

RE-EXAMINING FUNCTIONAL LOAD IN LIGHT OF RATERS' PERCEPTION OF ERROR
GRAVITY IN SECOND LANGUAGE SPEECH

By
Adam Pfau

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Teaching English to Speakers of Other Languages – Master of Arts

2020

ABSTRACT

RE-EXAMINING FUNCTIONAL LOAD IN LIGHT OF RATERS' PERCEPTION OF ERROR GRAVITY IN SECOND LANGUAGE SPEECH

By

Adam Pfau

The current study looks at the role FL values has concerning the perception that listeners have of the intelligibility and comprehensibility of unclear phonemes in second language speech. A listener's familiarity with accented speech is also considered. Native English listeners from two separate populations – a student population exposed to a variety of second language speech and a community member population with little exposure to accented speech – were presented with recorded speech samples from L1 Japanese speakers of English. The speech samples contained unclear, or non-native like, examples of two separate phoneme pairs: the /r-l/ consonant contrast and the /s-θ/ contrast. The first carries a very high functional load value, while the second carries a very low value. Listeners responded to a comprehensibility and intelligibility task that contained examples of both target contrasts. The results indicated that the student population found the speech samples more intelligible and easier to comprehend. The /r-l/ contrast, with a higher FL value, was more difficult for them to transcribe and comprehend than the /s-θ/ but these differences were less pronounced for the community member raters. This suggests that, while teachers may be wise to use FL values as a basis for a pronunciation syllabus or instruction, they should be aware that the FL values do not paint the whole picture concerning how listeners respond to errors made in second language speech.

TABLE OF CONTENTS

LIST OF TABLES	iv
1. Introduction.....	1
2. Literature Review.....	6
2.1 Intelligibility.....	6
2.2 Comprehensibility.....	9
2.3 The Current Study.....	13
3. Methodology.....	16
3.1 Participants.....	16
3.1.1 L1 Japanese L2 English Speakers.....	16
3.1.2 Native English-Speaking Listeners.....	17
3.1.2.1 Listener Population #1: Student Listeners.....	18
3.1.2.2 Listener Population #2: Community Member Listeners.....	18
3.2 Materials.....	19
3.2.1 Recorded Speech Samples.....	19
3.2.2 FL Calculations.....	20
3.3 Procedure.....	22
3.3.1 Task #1: Intelligibility.....	22
3.3.2 Task #2: Comprehensibility.....	23
4. Results.....	25
4.1 Data Analysis.....	25
4.1.1 Comprehensibility Task.....	25
4.1.3 Intelligibility Task.....	27
4.1.3 Correlation Between the Two Tasks.....	31
5. Discussion.....	33
5.1 Research Question #1.....	33
5.2 Research Question #2.....	35
5.3 Research Question #3.....	37
5.4 Pedagogical Implications.....	40
APPENDIX.....	43
BIBLIOGRAPHY	45

LIST OF TABLES

Table 1: Intelligibility Task. This table shows data collected from raters' on their ability to transcribe target sounds upon hearing them.....	26
Table 2: Comprehensibility Task, Section One. This table displays data collected from raters on a comprehensibility task using a nine-point scale.....	29
Table 3: Comprehensibility Task, Section Two. This table displays data collected from raters on their perception of difficulty in comprehending target sounds.....	30
Table 4: Comprehensibility and Intelligibility Correlation. This table displays correlation data between raters' intelligibility and comprehensibility results.....	31
Table 5: Functional Load Values For Consonant Contrasts. This table shows all of the calculated functional load values for all possible consonant contrasts, as calculated using Phonological CorpusTools.....	44

1. Introduction

Though Adam Brown wryly remarks that “no two linguists will agree on what functional load should measure, or how” (Brown, 1988, 594), the concept of *functional load*, shortened to FL for future reference, customarily refers to the degree of contrast that exists between units of language, such as phonemes (King, 1967). A phoneme is the abstract concept that represents a sound in a language that can serve a contrastive function; for example, in the minimal pair rock – lock, /r/ and /l/ are contrastive phonemes. In this regard, FL refers to the degrees of importance that various phonemes carry in creating meaningful distinctions within a language, traditionally calculated through a counting of minimal pairs. If FL has implications for speech processing, a high FL should make it difficult for a listener to identify a phoneme that is ambiguous due to either noise, its omission, or it being produced in a non-native like fashion.

Though FL needs to be considered alongside other context, such as a listener’s familiarity with different kinds of speech, some research analyzing the role of FL values and the assessment of comprehensibility has been performed that underscore its role as an influencing factor. Catford (1987) and Brown (1991) demonstrated that substitutions involving certain phonemes have more of an impact on a listener’s ability to comprehend the speech if the substitutions concern phonemes with higher FL values. Munro and Derwing (1995) similarly examined the role of FL in comprehensibility judgements in listeners. Their study analyzed how incorrect substitutions of phonemes impacted listeners’ perception of how easily they were able to understand the speech when the substitutions had high or low FL values. Their findings were that incorrect substitutions made with phonemes of higher FL values resulted in lower comprehensibility scores as more and more of the substitutions were integrated into speech samples.

The current study takes that concept and applies it to native English speakers' perceptions of ambiguous English phonemes in second language speech. If FL values truly do determine the difficulty a listener has in guessing an ambiguous phoneme, that would have clear implications for a second language speaker who may struggle to learn certain phonemes not present in their L1. That said, there is a paucity of previous empirical research on FL to demonstrate that the concept truly matters in a practical way, that is, in a way which impacts the everyday communication that second language speakers engage in. Why these implications are important will be discussed below, but first it is worth commenting on this scarcity of research. The reasons for this are two-fold. Firstly, Surendran and Niyogi's relatively recent 2003 study established how "researchers who want to measure FL often cannot" (p. 7) because of the limitations in concrete definitions of the term as well as the methodological complexity involved in doing so. Mi Oh (2015) touched on this issue, explaining that the lack of many large corpora and the difficulty in processing them contributed to the scarcity of FL research. She also succinctly explains the second reason for a lack of empirical research when she writes that the role of FL in research "has often been considered in an impressionistic way" (p. 154) that fails to provide either a clear practical definition of what FL means or any analysis of it that carries real-world implications. Put simply, the operationalization of FL involves calculations that were previously too complicated to easily research (e.g. Hockett, 1955), and the theory itself was rarely put under the lens of empirical studies. No one would dispute, for instance, the importance of analyzing phonemes from the perspective of phonological processes or grouping phonemes into natural classes based on those processes; how this relates to pronunciation, and the pedagogical implications from that analysis are very clear. Analyzing how phonemes operate within the framework of FL, however, does not supply as clear and obvious real-world implications. What

implications are there to our practical understanding of English, for instance, that the phonemic contrast between /p/ and /b/ carries a relatively high FL whereas the contrast between /s/ and /z/ carries a FL that is relatively low?

The current study is designed to address this question by situating the concept of FL in an empirical framework in which it can be observed. Multiple phonetic contrasts with widely differing FL values will be analyzed, which allows an evaluation of FL as an objective assessment of the importance of different contrasts by examining how listeners perceive those contrasts. To assess a listener's perception, the current study will examine the relationship that a phonetic contrast's FL values has with the *intelligibility* and *comprehensibility* of second language speech. The role that these two concepts play in relation to FL will be discussed below, but first one must define these two terms clearly. The current study assumes definitions of these two concepts that have been illustrated and utilized in research by Munro and Derwing (1995, 1997, 2001; Smith, 1992), among others. These studies have defined *intelligibility* as word or utterance comprehension, and *comprehensibility* as the effort involved in that comprehension. For the purpose of the current study, *intelligibility* is defined as the extent to which listeners can correctly decipher the sounds they hear. The current study will assess the intelligibility of second language speech to a listener through the use of a transcription task, as such a task clearly demonstrates whether or not a listener can correctly decipher the sounds they have heard. As for *comprehensibility*, the notion is defined as the listeners' perception of the ease, or how much effort, went into deciphering those sounds. Unlike intelligibility, which can be operationalized based on a binary scale in which a listener is either correct or incorrect in their deciphering, comprehensibility, which analyzes effort, allows listeners to perceive different degrees of effort. The current study will thus assess comprehensibility using a 9-point scale:

A study that examines intelligibility and comprehensibility based on two separate tasks will allow an analysis of how the two concepts weave into each other, and the role that both play when analyzing FL as an objective assessment of the importance of contrasts. As mentioned above, the notion of FL should theoretically have clear implications for a listener's ability to assess the comprehensibility and intelligibility of second language speech in a situation in which a learner of English may face difficulty in learning certain phonemes not present in their L1. If a native English listener was presented with any error or ambiguity in hearing a phoneme that carried a high FL value, whether the phoneme was not heard or pronounced wrongly, the FL implications for perception should indicate a greater chance for misidentification compared to a sound with a low FL. This is how one would expect a listener to react if the FL values attached to phonemes dictated the error gravity that a listener attached to phonemes with no additional factors impacting how the listener perceived the second language speech, and if FL actually can be used as an objective assessment of the importance of different contrasts for pedagogical purposes.

The current study will adopt the above-mentioned value-based methodological approach in order to assess how native English listeners react to second language speech when pronunciation is unclear. In this context, the present study focuses on two phonetic contrasts with widely different FL values, namely: the American English /r-l/ phonetic contrast and the /s-θ/ contrast produced by L1 Japanese speakers of English for whom /r/, /l/, and /θ/ are not present in their L1. Looking back at the role of FL in intelligibility and comprehensibility, one would assume that a native English listener faced with an ambiguous /r/ or /l/ would experience more difficulty deciphering (or, transcribing, using our methodology) the correct target sound because the high FL values for those phonemes reflect a large number of confusable words. One would

also expect that ambiguous examples of either the /s/ or /θ/ phonemes would not impede the intelligibility as much and a listener would have less difficulty in deciphering the correct sound if no other factors (such as contextual clues supplied by a sentence) had an influence. FL should also have implications for listeners' perception of comprehensibility. For the /r/ or /l/ phonemes, not only should the high FL value for that contrast greatly suggest a problem in terms of the intelligibility of speech, but it should also require the listener to use more effort into deciphering the target sound in that contrast. The ease of deciphering, that is, should be less than if a listener was tasked with deciphering an ambiguous /s/ or /θ/ compared to /r/ - /l/.

By focusing on the intersection of these three elements - FL, intelligibility, and comprehensibility – the current study, in sum, will assess the role of FL in how native English listeners perceive those four sounds in words produced by L2 speakers of English (L1 Japanese). Though the focus of this study does not explicitly examine other possible factors (e.g. familiarity with accented speech, attitudes towards accented speech), addressed below, that may have a role in impacting how a listener perceives different phonemes when ambiguity is present, it will allow an examination as to whether or not FL, as an objective assessment of the importance that contrasts play, truly represents the complete picture behind how a listener reacts to those sounds. The following literature review will consider these three concepts through the lens of previous research in order to examine factors important to how listeners perceive second language speech and how those factors, in addition to FL, may also influence the comprehensibility and intelligibility of second language speech.

2. Literature Review

The current study focuses on examining the real-world implications that FL values have concerning how listeners respond to unclear pronunciations of L2 sounds of varying FL values. That is, how much influence FL has on real-time speech processing, and consequently, how much influence it should have on pedagogical decisions. In doing so, the scarce research that has been done connecting FL to pedagogic or empirical evidence must also be examined, as well as research that analyzes other factors important to how listeners process speech, specifically the familiarity a listener has with accented speech, how meaningful that exposure to second language speech may be, and their attitudes towards accented speech. This is in order to demonstrate FL's place as one of many factors that influence how listeners respond to errors, or unclear phonemes, in second language speech and also to examine how factors related to familiarity need to be considered as influences on how listeners perceive sounds from accented speech. As the current study examines listener responses to speech through their perception of the comprehensibility and intelligibility of specific sounds, the above-mentioned factors need to be examined in the framework of those two terms.

2.1 Intelligibility

Intelligibility and comprehensibility are both factors that refer to how well a listener can understand the speech they are hearing, although the two terms refer to different things. Intelligibility refers to assessing whether or not a listener can understand what a speaker is saying. Unlike comprehensibility, which refers more to how difficult it is for a speaker to understand speech, intelligibility is an assessment that can be measured with a binary scale based

on whether a speaker understood speech or not. When looking at the intelligibility of speech, one must consider both the speaker and the listener as important influencing factors.

Research has looked at the variety of factors that influence a speaker's ability to produce intelligible speech (e.g. the speaker's L1, the age at which they began to acquire their L2) and has shown that these factors can play an influencing role in how intelligible their speech is to listeners (Flege, 2003; Flege, Munro, & MacKay, 1995). However, intelligibility of speech depends very much on a listener's perception of that speech, and listeners also present many factors that influence how intelligible they perceive different sounds to be.

One of these factors concerns a listener's *familiarity with accented speech*. Gass and Varonis (1994) examined how various familiarity factors (a listener's familiarity with a speaker, familiarity with an accent) influenced listeners' perception of speech intelligibility. Their findings illustrated that a listener's familiarity with a specific accent improved their ability to understand speech from speakers of that accent even if the listener was unfamiliar with the specific speakers themselves. This finding is in line with other studies (Wingstedt & Schulman, 1984) which produced similar results while examining different accents from speakers of different L1s.

Bradlow and Bent (2008) and Sidaris et al (2009) examined one's familiarity with non-standard accented speech, finding that listeners' transcriptions of speech samples can improve with repeated exposure to the kind of accented speech found within the samples, even when that task is extended to novel accents that are not realistically found in any language. This demonstrates how a listener's perception of the intelligibility of second language speech can be greatly influenced by allowing them to familiarize themselves with the accented speech. Even a brief familiarity to a novel accent can improve how intelligible a listener will find speech

samples containing that accent after they have been given prior exposure to it. Another study that demonstrates this relationship, Kennedy and Trofimovich (2008), illustrates how connecting certain target sounds to connected speech, thus giving target sounds some semantic context, can improve a listener's rating for the intelligibility of those target sounds when faced with accented examples of them. In their study, participants found target sounds more intelligible when those sounds were given brief semantic context or put into sentence frames. This shows how even the slightest context or added room for familiarity can influence a listener's perception of how intelligible speech, and how allowing a listener to even briefly familiarize themselves with an accent can influence how intelligible a listener feels that speech samples in that accent are.

If these studies illustrate how listeners, as well as speakers, bring factors with them that can influence their perception of speech intelligibility, one also must consider that a listener's attitude towards accented speech will influence how intelligible they perceive samples of accented input. Sheppard, Elliott, and Berk (2017) performed a study in which different listener populations listened to samples of speech from L2 English learners. In their study, one population had both more familiarity with accented speech than the other population and also indicated much more positive attitudes to accented speech than did the other population. The population who had a more positive view of accented speech not only was able to more easily understand the speech samples they heard, but their ability to transcribe the speech (thus, how intelligible the speech was to them) was far more accurate. Other, similar, research has been done that demonstrates a positive correlation between a listener's positive attitudes towards accented speech and their ability to more accurately transcribe it (Kang & Rubin, 2009; Lindemann, 2002). These studies demonstrate how it is not only a listener's familiarity with specific accented speech that can impact their ability to understand speech samples, but also their

familiarity with second language speech overall, and also their attitudes regarding accented speech.

The factors detailed above concerning familiarity are important to the current study because, when considering if the specific FL values of different phonemes play a central role in influencing how listeners perceive the intelligibility of those phonemes when they are presented with unclear examples of them, one also must consider the familiarity to accented speech and the attitudes of accented speech that different listeners bring with them. This previous research makes it clear that, when comparing the responses of listeners who are assessing the intelligibility of second language speech, one cannot compare the ratings of a listener with little meaningful exposure to accented speech with those of someone who has had considerable exposure. This is similarly true for comparing two listeners with widely different attitudes towards second language speech. The current study builds from this research by examining how two different listener populations perceive the intelligibility of target sounds in second language speech: listeners with a high degree of meaningful exposure to a variety of accents, and those with a far less meaningful exposure.

2.2 Comprehensibility

Just as research has analyzed the factors that listeners bring with them that can influence their perception of speech intelligibility, it has also looked at how those factors can influence a listener's perception of the comprehensibility of the speech, or, how much effort they put into that understanding. Grammatical accuracy, lexical richness, and rate of speech, correlated significantly with comprehensibility scores: Derwing and Munro (2001) and Kang (2010) found that extreme speaking rates (either too fast or slow) corresponded to reduced comprehensibility, while other studies focused on reduced accuracy in word stress placement (Hahn, 2004). Those

studies all demonstrate how multiple linguistic factors that speakers may struggle with can impede the ease of comprehension during a speech exchange; the familiarity factors that listeners bring with them, however, have also been examined.

Just as with the intelligibility scores, Sheppard, Elliott, and Berk (2017) found that raters of speech samples had indicated that they expended less effort in perceiving second language speech when they had either more exposure to accented speech or more positive attitudes towards second language speech. Gass and Varonis (1982) examined the roles of pronunciation and grammar in native speaker's listening comprehension of second language speech. The findings from that study indicated that native speaker listeners were unable to separate pronunciation from grammar when assessing the speech samples, and that the listeners had difficulty separating aspects of non-native discourse from one another when trying to assess how a speech sample is more easily comprehended than others. These findings are important to compare to the findings of Gass and Varonis' 1984 study, in which it was found that a listener's familiarity with accented speech, with a topic that a speaker was discussing, or familiarity with a specific accent or speaker, all had positive correlations with how easily the listener felt it was to comprehend second language speech. The findings of these two studies, taken together, demonstrate how the comprehensibility of second language speech, from a listener's perspective, does not depend on a single linguistic feature, and instead relies on a listener's wider *understanding* of a speech sample. That is, not all listeners may perceive the comprehensibility of speech samples in the same way because there are no specific linguistic features that determine how difficult listeners find it to understand second language speech. Instead, listeners' judgement of difficulty relies on a variety of factors, chief among them are the listener's experience with accented speech or with second language speakers. In the current study, which is

analyzing FL as a potential objective assessment of the difficulty of comprehending different phonemes when presented with unclear examples, these studies demonstrate how FL may just be one of multiple factors that impact listeners' perception.

The findings from studies, mentioned before, that suggest FL may be a crucial factor in comprehending L2 speech - Catford (1987), Brown (1991), and Munro & Derwing (1995) - create a compelling case for operationalizing FL values as a tool for instructing second language learners so that they could master the specific phonemes that were found to most hinder native-speaker comprehension of their speech. Derwing and Munro (2005) explain this succinctly when making the case that pronunciation instructors shouldn't "spend time on something that doesn't affect the intelligibility or comprehensibility" of second language speech in a listener, and that "the evidence is accumulating...[on] segmentals with a high functional load" (pg. 483). This would be a direct supportive argument for using FL values as an objective assessment for the importance of different phonemes and contrasts in teaching second language speakers, and their studies provide compelling evidence for taking FL seriously as a factor important to how listeners' perceive errors in second language speech. This idea is also supported by other studies, such as Pye, Ingram, and List (1987), which proposes using FL values and basing pronunciation instruction on the importance of contrastive oppositions instead of just focusing on the frequency with which isolated consonants appear across lexical types. In this view, FL values could potentially assist with focusing phonological development on the contrast. These ideas are supported by researchers A. Brown (1988) and G. Brown (1974), who looked at the relationship between FL and English language teaching to argue that teachers should turn to phonetic oppositions as an important element of pronunciation teaching. From their findings it seems helpful for teachers, instead of focusing on every pronunciation misstep, to make sure that their

students are substituting a sound for a difficult phoneme that won't result in a highly contrastive opposition and thus "are acoustically similar...and bear a low FL" (A. Brown, 1988, pp. 72). A teacher's attention, and the time saved, could be better spent elsewhere, they argue. Their view is that doing so will give the student an effective enough platform of pronunciation until the low contrastive substitutions that the student is making can be addressed in more advanced stages of teaching. This view is hand-in-hand with the idea that not all phonetic contrasts are equally important, and that there are particular phonetic contrasts that are more crucial when it comes to building a developing phonological system for speech production. The FL values theoretically place concrete numerical values on how important those different contrasts are and how disruptive errors made concerning those contrasts should be to a listener's comprehensibility.

That said, the previous research detailed above also makes it clear that FL may be only one factor out of many when it comes to how listeners perceive the difficulty in comprehending speech samples and how easily they are able to overcome disruptions made by errors with differing FL values. The current study seeks to look at this intersection: the role that FL has in predicting comprehensibility will be examined, as will the familiarity with accented speech that the present study's two listener populations possess. It will then be possible to examine if the FL values attached to the analyzed phonemes impact the listener's comprehensibility of speech samples as they should; that is, if a listener perceives spending less effort in comprehending speech involving a consonant contrast with a low FL value compared to a consonant contrast with a high FL value. It will also be possible to examine if a familiarity with accented speech may play a role in how listeners in the current study perceive the speech samples, as both listener populations have different likelihoods of holding frequent meaningful exchanges with second language speakers or speakers who have accents. Looking at how these two factors impact

comprehensibility will allow for an analysis of whether the impact of FL values on listener comprehension extends to both listeners with little familiarity and those with high familiarity.

2.3 The Current Study

It is within this theoretical framework, looking at the role FL and familiarity with accented speech has to play in a listeners' perception of the intelligibility and comprehensibility of second language speech, that the current study is operating.

For the current study, the English /r-l/ contrast and the /s-θ/ contrast will be examined for two different reasons. First, the FL values for both of their contrasts differ greatly. With a FL value of 0.015 for the /r-l/ contrast, and a 0.002 FL value for the /s-θ/ contrast (the process for obtaining and interpreting these figures will be explained in *Section 3.2.2*), the difference between the two could not be starker, with the /r-l/ FL value ranking among the highest of all possible consonant contrasts and the /s-θ/ FL value among the lowest. If one looks at those FL values, and assumes no additional factors, then previous research would suggest that listeners should not only have more difficulty in correctly transcribing unclear examples of target sounds featuring a /r/ or /l/ phoneme but should also spend more effort (or have a greater difficulty) in their perception of those target sounds. Choosing these two contrasts will allow an examination of FL as an objective assessment for the importance of different consonant contrasts when it comes to listener comprehensibility and intelligibility scores.

Secondly, the current study aims to examine the implications that these calculated FL values has for second language speakers to see to what extent it may be difficult for some listeners to identify the speakers' speech if they fail to master a contrast between two phonemes with a particularly high FL value. The study also is faced with the need to collect speech samples from participants who will pronounce the target sounds in a non-native like fashion. Previous

research has demonstrated the difficulty that Japanese speakers of English have when learning these contrasts (Bannister, Hazan & Iverson, 2005; Yamada, 1992). Research by Lado (1957) and Ritchie (1968) also notes that Japanese speakers struggle with fricatives, especially the /s-θ/ contrast. Lambacher et al. (1988) performed a study on this and found that 25% of their Japanese participants' mis-identified the /s/ for the /θ/, and vice versa, upon hearing them. Choosing the /r-l/ and /s-θ/ contrasts to analyze, thus, makes it possible for us to compare two different contrasts with widely different FL values to see if there is any difference in the perception of these contrasts among native listeners. It also allows us to see if the differences in those FL values have any implications for second language speakers who may struggle to learn certain contrasts not present in their L1 (in this case, native Japanese speakers). This will allow us to gain some insight into whether or not the FL theory has any implications for second language speakers, and whether or not FL as a theory is an effective measurement or scale for second language speakers to use in determining which phonetic contrasts are most important for them to pronounce comprehensibly. The two contrasts are quite different in regards to the importance that they play in separating utterances in English (as represented by their FL values) and in the differences that should exist between the two in how they impact a listener's perception of comprehensibility and intelligibility.

To determine if there is a real-world difference in the way native speakers respond to non-native like examples of both the /r-l/ and /s-θ/ contrasts, the current study takes an empirical approach to analyzing FL in order to analyze the role it has to play in influencing the perception of comprehensibility and intelligibility of second language speech samples in two different listener populations with varying degrees of familiarity with accented speech.

The study was motivated by the following research questions:

1. To what extent do functional load statistics realistically reflect native speakers' perception of error gravity, as represented by an assessment of their comprehensibility of learner productions of English /r/, /l/, /s/, /θ/?
2. How much is the intelligibility of the English produced by Japanese speakers reduced if they do not learn to produce /r/ and /l/ clearly, given the FL value of this contrast in English?
3. Is there a noticeable difference in the ratings between an on-campus student population exposed to a wide range of accents and a population from the general (or off-campus) community not as extensively exposed to accented English?

3. Methodology

A Qualtrics survey with two tasks was developed: an intelligibility transcription task and a two-part comprehensibility task. Creating this survey involved recruiting L1 Japanese speakers of English to supply speech samples, and two different population pools of native English speakers to assess those speech samples. These speaker populations are described below.

3.1 Participants

3.1.1 L1 Japanese L2 English Speakers

Fifteen native Japanese speakers (8 males and 7 females), all undergraduate university students, were recruited to provide the recorded speech samples for ratings. Speakers with lower English proficiency were targeted by seeking Japanese speakers from majors outside of TESOL, the languages, or linguistics. All were between the ages of 18 – 30 and had learned English in Japan during their secondary schooling and lived on campus.

Within these fifteen recorded participants, five were selected for use in the rating sessions. Three were chosen based on their lower proficiency and the presence of target errors in their speech concerning the /r/, /l/, /s/, and /θ/ phonemes. One of them was based on their slightly higher proficiency in which target word errors occurred but were not as frequent. A final participant was selected based on his native-like proficiency and lack of any such errors. This allowed the study to view how listeners would react to a range of oral proficiencies while keeping the number of samples to be rated at a reasonable level.

3.1.2 Native English-Speaking Raters

Native English raters were recruited for two separate population pools of raters: those who lived on a very culturally diverse university campus, and those who lived in a nearby community and were not associated with the university. The goal for this was to see if exposure to accented English produced any noticeable or interesting trend when looking at native English speakers' perception of second language speech. It would allow us to compare the results between those living in a situation where they were surrounded by second language speakers and accented speech and those who had far less exposure to these elements.

All raters were screened on several criteria for inclusion in this study. They identified themselves as speakers of “standard American English”, which this study defined for them as “the English commonly spoken by news anchors in the US”. No raters had taken courses in formal linguistics or phonetics.

The rating survey was distributed in person, through networking, and using some online resources. Most raters (in both populations) were recruited by me through face-to-face interaction. For listeners in the student population, this included approaching them in hallways, posting flyers through various campus buildings, and “walking in” classes to ask if anyone would like to participate. Many of them completed their rating sessions in a small room with my presence. Listeners were paid US \$10 for their participation. Most community member raters were collected through personal networking. Because I have had multiple part-time retail and food service jobs outside of the campus community, this made it easier for me to find raters in those industries who had no on-campus experience.

3.1.2.1 Listener Population #1: Student Raters

From the student population who lived on campus, all 28 were undergraduates who occupied either dorm rooms or on-campus apartments, and all were aged 18 – 30 years old. Most of them came from majors within business, science, or athletic health majors (those majors comprising 17 of the 28) and none of them had majors related to languages or linguistics. Fourteen of them were male, and sixteen were female. From the 28 raters in total, all of them spoke the upper Midwest variety of English, with no participants seeming to possess any differing accent. We did not ask them whether they also worked on campus or off-campus, something that could have impacted their exposure to second language speech, but that would also have been an element which could have been included.

3.1.2.2 Listener Population #2: Community Member Raters

From the population of adults who did not live on campus, and worked and lived within the outside community, all 26 of them were adults aged 18 – 60 who had lived in the United States their entire lives. Of those 26 selected for participation, 19 of them were aged 30 – 60. Many of them, known to me personally, either worked in the retail sector, the food service sector, or in warehouses (21 of the 26 worked in one of those industries). Their level of meaningful exposure that this population had to accented or second language speech in their day-to-day lives was significantly less than the university student listener population.

3.2 Materials

3.2.1 Recorded Speech Samples

The speech samples included on the survey for assessment were recorded English sentences spoken by native Japanese speakers. This is because, as mentioned before, the target phonetic contrasts being analyzed - /r - l/ and the /s - θ/ - are contrasts not present in Japanese speech and, thus, Japanese speakers would likely have a greater difficulty in producing these sounds spontaneously. Each Japanese speaker was recorded speaking nine sentences. In the sentences, two minimal pairs for each target contrast - /r/, /l/, /s/, and /θ/ - were featured. The sentences were designed so that either member of the given minimal pairs could reasonably and logically be placed within the sentence frame. This made it so that the target words were placed within connected speech, but also so that the listener had to rely purely on comprehending the sound and word that they heard from the recording. The listener would not be able to predict the target word based on context or logic.

The minimal pairs selected for the /r-l/ contrast, for example, were *rock/lock* and *writer/lighter*. The sentences used to frame these words were as follows: *Adam walked past the writer on his way to the bathroom/Paul saw the lighter when he sat at the table; Amy looked to see if the rock was where she left it/Ben sat the lock down on his chair when he got up*. The minimal pairs selected for the /s- θ/ contrast, *sick/thick* and *theme/seam*, were used in the sentences: *Paul wanted to know how sick the puppy was/Frank could tell that the plant was very thick; Greg pointed out the seam of Lucy's costume/Paul pointed out the theme of Suzy's costume*. Again, each minimal pair could fit logically within the two sentences provided, and it is not possible to guess which of the words belongs in the sentence by context alone. A ninth

sentence was used that did not feature any of the minimal pairs or target contrasts: *Bill ate some cake after dinner*. This sentence was used to familiarize the raters with each new speaker's voice.

The speech samples were recorded by presenting the native Japanese speakers with a delayed repetition task, a method used in previous research using similar tasks (Flege, Munro, & MacKay, 1995) because of its ability to elicit speech from participants without the participants simply imitating what had been said. I would play an audio recording of a woman's clear, crisp, voice, reading each sentence naturally. After a slight pause, a male's voice would prompt the speaker to recite what they had heard. This was done so that the speakers would follow the sentences, but the male's sudden vocal prompt and the slight pause before it would disrupt the Japanese speakers from rehearsal and imitation. The participants' productions were acceptable as long as the target word was recalled; other errors such as omissions were not relevant.

3.2.2 FL Calculations

The present study uses Phonological CorpusTools 1.4.1 (hereafter, PCT), a freely available and intuitive software developed by the Linguistic Department at the University of British Columbia, that gives users a simple graphical user interface by which they can perform phonological analysis on corpora of transcribed English. With little programming experience required for use, it is designed for phonologists interested in how frequency and usage play a role in phonology and allows for a multitude of related calculations to be run on a selected corpus. The current study will use the software's ability to calculate the FL of individual pairs of sounds and to supply a numerical value for the contrast that exists between those two phonemes.

The program calculates FL through its software by using a change in system entropy equation, a method of measurement used widely today (Surendran, 2003; Wedel, 2013) in FL research, by which a hypothetical merging of a pair of sounds takes place so that calculations can

be done on how much energy was lost during that merging. Under this method of measurement, the FL of any two sounds in the entire corpus can be calculated by looking first at the entropy of the system of the corpus. This includes all possible sounds. Next, the two target sounds in questions go through a hypothetical merger in which both sounds (say /p/ and /b/) are merged into a hypothetical /x/. The entropy of the system is re-calculated, showing the amount of energy that was lost in the system because of that hypothetical merger. If there is no change at all, there will be a zero value for the FL of that specific pairing.

In using this software for the current study's FL calculations of phonetic contrasts, the Irvine Phonotactic Online Dictionary (IPhOD) was applied as the corpus for analysis. This corpus, developed by the University of California, is a large collection of over 54,000 English words (with an emphasis on words pulled from spoken language) all written out in phonetic transcription. The entirety of the corpus is transcribed per the conventions of standard American English, is free to use and to download, and is fully functional alongside the PCT software for all of its calculations. This corpus, once loaded into the software, allows users to select any two phonemes at a time (as a pair), and then calculate the FL values that exist between those selected. The current study reports the FL calculations for every possible phonetic contrast between consonants (vowels were excluded from the scope of this study), so as to get a wide and complete look at how the FL calculations for the two target contrasts in focus (/r-l/ and /s-θ/) rank compared to the average FL contrast between pairs as well as all possible FL contrasts. These values can be seen below in Table 1, located in the Appendix. In the table, the FL is identified for every possible consonant contrast in English.

As can be seen by these results, the /r-l/ contrast, with a .015 FL value, ranks much higher than the /s-θ/ contrast at a .002. Not only that, but the /r-l/ contrast ranks at the very

highest among all possible FL values for consonant contrasts, whereas the /s-θ/ contrast ranks much lower.

If, then, FL does play a crucial role in determining how listeners perceive the intelligibility and comprehensibility of second language speech, these FL values should indicate different responses from the listeners for both phonetic contrasts. An unclear, or non-native like, /r/ or /l/, with its relatively high FL, should cause more disruption to a listener's ability to understand what is being said, which should impede the sound's intelligibility and cause the listener to use more effort in comprehending the sound. An unclear /s/ or /θ/, however, should not be as disruptive and should require less effort.

3.3 Procedure

3.3.1 Task #1 – Intelligibility

The first task for listeners was a transcription task which aimed to determine the intelligibility of specific sounds. In this task, listeners were presented with all nine sentences being spoken by all five Japanese speakers. Recordings were blocked by speaker. After each recording, the sentence frame appeared on the screen without the target word. Listeners were told, in instructions before this task, to view the sentence frames only as connected speech and not as meaningful to the task. This sentence frame was not the intended sentence, but instead whatever the speaker actually said in the audio. Thus, ellipses ([...]) or refrains such as “jumbled speech” were sometimes included within the transcription to represent mumbling. Raters then typed the missing target word that they heard. For instance, given the sentence *Paul saw the lighter when he sat at the table*, the audio recording of the speaker saying it was supplied in its

entirety, but the transcription of the text below that omitted the word *lighter* and included just a blank space.

Because of this study's focus on those two phonetic contrasts, we were not concerned with whether the entire word was correct, just if the listener was able to comprehend the target sound. The target sound was the sole focus.

3.3.2 Task #2 - Comprehensibility

The second task asked two things of the listeners. The listener would re-listen to each of the speech samples, only this time would have two sliding bars to respond to. The first sliding bar would ask listeners, upon hearing the target sound in context again, which sound – a /r/ or /l/, or a /s/ or /θ/ - they believed they heard. As a nine-point scale was used, it allowed listeners to represent the degrees to which they felt the sound represented one of those phonemes.

For instance, if listeners were presented with the audio recording of speech such as *Ben sat the lock down on his chair when he got up* and were given the sentence frame visually, they would be prompted to identify what best represented the initial sound in the missing word. This pointed the attention of the listeners to the target sound in question. On the scale, where an /l/ sound was a *one* and a /r/ sound was a *nine*, a *one* would count both as a perfectly target-like /l/ or a completely incorrect /r/. A *five* would indicate that the listener could not pick at all between the two sounds. For the /r/, /l/, and /s/, phonemes, a consonant letters *r*, *l*, and *s*, were used to represent the sounds for the raters. For the /θ/ phoneme, a *th* was used to represent it for the raters in a more understandable way.

After that task, listeners would then, for every audio sample rating, indicate how difficult it was for them to select between the two sounds. This was done using a sliding scale that ranged from *Very Easy* (a *one*) to *Very Difficult* (a *nine*).

The end of the survey included a space for listeners to write any comments on the survey, or questions or thoughts. It is unfortunate to indicate, however, that none of the participants, perhaps because most of them completed the task in my presence, included anything meaningful apart from contact information, gratitude for letting them participate, or wishes of luck.

4. Results

4.1 Data Analysis

Visual inspection of the intelligibility and comprehensibility data involved the following. First, data were removed for three raters because they completed the task in an unusually short amount of time. Most raters completed it in about 20 minutes. This left data from 61 raters for further analysis. Second, data from four raters were removed because they failed to use the entire rating scale. Third, data were removed for three raters as they exhibited numerous outliers. The final number of raters per group was 54: 28 in the student population and 26 community members.

4.1.1 Intelligibility Task

The binary rating system for the intelligibility task – whether listeners were able to correctly identify a target sound through their transcription – allowed us to code all listener responses as a *one* (for correct) or a *zero* (for incorrect) during analysis. Because we were looking specifically at the target /r/, /l/, /s/, and /θ/ sounds, a response was coded as correct (a one) so long as the listener correctly identified that target sound. The rest of the word they transcribed could have been incorrect, so long as the initial target sound was correctly deciphered.

Because of the dichotomous nature of the intelligibility data, the Kuder-Richardson test was used to assess reliability. Results indicated that all tests came back above a reasonable threshold of $KR = .7$ (.755 and .724 for the /s-θ/ and /r-l/ ratings of the community population, .770 and .834 for the /s-θ/ and /r-l/ ratings of the student population).

The number of correct answers was tabulated within each rater group. Table 4 displays the raw total of correct responses that each rater group provided during the transcription task, we well as the percentages that those correct responses represented for that task.

Table 1: Intelligibility Task

Raw Total (Percentage) of Correct Responses From the Student Population For the /r-l/ Contrast	250 (45%)
Raw Total (Percentage) of Correct Responses From the Student Population For the /s-θ/ Contrast	348 (62%)
Raw Total (Percentage) of Correct Responses From the Community Population For the /r-l/ Contrast	204 (39%)
Raw Total (Percentage) of Correct Responses From the Community Population For the /s-θ/ Contrast	213 (41%)

The implications from these results will be discussed further below, but it is clear that both sets of the community members' total correct answers fell below that of the student population. This falls in line with the previous results found of the comprehensibility scales, in which the student population out-performed the community members' ratings for both tasks – correctly comprehending the target sounds while also indicating that they did so with less effort. We find something similar here, in which the student population, for both target contrasts, were more likely than the community members to correctly transcribe the correct target sound upon hearing it.

Another result is that there is a difference between the gaps that exist within the two listener population's ratings for the /r-l/ and /s-θ/ comprehensibility ratings. In the student populations ratings, there exists a large twelve percentage point gap between the two, indicating that it was much easier for the student listeners to correctly transcribe the words containing the

/s-θ/ contrast. The community members have a smaller gap between their /r-l/ and /s-θ/ ratings – only two percentage points – though they similarly transcribed sentences featuring a /s/ or /θ/ sound correctly more often.

In short, the results here demonstrate that all listeners were more likely able to correctly transcribe a sound featuring the /s/ or /θ/ phoneme than they were able to do so with sounds featuring a /r/ or a /l/. Differences between the two rating groups, however, exist.

4.1.2 Comprehensibility Task

Due to the fact that the survey data for Task #2 was collected using a nine-point scale in which a low number (such as a one) would always indicate a target-like /l/ sound, or a very non-target-like /r/ sound, and similarly represent a target-like /s/ or non-target-like /θ/, it was necessary to reverse code the data. Because a one on the nine-point scale reflected a target-like /l/ and /s/ sound, and also represented “very easy” on the nine-point scale designed to gauge comprehension difficulty, the data results for sentences which were supposed to be a /r/ or /θ/ sound were reverse coded. Under this logic, a one on the nine-point scale would, for every sound, represent the best example, or most target-like example, of that sound. It would also represent the most ease participants had when comprehending those sounds. For instance, if listeners were presented with a sentence and had to fill in a missing word that included the /r/ sound, and they thought the sound heard represented a very target-like /r/, they would have selected an eight or nine on the nine-point scale to reflect this. Under the reverse coding, a nine would turn into a one, an eight into a two, and etcetera, down the scale. The data would then show that a one was the target-like /r/ sound that the participant indicated. This reverse coding was done for all sentences that featured a /r/ and /θ/ sound, as those sounds always occupied the right-hand side of the original nine-point scale and was done for both rater populations.

Since each speaker had four sentences per target phonetic contrast – four sentences with either a /r/ or /l/ sound, and four for the /s/ and /θ/ sounds – this meant that all 54 of the listeners had four numerical values to be totaled for all five of the speakers and for each of the target contrasts. This simplification resulted in the four spreadsheets mentioned above, with all listeners in each one having one large sum of their comprehension ratings per speaker.

The next step in analyzing the data was to determine if the inter-rater reliability for each of the spreadsheets justified averaging the results to a single number for easier analysis. SPSS was used to calculate the inter-rater reliability for each spreadsheet (that is, one for each of the two consonant contrasts for both population groups). Using the SPSS Reliability analysis tool with the intraclass correlation coefficient revealed .866 for students' ratings for the /r-l/ contrast, .939 for students' ratings of the /s-θ/ contrast, .816 for community member's ratings of the /r-l/ contrast, and .936 for community member's ratings of the /s-θ/ contrast. An interesting trend is that, in both populations, the coefficient was slightly lower than that for sentences featuring the /s-θ/ contrast, which indicates that both populations had more varied responses for the /r-l/ contrast, and could potentially indicate that both populations struggled more in coming to easy and uniform responses for ratings concerning the /r/ or /l/ sounds.

Because each of the inter-rater reliability statistics allowed for us to average the results of the rating spreadsheets to an easily analyzable number, the spreadsheets had all their ratings averaged into a single representative figure. This was done by averaging all the ratings for each target sound, by each population, producing the results in Table 2 below, which displays the mean comprehensibility results for each rating group. The standard deviation (hereafter, SI), as well as the 95% confidence interval (hereafter, CI), are given as well. The table displays the means for both rating populations and for both target sounds.

Table 2: Comprehensibility Task, Section One

Sound Contrast	Mean	Standard Deviation (SD)	95% Confidence Interval (CI)
Student Ratings of the /r-l/ Contrast	20.91	2.44	(17.6, 22)
Student Ratings of the /s-θ/ Contrast	16.06	2.11	(14.4, 18)
Community Member Ratings of the /r-l/ Contrast	23.30	2.77	(20.5,25.3)
Community Member Ratings of the /s-θ/ Contrast	17.73	1.94	(14.9,19.2)

As seen above the student comprehensibility ratings, for each target contrast, were lower than their respective contrasts in the community member ratings. Since lower ratings in this scale corresponded to “correctness” in listeners’ ability to correctly determine the target sound from the speech sample, this already demonstrates that the student population determined the correct target sound more often than the community member ratings. The community members, on the other hand, were less likely, concerning both target contrasts, to correctly comprehend the target sounds, with there being a greater gap in performance between the two rating populations concerning the /r-l/ contrast. This result would indicate that, while the community members under-performed the student population at correctly comprehending target sounds in both of the consonant contrasts, they struggled most when it concerned comprehending examples of the /r/ or the /l/.

For further analysis, the same process that had been done for the rating task above (averaging all of the ratings into a single analyzable figure after determining the inter-listener reliability) was done on the second task listeners completed in which they had to use a sliding nine-point scale to determine how much effort they put into identifying the target sounds they heard. Because this nine-point scale was uniform across all questions (*very easy* on the low end, *very difficult* on the high end), no reverse coding of this scale was required. The inter-listener

reliability was calculated for each spreadsheet (again, one per target sound for each population group). Since they all fell above a .700 threshold, it was determined it was appropriate to average these numbers. Averaging the ratings that every listener gave for each speaker on the second sliding nine-point scale in the comprehensibility task, produced the averages seen in Table 3 below. The table displays the average means for all rating groups' responses to the amount of effort they perceived using in comprehending the different sound contrasts.

Table 3: Comprehensibility Task, Section Two

Sound Contrasts	Means	SD	CI
Student Ratings, /r-l/ Contrast	18.84	2.25	(15, 20.1)
Student Ratings, /s-θ/ Contrast	15.31	2.03	(12.2, 18.2)
Community Members, /r-l/ Contrast	19.87	3.21	(16.2, 21.1)
Community Members, /s-θ/ Contrast	16.22	2.11	(14.2, 19.1)

As can be seen from these results, all of the student rating averages are, again, lower than their counterparts in the community member ratings. This would indicate that the student listeners, generally, indicated that they spent less effort in comprehending the target sounds they were given, whereas the community members had to put forth more effort in comprehending all target contrasts. That said, both ratings groups are consistent in indicating that they put forward less effort in comprehending the target sounds featuring the /s-θ/ contrast as opposed to the /r-l/ contrast. Both rating populations, based on these results, show they perceived putting less effort into comprehending the /s-θ/ sounds and that those sounds were less disruptive to them when faced with possible errors concerning it in second language speech.

Thus, the findings for the comprehensibility tasks find that, though differences in degrees exist between both listener groups, all raters put forth more effort in comprehending examples of the /r/ or /l/ and struggled to comprehend those examples more when compared to the /s/ and /θ/.

4.1.3 Correlation Between the Two Tasks

Further analysis was done to compare the intelligibility ratings alongside the comprehensibility ratings from above. Because this involved dealing with a binary variable – the intelligibility rating score that operated on a correct or incorrect basis – and a continuous scale – the nine-point comprehensibility scale – a point biserial correlation analysis was done on SPSS to determine the relationship that the listeners’ intelligibility ratings had with their comprehension of the two target contrasts. A positive correlation between the two factors can be seen below in Table 5. The table displays the correlation between the intelligibility and comprehensibility tasks, as seen in each rater group.

Table 4: Comprehensibility and Intelligibility Correlation

Correlation Between Intelligibility and Comprehensibility (Student Ratings of the /r-l/ contrast	.808
Correlation Between Intelligibility and Comprehensibility (Student Ratings of the /s-θ/ contrast	.630
Correlation Between Intelligibility and Comprehensibility (Community Ratings of the /r-l/ contrast	.788
Correlation Between Intelligibility and Comprehensibility (Community Ratings of the /s-θ/ contrast	.593

Looking at these results demonstrates that this correlation was stronger, for both populations, with the /r-l/ contrast. There is a clear trend that exists between both listener groups by which the correlation of their intelligibility and ease of comprehensibility are stronger with the /r-l/ contrast than with the /s-θ/. This indicates that, for all populations, their ability to correctly transcribe a target sound concerning the /r-l/ contrast correlated with a higher likelihood that they would be able to more easily identify a target-like example of that sound,

whereas with the /s-θ/ their ability to transcribe it correctly did not correlate as strongly with their perception of ease when it comes to comprehending the sound.

The correlation between the two tasks in the student listener population, however, is stronger, concerning both target contrasts, than that of the community member listeners. This indicates that the student listeners' ability to correctly transcribe one of the target sounds had a stronger correlation to their ability to comprehend it more easily. The community member listeners' ability to correctly transcribe a target sound, however, did not necessarily correlate to being able to comprehend it with greater ease. The implications of these findings, as well as other possible factors that need to be analyzed when noting the differences between the two listener populations, will be discussed below.

5. Discussion

Limitations in this study's design make it impossible to discuss the wide range of other factors that could play crucial roles in influencing how listeners respond to errors in second language speech. Among them concern the attitudes that listeners have concerning accented speech as well as the specific acoustic properties that each phoneme possesses that may impact its intelligibility to listeners. A further limitation is that the listeners did not provide any detailed qualitative responses about what they found most difficult when taking this survey. Those responses could have given more insight into what the listeners perceived themselves of having spent the most effort on. Though the results from this study are methodologically interesting, future studies can continue to look further at the role these other factors have on how listeners process errors in second language speech pertaining to different phonemes.

5.1 Research Question #1

RQ1: To what extent do functional load statistics realistically reflect native speakers' perception of error gravity, as represented by an assessment of their comprehensibility of learner productions of English /r/, /l/, /s/, /θ/?

When answering the question as to whether or not the FL values calculated from the phonological corpus realistically reflect a listeners' perception of second language speech errors, it would appear that the answer is nuanced.

When looking at the results of the comprehensibility ratings given for each community, it is noticeable that, for each population, the sentences containing the /s/ or /θ/ sounds were both easier for listeners to correctly identify the target sounds and also sentences that required less effort from the listeners in doing so. Since these results were seen in both rating populations, the students and the community members, this could support the idea that unclear sounds featuring

either a /r/ or /l/, a consonant contrast with a comparatively high FL value, were more disruptive to the native English listeners and required more effort from them as evidenced in their comprehensibility scores.

Where the results display some nuance is when looking at the differences that exist between the two rating populations. The results from the first section in the comprehensibility task demonstrates that the community member rating population found comprehending words featuring sounds from both target contrasts more difficult than did raters in the student population. They also indicated, in the second section of the comprehensibility task, that they perceived themselves as having spent more effort for both target contrasts. These results indicate that the disruptive weight that the FL values attach to the different phonemes being analyzed seem to have some observable influence on how listeners perceive second language speech, but that this influence needs to be considered alongside other factors. Because the ratings given for both populations had a consistent gap between them, with the students consistently over-performing the community member population, it is clear that FL's relevance for listener perception does not accurately portray the full picture of what shapes that perception.

One possible explanation for the differences in these results would be the increased exposure to second language speech that the student population had. As studies detailed earlier (Gass & Varonis, 1982; Kennedy & Trofimovich, 2008; Sheppard et al., 2017) have found, familiarity to second language speech patterns will increase the ease with which a listener can comprehend speech samples of accented speech. Seeing as the community member rating population had little meaningful exposure to this type of speech, this factor would explain why they faced greater difficulty in their comprehension of the speech samples and why they perceived themselves as having spent more energy in their comprehension. The increased

exposure to second language speech that the student population experienced could have made it easier for them to comprehend the speech samples. Multiple studies (Bradlow & Bent, 2008; Sidaris et al, 2009) have demonstrated how comprehensibility scores can increase for listeners when they have experienced a greater familiarity to second language speech, even when they are assessing speech samples from multiple different accents or L1s. It thus is possible that the student rating population was more easily able to comprehend the speech samples of L1 Japanese speakers of English even if the majority of their exposure to second language speech came from speakers with different L1's. Further studies need to be developed to explore this possibility in greater detail, but the potential of a familiarity influence could explain why the FL values did not represent the full scope of influence on the listeners' ratings.

5.2 Research Question #2

RQ2: How much does the intelligibility of the English produced by Japanese speakers suffer if they do not learn to produce /r/ and /l/ clearly, given the FL value of this contrast in English?

When looking at how much Japanese second language speakers of English would suffer if they did not learn a highly contrastive pair of sounds in English – like the /r-l/ - compared to a contrast with a much lower FL value, it would seem to depend. The intelligibility scores for words featuring an unclear /r/ or /l/ sounds demonstrate that both rating populations were unable to transcribe the words with a greater frequency than words featuring either a /s/ or a /θ/. These results indicate that the intelligibility of a Japanese speaker of English would suffer more if they failed to master learning a highly contrastive phonetic opposition – such as the /r-l/ contrast – than if they failed to master a phonetic contrast with lower FL values. A limitation of this study is that only two consonant contrasts were analyzed – though, two with very different FL values. It thus cannot be stated if these findings would play out for all possible consonant contrasts or if

there are differences among contrasts with high FL values. It is clear, however, concerning at least the /r-l/ and /s-θ/, as analyzed in the current study, that the Japanese speakers' intelligibility, as perceived by the native English speakers, suffered more when the speakers failed to supply target-like examples of the /r/ or /l/ sound. This finding would support FL's role as an influencing factor in intelligibility, as it would demonstrate how the unclear consonants with higher FL values impeded intelligibility more often than those with low FL values.

An important factor to consider alongside these results, though, is how both rating populations had significantly different gaps in their intelligibility scores between the two target contrasts. The results from the student rating population indicate that they were able transcribe words containing a /s/ or a /θ/ much more easily than words with a /r/ or /l/ - a twelve percentage point gap between the two sounds. This demonstrates that their perception of the intelligibility of those two sounds was clearly different, and that the /s-θ/ examples very clearly did not reduce the intelligibility of the speech as much to them as the other examples. With the community member ratings, however, there was only a small two-point gap between their intelligibility scores for the two contrasts. This demonstrates that, unlike the student rating population, these raters did not perceive as large of a difference in how the intelligibility of the speech samples was reduced when an unclear /r/ or /l/ was used versus an unclear /s/ or /θ/. These results could indicate that an increased familiarity with accented speech could make the influence of FL on listener perception more pronounced, whereas a lower degree of exposure to accented speech may perhaps make all accented sounds so difficult for the listener that the influence of FL is less pronounced.

Further results from the point biserial correlation demonstrate that both rating populations had stronger correlations between intelligibility and comprehensibility scores for the

/r-l/ sounds when compared to the /s/ or /θ/ sounds. That is, both rating groups' ability to accurately transcribe a target sound for the /r/ or /l/ correlated more strongly to their ability to more easily comprehend it later on and use less effort in doing so. This could also indicate that a native English speaker's perception of Japanese speakers' second language speech would suffer more if they failed to master a /r/ or a /l/ sound when compared to a /s/ or /θ/, as there would be a greater likelihood, as seen in the stronger correlation, between a decrease in intelligibility and a decrease in comprehensibility. That is, if a Japanese speaker of English's intelligibility suffered when failing to use a target-like example of a /r/ or /l/, then there would be a greater likelihood that the listeners' ability to easily comprehend the word they were saying would decrease and that the listener would perceive themselves as having spent more effort in that comprehension. The likelihood of this correlation is decreased for words containing unclear /s/ or /θ/ sounds. This is a trend that exists in both listening populations, though to a lesser extent for the student raters. Thus, the student listeners may, because of the potential benefits that they have when it concerns an increased familiarity to second language speech, not perceive as much reduction to the intelligibility of a Japanese speaker's speech when errors are made. That said, because the trend exists in both rating groups, the intelligibility of Japanese speakers' second language speech would seem to suffer more if they failed to learn a highly contrastive phonetic contrast, like the /r-l/, among listeners with both great and little exposure to second language speech.

5.3 Research Question #3

RQ3: Is there a noticeable difference in the ratings between an on-campus student population exposed to a wide range of accents and an older population from the general (or off-campus) community not as extensively exposed to accented English?

Looking at the differences that exist between the two rating populations helps to expand the discussion already proposed earlier – that FL values make up only a part of the complete

picture concerning how listeners comprehend speech – and also allows a discussion concerning the role that familiarity may play. The most glaring implication from the data is that the student population of listeners, in both comprehensibility rating tasks and in the intelligibility rating task, out-performed the community member population in that they were more likely to transcribe the correct target sounds, they were more likely to indicate the target sound upon hearing them in context, and they indicated that they spent less effort in comprehending those sounds. Without exception, the student population thus seemed to have an easier time undergoing all tasks when faced with the second language English speech they were presented.

One of the key factors that may possibly play a role in explaining how listeners responded to the speech samples is familiarity, as evidenced in previous studies such as Gass and Varonis (1984). Because the student population comes from an on-campus environment in which there is a large number of International students and diverse accented speech, this population has a much greater exposure to meaningful exchanges – not only sales encounters or brief encounters – with accented speech than does the community population. The community members, after all, live removed from any sort of campus environment in which International students converge and have far fewer meaningful exchanges with accented speech in their environment. It is true that most of the community members gathered were employed in service-type jobs – food service, retail service – and it is possible that they have frequent interaction with customers who have accented speech, but that is far different from the extended and more meaningful interaction that students on-campus at a diverse university are likely to have with accented speech.

If this factor did indeed play any role in influencing the ratings gathered, then this would indicate that, while FL certainly does seem to bear out in reality-based ratings, one also has to consider how familiarized the listener is with accented speech. It is possible that one's familiarity

with accented speech, or lack thereof, can influence how one perceives second language speech and the gravity of errors made concerning different consonants in a way that influences the impact that the FL values carry. The most obvious evidence for this from this study concerns how the student population, in all tasks, found the speech samples more intelligible and easier to comprehend. Their ratings still indicated that consonants with higher FL values were more difficult for them to transcribe and comprehend, but that phenomenon was less pronounced than in the results from the community member ratings. This could indicate that the influence that the FL values have on intelligibility and comprehensibility can be more noticeable when there is a lack of exposure to accented or second language speech. That exposure may make the influence of the FL values less noticeable for listeners with an increased familiarity with L2 speakers, though it does not erase its influence. This would explain why the FL values were shown to be related to the ratings of both listener groups, those with and those without that familiarity.

Of course, a large limitation to this study is that it only presented native English listeners with one sort of accented speech – Japanese second language speakers of English. Future studies concerning the same topic have much to explore regarding how much the type of accented speech matters; that is, if exposure to any accented speech will help improve one's ratings on an intelligibility and comprehensibility task (as evidenced in Bradlow & Bent, 2008), or if it is important that one is familiarized with the specific accent they are being asked to rate. It is impossible for us to say if the community members' ratings were lower, across the board, because they were lacking as much experience in meaningful exchanges with any sort of accented speech when compared to the student population, or if their ratings would have improved for this specific study only if those meaningful exchanges with accented speech were with Japanese second language speakers of English. Looking at different examples of accented

speech in future studies, and how native English listeners respond to a multitude of different accents, could help answer some of these questions in a way that this study cannot.

5.4 Pedagogical Implications

The pedagogical implications from these findings are limited by the fact that this study only chose to look at the differences that existed between two consonant contrasts - the /r-l/ and the /s-θ/ contrast – and that potential *other factors*, excluding familiarity with accented speech, that may potentially go into influencing comprehension were not considered in this study. Those factors would be necessary in truly looking at the pedagogical implications brought up here, but it is still worthy to consider them in the context of the limited results from this study.

The main implication from this study concerns FL as an objective assessment of the importance of consonant contrasts based solely on frequency, and whether that assessment alone is worth basing a syllabus on. The results from this study seem to indicate that building a syllabus for pronunciation based on this objective assessment of frequency may be useful for teachers as a way to determine which consonants are most important for the L2 speakers to learn, but that teachers need to be aware of the limitations of that method. The findings in this study do suggest that building a syllabus on the basis of FL values would aid Japanese speakers of English in improving the intelligibility of their speech – their intelligibility suffered more when they failed to use a target-like /r/ or /l/ than sounds with lower FL values. Native English speakers' ability to more easily comprehend examples of words with an unclear /s/ or /θ/, and spend less work in doing so, would also be evidence that a teacher's time would be better spent on the /r/ or /l/ sounds based on the higher FL values that those sounds carry. That said, the findings from this study also make it clear that other factors need to be considered as well.

There is also some evidence that creating a syllabus based solely on the objective assessment of the FL values may not adequately reflect the reality of how all native English speakers will comprehend a second language speaker's speech, and that doing so may deprive some contrasts with very low FL values from receiving attention in favor of contrasts with very high FL values, when the reality seems to indicate that doing so may not always be helpful. When it came to the intelligibility scores, after all, the student rating population found that unclear /r/ or /l/ sounds impacted intelligibility much more than unclear /s/ or /θ/ sounds, whereas the community member raters found almost no difference between the two contrasts. This would seem to indicate that, while building a syllabus on FL values may benefit Japanese speakers of English when they interact with students or listeners who have a greater familiarity with second language speech, the benefit might not be as strongly felt when interacting with listeners who have less familiarity. While teachers may, in some cases, spend their time more wisely by focusing on consonants with higher FL values, and thus consonants that may disrupt comprehensibility and intelligibility more starkly, they also need to consider that the benefits from doing so may only be seen strongly when their students interact with certain types of listeners. The fact that listeners with little exposure to second language speech may not be as forgiving when faced with errors concerning consonants with low FL values, when compared to listeners with greater familiarity, suggests that teachers cannot rely exclusively on the FL values in all cases. Though it is impossible to say, from this study, which factors a teacher should truly consider when developing a syllabus in addition to the FL values, the findings from this study still indicate that using FL as a sole methodology for developing a syllabus or a textbook ignores the reality of how different types of listeners will respond to second language speech.

A second pedagogical implication of this would be for teachers to be careful concerning how to select textbooks and paying attention to how those textbooks select which consonant contrasts to focus their attention on. A limitation of this study is that it looked only at English, one language. Not all languages have the same available corpora, and teachers would be wise to seek out how their textbooks chose to determine what to base their content on. As mentioned above, any textbook that focuses their attention to consonant instruction on frequency, or the objective assessment of that frequency through FL values, alone may not truly be developing their materials in a manner that is reflected in reality, especially the reality of speaking with accented speech to populations who do not have as much meaningful exposure with accented speech. This issue could matter more or less to different kind of learners, depending on the type of native English speaking populations that they foresee interacting the most with (a second language speaking college professor versus a service worker, for instance), but is worth considering when developing instructional materials for second language learners of English. It is also worth considering, when teaching, that the materials you are using to teach your students consonants may be devised in a way that assumes other comprehension factors will not be an influence on listeners' comprehension of certain contrasts or of the error gravity of errors made using those contrasts. If this study, in combination with studies such as Gass, S. & Varonis, E.M. (1984), can have any implications drawn from it, it would seem that other factors, such as a listeners' familiarity with second language speech, can influence listener comprehension and that FL alone as a device for developing instructional materials does not paint a complete picture of how listeners will respond to that speech.

APPENDIX

Table 5: Functional Load Values For Consonant Contrasts

p	b	t	d	k	g	tf	ds	f	v	θ	ð	s	z	ʃ	ʒ	m	n	ŋ	l	r	j	h	w	
0.007	0.009	0.013	0.008	0.015	0.005	0.004	0.003	0.007	0.003	0.015	0.001	0.009	0.003	0.004	0	0.009	0.008	0.001	0.013	0.006	0.001	0.005	0.004	p
		0.01	0.008	0.01	0.007	0.004	0.004	0.008	0.003	0.002	0	0.008	0.002	0.004	0	0.009	0.006	0.001	0.008	0.007	0.001	0.006	0.004	b
			0.013	0.015	0.006	0.005	0.004	0.007	0	0	0	0.004	0.004	0	0	0.005	0.008	0	0.002	0.002	0	0.005	0.004	t
				0.01	0.005	0.004	0	0.001	0	0	0	0.003	0.002	0	0	0.002	0.002	0	0.001	0.001	0	0.004	0.003	d
					0.008	0.005	0.002	0.001	0.001	0	0	0.002	0	0.001	0.001	0.002	0.001	0	0	0.001	0	0.006	0.004	k
						0.002	0.001	0	0	0	0	0.002	0	0	0	0.001	0.003	0	0.002	0.001	0.002	0.003	0.002	g
							0.002	0.003	0.002	0	0	0.001	0	0.002	0.002	0	0	0	0	0	0.001	0.002	0.001	tf
								0.003	0.003	0.003	0.001	0.002	0	0.003	0.002	0	0	0	0	0.001	0.002	0.002	0.001	ds
									0.003	0.002	0.002	0.001	0.001	0.002	0.001	0.002	0.003	0.001	0.001	0.001	0	0.004	0.004	f
										0.001	0.003	0.001	0.003	0.001	0.001	0.002	0.003	0.001	0.002	0	0	0.001	0.001	v
										0	0.002	0.002	0	0	0.001	0.002	0	0.002	0	0	0	0.001	0.001	θ
										0.001	0.001	0	0	0.001	0.001	0	0.002	0	0	0	0	0	0	ð
												0.002	0.002	0	0	0.001	0	0.001	0	0.001	0	0.005	0.003	s
													0.004	0.002	0	0	0.001	0	0.001	0	0.001	0.001	0	z
														0.001	0	0.003	0.006	0.006	0.004	0.002	0	0.001	0	j
															0	0.004	0.003	0.001	0.004	0.003	0.001	0.003	0.002	ʃ
																0	0	0	0	0	0	0	0	ʒ
																	0	0	0	0	0	0	0	m
																	0.009	0.001	0.011	0.006	0.001	0.005	0.003	n
																	0.002	0.013	0.007	0.001	0.003	0.002	0.002	ŋ
																		0.002	0.002	0	0	0	0	l
																			0.015	0.001	0.004	0.004	0.004	r
																				0.001	0.004	0.004	0.001	j
																					0.001	0.001	0.001	h
																								0.004

BIBLIOGRAPHY

BIBLIOGRAPHY

- Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to nonnative speech. *Cognition*, 106, 707-729.
- Brown, A. (1988). FL and the teaching of pronunciation. *TESOL Quarterly*, 22(4), 593-606. doi:10.2307/3587258
- Brown, A. (1991). Functional load and the teaching of pronunciation (pp. 379-397). *Teaching English pronunciation: A book of readings*. New York: Routledge.
- Brown, G. (1974). Practical phonetics and phonology. In J.P.B. Allen & S.P. Corder (Eds.) *The Edinburgh course in applied linguistics: Vol. 3. Techniques in applied linguistics* (pp. 24-58). Oxford: Oxford University Press.
- Catford, J.C (1987). Phonetics and the teaching of pronunciation: A systemic description of English phonology. In J. Morley (Ed.), *Current perspectives on pronunciation: Practices anchored in theory* (pp. 87-100). Alexandria, VA: TESOL.
- Derwing, T., and Munro, M. J. (1997) Accent, intelligibility, and comprehensibility. *Studies in Second Language Acquisition*, 45, 73-97.
- Derwing, T.M. and Munro, M.J. (2005), Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly*, 39, 379-397. doi:10.2307/3588486
- Flege, J. (2003) Assessing constraints on second-language segmental production and perception. A. Meyer & N. Schiller (Eds.), *Phonetics and phonology in language comprehension and production, differences and similarities*. Berlin: Mouton de Gruyter.
- Flege, J., Munro, M. J., & MacKay, I. R. A. (1995). Effects of age of second-language learning on the production of English consonants. *Speech Communication*, 40, 467-491.

- Flege, J.E., MacKay, I.R.A. & Munro, M. J. (1995). Factors affecting strength of perceived foreign accent in a second language. *The Journal of the Acoustical Society of America*, 97, 3125–3134.
- Gass, S. M. & Varonis, E.M. (1984). The effect of familiarity on the comprehensibility of nonnative speech. *Language Learning*, 34, 65-89.
- Gass, S. M., & Varonis, E.M. (1994). Input, interaction, and second language production. *Studies in Second Language Acquisition*, 16, 283-302.
- Hahn, L.D. (2004). Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly*, 38, 201-223. doi:10.2307/3588378
- Hall, Kathleen Currie, Blake Allen, Michael Fry, Khia Johnson, Roger Lo, Scott Mackie, and Michael McAuliffe. (2017). Phonological CorpusTools, Version 1.3. [Computer program]. Available from PCT GitHub page.
- Hockett, C.F. (1955). A manual of phonology. *Memoir of International Journal of American Linguistics*, No. 11. Baltimore: Waverly Press.
- Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulation: A comparison of methods for teaching English /r/-/l/ to Japanese adults. *Journal of the Acoustical Society of America*, 118, 3267-3278. doi:10.1121/1.2062307.
- Kang, O., Rubin, D., & Pickering, L. (2010). Suprasegmental measures of accentedness and judgments of language learner proficiency in oral English. *The Modern Language Journal*, 94, 554-566. doi:10.1111/j.1540-4781.2010.01091.x
- Kang, O., & Rubin, D. L. (2009). Reverse linguistic stereotyping: Measuring the effect of listener expectations on speech evaluation. *Journal of Language and Social Psychology*, 28(4), 441–456. <https://doi.org/10.1177/0261927X09341950>
- Kennedy, S., & Trofimovich, P. (2008). Intelligibility, comprehensibility, and accentedness of L2 speech: The role of listener experience and semantic context. *Canadian Modern Language Review*, 64(3), 459-489.
- King, R. D. (1967). A measure for FL. *Studia Linguistics*, 21, 1–14.
- Lado, R. (1957) *Linguistics across cultures: Applied linguistics and language teachers*. Ann Arbor: University of Michigan Press.
- Lambacher S., Martens W., Kakehi K., Marasinghe C., & Molholt G. (2005). The effects of identification training on the identification and production of English vowels by native speakers of Japanese. *Applied Psycholinguistics*, 26, 227–247.
- Lindemann, Stephanie. (2002). Listening with an attitude: A model of native-speaker

- comprehension of non-native speakers in the United States. *Language in Society* 31, 419-441.
- Mi Oh, Y., Coupé, C., Marsico, E., & Pellegrino, F. (2015). Bridging phonological system and lexicon: Insights from a corpus study of FL. *Journal of Phonetics*, 53, 153-176. 10.1016/j.wocn.2015.08.003.
- Munro, M. J., & Derwing, T.M. (1995). Foreign accent, comprehensibility and intelligibility in the speech of second language learners. *Language Learning*, 45, 73-97.
- Munro, M. J., & Derwing, T. M. (2001). Modeling perceptions of the accentedness and comprehensibility of L2 speech: The role of speaking rate. *Studies in Second Language Acquisition*, 23(4), 451-468.
- Pye, C., Ingram, D. & List, H. (1987). A comparison of initial consonant acquisition in English and Quiché. In K. Nelson & A. van Kleeck (Eds.), *Children's language*, Vol. 6, 175-190. Hillsdale, NJ: Erlbaum.
- Ritchie, William C. (1968). On the explanation of phonic interference. *Language Learning* 18, 183-197. <http://hdl.handle.net/2027.42/98385>
- Sheppard, B., Elliott, N., & Baese-Berk, M. (2017). Comprehensibility and intelligibility of international student speech: Comparing perceptions of university EAP instructors and content faculty. *Journal of English for Academic Purposes*, 26, 42-51. 10.1016/j.jeap.2017.01.006.
- Sidasaras, S. K., Alexander, J. E., & Nygaard, L. C. (2009). Perceptual learning of systematic variation in Spanish-accented speech. *The Journal of the Acoustical Society of America*, 125(5), 3306-3316.
- Smith, L.E. (1992). Spread of English and issues of intelligibility. In B.B. Kachru (Ed.), *The other tongue: English across cultures* (pp. 148-161). Urbana: University of Illinois Press.
- Surendran, D. & P, Niyogi. 2003. Measuring the FL of phonological contrasts. Tech. Rep. No. TR-2003-12. Chicago: University of Chicago.
- Vaden, K. I., Halpin, H. R., Hickok, G. S. (2009). Irvine Phonotactic Online Dictionary, Version 2.0. [Data file]. Available from <http://www.iphod.com>.
- Varonis, E., & Gass, S. M. (1982). The comprehensibility of non-native speech. *Studies in Second Language Acquisition*, 4(2), 114-136. doi:10.1017/S027226310000437X
- Wedel, A., Kaplan, A. & Jackson, S. (2013). High FL inhibits phonological contrast loss: A corpus study. *Cognition*, 128, 179-86.
- Wingstedt, M. & Schulman, R. (1984). Comprehension of foreign accents. *Phonologica* 1984:

Proceedings of the 5th International Phonology Meeting. Cambridge: Cambridge University Press.

Yamada, R.A., & Tohkura, Y (1992). The effects of experimental variables on the perception of American English /r/ and /l/ by Japanese listeners. *Perception & Psychophysics*, 52, 376–392. <https://doi.org/10.3758/BF03206698>