasts but non-problematic for L2 English speakers from a Chinese or Japanese L1 background (see Appendices 4 and 5 for the worksheets used during Test 2). In order to succeed on Test 2, Chinese-Japanese dyads have to create mutually intelligible phonemic contrasts between minimal pairs that include /k/ and /h/, /k/ and /p/, /d/ and /n/, /æ/ and /ɪ/, /p/ and /t/, /s/ and /h/, /t/ and /s/, /p/ and /f/, /k/ and /t/, /g/ and

/b/, /t/ and /d/, /ɪ/ and /ɛ/, as well as /ʊ/ and /ɔ/. Although the FL principle predicts that a Chinese-Japanese dyad will have difficulty adjusting these phoneme contrasts more intelligible variants during a segmental repair sequence, the L1 background principle predicts that it would be relatively easy to adjust these phonetic contrasts into more intelligible variants during a segmental repair sequence, and thus we expect that the intelligibility score on test 2 will higher than on test 1.

Test 3 assesses how well Chinese-Japanese dyads can create mutually intelligible phonemic contrasts during an interaction that are low FL phonemic contrasts but nonetheless problematic for L2 English speakers from a Chinese or Japanese L1 background (See Appendices 6 and 7 for the worksheets used during Test 3). In order to succeed on Test 3, Chinese-Japanese dyads have to create mutually intelligible phonemic contrasts between minimal pairs that include /f/ and /v/, /θ/ and /s/, /s/ and /z/, /u/ and /ʊ/, /ð/ and /z/, /d/ and /z/, /ð/ and /d/, /θ/ and /t/, /d/ and /d͡ʒ/, /θ/ and /s/, /n/ and /ŋ/, /ɛ/ and /ʌ/, as well as /ʌ/ and /ʊ/. The L1 background principle predicts that Chinese-Japanese dyad will have trouble adjusting these phonemic contrasts into more intelligible variants during a segmental repair sequence, but the FL principle, on the other hand, does not predict that Chinese-Japanese dyads will have problems adjusting these phonemic contrasts into more intelligible variants during a segmental repair sequence, and thus we expect that the intelligibility score on test 3 will higher than on test 1.

Test 4 assesses how well a Chinese-Japanese dyad can create mutually intelligible phonemic contrasts during an interaction that are both low FL contrasts and non-problematic for L2 English speakers from a Chinese or Japanese L1 background (see Appendices 8 and 9 for the worksheets used during Test 4). In order to succeed on Test 4, Chinese-Japanese dyads have to create mutually intelligible phonemic contrasts between minimal pairs that include /t/ and /t͡ʃ/, /d͡ʒ/ and /j/, /d͡ʒ/ and /t͡ʃ/, /ɛ/ and /ɑ/, /k/ and /g/, /f/ and /θ/, /v/ and /z/, /v/ and /θ/, /aɪ/ and /aʊ/, /aɪ/ and /ɔɪ/, /ʊ/ and /ɑ/, as well as /ɔ/ and /u/. Neither the FL principle nor the L1 background principle predict that a Chinese-Japanese dyad will have trouble adjusting these phonemic contrasts into more intelligible variants during a segmental repair sequence, and thus we expect that the intelligibility score on test 4 will be the highest of all four tests.

3.3. Experimental Procedure

One Japanese participant was assigned to one Chinese participant to create eleven dyads, and thus all dyads have different L1 backgrounds. The pairing of a Japanese participant to a Chinese participant was done by convenience (i.e., according to the times that the participants were available to do the experiment). However, each dyad was randomly assigned to their test order (which of the four tests was first) and role order (who was the instructor first and who was the listener first) through

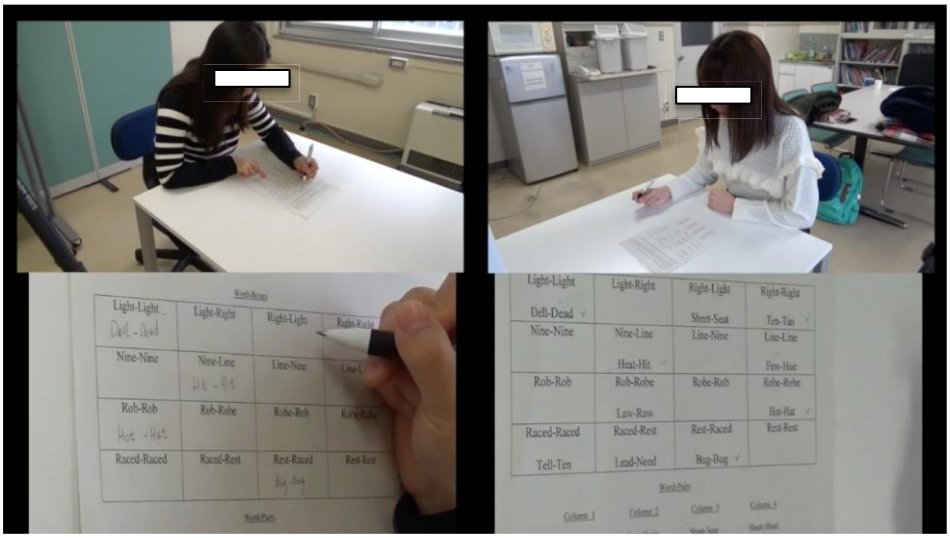
a lot-drawing.

Once the test and role orders were determined the Chinese-Japanese dyad was brought to the testing room and sat at opposite ends of a long table. A large cardboard box was placed in the middle to the table so as to obstruct the view of the participants and thus neither participant could see the other one, although they were close enough that they could easily hear each other. This was done in order to remove gestures and facial expressions as potential explanations for the results.

After the dyads sat at their respective positions at the table, the dyad practiced the experimental task twice with two practice tests: in the first practice test, one of the dyad was the instructor and the other was the listener; in the second practice test, the roles were reversed. After the practice tests were completed, the instructor and listener conducted the randomly determined first of the four tests. The instructor was handed the instructor worksheet for the designated test. The instructor was informed that he or she was to instruct the listener where to write ten minimal pairs on the listener's worksheet. The instructor was told to use the phrase: "Write (minimal pair) in (minimal pair in the word box)" to notify the listener where to write a minimal pair. For example, the instructor could state "Write 'Seat-Sheet' in 'Light-Right'". The instructor was informed that he or she can repeat himself or herself if he or she thinks that the listener does not understand. The instructor was also told that if the listener asks for clarification, then he or she can either repeat the phrase or just say "yes" and "no".

The listener, on the other hand, was handed the listener worksheet for the matching test. The listener was informed that he or she was to listen to the instructor, and write the minimal pairs into the minimal pair word boxes on his or her worksheet. The listener was also told that he or she can confirm the instructor's pronunciation by asking "Should I write (minimal pair) in (minimal pair word box)?" or just "please repeat." These instructions are specifically designed to limit the possible communication strategies to just a segmental repair sequence if there is a mutual intelligibility problem. There was no time limit nor was there any restriction on the number of times that a dyad could attempt to utilize a segmental repair sequence to overcome an intelligibility breakdown.

Figure 1: Experimental Procedure



After the dyad had completed the first test, the instructor and listener roles were exchanged for the next test. Once the instructions were repeated to the new instructor and the new listener, the dyad completed the second test. After the dyad completed the second test, the roles were again exchanged for the third test. The instructions were repeated and then the dyad completed the third test. After the dyad completed the third test, the roles were once again exchanged for the fourth test, after which the dyad completed the fourth test.

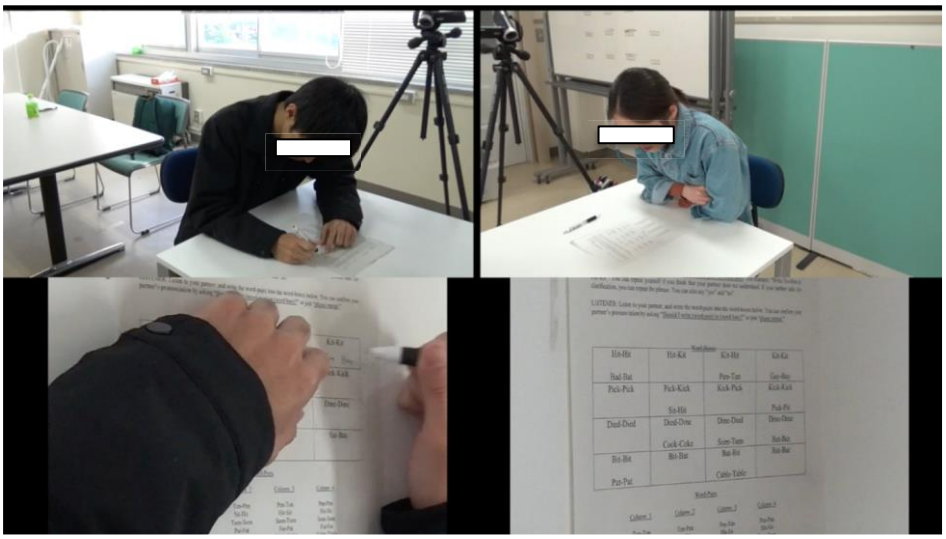
Examples of the segmental repair sequences that appeared during the tests are warranted here. In the following example, a Chinese-Japanese dyad is doing test 2. The Chinese instructor, Zhan, is attempting to convey that the "gay-bay" minimal pair is in the "kit-kit" minimal pair word box to the Japanese listener, Masaki. However, intelligibility falters, and a segmental repair sequence begins to restore it.

Transcript 2: Segmental Repair Sequence between Zhan and Masaki

1	Zhan:	please write [keɪ̲] [beɪ̲].
2		(0.3)
3		in [kɪt] [kɪt].
4		(4.5)
5	Masaki:	please repeat.
6		(2.1)
7	Zhan:	write [geɪ̲].

8		(0.4)
9		[beɪ̲].
10		(0.7)
11		in [kɪt].
12		(0.3)
13		[kɪt].
14		(5.3)
15	Masaki:	okay.

Figure 2: Zhan and Masaki doing test 2



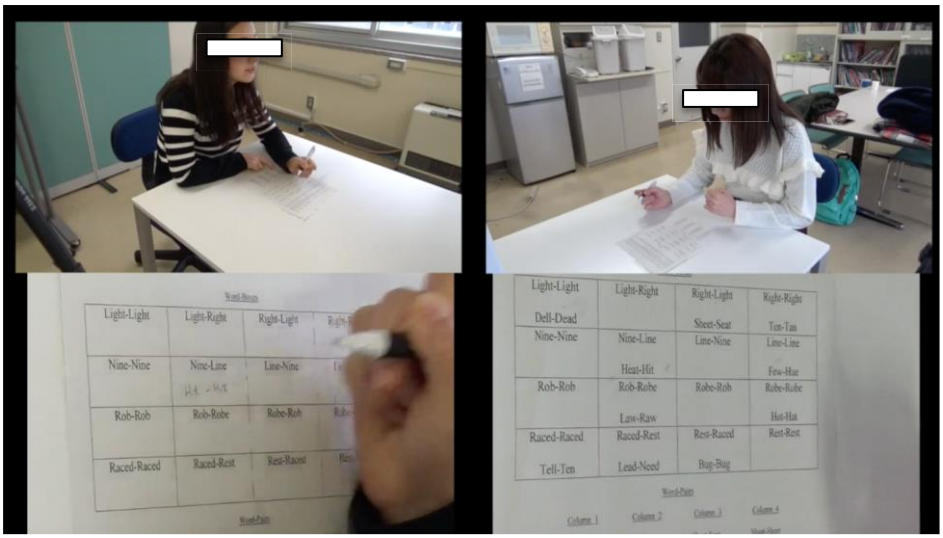
In line 5, Masaki provides negative evidence that something about Zhan's pronunciation is unintelligible. Between lines 7 and 13, Zhan adjusts her pronunciation. In response, in line 15, Masaki provides positive evidence that the new pronunciation is intelligible, and Masaki successfully writes "gay-bay" in the "kit-kit" box. Accordingly, it can be surmised that Zhan and Masaki were able to differentiate the /g/ and /b/ phonemic contrast and well as the /k/ and /g/ phonemic contrast, and that this was due to the segmental repair sequence.

However, not all segmental repair sequences led to successful outcomes, as the following example demonstrates. In the next example, Rei and Yan are doing test 1. Rei is attempting to convey that the "heat-hit" minimal pair is located in the "nine-line" minimal pair box to Yan. In spite of their efforts, however, the Yan writes the "hit-hit" minimal pair in the "nine-line" minimal pair box.

Transcript 3: Rei and Yan

1	Rei:	write [hi:t] [hit] in.
2		(0.3)
3		in: [naɪ̯]-
4		(0.4)
5		[naɪ̯n] [ɭaɪ̯n].
6		(4.1)
7	Yan:	should I write [hɪt] [hɪt].
8		(0.1)
9		in [naɪ̯n] [laɪ̯n] box?
10		(0.3)
11	Rei:	yes.
12	Yan:	okay.

Figure 3: Rei and Yan doing test 1



Between lines 1 to 5, Rei attempts to convey to Yan that the "heat-hit" minimal pair is located in the "nine-line" minimal pair box. But in line 7, Yan provides negative feedback as to the intelligibility of Rei's pronunciation. In line 12, Yan provides positive feedback and confirms Rei's pronunciation. However, although Yan does write a minimal pair in the correct location, she does not write the correct minimal pair, and thus it can thus be inferred that although this Rei and Yan were able to differentiate the /n/ and /l/ phonemic contrast, they were not able to differentiate the /i/ and

/ɪ/ phonemic contrast.

4. Results

The dependent variable is the number of minimal pairs that the listener writes on his or her worksheet that matches the locations of the minimal pairs on the instructor's worksheet. Mutual intelligibility is thus operationalized as the number of matching minimal pairs on both the instructor's and the listener's worksheets. Accordingly, a Chinese-Japanese dyad could be assigned a mutual intelligibility score from 0 (none of the minimal pairs are in the same boxes) to 10 (all of the minimal pairs are in the same boxes). This score is interpreted to be a ratio scale measurement of the mutual intelligibility of the dyad's pronunciation, the dyad's ability to create mutually intelligible phonemic contrasts, and the dyad's ability to adjust their pronunciation during the interaction.

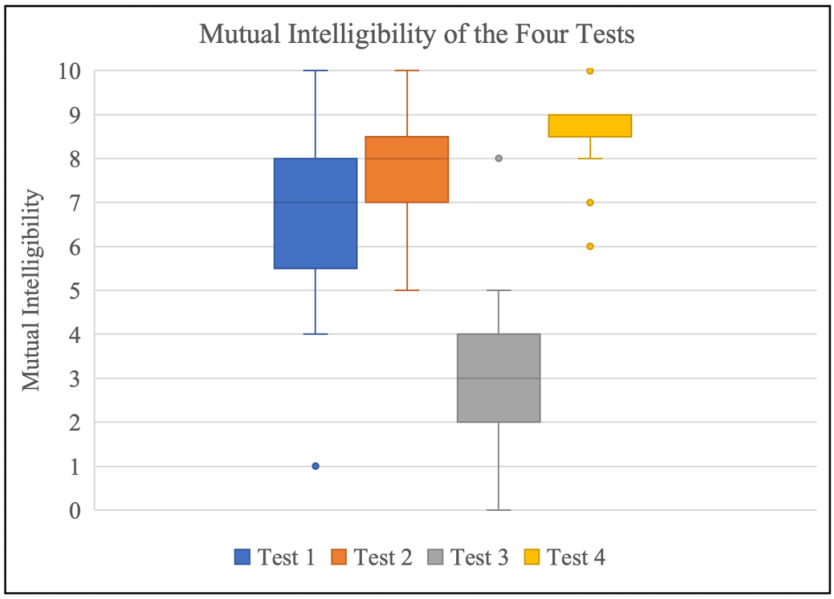
4.1. Descriptive Statistics

Dyads doing Test 4 had the highest average mutual intelligibility scores (M = 8.55; Mdn = 9). Dyads doing Test 2 had the second highest average mutual intelligibility scores (M = 7.82; Mdn = 8). Dyads doing Test 1 had the third highest average mutual intelligibility scores (M = 6.55; Mdn = 7). Dyads doing Test 3 had the lowest average mutual intelligibility scores (M = 3.08; Mdn = 3). These scores are reflected in Table 1 and Figure 1 below.

Table 1: Descriptive Statistics

	Test 1	Test 2	Test 3	Test 4
Minimum	1	5	0	6
Quartile 1	5.5	7	2	8.5
Median (Quartile 2)	7	8	3	9
Quartile 3	8	8.5	4	9
Maximum	10	10	8	10
Mean	6.55	7.82	3.09	8.55
Range	9	5	8	4
Interquartile Range	2.5	1.5	2	0.5

Figure 4: Box-whisker plots



4.2. Statistical Analyses

Initial omnibus statistical analysis and subsequent post-hoc analyses were conducted via SPSS v.25, with the intention of conducting a one-way repeated measures ANOVA on the data. However, because the sample size of Chinese-Japanese dyads is small (N = 11), the assumptions of a parametric one-way repeated measures ANOVA test cannot be met. The initial plan for this experiment was to gather data until 30 samples of data were procured so that the data could be examined through a repeated measures ANOVA. However, after the COVID-19 pandemic began, the researchers made the principled decision to end the experiment because the experimental procedure required putting participants in close proximity with each other in a small room. The researchers further decided to not wait until after the COVID-19 pandemic ended to restart data collection because it could be many years before the pandemic truly ends.

Accordingly, a non-parametric Friedman's test was used to examine the data set instead. The data meets the assumptions of a Friedman's test: 1) the data represents the scores of the same dyads four times, and thus the data are related; 2) although the sample of Chinese-Japanese dyads is not completely random, it is roughly representative of the Chinese and Japanese student population at the public Japanese university; 3) the Median (*Mdn*) mutual intelligibility scores are being used to compare the differences among the four tests rather than the Mean scores, so the data is ordinal rather than interval.

The initial Friedman's test indicated that there was a statistically significant difference in median

(*Mdn*) mutual intelligibility among the four tests, χ^2 (3, N = 11) = 20.43, $p < .001$. Six Wilcoxon signed-rank tests were conducted with a Bonferroni correction applied to the alpha level, which results in a post-hoc test alpha level of .0083. The first Wilcoxon signed-rank test indicated that there was not a significant difference between mutual intelligibility in test 1 (*Mdn* = 7.0) and test 2 (*Mdn* = 8.0), *z* = -1.61, *p* = .107, *r* = .34. The second Wilcoxon signed-rank test indicated that mutual intelligibility in test 1 (*Mdn* = 7.0) was significantly higher in test 3 (*Mdn* = 3.0), *z* = -2.69, *p* = .007, *r* = .57. Furthermore, the effect size indicates that this difference in mutual intelligibility is large. For interpreting the effect size *r*, we refer to the following scale: *r* = .10~.29 (a small difference in mutual intelligibility), *r* = .30~.49 (a medium difference in mutual intelligibility), *r* = .50~1.0 (a large difference in mutual intelligibility). The third Wilcoxon signed-rank test indicated that there was not a significant difference between mutual intelligibility in test 1 (*Mdn* = 7.0) and test 4 (*Mdn* = 9.0), *z* = -2.09, *p* = .036, *r* = .44. The fourth Wilcoxon signed-rank test indicated that mutual intelligibility in test 2 (*Mdn* = 8.0) was significantly higher in test 3 (*Mdn* = 3.0), *z* = -2.94, *p* = .003, *r* = .62. Furthermore, the effect size indicates that this difference in mutual intelligibility is large. The fifth Wilcoxon signed-rank test indicated that there was not a significant difference between mutual intelligibility in test 2 (*Mdn* = 8.0) and test 4 (*Mdn* = 9.0), *z* = -1.14, *p* = .253, *r* = .24. The sixth Wilcoxon signed-rank test indicated that mutual intelligibility in test 4 (*Mdn* = 9.0) was significantly higher in test 3 (*Mdn* = 3.0), *z* = -2.80, *p* = .005, *r* = .59. Furthermore, the effect size indicates that this difference in mutual intelligibility is large. These results are displayed below in Table 3.

Table 2: Post Hoc Test Statistical Results

Post Hoc Test	Comparison	*Mdn*	*z*	*p*	*r*
1	Test 1	7	-1.61	0.107	0.34
	Test 2	8			
2	Test 1	7	-2.69	0.007	0.57
	Test 3	3			
3	Test 1	7	-2.09	0.036	0.44
	Test 4	9			
4	Test 2	8	-2.94	0.003	0.62
	Test 3	3			
5	Test 2	8	-1.14	0.253	0.24
	Test 4	9			
6	Test 3	3	-2.8	0.005	0.59
	Test 4	9			

5. Discussion

The results of this experiment offer tentative support for some of the hypotheses. First, the results support H_1. The phonemic contrasts in Test 4, which were low FL phonemic contrasts that were also non-problematic for L2 English speakers from a Chinese or Japanese L1 background, were the most intelligible, and so this result supports H_1. The results show mixed support for H_2. This hypothesis stated that phonemic contrasts that were mixes of low/high FL phonemic contrasts and problematic/non-problematic phonemic contrasts would be the next most mutually intelligible. Although this was true of Test 2, it was not true of Test 3. Last, the results do not support H_3. This hypothesis predicted that the phonemic contrasts in Test 1 would be the least intelligible, but this was not true. The phonemic contrasts in Test 3 were the least mutually intelligible.

However, the results of this study have interesting implications beyond support for the hypotheses. The results of this study suggest that the intelligibility of some phonemic contrasts is enhanced through interaction while other phonemic contrasts seem to be resistant to it. First, the results seem to indicate that interaction facilitates the development of intelligibility among the phonemic contrasts present in Test 1, 2, and 3. This interpretation is based on the fact that there was no statistically significant difference in intelligibility in Test 1, Test 2, and Test 4 and because the average intelligibility scores in all three of these tests approximate the ceiling of possible scores. Interaction increased the intelligibility of these three sets of phonemic contrasts in a similar way. Second, the results seem to indicate that the phonemic contrasts in Test 3 are resistant to the ameliorative effect of interaction. The intelligibility of the phonemic contrasts in Test 3 were consistently lower than the mutual intelligibility of the phonemic contrasts in Test 1, test 1, and Test 4 to a statistically significant degree. Furthermore, the effect size is consistently large, which indicates that the mutual intelligibility of Test 3 is dramatically lower than the other three tests. These facts suggest that low FL phonemic contrasts that are also problematic for L2 English speakers from a Chinese or Japanese L1 background are resistant to interaction.

These results have interesting implications for the interaction hypothesis. The interaction hypothesis predicts that intelligibility should increase due to interaction in all four tests, but these results do not support that prediction. Indeed, the results suggest that interaction works in different ways across phonemic contrasts. The predictions of the interaction hypothesis seem to be valid in the results of those Tests 1, 2 and 4. However, interaction does not seem to facilitate the development of intelligible phonemic contrasts in Test 3; interaction alone was not enough to facilitate the distinction between the minimal pairs in this test. Accordingly, researchers interested in the interaction hypothesis need to be open to the possibility that the beneficial effects of interaction do not manifest

equally across all aspects of language. Indeed, the results of this experiment suggest that L2 English phonemic contrasts that could be conflated with L1 phonemes are likely somewhat resistant to the ameliorative effect of interaction.

These results have interesting implications for the Functional Load and L1 Background Principles as well. The results suggest that the Functional Load Principle does not tell us much about how interaction would affect pronunciation development. The Functional Load Principle is a universal principle, and predicts that mutual intelligibility is affected by the frequency of phonemic contrasts in English vocabulary. However, the results of this study suggest that the L1 backgrounds of the speakers matter a lot to how interaction affects the development of intelligible phonemic contrasts: low functional load phonemes that were also problematic for L2 English speakers from a Chinese or Japanese L1 background were the most intractable to the ameliorative effect of interaction.

Last, every study has limitations, and this one in particular has several that could be amended in a replication study. First, and most obviously, the sample size of the experiment is small, and a larger sample size could reveal more about the complex relationship between intelligibility and interaction, but data collection for this experiment was cancelled due to the COVID-19 pandemic. Thus, although the researchers believe that cancelling data collection was a socially responsible decision, it is nonetheless true that a larger sample size of Chinese-Japanese dyads might have revealed something different. Second, because the sample size is small, the researchers had to use a less powerful non-parametric test to analyze the data set. It is more than fair to say that a more powerful statistical test, like the repeated measures ANOVA, might have detected differences that the non-parametric statistical tests did not. Third, this experiment only used Chinese-Japanese dyads, and thus it could be that different pairings of L1 backgrounds might reveal different results. In spite of these limitations, however, this experiment is still the first to demonstrate that interaction does not automatically lead to more intelligible pronunciation across all phonemic contrasts.

6. Conclusion

L2 phonology and its interactional development have a fascinating relationship. Although it is true that interaction leads to more mutually intelligible pronunciation, this does not mean that interaction affects the mutual intelligibility of all phonemic contrasts the same way. This study is the first to suggest that interaction's positive effects on mutual intelligibility operate differently across different kinds of phonemic contrasts, and it is hoped that this will open up a new research avenue within the interaction hypothesis paradigm.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number JP19K13220.

References

Baker, W., & Trofimovich, P. (2005). Interaction of native- and second-language vowel system(s) in early and late bilinguals. *Language and Speech, 48*(1), 1-27.

Baker, W., Trofimovich, P., Flege, J. E., Mack, M., & Halter, R. (2008). Child-adult differences in second-language phonological learning: The role of cross-language similarity. *Language and Speech*, *51*(4), 317-342.

Bradlow, A. R., & Bent, T. (2008). Listener adaptation to non-native speech. *Cognition*, 106(2), 707-729.

Brown, A. (1988). Functional load and the teaching of pronunciation. *TESOL Quarterly*, 22(4), 593–606.

Catford, J. C. (1987). Phonetics and the Teaching of Pronunciation. In J. Morley (Ed.), *Current perspectives on pronunciation: Practices anchored in principle* (pp. 83–100). Washington, DC: TESOL.

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.

D'Angelo, J. (2017). The status of ELF in Japan. In J. Jenkins, W. Baker, & M. Dewey (Eds.), *The Routledge handbook of English as a lingua franca* (pp. 165-175). London/New York: Routledge.

Derwing, T., & Munro, M. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Ellis, R., & Sheen, Y. (2006). Reexamining the role of recasts in second language acquisition. *Studies in Second Language Acquisition, 28*, 575-600.

Flege, J. E. (1987). The production of "new" and "similar" phones in a foreign language: evidence for the effect of equivalence classification. *Journal of Phonetics*, *15*, 47-65.

Gass, S. & Mackey, A. (2015). Input, interaction, and output in second language acquisition. In B. VanPatten & J. Williams (eds.), *Theories in second language acquisition: An introduction*, 2nd ed. (pp. 180-206). New York: Routledge.

Gass, S., & Varonis, E. (1994). Input, Interaction, and Second Language Production. *Studies in Second Language Acquisition, 16*, 283-302.

Gilner, L. (2016). Identification of a dominant vocabulary in ELF interactions. *Journal of English as a Lingua Franca, 5*(1), 27-51.

Gilner, L., & Morales, F. (2008). Elicitation and application of phonetic description of the General Service List. *System, 36*, 517-533.

Gilner, L., & Morales, F. (2010). Functional load: Transcription and analysis of the 10,000 most frequent words in spoken English. *The Buckingham Journal of Language and Linguistics, 3*, 135–162.

Goo, J., & Mackey, A. (2013). The case against the case against recasts. *Studies in second language acquisition, 35*, 127-165.

Holliday, A. (2006). Native-speakerism. *ELT Journal, 60*(4), 385-387.

Levis, J. (2018). *Intelligibility, Oral Communication, and the Teaching of Pronunciation*. Cambridge: Cambridge University Press.

Long, M. (1983). Native speaker/non-native speaker conversation and the negotiation of comprehensible input. *Applied Linguistics, 4*, 126-141.

Long, M. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of language acquisition, Vol. 2: Second language acquisition* (pp. 413-468). New York: Academic Press.

Mackey, A. (1999). Input, interaction, and second language development. *Studies in Second Language Acquisition, 21*, 557-587.

Mackey, A., & Philp, J. (1998). Conversational interaction and second language development: Recasts, responses, and red herrings? *Modern Language Journal, 82*, 338-356.

McDonough, K. (2006). Interaction and syntactic priming. *Studies in Second Language Acquisition, 28*, 179-207.

Matsumoto, Y. (2011). Successful ELF communications and implications for ELT: Sequential analysis of ELF pronunciation negotiation strategies. *The Modern Language Journal, 95*, 97–114.

Munro, M. J., & Derwing, T. M. (2006). The functional load principle in ESL pronunciation instruction: An exploratory study. *System, 34*(4), 520–531.

Nilsen, D., & Nilsen, A. (2010). *Pronunciation contrasts in English*, 2nd ed. Long Grove: Waveland Press.

O'Neal, G. (2015). Segmental repair and interactional intelligibility: The relationship between consonant deletion, consonant insertion, and pronunciation intelligibility in English as a Lingua Franca in Japan. *Journal of Pragmatics, 85*, 122-134.

O'Neal, G. (2019). The accommodation of intelligible segmental pronunciation: Segmental repairs and adjustments in English as a lingua franca interactions. *Journal of Second Language Pronunciation, 5*(1), 119-138.

O'Neal, G. (2020). Does an ELF phonology exist? *Asian Englishes, 22*(3), 282-296.

O'Neal, G., & Matsumoto, Y. (2019). Beyond intelligibility: Negotiating transintelligibility in English as a lingua franca interactions. *International Journal of Applied Linguistics, 29*, 44-60.

Polio, C., & Gass, S. (1998). The role of interaction in NS comprehension of non-NS speech. *Modern Language Journal, 82*, 308-319.

Saito, K., & Akiyama, Y. (2017). Video-based interaction, negotiation for comprehensibility, and second language speech learning: A longitudinal study. *Language Learning, 67*(1), 43-74.

Sewell, A. (2017). Functional load revisited: Reinterpreting the findings of 'lingua franca' intelligibility studies. *Journal of Second Language Pronunciation, 3*(1), 57-79.

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*. Early Access.

Zhang, L. (2015). An empirical study of the intelligibility of English spoken by Chinese university students. *Chinese Journal of Applied Linguistics, 38*(1), 36-54.

Zoghbor, W. S. (2018). Teaching pronunciation to multi-dialect first language learners: The revival of the Lingua Franca Core (LFC). *System, 78*, 1-14.

Appendix 1: Transcription Symbols

Symbol	Represents
{ }	simultaneous speech
[ ]	speech in IPA phonetic transcription
/ /	speech in IPA phonemic transcription
(.)	micro silence (e.g., less than one tenth of a second of silence)
(1.5)	a timed silence (e.g., a one and a half seconds of silence)
:	elongated sound
.	falling intonation
,	slightly rising intonation
?	rising intonation
↑	sudden volume and/or pitch increase