Intelligibility and Comprehensibility of Japanese Speakers of English from a Global Englishes Perspective



Naosuke Amano

Kellogg College

Candidate number: 1029481

Word count: 18,597

Dissertation submitted to the University of Oxford in partial fulfilment of the degree of

Master of Science in Applied Linguistics and Second Language Acquisition

Trinity Term, 2019

Abstract

This study investigated how different listener factors are associated with the intelligibility and comprehensibility of Japanese speakers of English. Specifically, the association between the two dimensions of speech understanding (intelligibility and comprehensibility) and three listener factors (listeners' first language, familiarity with English spoken by Japanese people, and listeners' English level) was examined. A total of 67 listeners from three different L1 groups (Japanese, Mandarin, and English) participated in this study. They listened to thirty-two English sentences spoken by two different Japanese groups (high proficiency and low proficiency) and rated the intelligibility and comprehensibility of those speech acts with a transcription task and a Likert scale, respectively. Their familiarity with Japanese accented English and English proficiency was measured with a questionnaire. Statistical analyses revealed that when the speaker and listener shared the same L1, both intelligibility and comprehensibility were high, indicating there was a shared L1 effect. The effect was greater in the case where speakers' English level was low. Similarly, listeners' familiarity with English spoken by Japanese people was significantly correlated with both intelligibility and comprehensibility. These correlations tended to be stronger when the speaking participants' English level was low. By contrast, there were no significant correlations between listeners' English proficiency and intelligibility, nor between listeners' English proficiency and comprehensibility. An implication of these findings is that sharing a L1 and familiarity with a speakers' L1 accented English could enhance intelligibility and comprehensibility. Additionally, these results can be interpreted as showing that a high level of English does not necessarily lead to a better understanding of English speech by non-native English speakers.

Acknowledgements

First of all, I should like to express my deepest gratitude to Dr. Hamish Chalmers, my supervisor, for his constant encouragement and guidance throughout the course. Without his support, I could not have finished my dissertation.

I am grateful to Dr. Robert Vanderplank and Dr. Heath Rose. Their valuable comments on my research proposal helped me to develop the idea of this study. I wish to express my gratitude to the following instructors of the ALSLA degree as well: Professor Victoria Murphy, Dr. Jessica Briggs Baffoe-Djan, and Dr. Robert Woore. They taught me the charms of applied linguistics.

I would also like to acknowledge Dr. John Lumsden for his sophisticated proofreading of the draft of this dissertation. He taught me the importance of clear writing and academic conventions. Of course, all errors are my own.

I was blessed with good course mates. I was lucky to study with such wonderful people. The time spent with them will always be one of the most valuable memories of my life.

This paper also owes much to the participants who gave their time to take part in my study. The data they provided me with was invaluable.

Finally, I would like to thank my family, especially my parents. Quitting my job and returning to academia was the most challenging decision of my life, but they respected my choices and supported me throughout.

Table of contents

Abstract i
Acknowledgements ii
Table of contents iii
List of tables vii
List of figures viii
List of abbreviations ix
Chapter 1. Introduction 1
1.1 Global Spread of English and Intelligibility 1
1.2 Aim of Study 1
1.3 Rationale of Study 2
1.4 Definition of Key Terms 3
1.5 Structure of Study 4
Chapter 2. Literature Review 5
2.1 Definition of Intelligibility and Comprehensibility 5
2.2 Measurements of Intelligibility, Comprehensibility, and Accentedness 8
2.2.1 Intelligibility Measurements 8
2.2.2 Comprehensibility Measurements 9
2.2.3 Accentedness Measurements 12
2.3 Relationships between Intelligibility, Comprehensibility, and Accentedness 13
2.4 Factors Affecting Intelligibility and Comprehensibility 14
2.4.1 Speaker Factors 14
2.4.2 Listener Factors 16
2.4.2.1 Shared L1 Effect 16
2.4.2.2 Familiarity with Foreign Accent Effect 21
2.5 Summary 24
2.6 Research Questions 25

Chapter 3. Methodology 27
3.1 Summary of Study Design 27
3.2 Materials 28
3.3 Speakers 29
3.4 Listening Participants 31
3.5 Data Generation Tools 33
3.5.1 Questionnaire 34
3.5.2 Intelligibility Measurement 34
3.5.3 Comprehensibility Measurement 36
3.6. Online Task Procedure 36
3.7 Pilot Study 37
3.8 Data Analysis Strategy 38
3.8.1 Research Question 1. Are listeners' L1 differences associated with the intelligibility and comprehensibility of non-native English speech? (i.e., Is there a shared L1 effect?) 38
3.8.2. Research Question 2. Is a listeners' familiarity with speakers' L1 accented English associated with the intelligibility and comprehensibility of non-native English speech? 39
3.8.3 Research Question 3: Is a listeners' English proficiency associated with the intelligibility and comprehensibility of non-native English speech? 39
3.9 Ethics 40
3.9.1 Ethical Approval 40
3.9.2 Consent Form 40
3.9.3 Data Protection and Confidentiality 40
Chapter 4. Results 41
4.1 Results of Questionnaire 41
4.1.1 Familiarity with Japanese Accented English 41
4.1.2 English Level of Non-Native Listening Participants 44

4.2 Reliability of Intelligibility Scores and Comprehensibility Ratings 46
4.3 Influence of Listeners' L1 on Intelligibility and Comprehensibility 47
4.3.1 Influence of Listeners' L1 on Intelligibility 47
4.3.2 Influence of Listeners' L1 and Speakers' Proficiency on Comprehensibility 53
4.4 Influence of Listeners' Familiarity with Japanese Accented English on Intelligibility and Comprehensibility 59
4.5 Influence of Listeners' English Proficiency on Intelligibility and Comprehensibility 62
Chapter 5. Discussion 63
5.1 Research Question 1. Are listeners' L1 differences associated with the intelligibility and comprehensibility of non-native English speech? (i.e., Is there a shared L1 effect?) 63
5.1.1 Influence of Listeners' L1 on Intelligibility 63
5.1.2 Influence of Listeners' L1 on Comprehensibility 65
5.2 Research Question 2. Is a listeners' familiarity with speakers' L1 accented English associated with the intelligibility and comprehensibility of non-native English speech? 66
5.2.1 Influence of Listeners' Familiarity with Japanese English on Intelligibility 66
5.2.2 Influence of Listeners' Familiarity with Japanese English on Comprehensibility 68
5.3 Research Question 3: Is a listeners' English proficiency associated with the intelligibility and comprehensibility of non-native English speech? 69
5.3.1 Influence of Listeners' English Proficiency on Intelligibility 69
5.3.2 Influence of Listeners' English Proficiency on Comprehensibility 71
5.4 Pedagogical Implications 71
Chapter 6. Conclusion 73
6.1 Limitations of Current Study 73
6.2 Suggestions for Future Studies 77
References 79

Appendix A. Speech Material 86
Appendix B. Invitation Letter 87
Appendix C. Online Task Form (English) 88
Appendix D. Online Task Form (Japanese) 97
Appendix E. Online Task Form (Mandarin) 106
Appendix F. Results of Pilot Study 114
Appendix G. Confirmation of Ethical Approval 115
Appendix H. Output of SPSS for The Relationships between L1 Differences and Intelligibility 116
H.1. Results of The Normality Test of Intelligibly 116
H.2. Results of The Brown-Forsythe Test of Intelligibly 116
H.3. Results of The Kruskal-Wallis Test of Intelligibility 116
H.4. Results of The Normality Test of Intelligibly (HP and LP) 117
Appendix I. Output of SPSS for The Relationships between L1 Differences and Comprehensibility 118
I.1. Results of The Normality Test of Comprehensibility 118
I.2. Results of The Kruskal-Wallis Test of Comprehensibility 118
I.3. Results of The Normality Test of Comprehensibility (HP and LP) 118
Appendix J. Output of SPSS for The Relationships between Familiarity with Japanese Accented English and Intelligibility 119
J.1. Results of The Normality Test 119
J.2. Scatter Plots (Familiarity with Japanese Accented English vs. Intelligibility) 119
J.3. Scatter Plots (Familiarity with Japanese Accented English vs. Comprehensibility) 121
Appendix K. Output of SPSS for The Relationships between English Proficiency and Two Speech Understanding (Intelligibility and Comprehensibility) 122
K.1. Results of The Normality Test 122
K.2. Scatter Plots (English Proficiency vs. Intelligibility) 122
K.3. Scatter Plots (English Proficiency vs. Comprehensibility) 123
Appendix L. Dissertation Title Change Form 124

List of tables

Table 1. Measurement methods of intelligibility 10
Table 2. Pronunciation scores of speaking data (40 speech data) 30
Table 3. The number of words in a sentence (40 speech data) 30
Table 4. Pronunciation scores of speaking data for the main study (32 speech data) 30
Table 5. The number of words in a sentence in the main study (32 speech data) 30
Table 6. Basic demographic information of listening participants 31
Table 7. Basic demographic information of listening participants in the pilot study 37
Table 8. Results of intraclass reliability of intelligibility scores and comprehensibility ratings 47
Table 9. Raw intelligibility scores 47
Table 10. Result of one-way independent ANOVA of intelligibility scoring 49
Table 11. Raw intelligibility scores in two different speaker groups 50
Table 12. Result of two-way repeated measures ANOVA: Listeners' L1×Speakers' proficiency in Intelligibility scores 51
Table 13. Results of Games-Howell test for intelligibility 52
Table 14. Results of t-test for intelligibility 53
Table 15. Raw comprehensibility ratings 54
Table 16. Result of one-way independent ANOVA of comprehensibility ratings 54
Table 17. Raw comprehensibility ratings in two different speaker groups 56
Table 18. Result of two-way repeated measures ANOVA: Listeners' L1×Speakers' proficiency in comprehensibility ratings 56
Table 19. Results of Games-Howell test for comprehensibility 58
Table 20. Results of t-test for comprehensibility 59
Table 21. Results of Correlation Analysis for Familiarity with Japanese Accented Speech and Intelligibility Scores 60
Table 22. Results of Correlation Analysis for Familiarity with Japanese Accented Speech and Comprehensibility Ratings 61
Table 23. Results of Correlation Analysis between Listeners' English level and Intelligibility and Comprehensibility 62

List of figures

Figure 1. Education background of listening participants 32

Figure 2. Current occupational status of listening participants 33

Figure 3. Self-evaluated familiarity with Japanese accented English 42

Figure 4. Frequency of interactions with Japanese people in English 43

Figure 5. Self-reported English proficiency (Beginner is A1 level and Master is C2 level in the CEFR) 46

Figure 6. Three different L1 groups' intelligibility scores 49

Figure 7. Three different L1 groups' intelligibility scores for each of the two speaker groups 51

Figure 8. Three different L1 groups' comprehensibility ratings 55

Figure 9. Three different L1 groups' comprehensibility scores for each of the two speaker groups 57

List of abbreviations

The following abbreviations are used in this study.

ANOVA	Analysis of Variance
CEFR	Common European Framework of Reference
IELTS	International English Language Test System
CET	College English Test
EFL	English as a Foreign Language
ESL	English as a Second Language
HP	High proficient
NS	Native speaker of English
NNS	Non-native speaker of English
L1	First language
L2	Second language
LP	Low proficient
TOEFL iBT	Test of English as a Foreign Language Internet-Based Test
TOEFL ITP	Test of English as a Foreign Language Institutional Testing Program
TOEIC	Test of English for International Communication

Chapter 1. Introduction

1.1 Global Spread of English and Intelligibility

English is now regarded as a global language and the number of English speakers has increased exponentially (Burns, 2005; Galloway & Rose, 2015). According to Crystal (2003), while the estimated number of native speakers of English (NS) is 320 to 380 million, the number of speakers of English as a second language (ESL) and English as a foreign language is about 300 to 500 million and almost one billion, respectively. Since English is used in many different countries, with many different native languages, it is likely that English is spoken in many different ways (Rose & Galloway, 2019; Smith, 1992). In this context, what makes English speech intelligible is an important question in teaching and learning. Many studies in the field of applied linguistics have attempted to answer the question by focusing on speaker factors (e.g., pronunciation) although both speakers and listeners have a critical role in communication (Kennedy & Trofimovich, 2008; Munro & Derwing, 2015).

1.2 Aim of Study

The aim of this study is to focus on a factor that has not received as much focus as speaker factors, namely listener factors. More specifically, the influence of the latter on the perception of intelligibility and comprehensibility of non-native speakers of English (NNS) in the Global Englishes context is investigated. This study investigates the case where three different groups of listeners (native Japanese, Mandarin, and English users) listen to English produced by Japanese learners of English. The listener factors that this study pays attention to are as follows: first language (L1) differences (i.e., whether there is a shared L1 effect), familiarity with accented English, and listeners' English proficiency.

1.3 Rationale of Study

It would be beneficial to clarify the rationale of focusing on the three specific listener factors just specified.

The main focus on the influence of L1 differences in this study is whether there is a shared L1 effect. Here, a shared L1 effect is where people who share the same L1 understand each other better than people who do not share that L1. It would be assumed that there are very few opportunities for NNS to interact with people in English when they share the same L1 since it would be easier for them to communicate with each other in their L1. However, due to the global spread of English, nowadays NNS have many opportunities to interact with people that share the same L1 in English in classroom and business settings. Taking the examples of Japan, many Japanese universities teach some courses in English (i.e., English medium instruction). The lecturers are not only NS but also Japanese L1 lecturers. Additionally, some Japanese companies such as Rakuten and Uniqlo ask their employees to use English in the company.

As for familiarity with accented English, as discussed in 1.1, there are many variations of English that now exist and many occasions to listen to them. Carter (1998) notes that roughly 80 percent of English communication occurs between NNSs. This indicates the vast majority of English communication is shifting from NSs' interactions and interactions between NS and NNS, to interactions in English among NNSs. Therefore, it is meaningful to investigate whether familiarity with accented English is associated with intelligibility and comprehensibility because the results could suggest whether language teachers need to teach various forms of accented English in their classes.

Concerning the third factor, listeners' English level, it would be natural to presume that as

listeners' English proficiency increases, their comprehensibility of native English speech increases. However, does this also apply to NNS? If the listener's proficiency effect is weak in NNS-NNS interactions, it could suggest the necessity of learning varieties of accented English, which would be an important pedagogical implication.

1.4 Definition of Key Terms

A definition of two dimensions of speech understanding (intelligibility and comprehensibility) and Global Englishes is necessary before moving onto the rest of the study. The definition of intelligibility and comprehensibility used in this study follows the definition provided by Munro and Derwing (1995b). They defined the terms as follows: "intelligibility refers to the extent to which a listener actually understands an utterance or message" and "comprehensibility refers to listeners' perceptions of difficulty in understanding particular utterances" (p. 291). The reason this paper applies their definitions is that they are broader than other definitions (e.g. Field, 2005, Smith & Nelson 1985) and suitable for empirical studies. Since the terms are controversial, further discussion of them will be presented in the literature review chapter.

For the definition of Global Englishes, this study uses the definition by Galloway (2013). According to her, it is an umbrella term of several different paradigms of the global spread of English. Specifically, she states Global Englishes

> bring(s) together World Englishes (WE), which focuses on the identification and codification of national varieties of English, and English as a Lingua Franca (ELF), which is 'any use of English among speakers of different first languages for whom English is the communicative medium of choice, and often the only option' (Seidlhofer, 2011, p. 7)" (p. 787).

1.5 Structure of Study

This dissertation is organized as follows: following the introductory chapter, Chapter 2 discusses previous studies of intelligibility and comprehensibility and research questions are proposed. Chapter 3 explains the methods that were used for conducting the current research. Then chapters 4 and 5 present the results and clarify the findings through discussion. Finally, chapter 6 summarizes the research findings and describes the limitations of the study and offers suggestions for further research.

Chapter 2. Literature Review

The purpose of this chapter is to review intelligibility and comprehensibility studies, discuss their limitations, and address how the current study contributes to the field of applied linguistics. This chapter proceeds in the following way. Firstly, different approaches to the conceptualization of intelligibility and comprehensibility are discussed, especially the two widely used definitions by Smith and Nelson (1985) and Munro and Derwing (1995b). Secondly, how intelligibility and comprehensibility have been measured is discussed and thirdly, the relationships between intelligibility and comprehensibility are explored. Following this, factors that influence intelligibility and comprehensibility are discussed. Broadly speaking, the factors can be categorized into speaker factors and listener factors, thus these factors is discussed separately. Finally, the contribution of this study to the current research is shown.

2.1 Definition of Intelligibility and Comprehensibility

Numerous researchers have attempted to conceptualize the different ways that people understand speech (e.g., Field, 2005; Munro and Derwing, 1995b; Smith & Nelson, 1985). In particular, two definitions became the most widely used: the definition by Smith and Nelson (1985), and the definition by Munro and Derwing (1995b).

Smith and Nelson (1985) reviewed intelligibility studies from 1950 to 1985 and found that “the terms intelligibility, comprehensibility and interpretability are often used interchangeably” (p. 334). To clarify the concepts, they classified the terms as follows:

> (1) intelligibility: word / utterance recognition,
>
> (2) comprehensibility: word / utterance meaning (locutionary force),
>
> (3) interpretability: meaning behind word / utterance (illocutionary force).
>
> (Smith & Nelson, 1985, p. 334)

Smith (1992) stated that these mark different levels of understanding, with intelligibility being the lowest and interpretability being the highest. Nelson (1995) further explained each of the levels by using the sentence ‘Please open the window’ as follows:

> Briefly, intelligibility refers to recognizing words and other sentence-level elements of utterances. In ‘Please open the window,’ for example, you recognize the elements ‘please,’ ‘open,’ and so on. Comprehensibility refers to the category of understanding meaning; you understand ‘please’ as an element that introduces requests, ‘window’ as having a certain referential meaning to an object in the immediate environment, and so on. Further, you understand the utterance ‘Please open the window’ as a speech act of request for a particular activity which you may carry out, ignore, object to, or otherwise react to in ways that will, in their turn, elicit another round of response from the other participants in the situation. Interpretability refers to the apprehension of intent, purpose, or ‘meaning behind’ an utterance. It is the capacity to take ‘Gee, it’s hot in here’ as the equivalent, as far as appropriate response is concerned, of the direct request ‘Please open the window’. (Nelson, 1995, p. 274)

Their definition has been used by several World Englishes and English as a Lingua Franca researchers (e.g. Jenkins, 2000; Matsuura, Chiba, & Fujieda, 1999; Smith, 1992).

Another widely utilized conceptualization of speech understanding is the definition by Munro and Derwing (1995b). They conceptualized speech understanding from a psycholinguistic perspective (Yamato, 2011). They classified it into three dimensions: intelligibility, comprehensibility, and accentedness. The definition of each dimension is as follows: “intelligibility refers to the extent to which a listener actually understands an utterance or message”, “(c)omprehensibility refers to listeners’ perceptions of difficulty in understanding particular utterances”, and “(a)ccentedness refers to how strong the talker’s

foreign accent is perceived to be" (p. 291). Their interpretations are operational definitions for experimental measurability (Yamato, 2011) and the measurement methods of each dimension are proposed as follows: intelligibility is measured with a transcription task and comprehensibility and accentedness are measured with Likert scales. Further discussion of measurement methods will be presented in the next section.

Many scholars have commented on Munro and Derwing's three dimensions of speech understanding. It has been supported by Munro (2017), who argues that, in terms of practicality, "intelligibility is the most important requirement in an interaction" (p. 414). By contrast, Levis (2006) claims, broadly speaking, intelligibility and comprehensibility are almost the same thing. Saito, Trofimovich, and Isaacs (2015) further point out that "outside research, most real-world applications of intelligibility, such as high-stakes assessment instruments (e.g. TOEFL, IELTS, CEFR), involve scalar ratings [i.e. Likert scales], which implies that the targeted construct is in fact comprehensibility" (p. 2).

These two definitions – Smith and Nelson's and Munro and Derwing's – have been thoroughly compared in terms of their scope and their use in empirical research. Nelson (2008) argued that Munro and Derwing's definition of intelligibility incorporates both the definition of intelligibility and comprehensibility used by Smith and Nelson. Munro and Derwing (2015) state their definition is even more inclusive saying their definition "encompass[es] all three of Smith and Nelson's concepts" (p. 379). Furthermore, Munro and Derwing (2015) argue that Smith and Nelson's definition "does not appear to have motivated much work on L2 communication, perhaps because it is unclear how to apply it empirically" and Munro and Derwing's (1995b) definition "offers the advantage of relatively straightforward empirical assessment" (pp. 378-379).

2.2 Measurements of Intelligibility, Comprehensibility, and Accentedness

As noted above, Munro and Derwing's (1995b) definition is said to be more suitable for empirical studies than Smith and Nelson's (1985). In fact, it appears that Munro and Derwing's classification is most widely used in research on the topic and even those studies that applied Smith and Nelson's definition incorporate the measurement methods proposed by Munro and Derwing (e.g., Matsuura et al., 1999; Matsuura, 2007). To my knowledge, the only study which measured all three levels of speech understanding suggested by Smith and Nelson is Smith (1992). He measured intelligibility, comprehensibility, and interpretability with a cloze transcription task, multiple choice questions, and a summarization of speech, respectively. Since a number of studies apply the Munro and Derwing paradigm, the following sub-sections will provide an overview how their three dimensions of speech - intelligibility, comprehensibility, and accentedness - have been measured.

2.2.1 Intelligibility Measurements

Scholars have used a variety of tasks to measure intelligibility, including transcription tasks, comprehension questions, and summarization tasks. See Table 1 for a fuller breakdown of these methods. Among these methods, transcription tasks are the most widely used (Munro & Derwing, 2015). However, transcription tasks themselves have many variations as shown in Table 1. Kang, Thomson, and Moran (2018a) investigated the reliability of a variety of methods to measure intelligibility. Specifically, they examined the relationships between intelligibility scores measured with five different methods and the scores for TOEFL listening tasks, the audio for which was produced by participants (i.e., the TOEFL listening material were recordings of speaking participants). The five intelligibility measurements were: transcription of semantically meaningful sentences, transcription of nonsense sentences, transcription of filtered sentences,

true/false questions, and a Likert scale to measure the intelligibility of semantically meaningful sentences. The result showed that the transcription of nonsense sentences like "A clean soul tapes the keys", has the highest reliability for predicting the score of the TOEFL listening.

However, it should be noted that this does not mean transcription tasks are a perfect measurement method. Zielski (2004, as cited in Munro, Derwing, & Morton, 2006) showed that the scores of transcription tasks were not necessarily correlated with the understanding of the whole message and provided examples where "all the words were correctly identified, with the listener still puzzling over what the speaker was trying to communicate" (p. 113). Additionally, one specific weakness of using semantically meaningless sentences is that they are less authentic speech material. We do not hear semantically meaningless sentences in our daily lives. Though it is also true that if we use transcription of semantically meaningful sentences, there is a possibility that participants can infer the words by using the contextual information of the sentence (Thomson, 2018). In fact, Kennedy and Trofimovich (2008) report that the intelligibility scores get higher when the speech material contains more contextual information. Therefore, it is necessary to choose the measurement method based on the purpose of research.

2.2.2 Comprehensibility Measurements

In contrast to intelligibility measurements, only two methods for measuring comprehensibility have been proposed: a Likert scale and native shadowing. The most widely used tool for measuring comprehensibility is a Likert scale. Here, raters are asked to judge how easy it is to understand an utterance, placing their judgment on a Likert scale. The weakness of this measurement is that the rating relies on listeners' subjective perception (Trisitichoke, Ando, Inoue, Saito, & Minematsu, 2019). Nelson (2011) states it

Table1

Measurement methods of intelligibility

Name of methods	Explanation of methods	Example studies
Transcription tasks (semantically meaningful sentence)	Participants listen to semantically meaningful sentences and are asked to transcribe them. Transcriptions are scored on the number of correctly transcribed words.	Munro & Derwing (1995b) Munro, Derwing, & Morton (2006) Osimk (2009)
Transcription tasks (semantically nonsense sentence)	Participants listen to semantically nonsense sentences and are asked to transcribe them. Transcriptions are scored on the number of correctly transcribed words.	Kang, Thomson, & Moran (2018b) Kennedy & Trofimovich (2008)
Transcription tasks (word)	Participants listen to words and are asked to transcribe them. Transcriptions are scored on the number of correctly transcribed words.	Field (2005)
Cloze dictation	Participants are given a passage and asked to fill in the blanks while listening to the passage being read aloud.	Matsuura (2007) Matsuura et al. (1999)
True/False judgement	Participants listen to sentences and choose whether an utterance is true or not. Example true sentence: Elephants are big animals. Example false sentence: Spaghetti grows on tall trees.	Derwing & Munro (1997)
Forced choice questions	Participants listen to a word and choose an answer from several choices.	Hayes-Hard, Smith, Bent, & Bradlow (2008)
Comprehension question	Participants listen to a speech and choose an appropriate answer from a set of options (multiple choice).	Major, Fitzmaurice, Bunta, & Balasubramanian (2002)
Summaries	Participants listen to a speech and summarize the ideas of the utterances. The task is scored by adding up the correctly described ideas.	Hahn (2004) Perlmutter (1989)
Oral translation	Participants listen to sentences and orally translate the message into their L1. The task is scored by adding up the correctly described ideas.	Amano (2012)
Likert scale	Participants listen to a speech and rated with a five-point scale.	Fayer & Krasinski (1987)

is doubtable that the word “understanding” in the instruction is well contextualized for the participants. Despite this weakness, the measurement is correlated with other objective measurements, which measure the cognitive burden on listening participants. For example, Munro and Derwing (1995b) investigated the relationship between Likert scale scores and reaction times to stimuli (i.e., the amount of time that the listeners take to respond to a speech act after it is presented to the them). In the case of NS listening to NNS speech, they found that speech with low comprehensibility ratings had significantly longer reaction times than those with high comprehensibility ratings. Furthermore, it also has a high correlation with the native shadowing method, which will be explained next.

Another method of measuring comprehensibility is native shadowing which was proposed by Inoue et al. in 2018. In this measurement, native listeners of the target L1 shadow non-native speakers' utterances, which Minematsu (2019) defined as the “simultaneous reproduction of words intended by the learner, in a native pronunciation, not imitation of accented pronunciations” (slide 47). Smooth shadowing indicates that the utterance is highly comprehensible. The smoothness of raters’ shadowing is evaluated in terms of accuracy of pronunciation in shadowing and/or delay of shadowing. The accuracy of shadowing is calculated with Deep Learning technology that identifies how accurate a speech is in relation to a certain criterion of pronunciation. The advantage of this method is that it can measure comprehensibility objectively. The drawback of shadowing is the necessity of training raters before the experiment and the greater time that is needed for data analysis (Minematsu, 2019). Furthermore, since this measurement method has only been used in studies that focus on native listeners, it is not clear if it would be appropriate for studies that focus on non-native listeners.

2.2.3 Accentedness Measurements

Although this paper's focus is on intelligibility and comprehensibility, it is worthwhile to briefly mention the operationalization of the third dimension of speech understanding, accentedness, since many studies measure all three dimensions (e.g., Munro & Derwing 1995b; Munro et al., 2006) and one of the concerns in the field is, as will be discussed in the next section, the inter-relationships among the dimensions. As with comprehensibility, one of the measurement methods of accentedness uses a Likert scale (Munro, 2017), where participants listen to a recording of a speech and rate how heavily accented they think the speech is. Another option for measuring accentedness is direct magnitude estimation (DME). In this method of measurement, participants listen to recordings by the speaking participants and evaluate the level of accentedness compared with a recording that functions as a standard for the listening participants (referred to as a modulus). For example, if a participant perceives the presented speech as half as accented as the modulus, then the participant gives it a rating of 50. Evidence for the reliability of the measurement was reported by Brennan, Ryan, and Dawson (1975). However, Munro (2017) states a Likert scale is "more convenient and ha(s) been used more extensively" than the DME (p. 418). One of the limitations of measuring accentedness with a Likert scale or the DME is the lack of a clear and uncontested definition of accent. According to Munro, generally speaking, accentedness "denotes the degree of differences between the pronunciation of an utterance and a listener's internalized representation of it" (p. 413). Given the existence of a wide variety of English pronunciation (even NS speech has a huge variety of accents), it would be difficult to set the standard for an accent, as all speakers may have different internal representation of what the accent *should* sound like.

2.3 Relationships between Intelligibility, Comprehensibility, and Accentedness

One area of focus in the study of speech understanding is interrelationships between the three different dimensions of speech understanding proposed by Munro and Derwing (1995b): intelligibility, comprehensibility, and accentedness.

One of the first studies to investigate these relationships was by Munro and Derwing (1995a). In their study, 25 NS listened to English speech produced by ten proficient Mandarin learners of English and two NS. The listening participants rated intelligibility, comprehensibility, and accentedness with a transcription task and Likert scales, respectively. The results showed that only five of the 25 participants' intelligibility scores and accentedness ratings were significantly negatively correlated with low to moderate correlation coefficient value. Here, a negative correlation suggests the intelligibility lowered as the degree of accent increased. On the other hand, most participants' intelligibility scores and comprehensibility ratings were negatively correlated, indicating that as intelligibility increased, comprehensibility increased. Similarly, most participants' comprehensibility and accentedness ratings were significantly positively correlated. This means less accented speech was easier to understand. However, the co-efficiencies of comprehensibility and the other two varied greatly depending on the participants (r = ranging from -0.44 to -0.90 for comprehensibility-intelligibility and ranging from 0.41 to 0.82 for comprehensibility-accentedness). They concluded the three relationships were "partially independent" (p. 90). Similar relationships between the three dimensions of speech understanding were shown by Derwing and Munro (1997), Matsuura et al. (1999), and Matsuura (2007).

Furthermore, the three dimensions of speech understanding's partial independence from

one another was also shown in a teaching intervention study. Derwing, Munro, and Wiebe (1997) taught English pronunciation to 13 NNS for 12 weeks and compared intelligibility, comprehensibility, and accentedness scores before and after the intervention. The participants recordings of true/false statements were presented to the raters, 37 NSs. The example of a true statement is 'Elephants are big animals' and false statement is 'Spaghetti grows on tall trees'. The listeners were asked to transcribe the sentences and rate the ease of understanding of the recordings and the degree of accentedness with Likert scales. The finding was that while intelligibility increased in both true and false statements, comprehensibility and accentedness increased only in true statements from the beginning of the course to the end of course. The findings of the partial independence between accentedness and the other dimensions of speech understanding (intelligibility and comprehensibility) have a strong pedagogical implication, in that learners of English do not necessarily need to aim to have native like pronunciation in order to be understood.

2.4 Factors Affecting Intelligibility and Comprehensibility

This section reviews what kind of factors influence intelligibility and comprehensibility. Broadly speaking, the factors can be divided into two categories: speaker factors such as pronunciation and speech rate and listener factors such as L1 differences and familiarity with accented English. According to Kennedy and Trofimovich (2008), the number of studies that deal with speaker factors is larger than that of listener factors. However, since this paper is chiefly concerned with listener factors, the main focus will be on the latter.

2.4.1 Speaker Factors

Many researchers have attempted to find what speech factors make an utterance

intelligible or comprehensible. For example, Anderson-Hsieh and Koehler (1988) found that speech rate affects comprehensibility when NS listen to NNS speech acts. Isaacs and Trofimovich (2012) conducted more extensive research. They investigated which linguistics features – phonology, fluency, lexicon, grammar, and discourse – most strongly correlate with comprehension and accentedness. Both dimensions of speech were evaluated with a Likert scale. The speaking participants were French learners of English and the listening participants were novice NS raters and ESL teachers. As for the novice raters' judgements, all of the language features except pitch range were correlated with comprehensibility. Particularly, wide range of vocabulary usage, word stress, vowel reduction, and story breadth had high correlations. On the other hand, ESL teachers commented that grammar, vocabulary, and frequency (pause and pacing etc.) affected comprehensibility.

As shown above, there are many speaker factors. Among them, pronunciation has received the most attention because, in contrast to other linguistic features, "pronunciation is [an] immediately salient" feature (Nelson, 2008, p.299). Specifically, the pronunciation features that contribute most to intelligibility and comprehensibility have been widely examined. It should be noted that, broadly speaking, pronunciation features can be divided into two categories: segmental and suprasegmental features. According to Kang, Thomson, and Murphy (2017), segmental features include vowel and consonant sounds, and suprasegmental features include stress, rhythm and intonation. Kang, Thomson, and Moran (2018b) investigated which pronunciation features are critical for intelligibility in both NS-NNS and NNSs' interactions. In their study, there were 18 speakers consisting of NS (British and American) and highly proficient NNS (South African, Indian, Mexican, and Chinese). Thirty listening participants who were

from the same countries as the speakers listened to recordings of semantically nonsensical sentences produced by the speaking participants. They were instructed to transcribe the speech acts. The results showed that segmental deviations were the strongest predictor of lack of intelligibility. Their findings are in line with other studies such as Jenkins (2000, 2002) in that segmental features are more important than suprasegmental features for intelligibility.

While segmental features contribute more to intelligibility, suprasegmental features affect comprehensibility more. In a study by Anderson-Hsieh, Johnson, and Koehler (1992), 60 NNSs from 11 different language groups were asked to read a passage aloud, which was evaluated subjectively by three English pronunciation specialists in terms of comprehensibility and segmental and suprasegmental features of pronunciation. The result showed that suprasegmental feature scores were a better predictor of comprehensibility than segmental scores. Similar findings were also shown in Derwing, Munro, and Wiebe (1998) whose results also indicated the primacy of suprasegmental features in comprehensibility.

2.4.2 Listener Factors

A decade has passed since Kennedy and Trofimovich (2008) noted the relative lack of listener' effect studies. The number of studies that focus on the contribution of listeners factors to intelligibility and comprehensibility have increased since then, but still the number is not significant and the results are not consistent. There are many potential listener factors that are considered to be associated with intelligibility and comprehensibility. For example, a recent study by Saito et al. (2019) listed the following six listener variables: L1 (distance from the target language), L2 proficiency, age,

experience (e.g., TESOL and L2 learning), familiarity with accented speech, and metacognition of L2 speech. Among the factors, the effect of sharing an L1 and familiarity with accented speech seem to be two of the most studied factors in the field. The following subsections' focus will be on these two factors.

2.4.2.1 Shared L1 Effect

Investigating the existence of the effect of a shared L1 is one of the biggest concerns in intelligibility and comprehensibility studies. It is generally assumed that sharing an L1 affects the perception of intelligibility and comprehensibility positively when they interact in a non-L1 language. This would be because "L2 accents are primarily characterized by transfer from the L1, and that listeners who share a speaker's L1 will have an intimate familiarity with the phonological patterns of that speaker's L2 accent" (Harding, 2012, p. 165). However, there seems to be disagreement about the nature and extent of this effect. Some studies have shown that there is a limited or no shared L1 effect (e.g., Crowther, Trofimovich, & Isaacs, 2016; Foote & Trofimovich, 2017; Major, Fitzmaurice, Bunta, and Balasubramanian, 2002). Others, for example, Bent and Bradlow (2003) claim that the effect of sharing an L1 is large.

One large-scale study, which researched the effect of L1 differences on listeners' utterance understanding is Major, et al. (2002). In their study, there were eight speakers that consist of native speakers of American English, Japanese, Chinese, and Spanish (n = 2 per group). The selection criterion of NNSs was that they had a 'moderate foreign accent'. This was judged by the researchers and 76 American NS. The speaking participants read aloud TOEFL scripts that were about two minutes long. A total of four hundred listening participants who were also from the same selection of countries as the

speakers listened to the recording of the TOEFL speech acts and answered the corresponding questions. The research result was that while the Spanish listening group showed a shared L1 effect, no such effect was observed for the Japanese groups, and shared L1 between Chinese groups actually impeded intelligibility.

Munro, et al. (2006) showed that the influence of L1 differences on intelligibility, comprehensibility, and accentedness is very limited. They recruited 48 speaking participants whose native languages were English, Cantonese, Japanese, Polish, and Spanish (n = 12 per group). The NNS participants' English proficiency was high intermediate. They described the same cartoon story. The listening participants were NS, Cantonese, Mandarin, and Japanese learners of English (ten per group). All the NNS listening participants were highly proficient English speakers. They evaluated 48 speech acts in terms of intelligibility using a transcription task, and in terms of comprehensibility and accentedness using Likert scales. Only a few perception differences between the different listeners' L1 groups were observed and the effect size of the Listener × Speaker interaction effect was small. For example, Japanese listeners' intelligibility scores of Japanese speakers' speech acts were higher than NS listeners. However, there were no perception differences of Japanese English among Japanese, Cantonese, and Mandarin L1 listeners. Additionally, all listener groups perceived Cantonese speakers' English similarly. The researchers concluded that the listeners' L1 differences have a limited influence on intelligibility and comprehensibility. One weakness of this study would be the use of the same cartoon story for speech material. Although each response would differ and the order of presenting to the listening participants were randomized, there is a possibility that the listening participants got used to the contents of the story since they heard speakers attempt to describe the same story 48 times. Thus, they may have inferred

the contents from the context, which might reduce the score differences of intelligibility and comprehensibility among groups.

Crowther, et al. (2016) also demonstrated that listeners' L1 differences do not influence comprehensibility and accentedness. In their study, 40 French learners of English engaged in an oral picture narrative task, in which the speakers described a story based on eight pictures. The listening participants were 26 high intermediate French and Chinese learners of English (13 for each). Both comprehensibility and accentedness of the speech acts were evaluated with Likert scales. Furthermore, the data were compared with Trofimovich and Isaacs (2012) which examined how 63 NS rate the comprehensibility and accentedness of the same speaking samples. The results showed that it did not matter whether the listener was French, English or Chinese, they tended to rate the comprehensibility and accentedness of the speech acts similarly. The limitation of this study would be similar to that of Munro, et al. (2006), namely that the same picture was used for all the tasks.

In contrast, a study by Bent and Bradlow (2003) showed that sharing an L1 has a strong effect on intelligibility. Unlike other studies described above on the effects of a shared L1 on intelligibility, their study included low proficiency speakers as well as high proficiency speakers. The speech data from the Northwestern University Foreign Accented English Speech Database were used in their study. The speech data contained intelligibility scores of each speaker evaluated by a total of 40 NSs of English with a transcription task. The researchers regard this score as the criterion for proficiency level. There was a group of two Chinese learners of English and a group of two Korean learners of English (there was one high proficiency and one low proficiency in each group) and

one NS's recordings were selected. A total of 60 speech acts, which had white noise added, were presented to listeners with a fixed order. The listening participants transcribed the words they listened to. The results showed that speech by highly proficient NNS had high intelligibility whether or not the speakers and listeners shared the same L1. Another interesting finding was that Korean listeners rated low proficiency Korean speakers' speech as highly as the speech by NS.

Similar findings were shown in the replication study done by Stibbard and Lee (2006). They address methodological limitations found in Bent and Bradlow (2003) while using different populations. There were ten potential speaking participants whose L1s were Korean and Arabic (five for each). Four of them (one high and low intelligibility participants per different L1) were selected based on an evaluation by five NSs with a five-point Likert scale (1 = heavy foreign accent; very difficult to understand to 5 = no foreign accent; very easy to understand). The same script as Bent and Bradlow was used for the recording material. Similar to Bent and Bradlow, there were four listener groups: NS, Korean, Saudi Arabic and NNS other than Korean and Arabic speakers. Contrary to Bent and Bradlow, the recordings were randomly presented to each participant and white noise was not used in order to avoid order effect and prioritize the authenticity of the listening material, respectively. The results revealed that the low proficiency speakers' speech had significantly lower scores compared to highly proficient speakers when the speakers and listeners did not share their L1. In contrast, the highly proficient NNS's speech was almost equally intelligible as native English to every listener group. Another finding was that Korean and Arabic listening participants evaluated the speech acts produced by the low proficient speakers who shared the same L1 higher than those produced by high proficient speakers that did not share the same L1, but they were not

statistically significant. One possible critique of this study would be the ways of selecting the speaking participants. As has been discussed in 2.2.1, intelligibility and accentedness are partially independent, thus, it would be problematic to measure both dimensions of speech understanding at the same time.

Ludwig and Mora (2017) showed that shared L1 contributes significantly to comprehensibility. Comprehensibility was measured with both reaction time and a seven-point Likert scale. The English recordings of Catalan, German, and English (two per group) were presented to 20 Catalan, 20 German, and ten NS. Ten of each NNS listening groups were high proficient and the other ten were low proficient. The results indicated that the overall comprehensibility ratings of listeners displayed a clear shared L1 effect, i.e., Catalan L1 listeners rated the speech acts by Catalan L1 speakers as easier to understand than did German L1 speakers; and German L1 listeners rated the speech acts by German L1 as highly comprehensible than the speech by Catalan L1 speakers. However, what is to be noted here is that the influence of shared L1 differed when taking into account listeners' English proficiency. While both low- and high-level German L1 users of English showed significant shared L1 effect, only low proficient Catalan listeners showed significant shared L1 effect. In terms of the reaction time, there was also shared L1 effect, i.e., when the speaker and listener shared L1, the reaction time was shorter than when not sharing the L1.

2.4.2.2 Familiarity with Foreign Accent Effect

Another listening factor that has been widely discussed is familiarity with accented speech. Catford (1950) states that those who are familiar with foreign accented speech reduce their "threshold of intelligibility for foreign speech" (p. 13). However, similar to

the studies on the effect of sharing an L1, conflicting results have been reported about its influence on intelligibility and comprehensibility. While some studies reported that familiarity with accented speech only affects comprehensibility (Matsuura, et al., 1999), others, for example, Kennedy and Trofimovich (2008) claimed its effect is limited to intelligibility.

Matsuura et al. (1999) reported familiarity with a particular accent positively affects only comprehensibility in cases where the speakers are NS and the listeners are NNS. The speaking participants were three American and three Irish subjects. The listening participants were 106 Japanese university students, 36 of whom learned English from an Irish university lecturer. The intelligibility and comprehensibility were measured with a transcription task and a Likert scale. The results showed that the group with high familiarity with Irish English evaluated Irish speech comprehensibility higher than the group with low familiarity with Irish English. However, the group who were unfamiliar with Irish English found it more intelligible than the high familiarity group. Another finding was that while all six participants' intelligibility scores were significantly correlated with the listening participants' TOEFL ITP score (r ranged from 0.26 to 0.48), only three out of six participants' comprehensibility scores were correlated with the listening participants' TOEFL ITP score (r ranged from 0.27 to 0.32). A weakness of this study is that it did not take into account the English level of listening participants when comparing the scores of those who are familiar with Irish English and those who are not. Since all the speakers' intelligibility and listeners' TOEFL ITP scores were correlated, there is a possibility that the group unfamiliar with Irish English had higher TOEFL ITP scores than the group familiar with Irish English.

Kennedy and Trofimovich's (2008) results conflict with those of Matsuura et al. (1999). There were two different listener groups, 12 experienced raters (ESL teachers) and 12 novice raters who had none or very little interaction with NNS in English. Both groups' raters were NS and the 12 experienced raters had at least three years teaching experience. The speakers were four Canadian NS, two American NS, and six Chinese learners of English. A total of 48 sentence recordings which consisted of true/false, semantically meaningful, and semantically anomalous sentences were presented to the listening participants. Intelligibility and comprehensibility were measured with a transcription task and a Likert scale respectively. The result indicated that while the intelligibility score was significantly higher when listeners were experienced raters, there were no rater effects in comprehensibility and accentedness scores. The finding is in line with Bradlow and Bent (2008).

Matsuura (2007) also showed the effect of familiarity with accented English on intelligibility. In her study, 106 Japanese university students with intermediate English ability listened to speech by an American NS and a Hong Kong NNS. Both speakers' recordings were evaluated as very understandable speech by three Japanese EFL teachers. Intelligibility and comprehensibility were measured with a transcription task and a Likert scale respectively. The listeners familiarity with foreign accented speech, English proficiency, and language anxiety were also measured. The familiarity was scored based on the varieties of accented speech that the participants had encountered. A regression analysis indicated that among the variables, proficiency was the best predictor of intelligibility in both speakers and familiarity was only a good predictor of Hong Kong English. There are limitations in this study. Firstly, this study lacks an analysis of the relationships between comprehensibly scores and familiarity scores. Secondly, the study

analyzed the listeners' familiarity with varieties of English and not a particular accented speech. It would be necessary to see how familiarity with a specific accented speech can affect intelligibility and comprehensibility.

2.5 Summary

Although the number of studies dealing with listener factors is increasing, the number of studies which focused on shared L1 effects and familiarity with L1 is still limited, and the results are not necessarily consistent. There are three limitations in particular that should be addressed. First, there is an apparent lack of studies that investigate the influence of shared L1 effect on listeners' perception of comprehensibility while taking into account speakers' English proficiency. Second, similarly, there is a lack of studies that investigate the influence of familiarity with accented English on listeners' perception of both intelligibility and comprehensibility while taking into account speakers' English proficiency. Third, Matsuura et al. (1999) and Mastuura (2007) showed the positive association of listeners' English proficiency and intelligibility, but it was not clear if this is also the case with comprehensibility. Therefore, it is necessary to conduct further research which focuses on these three effects. In addition to these limitations in the field, previous studies suffer from the following problems.

Concerning the issues around the effect of shared L1, firstly, Bent and Bradlow (2003) did not consider the order effect or authenticity of listening materials. Though Stibbard and Lee (2006) addressed this limitation, their studies were also flawed with regards to the selection of speakers. Secondly, there seems to be only one study which showed a strong contribution of sharing an L1 to comprehensibility, Ludwig and Mora (2017). Other studies which indicated no significant effect of sharing an L1 (e.g. Munro et al.,

2006; Crowther et al., 2016) may not be reliable due to the fact that they used the same speech materials for all the tasks. Thus, it would be necessary to conduct further research to see whether a strong contribution of sharing an L1 is present in different populations.

Previous studies about the effects of familiarity with accented speech also have limitations. Firstly, there seems to be only one study which focused on NNS' interactions, Matsuura (2007). However, her study did not analyze the relationships between comprehensibility and familiarity effects. Additionally, the study focused on the listeners' experiences in encountering varieties of English and not a target accented speech. Secondly, the ways of quantifying the familiarity with accented speech were limited and it would be necessary to quantify it from multiple perspectives such as self-evaluated familiarity with accented English and frequency of interacting with accented speech.

2.6 Research Questions

Based on the above discussion, the current study focuses on a non-native English variation, Japanese English and whether Japanese accented English is perceived differently depending on listeners' characteristic differences. The followings are the research questions:

1. Are listeners' L1 differences associated with the intelligibility and comprehensibility of non-native English speech? (i.e., Is there a shared L1 effect?)
2. Is a listeners' familiarity with speakers' L1 accented English associated with the intelligibility and comprehensibility of non-native English speech?
3. Is a listeners' English proficiency associated with the intelligibility and comprehensibility of non-native English speech?

Each research question also took into account the speakers' proficiency influence – i.e., whether high and low proficiency speakers enhance or reduce the association of each factor.

Chapter 3. Methodology

This chapter describes the methodological approach applied in this study. Firstly, an overview of the research design is shown. Next, the information about the speech material and participants is provided. This is followed by the descriptions of three data gathering tools used in this study: questionnaire, intelligibility measurement, and comprehensibility measurement. The subsequent section provides the procedure participants followed during the tasks. In addition, the pilot study is described. Following this, data analysis methods of this study is shown. Finally, ethical issues concerning the current study are described.

3.1 Summary of Study Design

The aim of this research was to examine how listener characteristics are associated with the intelligibility and comprehensibility of NNS in a Global Englishes context. As introduced in the previous chapter, the research questions of this study were as follows:

1. Are listeners' L1 differences associated with the intelligibility and comprehensibility of non-native English speech? (i.e., Is there a shared L1 effect?)
2. Is a listeners' familiarity with speakers' L1 accented English associated with the intelligibility and comprehensibility of non-native English speech?
3. Is a listeners' English proficiency associated with the intelligibility and comprehensibility of non-native English speech?

Each research question also took into account speakers' proficiency influence. To answer the above research questions, recordings of Japanese people speaking English were played to listeners whose L1s were either Japanese, Mandarin, or English. The listeners rated each speech act in terms of intelligibility and comprehensibility. Their familiarity

with Japanese accented English and English proficiency were measured with a questionnaire.

This study applied quantitative analysis. The relationships between the two dimensions of speech understanding (intelligibility and comprehensibility) and listener factors were analysed statistically (see 3.8).

3.2 Materials

Originally, forty target sentences were selected from a speech corpus (English Speech Database Read by Japanese Students), which was provided by the National Institute of Informatics (NII) in Japan. Following the results of the pilot study, 32 sentences were selected in order to reduce the burden on listening participants (see 3.7). The speech corpus contained recordings of 806 sentences and 711 individual words. They were read aloud by 202 university students (100 men and 102 women) from 20 different universities in Japan. Each speaker read aloud about 120 sentences and 220 individual words.

There were three criteria for choosing the target material from this corpus. The first criterion was that all the vocabulary in the sentences consisted of the most frequent 2,000 words. The purpose of this was to reduce the influence of vocabulary knowledge on intelligibility scores. The vocabulary frequency level was judged with Cobb's (n.d.) VocabProfiler. The second criterion was that the number of words in a sentence was less than ten. The purpose of this was to reduce the influence of participants' working memory. The third criterion was that the sentences were not old fashioned. This was judged by a native British English speaker.

The corpus also contains pronunciation evaluations of the 190 out of 202 speakers which were done by five native American English speakers. All the evaluators were English teachers who had experience in teaching English to Japanese people and had knowledge of phonetics. The sentences were evaluated with three categories of pronunciation: segmental, rhythm, and intonation. All were measured with a five-point holistic scale from 1 = very poor: Inaccurate in pronouncing sentences, and apt to be misunderstood to 5 = excellent: Good in pronouncing sentences, and very good in intelligibility, near-native speaker level. It should be noted that not all speech data were evaluated. Twenty-five sentences of each speaker were evaluated.

3.3 Speakers

Forty speakers were selected from the speech corpus. Half of them had a segmental rating score above 3.0. They can be seen as having high proficient participants because they could pronounce the sentences relatively accurately and intelligible to American NS. Hereafter this group will be referred to as Japanese HP. The rest of them had rating scores of 3.0 or less. They can be regarded as low proficient participants. They will be referred to as Japanese LP hereafter. Gender balance was also considered and an equal number of men and women speakers were selected. Therefore, the participants constitute ten HP male, ten HP female, ten LP male, and ten LP female. The mean segmental evaluation of the HP and LP groups and the number of words in a sentence of each group are summarized in Tables 2 and 3. As briefly mentioned in 3.2, the speech data were reduced to 32 based on the result of the pilot study (see 3.7). The summary of the speech data for the main study (segmental scores and the number of words in sentence) is described in Table 4 and 5. All the sentences and each speaker' pronunciation scores are shown in Appendix A.

The volume of the speech data varied. Therefore, audio editing and recording software, Audacity, was used to normalize the volume of each speech data. Additionally, a beep sound was added to the initial part of each sound file. The sound file was played 0.5 seconds after the beep sound.

Table 2

Pronunciation scores of speaking data (40 speech data)

	M	*SD*	*Min*	*Max*
Japanese HP	3.71	0.54	3.10	4.98
Japanese LP	2.42	0.22	1.86	2.65

Table 3

The number of words in a sentence (40 speech data)

	M	*SD*	*Min*	*Max*
Japanese HP	6.8	1.44	4	9
Japanese LP	6.75	1.29	5	9

Table 4

Pronunciation scores of speaking data for the main study (32 speech data)

	M	*SD*	*Min*	*Max*
Japanese HP	3.6	0.48	3.10	4.76
Japanese LP	2.42	0.23	1.86	2.65

Table 5

The number of words in a sentence in the main study (32 speech data)

	M	*SD*	*Min*	*Max*
Japanese HP	7.25	1.06	6	9
Japanese LP	6.69	1.30	5	9

3.4 Listening Participants

Listeners of this study were recruited through e-mails and social media. Invitation letters (Appendix B) for this study were sent to the author's friends and acquaintances. The receivers of the e-mail were not only asked to participate in the experiment but also to forward the e-mail to their friends or family (i.e., snowball sampling). Additionally, the research information was posted on social media services such as Facebook and Twitter. The participation was on a voluntary basis.

A total of 67 eligible participants completed the listening tasks. They were 20 native Japanese, 22 native Mandarin, and 25 British native English listeners. Their basic demographic information (age and gender) is summarized in Table 6. The gender balance of Japanese listeners was almost equal (male = 9 and female = 11). Their age ranged from 20 to 71 (M = 31). For Mandarin L1 listening participants, their gender balance was not balanced. While the number of male participants was four, that of female was 18. Their age ranged from 20 to 32 (M = 25.4). Similar to the Mandarin L1 listeners, British English L1 listening participants' gender balance was not ideal (male = 8, female = 16, and other = 1). Their age ranged from 19 to 68 (M = 37.28).

Table 6

Basic demographic information of listening participants

L1	N	Gender	Age
Japanese	20	9 male; 11 female	M = 31 (range = 20-71)
Mandarin	22	4 male; 18 female	M = 25.4 (range = 20-32)
English	25	8 male; 16 female; 1 other	M = 37.28 (range = 19-68)

As shown in Figure 1, the participants' educational level varied. For Japanese L1 participants, six finished a bachelor's degree, six finished a master's degree, two finished a doctoral degree, and six did not have any degree. Mandarin L1 participants had overall higher educational background compared to Japanese participants. Thirteen of them had a bachelor's degree and seven had a master's degree, and two had a further higher degree. English L1 participants' educational background was also diverse: five received college certificates, seven had a bachelor's degree, eleven had a master's degree, one had a doctoral degree, and one did not have any degree.

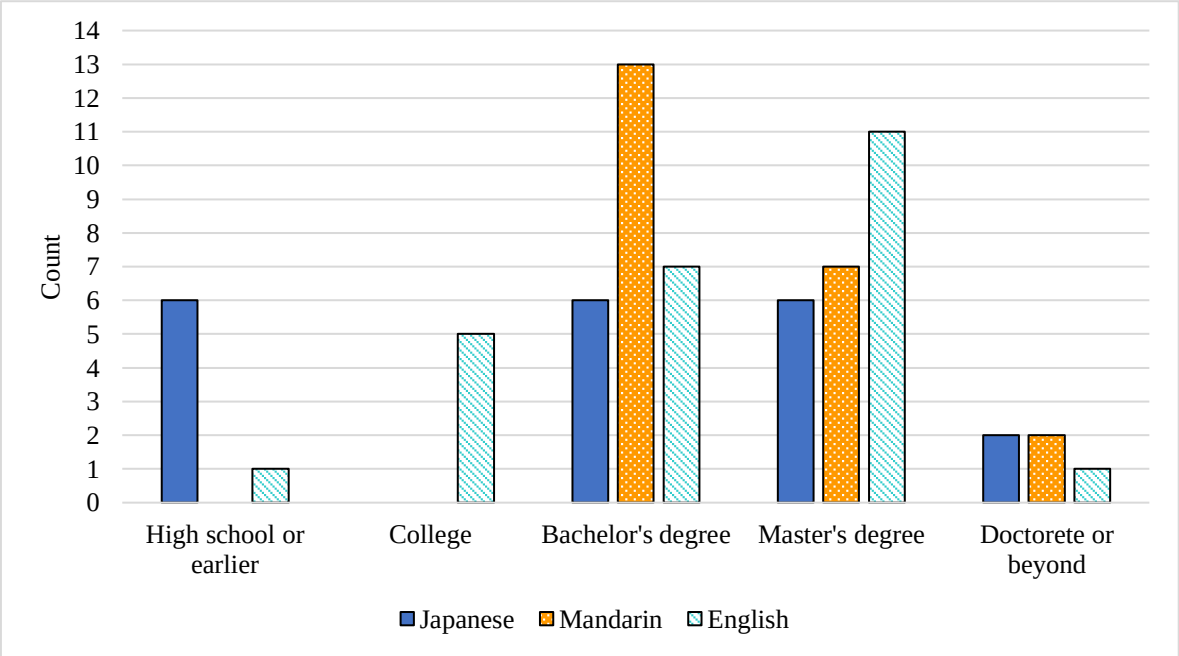


Figure 1. Education background of listening participants

Figure 2 displays the participants' current occupational status. Japanese L1 participants reported that while eight were undergraduate students and three were masters level students, eight had occupations (e.g., office worker and university lecturer) and one was retired. For Mandarin L1 participants, twelve were students (four undergraduate, seven master's, and one doctoral level) and nine had occupations (e.g., teacher and hotel staff).

Concerning English L1 participants, eleven reported they were currently students (five bachelor level, four master level, and two doctoral level). Four of the students also reported they had occupations. In total, 15 answered they had occupations and many of them were teachers. The final three reported that they were retired.

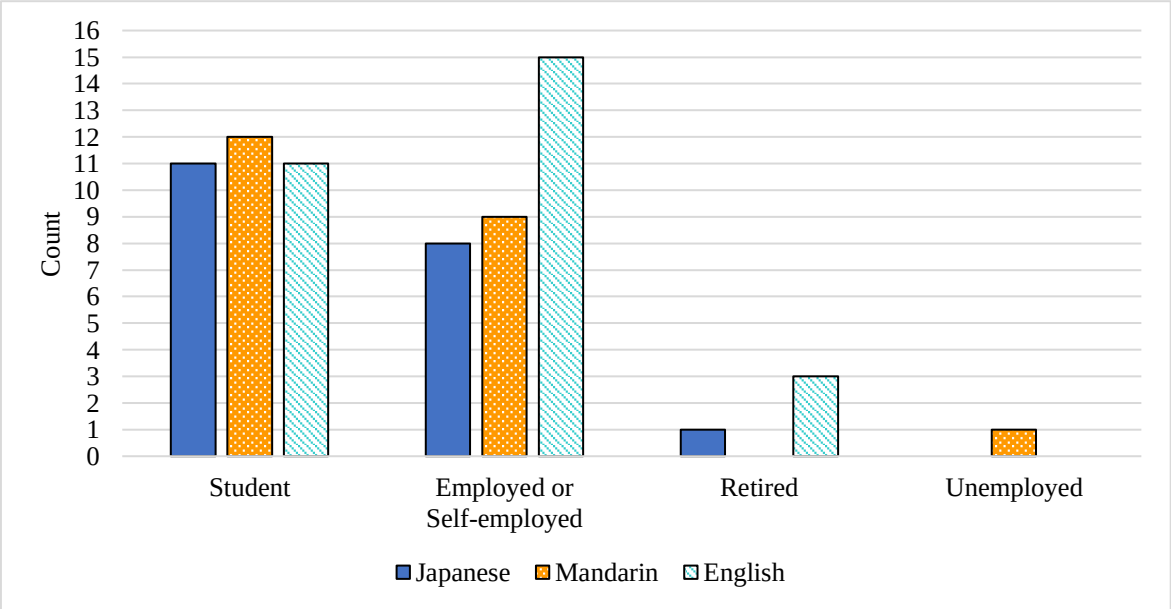


Figure 2. Current occupational status of listening participants

3.5 Data Generation Tools

In this study, questionnaires and intelligibility and comprehensibility measurement tasks were used to generate data. All the tools were presented to the participants through an online survey tool, Qualtrics. The online task form included the following: explanation of the research, consent form, questionnaire, and intelligibility and comprehensibility measurement tasks. The online form was translated into Japanese and Mandarin and the participants could choose either English, Japanese, or Mandarin (see Appendix C for English, D for Japanese translation, and E for Mandarin translation). The translation to Japanese was done by the researcher. The translation to Mandarin was done by a professional translator who had more than six years' translation experience. Additionally,

the Mandarin translation was checked by a highly proficient Mandarin learner of English who was an MSc student majoring in applied linguistics at the University of Oxford.

3.5.1 Questionnaire

In order to collect participants' background information and assess their familiarity with Japanese English, a questionnaire was created. The questionnaire consisted of three sections. The first section was regarding demographic information of participants (e.g., age, gender, educational background, and L1, etc.). The next section was about their language learning experiences. If their L1 was not English, they were asked questions concerning their English level and English learning experiences. They were asked to evaluate their English level with a six-point Likert scale (1 = beginner to 6 = Master). They were provided a language evaluation criteria chart which was created based on the CEFR and asked to refer it when they rated the English proficiency in order to increase the reliability of the self-evaluation. Additionally, if their L1 was not Japanese, they were asked whether they had learned or studied Japanese before. Finally, they were asked to report their familiarity with English spoken by Japanese people with a Likert scale: the frequency of their interactions with Japanese people in English (1= never to 5 = often) and their self-evaluated familiarity with Japanese accented English (1 = I am very familiar with English spoken by native Japanese speakers to 9 = I am not familiar with English spoken by native Japanese speakers at all).

3.5.2 Intelligibility Measurement

The intelligibility of the speech acts was assessed by having listeners transcribe what they heard when they listened to the recordings. As Munro and Derwing (2015) state, transcription tasks have been the most widely used measurement methods for

intelligibility. It should be noted that a transcription task of semantically nonsense sentences was the best predictor of TOEFL listening task performance (Kang et al., 2018a). However, it is less authentic listening material since people hardly listen to the semantically nonsense sentences in their daily lives. This study prioritized ecological validity, and so only semantically meaningful sentences were chosen. The drawback of using semantically meaningful sentences is that participants may infer the meaning from the context (Thomson, 2018). To minimize the effect of context, as discussed in 3.2, short sentences – where the number of words is fewer than ten – were selected.

As for scoring, intelligibility scoring methods varied in past studies. For example, Munro and Derwing (1995a) and Munro et al. (2006) calculated the percentage of correctly transcribed words in the sentences. They evaluated all the words in sentences and minor spelling mistakes counted as errors. In comparison with this, Bent and Bradlow (2003) only evaluated the three key words which were content words in each sentence. In their study, apparent spelling mistakes did not lead to a deduction of scores. However, transcription errors of morphology did count as errors.

In this study, intelligibility scores were determined by calculating the percentage of correctly transcribed words. All the words in the sentence were evaluated. Minor spelling mistakes were accepted. However, in the case where the word meaning changed, minor spelling errors led to a deduction. For example, if a participant transcribed the word ‘worm’ in the sentence ‘The small boy put the worm on the hook’ as ‘warm’ then there was a deduction. Additionally, words with morphological errors did not lead to a deduction of score. The reason minor spelling mistakes and morphological errors were admitted was to reduce the advantage of NS. NS would have an advantage in

morphological knowledge and vocabulary knowledge including spelling.

3.5.3 Comprehensibility Measurement

This study used a Likert scale to measure comprehensibility. It appears to be the most widely used measuring method for comprehensibility. One weakness of the measurement method is that it is a relatively subjective measure compared to other proposed measurement methods: native shadowing and measurement of reaction time. However, the results of the measurement had high correlations with the other two methods (Inoue et al., 2018; Munro & Derwing, 1995b).

One thing that should be considered when using a Likert scale is what scale point to use. The scale varies depending on studies. For example, Barrass (2017) applied a seven-point scale, Munro, et al. (2006) used a nine-point scale, and Saito, et al. (2015) adapted a rating scale of 0 to 1,000. Although rating scales vary in past studies, according to Thomson (2018), there seems to be "little to no impact of scale length" (p. 20). Therefore, this study used the most widely used scale, a nine-point scale (1 = very easy to understand, 9 = very difficult to understand).

3.6. Online Task Procedure

There were two main steps that the listening participants completed after accessing the online task form and providing their consent: answering a questionnaire and evaluating the intelligibility and comprehensibility of English produced by Japanese speakers. Firstly, the participants were asked to fill in a questionnaire form. Next, they were instructed to work on the next task with a headset in a quiet environment. Then, they listened to the speech data and were asked to transcribe the sentences that they had

listened to. Following this, they listened to the recording again and were asked to rate the easiness of understanding of speech (comprehensibility) using a nine-point Likert scale. They were asked to listen to each recording only once per task. Recordings were presented to each participant in unique random order. Finally, they were offered the option to leave their email address if they wanted to receive the results of this study.

3.7 Pilot Study

Before the main study, a pilot study was conducted to check whether there were any problems in the design of the online task and whether there were any technical issues. Ten people participated in the pilot test. They were composed of three native Japanese, four Mandarin, and three British English listeners. All the participants were acquaintances of the researcher and were asked to participate in the research by e-mails. Their basic demographic information is summarized in Table 7. All NNS participants were upper intermediate to advanced level English speakers. They were asked to give feedback about the online task form design and whether they suffered any technical issues in the final part of the online task form. (See the results of intelligibility scores and comprehensibility ratings of each sentence in the pilot study in Appendix F).

Table 7

Basic demographic information of listening participants in the pilot study

L1	N	Gender	Age
Japanese	3	2 male; 1 female	*M* = 27.3 (range = 21-34)
Mandarin	4	1 male; 3 female	*M* = 46.3 (range = 34-70)
English	3	3 male	*M* = 37.28 (range = 22-33)

One technical issue reported was that the audio files embedded in the online task form could not be played in a specific web browser (Internet Explorer). To deal with this problem, an explanation about the issue was added in the initial part of the form. The amount of time for completing the task also seemed to be a problem. The average time of the participant completing the task was 32.69 minutes (Median = 30.7, Min = 14.96, and Max = 65.43). According to Galesic and Bosnjak (2009), the response rate of the questionnaire decreases when the expected time of completion increases. Therefore, eight speech data were removed from the online task form for the main study to reduce the burden on participants and increase the response rate. The criteria for removing speech data was those which all participants in the pilot perfectly or near-perfectly transcribed. Additionally, since one of the Mandarin L1 participants reported that the word, 'pitcher' in a sentence 'He is a good pitcher' was not common in China, this speech act was deleted from this study. The deleted speech acts were distributed fairly across the speakers. That is, two speech acts from the HP male, HP female, LP male, and LP female speakers were deleted.

3.8 Data Analysis Strategy

This study applied a quantitative method. The statistical analysis of this research was done with statistical software, SPSS 25. In the following sub-sections, how data were analyzed to answer each research question is described.

3.8.1 Research Question 1. Are listeners' L1 differences associated with the intelligibility and comprehensibility of non-native English speech? (i.e., Is there a shared L1 effect?)

To investigate whether listeners L1 differences are associated with intelligibility and

comprehensibility, the mean of intelligibility scores and comprehensibility ratings of Japanese, Mandarin, and British English L1 listeners were compared. The dependent variables were the intelligibility scores and comprehensibility ratings. The independent variables were L1 differences (Japanese, Mandarin, and English).

3.8.2. Research Question 2. Is a listeners' familiarity with speakers' L1 accented English associated with the intelligibility and comprehensibility of non-native English speech?

The second research question was about whether the familiarity with Japanese accented English was associated with intelligibility and comprehensibility. To address this question, the research analyzed whether there was a correlation between two dimensions of speech understanding ratings (intelligibility scores and comprehensibility ratings) of the non-Japanese participants and the results of two familiarity measurements (self-evaluated familiarity with Japanese accented English and the frequency of interacting with Japanese people in English).

3.8.3 Research Question 3: Is a listeners' English proficiency associated with the intelligibility and comprehensibility of non-native English speech?

The third research question was about the relationships between listeners' English proficiency and two dimensions of speech understanding: intelligibility and comprehensibility. To address this question, the research analyzed whether there was a correlation between two dimensions of speech understanding ratings (intelligibility scores and comprehensibility ratings) and the listening participants' English level.

3.9 Ethics

3.9.1 Ethical Approval

Due to the involvement of human participants, this study needed to obtain an ethical approval from the Central University Research Ethics Committee (CUREC) at the University of Oxford before conducting an experiment. The CUREC form was submitted to the CUREC in February 2019 and received approval (Appendix G) in March 2019.

3.9.2 Consent Form

All the participants in this study were age 18 or above, therefore parents' consent was not necessary. The consent forms were collected before the participants answered the questionnaire and worked on the rating tasks. All participants read and agreed to the purpose, the process, and the data protection policy of this study.

3.9.3 Data Protection and Confidentiality

There was no automatic collection of personal identifiable data in this research. However, if participants select to leave their email address in order to be sent a summary of the research, the e-mail addresses were collected. All research data, consent forms, and participants' e-mail addresses were collected using an online survey tool, Qualtrics. The survey tool has strong data security and the data was hosted within the EU adhering to GDPR (Patrick, 2019). The data downloaded from the survey tool was secured on the author's local drive and on the University's encrypt

Chapter 4. Results

This chapter describes the analysis of data gathered in this study. Firstly, section 4.1 reports the results of the questionnaire: the listening participants' familiarity with English spoken by Japanese people and non-native listening participants' English level. Next, 4.2 demonstrates the reliability of intelligibility scores and comprehensibility ratings. The subsequent section, 4.3 presents the results of the analysis that investigated the influence of L1 differences on intelligibility and comprehensibility. Following this, section 4.4 shows the results concerning the ways that familiarity with Japanese accented English is associated with intelligibility and comprehensibility. Finally, in section 4.5, the analysis of the relationships between listeners' English level and two factors (intelligibility and comprehensibility) is presented.

4.1 Results of Questionnaire

4.1.1 Familiarity with Japanese Accented English

As shown in Figure 3, both Mandarin and English L1 listeners' self-evaluated familiarity with Japanese accented English was diverse. The mean ratings of Mandarin L1 listeners was 5.5 (*SD* = 2.5). As for the English L1 listeners, their mean ratings of familiarity with English spoken by Japanese people were slightly lower than Mandarin (*M* = 4.48 and *SD* = 2.48), indicating they were more familiar with Japanese accented English than Mandarin L1 participants. It should be noted that seven of the English L1 listeners reported that they were very familiar with English spoken by Japanese people. This seems to be the main cause of the discrepancy between Mandarin and English L1 listeners' mean ratings.

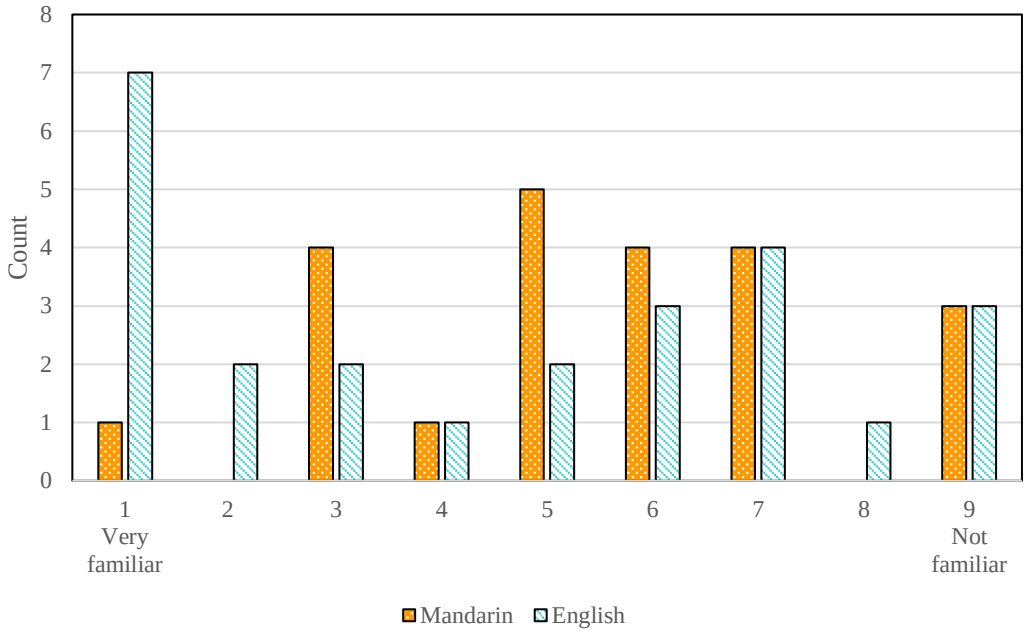


Figure 3. Self-evaluated familiarity with Japanese accented English

Figure 4 displays the Mandarin and English L1 listeners' frequency of interactions with Japanese people in English. For Mandarin L1 listeners, about half of them reported that they had never interacted with Japanese people in English. Additionally, none of them said that they frequently interacted with Japanese people in English. Their mean ratings were 1.86 (*SD* = 1.47), indicating they did not have many chances to interact with Japanese people in English. By contrast, English L1 listeners' ratings of interaction frequency seemed more diverse, with six people answering they had never interacted with Japanese people in English and eight reported they often interacted with Japanese people in English. Their mean ratings were 3.16 (*SD* = 1.48) suggesting they had more chances to interact with Japanese people in English than Mandarin L1 participants.

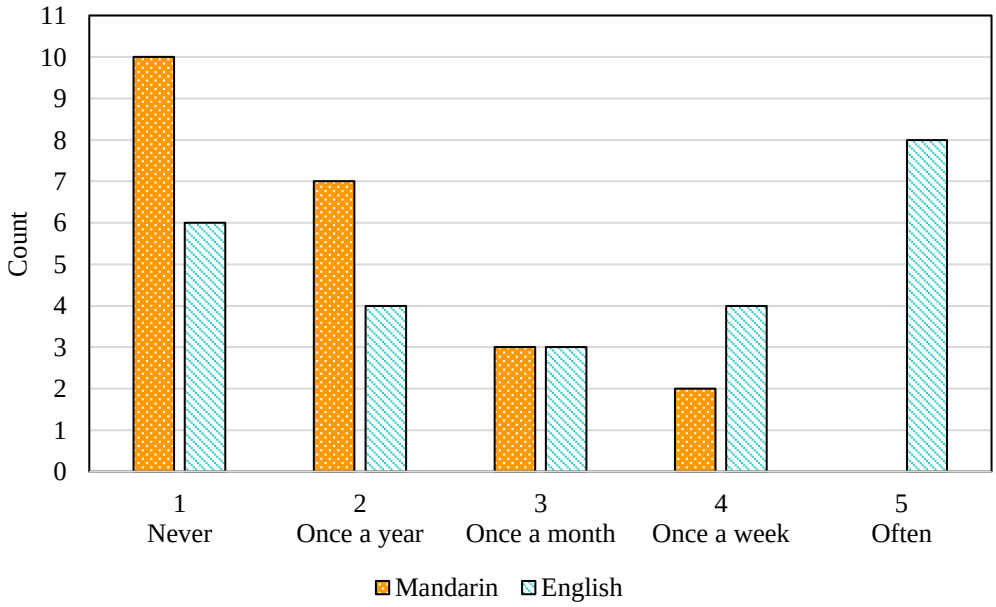


Figure 4. Frequency of interactions with Japanese people in English

The above results of two measurements of familiarity with Japanese accented English were used for the analysis in the following section.

Although this study did not analyze the relationships between the experience of learning or studying Japanese and two dimensions of speech understanding (intelligibility and comprehensibility), it should be reported that some of the participants stated that they had learned or studied Japanese before. As for the Mandarin L1 participants, nine answered they learned or studied Japanese. Their learning or studying period varied: less than six months (n = 3), six months to 11 months (n = 1), one to three years (n = 2), five to ten years (n = 1), and more than ten years (n = 2). Four took the Japanese Language Proficiency Test. Three of them had N1 level and one had N3 level. Their self-evaluated Japanese level was also diverse, ranging from 1 = beginner to 6 = master (M = 2.5).

For the English L1 participants, seven answered they learned or studied Japanese. Their

learning or studying period was diverse: six months to 11 months (n = 1), one year to three years (n = 3), three years to five years (n = 1), five to ten years (n = 1), and more than ten years (n = 1). One reported passing N2 level Japanese Language Proficiency Test. Their self-evaluated Japanese level ranged from 2 = elementary to 5 = advanced (*M* = 3.42).

4.1.2 English Level of Non-Native English Listening Participants

About half the Japanese L1 participants (n = 9) learned or studied English for more than 10 years, about another half of the participants (n = 9) had experience in learning or studying English for five to 9 years, and the final two participants' experience was between three and five. Similarly, many Mandarin L1 participants learned or studied English for more than ten years (n = 17). Four of them learned or studied English for five to nine years and one participant learned or studied English for three to four years.

In terms of English proficiency, 15 Japanese participants took TOEIC and their scores ranged from 535 to 970 (*M* = 797.33). Two took TOEFL iBT and their scores were 83 and 85. Six had IELTS scores ranging from 5.5 to 8 (*M* = 7). Only two participants did not take any standardized English exams. Taken together, the Japanese L1 listening participants standardized English exam scores suggest that most Japanese L1 listening participants' English level was high intermediate to advanced.

For Mandarin L1 participants, three took TOEIC and their scores ranged 495 to 845 (*M* = 675). Four reported they had TOEFL iBT scores ranging from 100 to 112 (*M* = 107.75). Seven had IELTS scores ranging from 6.5 to 8 (*M* = 7.28). One reported passing CET 6. Three other participants also reported that they took CET, but they did not mention which

level they took: either CET 4 or 6. Four participants reported they did not take any standardized English tests. The results of their English exams suggest that many of the Mandarin L1 participants' English level was high.

Based on the results of standardized English exams noted above, it was presumed that many of the non-native participants were high intermediate to high level English learners (B2 to C1 level in CEFR). The results were in line with their self-reported English proficiency. Figure 5 presents the self-reported English proficiency of Japanese and Mandarin L1 listeners. For Japanese L1 listeners, most ratings were concentrated on intermediate and upper intermediate. Only two rated their level as elementary and only one rated him- or herself as an advanced level English speaker. Their average ratings were 3.5 (*SD* = 1.02). Most Mandarin L1 listening participants rated themselves as an intermediate or above. In comparison with Japanese L1 participants, more participants rated themselves as at the advanced or master level. The Mandarin L1 participants' average rating was 4.05 (*SD* = 1). The ratings were mostly in accord with the results of their English standardized test scores. Based on this consistency, this self-evaluated rating was regarded as their English proficiency and used for the analysis in the following section.

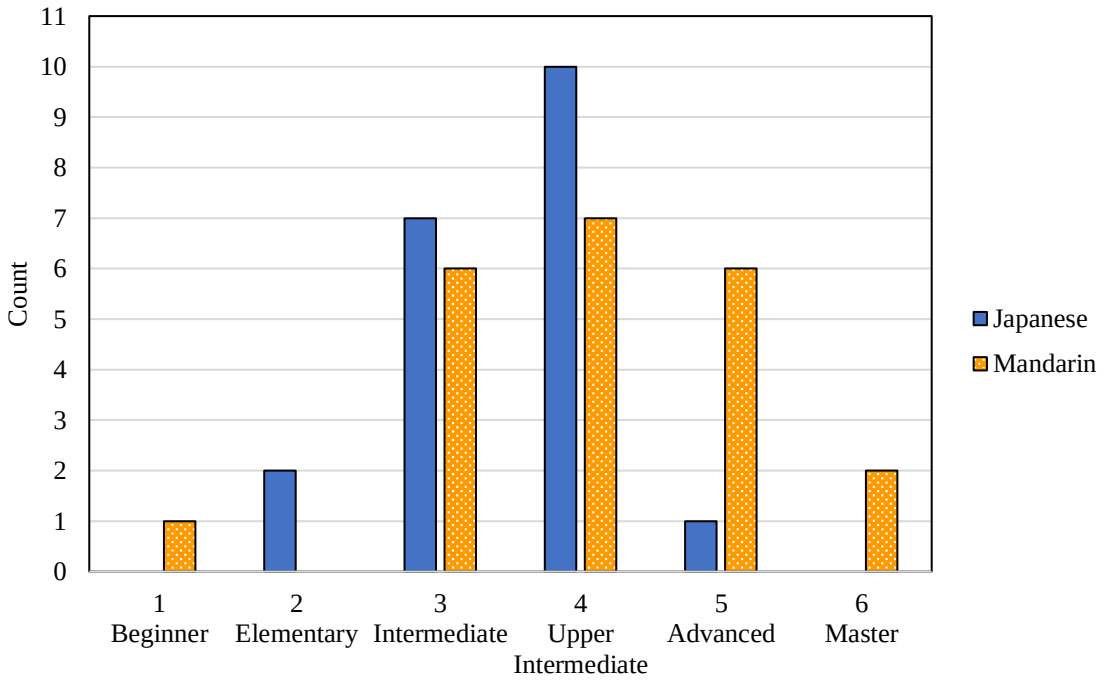


Figure 5. Self-reported English proficiency (Beginner is A1 level and Master is C2 level in the CEFR)

4.2 Reliability of Intelligibility Scores and Comprehensibility Ratings

For both intelligibility scores and comprehensibility ratings, intraclass reliability (Cronbach's alpha) was calculated with all 32 speech samples' intelligibility scores and the comprehensibility ratings given by all listening participants. The results are summarized in Table 8. For intelligibility, all listening groups' interclass reliability (α) was high, ranging from .875 to .944. Similarly, interclass reliability of comprehensibility ratings was strong, with Cronbach's alpha ranging from .897 to .965.

Each listeners' intelligibility scores and comprehensibility ratings were averaged and used for the analysis in the following section.

Table 8

Results of intraclass reliability of intelligibility scores and comprehensibility ratings

Listener group	Intelligibility scores (α)	Comprehensibility ratings (α)
Japanese	.875	.897
Mandarin	.918	.942
English	.944	.965

4.3 Influence of Listeners' L1 on Intelligibility and Comprehensibility

4.3.1 Influence of Listeners' L1 on Intelligibility

Table 9 describes raw scores of the three different L1 groups' intelligibility scores. It showed that English L1 listeners had the highest intelligibility score ($M = 0.89$), followed by Japanese L1 listeners with a slightly lower score ($M = 0.85$), and Mandarin L1 listeners with a further lower score ($M = 0.70$).

Table 9

Raw intelligibility scores

Listeners	*M*	*SD*
Japanese	0.85	0.05
Mandarin	0.70	0.14
English	0.89	0.04

To see whether their scores were significantly different from each other, firstly, a one-way independent analysis of variance (ANOVA) was conducted. Several preliminary assumptions need to be met to conduct a one-way independent ANOVA: random sampling, normal distribution of data, and homogeneity of variance (Takeuchi & Mizumoto, 2014). Since random sampling is difficult to conduct, the violation of this assumption is tolerated in the field of applied linguistics and psychology (Takeuchi & Mizumoto 2014). Concerning homogeneity of variance, in this study, the number of

participants in each group was different and a Levene's test showed that the p-value was less than 0.05 ($p = 0.008$). This indicated that the null hypothesis (i.e. that the variance was homogenous) was rejected. In terms of normal distribution of data, the following two variables were not normally distributed: Mandarin L1 listeners' intelligibility scores and English L1 listeners' intelligibility scores (see Appendix H for the results of normality tests). According to Takeuchi and Mizumoto, ANOVA is robust in terms of homogeneity and normality of variance. Accordingly, although the homogeneity of variance and normal distribution of data were not met, a one-way independent ANOVA was conducted. However, to be cautious, a Brown-Forsythe test which is used when homogeneity of variance is not met and a Kruskal-Wallis test which is a non-parametric equivalent of a one-way ANOVA, were also conducted. The results of these two tests were similar to that of the one-way ANOVA, thus here only the result of the one-way ANOVA is presented (see the results of the Brown-Forsythe test and the Kruskal-Wallis test in Appendix H).

The result of one-way independent ANOVA is described in Table 10 and Figure 6. The result was $F\ (2, 64) = 27.93$, $p < .001$, $\eta 2 = .466$. This means that there were significant group differences in intelligibility scores and strong effects of L1 differences. As a post hoc analysis, a Games-Howell test was conducted to see which variables had significant differences. The reason a Games-Howell test was chosen was because it is "accurate when sample sizes are unequal" (Field, 2018, p. 716). It was revealed that all groups had significant intelligibility score differences with a high effect size: Japanese and Mandarin L1 ($p < .001$, $d = 1.38$), Japanese and English L1 ($p = .025$, $d = 0.83$), and Mandarin and English L1 ($p < .001$, $d = 1.86$).

Table 10

Result of one-way independent ANOVA of intelligibility scoring

Source	*SS*	*df*	*MS*	*F*	*p*	η2
L1	.45	2	.23	27.93	<.001	.466
Error	.52	64	.01			
Total	.97	66				

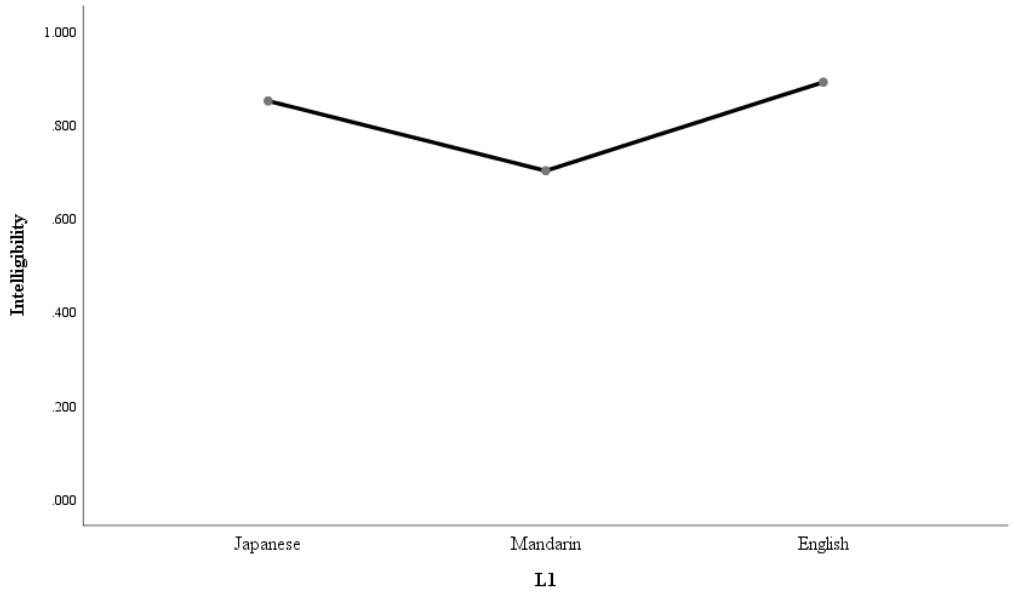


Figure 6. Three different L1 groups' intelligibility scores

Next, this study examined whether the influence of L1 differences differs depending on speakers' English proficiency. Table 11 displays a summary of the three different L1 groups' intelligibility scores for each of the two speaker groups. A two-way repeated measures ANOVA test was conducted to see whether the listeners' L1 and the speakers' proficiency had an influence on intelligibility. The between-subjects factor was listeners' L1 (Japanese, Mandarin, and English) and the within-subjects factor was proficiency level (high and low). It should be noted that similar to one-way independent ANOVA, there are several preliminary assumptions for conducting a two-way repeated measures ANOVA. In addition to random sampling and normal distribution of data, the data should

have sphericity (Takeuchi & Mizumoto, 2014). Sphericity was automatically met in this analysis since sphericity is an issue only when levels are more than two (Field, 2018). The following variables were not normally distributed: English listeners' intelligibility scores of Japanese HP, and Mandarin listeners' intelligibility scores of both Japanese HP and LP. (See Appendix H for the results of normality tests). However, as discussed above, ANOVA is robust in terms of normality and homogeneity of variance (Field 2018; Takeuchi & Mizumoto, 2014). Furthermore, as Field notes, there is no equivalent model for the two-way repeated measures ANOVA for non-parametric tests. Based on this fact, Tsushima (2010) argues that ANOVA is acceptable even when the relevant requirements are not satisfied. Accordingly, this study applied a two-way repeated measures ANOVA.

Table 11

Raw intelligibility scores in two different speaker groups

	Speakers			
	Japanese HP		Japanese LP	
Listeners	*M*	*SD*	*M*	*SD*
Japanese	0.85	0.05	0.85	0.06
Mandarin	0.76	0.14	0.64	0.16
English	0.93	0.04	0.85	0.06

The results of the two-way repeated measures ANOVA are displayed in Table 12 and Figure 7. It was revealed that the main effect of both listeners and speakers' proficiency were significant, F (2, 64) = 28.10, $p < .001$, $\eta 2 = .468$ and F (1, 64) = 60.34, $p < .001$, $\eta 2 = .485$. The interaction effect of Speakers' proficiency×Listeners' L1 was also significant, F (2, 64) = 15.23, $p < .001$, $\eta 2 = .323$.

Table 12

Result of two-way repeated measures ANOVA: Listeners' L1×Speakers' proficiency in Intelligibility scores

Source	*SS*	*df*	*MS*	*F*	*p*	η2
			Between Subjects			
Listeners' L1	.92	2	.459	28.10	<.001	.468
Error	1.05	64	.02			
			Within Subjects			
Speakers' proficiency	.16	1	.16	60.34	<.001	.485
Speakers' proficiency×Listeners' L1	.08	2	.04	15.23	<.001	.323
Error (Speakers' proficiency)	.17	64	.00			
Total	2.38	133				

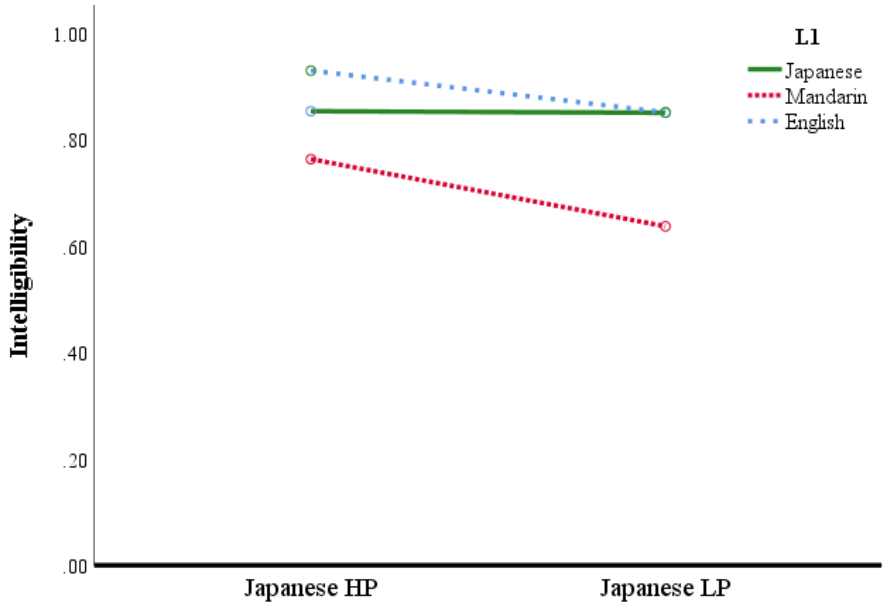


Figure 7. Three different L1 groups' intelligibility scores for each of the two speaker groups

Two post hoc tests were performed to interpret the interaction further: a Games-Howell test and a paired t-test. A Games-Howell test was used to conduct multiple comparisons

of speaker intelligibility within each listener group. The result is summarized in Table 13. As can be interpreted from Table 11 and 13, the intelligibility scores were significantly different among all the listeners when speakers' proficiency was high, with English L1 achieving the highest scores, followed by Japanese, and Mandarin. When the speakers' proficiency was low, the order of scores were the same, however, Japanese and English listeners' intelligibility scores were not significantly different (p = .999). What should also be noted here is that the effect size of Japanese and Mandarin L1 in LP speakers was higher than that in HP speakers (d = 1.71 for Japanese LP and d = 0.83 for Japanese HP). These results suggest that a shared L1 effect was stronger when speakers' proficiency was low.

Table 13

Results of Games-Howell test for intelligibility

Speaker	Listener Pairs	*p*	*d*
Japanese HP	Japanese-Mandarin	.024	0.83
	Japanese-English	<.001	1.7
	Mandarin-English	<.001	1.68
Japanese LP	Japanese-Mandarin	<.001	1.71
	Japanese-English	.999	0.01
	Mandarin-English	<.001	1.79

A paired t-test was run to compare intelligibility scores for a speaker across listener groups. Table 14 provides the result of the statistical analysis. Based on Table 11 and 14, it is clear that when the listening participants were Japanese, there were no statistically significant differences in intelligibility scores between high and low proficiency speaking groups. On the other hand, speakers' proficiency differences were significantly associated with intelligibility scores in the case where participants were Mandarin or English L1 listeners.

Table 14

Results of t-test for intelligibility

Listeners	Speaker Pair	*p*	*d*
Japanese	Japanese HP-LP	.788	0.05
Mandarin	Japanese HP-LP	<.001	0.83
English	Japanese HP-LP	<.001	1.55

4.3.2 Influence of Listeners' L1 and Speakers' Proficiency on Comprehensibility

Table 15 describes raw scores of the three different L1 groups' comprehensibility ratings. It is clear that Japanese L1 listeners rated the utterances easier to understand (M = 3.97) than the other two L1 groups: Mandarin (M = 5.25) and English (M = 4.59). To investigate whether their scores were significantly different from each other, similar to intelligibility, a one-way independent ANOVA was conducted. A Levene's test revealed that the p-value was more than 0.05 (p = .238), indicating the variance was homogenous. To test normal distribution of data, two normality tests, Kolmogorov-Smirnov and Shapiro-Wilk tests, were conducted. While the results of the Shapiro-Wilk test indicated that all the variables were normally distributed, that of the Kolmogorov-Smirnov test showed English L1 listeners' comprehensibility ratings were not normally distributed (See Appendix I for the results of normality tests). Although most of the assumptions seem to be met, to be cautious, the Kruskal-Wallis test was also conducted. The result was similar to that of the one-way independent ANOVA, thus here only the result of the one-way independent ANOVA is presented (see the result of the Kruskal-Wallis test in Appendix I).

Table 15

Raw comprehensibility ratings

Listeners	*M*	*SD*
Japanese	3.97	1.25
Mandarin	5.25	1.05
English	4.59	0.91

Note: Comprehensibility was measured with a nine-point Likert scale, with the lower ratings indicating the more comprehensible speech

The result of the one-way independent ANOVA is described in Table 16 and Figure 8. The result was F (2, 64) = 7.51, p = .001, $\eta 2$ = .196. This suggests that there were significant group differences in comprehensibility ratings and strong effects of L1 differences. As a post hoc analysis, a Games-Howell test was conducted to see which variables had significant differences. It was revealed that only Japanese and Mandarin L1 had significant comprehensibility rating differences (p = .003, d = 1.11). There were no significant comprehensibility rating differences between Japanese and English L1 (p = .171, d = 0.57), and Mandarin and English L1 (p = .068, d = 0.68).

Table 16

Result of one-way independent ANOVA of comprehensibility ratings

Source	*SS*	*df*	*MS*	*F*	*p*	η2
L1	17.09	2	8.55	7.51	.001	.190
Error	72.81	64	1.14			
Total	89.90	66				

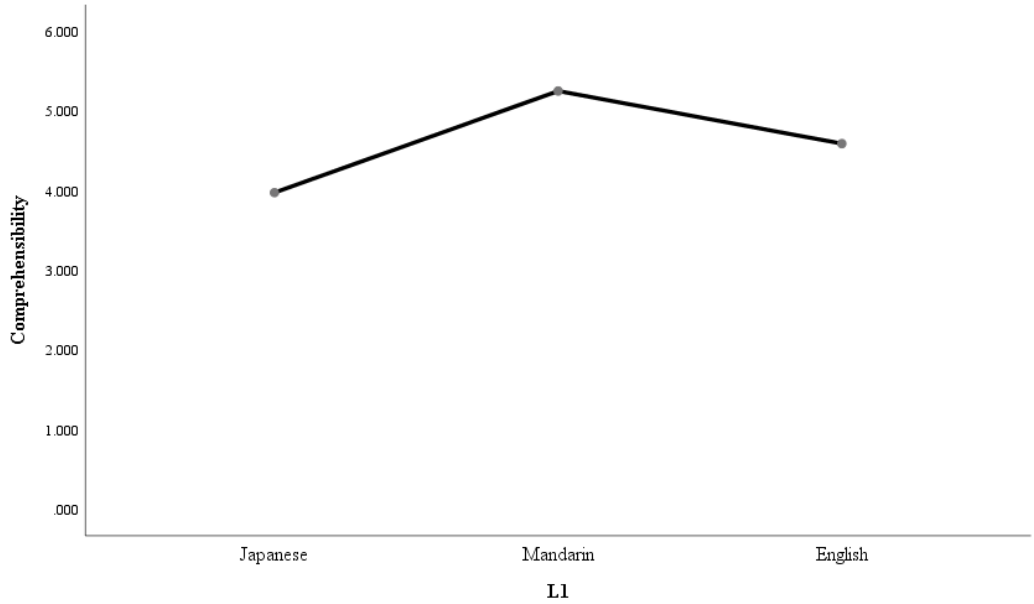


Figure 8. Three different L1 groups' comprehensibility ratings

Similar to intelligibility, next, whether the effects of L1 differences differs depending on speakers' English proficiency was examined. Table 17 summarizes three different L1 groups' comprehensibility ratings for each of the two speaker groups. Before two-way repeated measures ANOVA was conducted, the distribution of data was analyzed. While the results of the Shapiro-Wilk tests indicated that all variables were normally distributed, the results of the Kolmogorov-Smirnov tests showed that English L1 listeners' comprehensibility ratings of Japanese HP were not normally distributed. (See Appendix I for the results of normality tests). However, similar to the analysis of intelligibility, two-way repeated measures ANOVA was carried out since, as discussed in 4.3.1, ANOVA is a robust analysis method and there are no alternative analysis methods. Table 18 and Figure 9 present the results. It was found that both listeners and speakers' proficiency effects were statistically significant, F (2, 64) = 13.06, $p < .001$, $\eta 2 = .290$ and F (1, 64) = 127.34, $p < .001$, $\eta 2 = .666$. The interaction of Speakers' proficiency×Listeners' L1 was also statistically significant, F (2, 64) = 35.87, $p < .001$, $\eta 2 = .528$.

Table 17

Raw comprehensibility ratings in two different speaker groups

	Speakers			
	Japanese HP		Japanese LP	
Listeners	*M*	*SD*	*M*	*SD*
Japanese	3.97	1.25	3.93	1.26
Mandarin	5.22	1.06	6.01	1.08
English	4.59	0.91	5.40	0.94

Note: Comprehensibility was measured with a nine-point Likert scale, with the lower ratings indicating the more comprehensible speech

Table 18

Result of two-way repeated measures ANOVA: Listeners' L1×Speakers' proficiency in comprehensibility ratings

Source	*SS*	*df*	*MS*	*F*	*p*	η2
			Between Subjects			
Listeners' L1	58.90	2	29.45	13.06	<.001	.290
Error	144.29	64	2.26			
			Within Subjects			
Speakers' proficiency	8.94	1	8.94	127.34	<.001	.666
Speakers' proficiency×Listeners' L1	5.03	2	2.52	35.87	<.001	.528
Error (Speakers' proficiency)	4.49	64	.07			
Total	221.65	133				

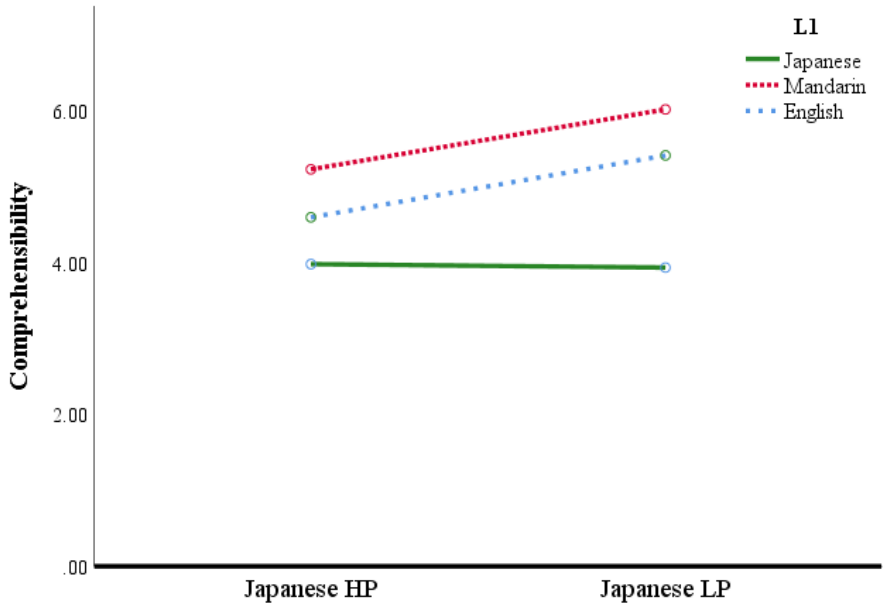


Figure 9. Three different L1 groups' comprehensibility scores for each of the two speaker groups

Since a significant interaction was observed, two post hoc tests were performed: a Games-Howell test and a paired t-test. The Games-Howell test was performed to assess the relationship between each speaker's comprehensibility and each listener L1 group. The results are summarized in Table 19. Based on Table 17 and 19, it can be interpreted that there were only statistically significant differences between Japanese and Mandarin L1 listeners in the case where speakers were Japanese HP. The discrepancy between Japanese listeners and other L1 listeners increased when the speakers were Japanese LP. While there were no statistically significant differences between ratings given by English and Mandarin L1 listeners (p = .087), there were statistically significant differences between Japanese and the other two listener groups. This is clearly seen in Figure 9 as well. It should also be noted that, in contrast to intelligibility scores, Japanese L1 listeners' comprehensibility ratings were the highest compared to the other two L1 groups

regardless of speakers' English level. Additionally, the effect size suggests the effect of a shared L1 increases when the speakers' English proficiency was low. While the effect size of Japanese and Mandarin L1 in LP speakers was 1.78, that in HP was 1.08. Similarly, the effect size of Japanese and English L1 in LP speakers was 1.35. By contrast, that in HP was 0.57.

Paired t-tests were calculated to compare comprehensibility ratings for a speaker across listener groups. The results are reported in Table 20. Based on Table 17 and 20, it is clear that when the listening participants were Japanese, there were no statistically significant comprehensibility rating differences between Japanese HP and LP. In comparison with this, speakers' proficiency differences were significantly associated with comprehensibility ratings in the case where participants were Mandarin or English L1 listeners.

Table 19

Results of Games-Howell test for comprehensibility

Speaker	Listener Pairs	*p*	*d*
Japanese HP	Japanese-Mandarin	.004	1.08
	Japanese-English	.171	0.57
	Mandarin-English	.087	0.64
Japanese LP	Japanese-Mandarin	<.001	1.78
	Japanese-English	<.001	1.35
	Mandarin-English	.115	0.6

Table 20

Results of t-test for comprehensibility

Listeners	Speaker Pair	*p*	*d*
Japanese	Japanese HP-LP	.587	0.04
Mandarin	Japanese HP-LP	<.001	0.74
English	Japanese HP-LP	<.001	0.88

4.4 Influence of Listeners' Familiarity with Japanese Accented English on Intelligibility and Comprehensibility

To address the second research question, correlation analyses were performed to examine the relationship between familiarity with Japanese accented English (self-evaluation of familiarity with Japanese accented English and frequency of interaction with Japanese) and two other factors: intelligibility scores and comprehensibility ratings of Mandarin and English L1 listeners. According to Hirai (2017), one of the preliminary assumptions of the parametric testing of correlation analysis is the normal distribution of data. Tests of normality, Kolmogorov-Smirnov and Shapiro-Wilk were run and it was found that some variables were not normally distributed: frequency of interactions with Japanese people in English, intelligibility (combined HP and LP), intelligibility (Japanese HP), and intelligibility (Japanese LP). (See Appendix J for the results of normality tests). Thus, a nonparametric test, the Spearman's rank correlation, was applied in this study.

Table 21 and 22 describe the results of correlation analysis of intelligibility scores and comprehensibility ratings, respectively. The first thing to report here is that the self-reported familiarity with Japanese accented speech strongly correlated with frequency of interactions at a significant level (ρ = -.725, p < .01). For the relationships between intelligibility scores and familiarity with Japanese accented speech, both combined intelligibility scores and LP intelligibility scores were significantly correlated with the

self-reported familiarity with Japanese accented English (ρ = -.315, p < .05 for combined HP and LP, ρ = .351, p < .05 for Japanese LP). Similarly, all the intelligibility scores and frequency of interactions had positive correlations (r = .496, p < .01 for combined HP and LP, ρ = .438, p < .01 for Japanese HP, and ρ = .507, p < .01 for Japanese LP). These results indicated that the participants who were very familiar with Japanese English or interacted with Japanese people in English often had higher intelligibility scores than participants less familiar with Japanese accented English and who had fewer interactions with Japanese people.

Table 21

Results of Correlation Analysis for Familiarity with Japanese Accented Speech and Intelligibility Scores

		1	2	3	4	5
1.	Self-reported familiarity with Japanese English	—				
2.	Frequency of interaction with Japanese	-.725**	—			
3.	Intelligibility (combined HP and LP)	-.315*	.496**	—		
4.	Intelligibility (Japanese HP)	-.248	.438**	.956**	—	
5.	Intelligibility (Japanese LP)	-.351*	.507**	.978**	.891**	—

Note. **p<.01, *p<.05.

By contrast, the statistical analysis yielded weaker associations of the comprehensibility ratings with familiarity with Japanese accented English overall. The self-reported familiarity with Japanese accented English was only linked to comprehensibility in the case of low proficiency speakers (ρ = .347, p < .05). There was no significant connection

between the frequency of interactions with Japanese people in English with each of the comprehensibility groups. Overall, the findings indicated that familiarity with accented speech had stronger relationships with intelligibility than comprehensibility. Additionally, there was a tendency for familiarity with accented speech to correlate with the ratings of the speech acts of lower proficiency speakers more than that of the high proficiency speakers. It is also interesting to note that although a strong and positive correlation existed between frequency of interactions and familiarity with Japanese English, their correlations with intelligibility and comprehensibility were different.

Table 22

Results of Correlation Analysis for Familiarity with Japanese Accented Speech and Comprehensibility Ratings

		1	2	3	4	5
1.	Self-reported Familiarity with Japanese English	—				
2.	Frequency of interaction with Japanese	-.725**	—			
3.	Comprehensibility (combined HP and LP)	.239	-.211	—		
4.	Comprehensibility (Japanese HP)	.063	-.142	.910**	—	
5.	Comprehensibility (Japanese LP)	.347*	-.245	.925**	.712**	—

Note. $^{**}p<.01$, $^{*}p<.05$.

Scatter plots are presented in Appendix J for the relationships between familiarity with Japanese accented English (self-evaluation of familiarity with Japanese accented English and frequency of interaction with Japanese people) and two other factors: intelligibility scores and comprehensibility ratings of Mandarin and English L1 listeners.

4.5 Influence of Listeners' English Proficiency on Intelligibility and Comprehensibility

To answer the third research question, correlation analyses were performed to examine the relationship between listeners' English proficiency and two other factors: intelligibility scores and comprehensibility ratings of Japanese and Mandarin L1 listeners. Some variables were not normally distributed: English proficiency, intelligibility (combined HP and LP), intelligibility (Japanese HP), and intelligibility (Japanese LP). (See Appendix K for the results of normality tests). Thus, a nonparametric test, the Spearman's rank correlation, was applied. As seen in Table 23, there were no significant correlations between listeners' English proficiency and any of the intelligibility scores and comprehensibility ratings. This means listeners' English level was not correlated with either intelligibility or comprehensibility in this study.

Scatter plots are presented in Appendix K for the relationships between listeners' English proficiency and two other factors: intelligibility scores and comprehensibility ratings of Japanese and Mandarin L1 listeners.

Table 23

Results of Correlation Analysis between Listeners' English level and Intelligibility and Comprehensibility

	Listeners' English level
Intelligibility (combined HP and LP)	.040
Intelligibility (Japanese HP)	.170
Intelligibility (Japanese LP)	-.062
Comprehensibility (combined HP and LP)	.067
Comprehensibility (Japanese HP)	-.002
Comprehensibility (Japanese LP)	.157

Note. All results were not statistically significant.

Chapter 5. Discussion

The main purpose of this chapter is to further discuss the results presented in chapter 4 in relation to each research question. The results of the current study are also discussed in reference to previous related studies. In the first section, the influence of L1 differences on intelligibility and comprehensibility are discussed. Next, the associations between familiarity with an accented English and two dimensions of speech understanding (intelligibility and comprehensibility) are shown. Following this, the association between listeners' English proficiency and both intelligibility and comprehensibility is considered. Finally, pedagogical implications based on the findings are explored.

5.1 Research Question 1. Are listeners' L1 differences associated with the intelligibility and comprehensibility of non-native English speech? (i.e., Is there a shared L1 effect?)

5.1.1 Influence of Listeners' L1 on Intelligibility

In this study, firstly, the influence of the listeners' L1 on intelligibility was analyzed. The results of the one-way independent ANOVA showed that there were significant intelligibility score differences between the listeners with different L1s, with English being the highest, followed by Japanese and Mandarin. The significant intelligibility score differences between Japanese and Mandarin listeners suggested the existence of a shared L1 effect.

This result moderately changed when speakers' proficiency was added as a factor to the analysis. The results of the two-way repeated measures ANOVA showed that there was a significant Speakers' proficiency×Listeners' L1 interaction effect. Furthermore, a post hoc analysis revealed that while Mandarin and English L1 listeners' intelligibility scores

significantly lowered in the case where speakers' English level was low, Japanese listeners maintained high intelligibility scores regardless of speakers' English proficiency. This also suggests that there was a shared L1 effect. Additionally, the effect size of the relationships between Japanese and Chinese L1 listeners was stronger in the case where speakers were LP rather than HP. This suggests the shared L1 effect increased when speakers English level was low.

The current study's findings support Bent and Bradlow (2003) and Stibbard and Lee (2006). Their studies suggest that while the utterances by high proficient speakers tend to be highly intelligible to all listeners regardless of L1, the speech acts by low proficient speakers tend to be as intelligible as NS speech to listeners who share the same L1 as the speaker.

By contrast, the results of this study are not in line with those reported in Major et al. (2002), Munro et al. (2006), and Harding (2012). In the study by Major et al., a shared L1 effect was observed only in native Spanish speakers and listeners interactions. The effect was not observed in Japanese and Cantonese L1 participants. The study by Munro et al. (2006) showed some shared L1 effect in Japanese participants, but the effect was very limited. A study by Harding (2012) showed that, while Mandarin L1 participants showed a shared L1 effect, it was not seen in Japanese L1 participants. One of the differences between the studies which show a strong shared L1 effect and those that show a partial effect is whether they take into account the speakers' English level. This might be one explanation for why the discrepancy occurred. Another possible explanation could be the influence of the kind of task used in the studies. The results of Harding pointed out that a shared L1 effect in Mandarin L1 speakers' interactions depended on certain questions,

suggesting the type of task affects the results. Therefore, this could be another possible factor which influenced the discrepancy of the results.

5.1.2 Influence of Listeners' L1 on Comprehensibility

Similar to intelligibility, the statistical results in this study indicated the existence of a shared L1 effect in comprehensibility. The relationships between listeners' L1 differences and comprehensibility was analyzed with a one-way independent ANOVA. The results showed that while Japanese listeners' comprehensibility ratings were significantly higher than Mandarin L1 listeners, there were no significant differences between Japanese and English L1 listener groups. The significant comprehensibility rating differences between Japanese and Mandarin L1 listeners suggest that there was a shared L1 effect.

When speakers' proficiency was added as a factor to the analysis, the effects of L1 differences were clearer. The results of two-way repeated measures ANOVA showed that there was a significant Speakers' proficiency×Listeners' L1 interaction effect. A post hoc analysis revealed that when the speakers' proficiency was low, Japanese listener groups considered it to significantly easier to understand the speech acts compared to the other two L1 groups. Additionally, while Mandarin and English L1 listeners' comprehensibility ratings significantly lowered in the case where the speakers' English level was low, Japanese listeners maintained high comprehensibility ratings regardless of speakers' English proficiency. Furthermore, the effect size of the relationships between Japanese listeners and the other two L1 listener groups were stronger in the case where the speakers were LP rather than HP. Taken together, these results suggest that there is a shared L1 effect and that the advantage was greater when the speakers' English proficiency was low.

The finding of the current study accords with Ludwig and Mora (2017). Their overall conclusion was that sharing an L1 could enhance comprehensibility. However, it should be noted that, when listeners' English proficiency was taken into account, the result was slightly changed in their study. While both low- and high- English level German L1 listeners showed a significant shared L1 effect, only low proficient Catalan listeners showed a significant shared L1 effect.

By contrast, the result of the current study did not support Munro et al. (2006) and Crowther et al. (2016). Munro et al., showed that, while Japanese listeners' comprehensibility ratings of Japanese speakers' speech acts were higher than Cantonese and Mandarin listeners, there were no perception differences of Japanese English between Japanese and native English listeners. Additionally, all listener groups perceived Cantonese speakers' English similarly in their study. Similar to intelligibility, the speaker's English proficiency differences and task differences could be the underlying reasons that caused the divergent results.

5.2 Research Question 2. Is a listeners' familiarity with speakers' L1 accented English associated with the intelligibility and comprehensibility of non-native English speech?

5.2.1 Influence of Listeners' Familiarity with Japanese English on Intelligibility

This study measured listeners' familiarity with speakers' L1 accented speech from two different perspectives: self-evaluated familiarity with Japanese accented English and frequency of interactions with Japanese people in English. Correlation analyses indicated that self-evaluated familiarity with Japanese accented English correlated weakly with the intelligibility scores of the combined HP and LP (-.315, $p < .05$) and the Japanese LP (ρ

= .-351, $p < .05$). The frequency of interactions with Japanese people in English correlated with intelligibility scores of combined HP and LP, Japanese HP, and Japanese LP. It should be noted here that Japanese LP had a stronger correlation coefficient than Japanese HP ($\rho = .438$, $p <. 01$ for LP and $\rho = .507$, $p < .01$ for HP).

These results corroborate Kennedy and Trofimovich (2008) and Matsuura (2007). Both studies reported that familiarity with accented English could enhance intelligibility. However, it should be noted that their studies were different from the current study in that they did not take into account speakers' proficiency and they were focusing on general accented English and not limited to speakers' L1 accented English.

By contrast, the results of the current findings were different from Matsuura et al. (1999). In their study, the listening participants who were not familiar with Irish English had significantly higher intelligibility scores than the listening participants who were familiar with Irish English. One possible reason for this discrepancy would be, as discussed in chapter 2, that the listening participants who were not familiar with Irish English had a higher English level than those who were familiar with Irish English.

Another possible reason could be the influence of speakers' English pronunciation differences. The speaking participants in Matsuura et al. (1999) were Irish native English speakers. Their English pronunciation deviates from General American English, but the degree of deviation must have been smaller than NNS. Nihalani (2010) notes "the consonant system of English is relatively uniform throughout the English-speaking country" (p. 31). In the current study, the results suggested that when speakers' English proficiency in pronunciation is high, the influence of familiarity with accented English is

smaller. Therefore, Irish people's relatively small deviation from General American English might have reduced the influence of listeners' familiarity with accented English speech.

5.2.2 Influence of Listeners' Familiarity with Japanese English on Comprehensibility

On the other hand, as for the relationships between comprehensibility ratings and familiarity with Japanese English, there is only a positive significant correlation between the self-evaluated familiarity with Japanese accented English and comprehensibility ratings of Japanese LP ($\rho = .347$, $p < .05$). There was no significant correlation between frequency of interactions with Japanese people in English and comprehensibility ratings. One simple interpretation of this result is that familiarity with Japanese accented English is associated with comprehensibility when speakers' English level was low. However, unlike the results of intelligibility scores, the frequency of interaction with Japanese people in English and comprehensibility scores were not correlated. This would make it difficult to reach a definite conclusion.

Past studies also reported conflicting results. While the findings of Kennedy and Trofimovich (2008) showed that familiarity with accented speech had no effect on comprehensibility, the study by Matsuura et al. (1999) indicated there was a significant influence of familiarity with accented English on comprehensibility. As mentioned above, both studies were different from this study in terms of the types of interactions and speech materials. Additionally, they did not take into account speakers' English proficiency. These differences might create different results, but this cannot be established with certainty.

One unexpected result regards the two following variables: self-reported familiarity with Japanese accented English and the frequency of interaction with Japanese people in English. While they were strongly correlated (ρ = -.725, p < .05), their relationships with both intelligibility and comprehensibility were different. It would be difficult to discuss this matter further since past studies have not discussed the methods for measuring familiarity with accented English.

5.3 Research Question 3: Is a listeners' English proficiency associated with the intelligibility and comprehensibility of non-native English speech?

5.3.1 Influence of Listeners' English Proficiency on Intelligibility

The results of correlation analyses showed that, regardless of a speakers' English proficiency, intelligibility did not significantly correlate with listening participants' (Japanese and Mandarin L1) English level. However, it should be noted that intelligibility scores of English L1 listeners was higher than the other two listener groups. This might suggest English proficiency influences intelligibility scores.

The results of this study contradict Matsuura et al. (1999) and Matsuura (2007). In Matsuura et al., Japanese listening participants' intelligibility scores (in relation to all six speaking participants) were significantly correlated with their TOEFL ITP scores. Similarly, in Matsuura, stepwise regression analysis revealed that listeners' self-reported English proficiency predicted intelligibility scores when Japanese L1 listeners listened to both American English and Hong Kong L1 speakers' English.

One possible explanation for this discrepancy could be that the listening participants' English level in this study was generally higher compared to both Matsuura et al. (1999)

and Matsuura (2007). Both listening participants in Matsuura et al. and Matsuura were low to intermediate. In Matsuura et al., the listening participants' TOEFL ITP scores ranged from 350 to 540 (M = 440), suggesting their English level would be between A2 to B2 in the CEFR. Similarly, Matsuura's listener's TOEIC scores were from 360 to 775 (M = 528.75), indicating their English level would be from A2 to B2 in the CEFR. By contrast, in the current study, participants' scores on standardized English tests and their self-evaluated English proficiency indicated their English level to be approximately from B1 to C2. The high proportion of high-level English listeners could reduce the influence of English proficiency on intelligibility.

Another possible explanation could be that shared L1 and familiarity with the speakers' L1 accented English influenced intelligibility to a greater extent. As discussed above, shared L1 and familiarity with the speakers' L1 accented English is strongly associated with intelligibility. The condition of this study was different from the other two studies, Matsuura et al. (1999) and Matsuura (2007). Both studies' speaking and listening participants did not share an L1. Additionally, although Matsuura et al.'s study was similar to the current study, in that it took into account familiarity with speakers' English accent, their study's speaking participants were American and Irish NS and not NNS. Furthermore, Matsuura only measured overall familiarity with accented English and not familiarity with speaking participants' L1 accented English. She reported that listening participants were not familiar with Hong Kong English. Therefore, there is a possibility that these study-design differences affected the results.

The fairly low-level speech material might also reduce the influence of proficiency on intelligibility. In the current study all the vocabulary in the speech data consisted of the

most frequent 2,000 words. There is a possibility that this reduced the influence of proficiency on intelligibility. However, both Matsuura et al. (1999) and Matsuura (2007) used fairly easy words in their speech material as well. Therefore, the influence would be limited compared to the other two possible explanations.

5.3.2 Influence of Listeners' English Proficiency on Comprehensibility

Similar to the relationships between the listeners' English proficiency and intelligibility score, the current research showed that there was no significant correlation between the listeners' English proficiency and comprehensibility ratings regardless of a speakers' English proficiency. The result is partially similar to Matsuura et al. (1999). Their study showed only three out of six speaking participants' comprehensibility scores were correlated with the listening participants' TOEFL ITP scores, with relative low correlation coefficient ranging from 0.27 to 0.32. Since the study by Matsuura et al. focused on NNS's perception of NS's speech, and not on NNS and NS's perception of NNS's speech (as was the case in the current study), we cannot directly compare the results. However, these results suggest that listeners' English proficiency on comprehensibility would be limited. It should be also noted that contrary to intelligibility scores, English L1 listeners' comprehensibility ratings were lower than Japanese L1 listeners in this study. This fact could also support the above conclusion.

5.4 Pedagogical Implications

The results of the current research indicate that sharing an L1 and familiarity with the speakers' L1 accented English could positively influence intelligibility and comprehensibility. Additionally, it was revealed that having a high level of English does not necessarily lead to a better understanding of NNS's English.

One pedagogical implication based on the above findings is the necessity of not relying only on archetypal ‘Native Speaker’ models in the classroom. Instead English language teachers might be encouraged to provide models of English representing a range of different accents and levels of proficiency. As discussed in Chapter 1, English is now regarded as a global language and many variations of it now exist. Additionally, the vast majority of English communication is shifting from NSs’ interactions and interactions between NS and NNS, to interactions among NNSs. Consequently, it is crucial to be familiar with various types of accented English.

However, it should be noted that the effect would be weaker when the speakers’ English level is high. This suggests that if the speakers’ English level is high, the listeners do not necessarily need to be familiar with accented English. In fact, the results of Bent and Bradlow (2003) and Stibbard and Lee (2006) showed that there was a tendency for English produced by high proficient NNS to be almost equally intelligible as that produced by NS even when the speakers and the listeners did not share the same L1. This suggests that if the learners’ potential interlocutors are NS or highly proficient speakers, the necessity of familiarizing themselves with accented English would decrease. Therefore, English language teachers need to take into consideration the demand of learners when they teach them.

Chapter 6. Conclusion

The purpose of this study was to investigate how the listeners' background is associated with their perceptions of intelligibility and comprehensibility of English spoken by Japanese learners of English. Sixty-seven listeners from three different L1 backgrounds listened to English speech acts produced by 32 Japanese learners of English. The listeners were asked to transcribe the speech acts and rate the ease of understanding in order to measure intelligibility and comprehensibility, respectively. There were three main findings in this study. Firstly, L1 differences influence both intelligibility and comprehensibility. A strong positive shared L1 effect was observed. This tendency was especially strong when speakers' English proficiency was low. Secondly, familiarity with speakers' L1 accented English is associated with comprehensibility. The participants who rated themselves very familiar with Japanese accented English and who reported having frequent interactions with Japanese people in English rated intelligibility more highly. This was especially seen when the speakers' English level was low. By contrast, it seemed that familiarity with accented speech was less associated with comprehensibility. Thirdly, listeners' English level seemed not to be associated with either intelligibility or comprehensibility. In the following sections, limitations of this study and suggestions for future studies will be explored.

6.1 Limitations of Current Study

The first limitation of the current study is the inclusion of potentially unrepresentative samples. Many of the non-native English listening participants in this study had either an upper intermediate or higher level of English. It would have been ideal to have an almost equal proportion of low proficient non-native English listening participants to categorically prove there is no link between listeners' English proficiency and

intelligibility and comprehensibility. Additionally, Mandarin L1 listening participants' familiarity with Japanese accented English is limited. Many of the Mandarin L1 listening participants reported that they hardly interact with Japanese people in English. Similar to English proficiency, it would be ideal to have a greater number of participants who frequently interacted with Japanese people in English.

The next limitation would be the limited L1 group of speakers. In this study, only Japanese speakers of English were used. It would have been ideal to use more than one L1 group in order to clarify a shared L1 effect.

Another limitation would be the criteria for defining English proficiency. The current study divided the speakers into high and low proficiency levels by following the evaluation of the speakers provided by experienced American NS EFL teachers. Similarly, Bent and Bradlow (2003) and Stibbard and Lee (2006) used judgments of NS as a criterion for dividing their participants into high and low proficiency levels. However, it is questionable whether NS evaluation is completely reliable in the global Englishes context. Jenkins (2000) reported that not all pronunciation deviation from standardized English affects intelligibility in NNS-NNS interactions. For example, a deviation of dental fricative sounds, /θ/ and /ð/, would be marked as a deviation, but does not necessarily hinder communication in NNSs' interactions. Additionally, some studies show native English is not necessarily intelligible to NNSs (e.g., Smith, 1992; Smith & Rafiqzad, 1979). These studies suggest that using NS norm would not be necessarily reliable for pronunciation evaluation. However, it is also true that since non-native English has not been codified (Dimova, 2018; Jenkins, 2006), using the native norm is one of the few options available to evaluate NNS's proficiency in terms of pronunciation.

An additional limitation is that this study only measured familiarity with Japanese accented English. Obviously, people interact with a variety of forms of accented English. Therefore, taking into account the practicality of this study, it would be necessary to measure listeners' familiarity with accented English in general as well. However, it should be noted that the measuring methods of familiarity with accented English do not seem to be established. This study measured it with two criteria: the self-evaluation of familiarity with Japanese accented English and frequency of interactions with Japanese people in English. Interestingly, the relationships of these two criteria and intelligibility were different from each other, which raises the question about how to measure familiarity with accented English.

A further limitation would be the scoring methods used for the transcription task. In this study, intelligibility scores were determined by calculating the number of words correctly transcribed. Each word has the same allocation of marks. In the task, some examples of mis-transcribed functional words such as articles and pronouns were observed. These words would have less effect on speech understanding compared to content words (i.e., nouns, verbs, adjectives, and adverbs). It would be necessary to consider the allocation of marks to content words and functional words (e.g. to give more marks to content words or only give marks to content words).

It should also be noted that the transcription task itself has limitations. As noted in chapter 2, Zielnski (2004, as cited in Munro et al. 2006) shows that the scores of transcription tasks were not necessarily correlated with the understanding of the whole message and provided examples where "all the words were correctly identified, with the listener still puzzling over what the speaker was trying to communicate" (p. 113). In the current study,

an example of this phenomenon was observed. One of the listening participants transcribed a sentence, ‘They need cars for their performance’ as ‘Did you need cars for their performance?’. Although the participant almost perfectly transcribed the sentence, the participants understanding of the utterance would be totally different from what the speaker actually intended to convey.

In addition to this, a transcription task has a limitation in terms of ecological validity. Gass and Varonis (1984) applied a transcription task to measure intelligibility similar to this study. One of the participants in their study reported that “I could understand the man speaking because I had a few seconds to think about what he said. I think it would be very difficult to carry on a conversation with him” (p. 84). This clearly indicates the limitation of a transcription task.

There are several possible solutions for these intelligibility measurement method issues. The first solution would be to conduct more than one measurement method of intelligibility and compare it. In fact, Munro and Derwing (2015) state each intelligibility measurement method has limitations and they recommend conducting more than one measuring method. Another solution would be to conduct research in a more naturalistic setting, such as observing real English communication, to enhance ecological validity.

The final limitation of this study is due to the usage of an online task format. The issue here is that the researcher cannot control the listening environment. Another weakness of using an online task format is the lack of strict control of listening participants. Although the participants were instructed to listen to each sound file once per task, there is a possibility that they listened to the recordings more than once per task.

6.2 Suggestions for Future Studies

The first suggestion for future studies is to replicate the current study with different methods and populations to assess whether the findings can be reproduced in different contexts. One replication method would be conducting the research with different intelligibility and comprehensibility measurement methods such as summarization and shadowing. Additionally, combining several intelligibility measurement methods, such as using both a transcription task and summarizations, could also clarify the nature of intelligibility.

The next suggestion is to see the relationships between the two dimensions of speech understanding (intelligibility and comprehensibility) and other listener factors. There are still many potential listener factors involved in intelligibility and comprehensibility such as working memory, educational backgrounds, gender, and metacognition. Future research should take into account these factors.

A further suggestion is to conduct experimental classroom research. For example, divide students into two groups: one receives English lessons where language input includes a variety of types of accented English and the other receives English lessons where the language input is only from native English speakers. Then, compare the changes in their intelligibility scores and comprehensibility ratings of NNS speech acts before and after the lesson. This could suggest further practical pedagogical implications.

The final suggestion is to investigate the effect of listener factors on intelligibility and comprehensibility in a naturalistic setting with qualitative analysis. According to Kaur (2017), speakers and listeners use various interaction practices such as “repetition,

paraphrase, comprehension checks and confirmation and clarification requests" (p. 551). It would be meaningful to investigate how these factors and listeners' characteristics contribute to intelligibility and comprehensibility in terms of ecological validity.

References

Amano, N. (2012). *Contribution of pronunciation features to intelligibility and comprehensibility in non-native interaction: A case of Japanese and Chinese learners of English* (Unpublished master's thesis). Nagoya University, Nagoya.

Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning, 42,* 529-555.

Barrass, J. (2017). *The Intelligibility of Korean English pronunciation from a lingua franca perspective* (Master's thesis). Retrieved from British Council ELT Masters dissertations.

Bent, T., & Bradlow, A. R. (2003). The interlanguage speech intelligibility benefit. *The Journal of the Acoustical Society of America, 114,* 1600-1610.

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

Brennan, E. M., Ryan, E. B., & Dawson, W. E. (1975). Scaling of apparent accentedness by magnitude estimation and sensory modality matching. *Journal of Psycholinguistic Research, 4,* 27–36.

Burns, A. (2005). Interrogating new worlds of English language teaching. In A. Burns (Ed), *Teaching English from a global perspective.* Maryland: Teachers of English to Speakers of Other Languages.

Catford, J. C. (1950). Intelligibility. *ELT Journal, 5*(1), 7-15.

Cobb, T. M. (n.d.). Compleat web VP! [Computer program]. Compleat Lexical Tutor. Montréal: Université du Québec à Montréal. Retrieved from http://www.lextutor.ca/vp/comp/

Crowther, D., Trofimovich, P., & Isaacs. (2016). Linguistic dimensions of second language accent and comprehensibility. *Journal of Second Language Pronunciation, 2*(2), 160-182.

Crystal, D. (2003). *English as a global language.* Cambridge: Cambridge University Press.

Carter, R. (1998). Orders of reality: Cancode, communication and culture. *ELT Journal, 52,* 43-56.

Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition, 19,* 1-16.

Derwing, T., Munro, M., & Wiebe, G. (1998). Evidence in Favor of a Broad Framework for Pronunciation Instruction. *Language Learning, 48(3),* 393-410.

Dimova, S. (2018). Pronunciation assessment in the context of world Englishes. In O. Kang & A. Ginther (Eds.), *Assessment in second language pronunciation* (pp. 49-66). Oxon: Routledge.

English Speech Database Read by Japanese Students. Distributed by National Institute of Informatics. URL: http://research.nii.ac.jp/src/en/UME-ERJ.html

Fayer, J. M., & Krasinski, E. (1987). Native and nonnative judgments of intelligibility and irritation. *Language Learning, 37,* 313–326.

Field, A. (2009). *Discovering statistics using SPSS* (5th ed.). London: Sage.

Field, J. (2005). Intelligibility and the Listener: The role of lexical stress. *TESOL Quarterly, 39,* 399-423.

Foote, J. A., & Trofimovich, P. (2018). Is it because of my language background? A study of language background influence on comprehensibility judgments. *Canadian Modern Language Review, 74*(2), 253-278.

Galesic, M. & Bosnjak, M. (2009). Effects of questionnaire length on participation and indicators of response quality in a web survey, *Public Opinion Quarterly, 73*(2),

349-360.

Galloway, N. (2013). Global Englishes and English language teaching (ELT) – Bridging the gap between theory and practice in a Japanese context. *System, 41*(3), 786-803.

Galloway, N., & Rose, H. (2015). *Introducing global Englishes.* Abingdon: Routledge.

Gass, S., & Varonis, E. (1984). The effect of familiarity on the comprehensibility of nonnative speech. *Language Learning, 34,* 65–89.

Hahn, L. D. (2004). Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly, 38,* 201–223.

Harding, L. (2012). Accent, listening assessment and the potential for a shared-L1 advantage: A DIF perspective. *Language Testing, 29*(2), 163-180.

Hayes-Harb, Smith, Bent, & Bradlow. (2008). The interlanguage speech intelligibility benefit for native speakers of Mandarin: Production and perception of English word-final voicing contrasts. *Journal of Phonetics, 36*(4), 664-679.

Hirai, A. (2017). Kyoiku Shinrikeikenkyu no tameno deta bunseki nyumon [Introduction to data analysis for education and psychology research (2nd ed.)]. Tokyo: Tokyotosho.

Inoue, Y., Kabashima, S., Saito, D., Minematsu, N., Kanamura, K., Yamauchi, Y. (2018). A study of objective measurement of comprehensibility through native speakers' shadowing of learners' utterances. *Proc. Interspeech 2018,* 1651-1655.

Isaacs, T., &Trofimovich, P. (2012). Deconstructing comprehensibility: Identifying the linguistic influences on listeners' L2 comprehensibility ratings. *Studies in Second Language Acquisition, 34*(3), 475-505.

Jenkins, J. (2000). *The phonology of English as an international language*. Oxford: Oxford University Press.

Jenkins, J. (2002). A sociolinguistically based, empirically researched pronunciation

syllabus for English as an international language. *Applied Linguistics, 23*, 83-103.

Jenkins, J. (2006). The spread of EIL: A testing time for testers. *ELT Journal, 60*(1), 42-50.

Kang, O., Thomson, R. I., & Moran, M. (2018a). Empirical approaches to measuring the intelligibility of different varieties of English in predicting listener comprehension. *Language Learning, 68*(1), 115-146.

Kang, O., Thomson, R. I., & Moran, M. (2018b). Which features of accent affect understanding? Exploring the intelligibility threshold diverse accent varieties. Applied Linguistics, 2018, 1-29.

Kang, O., Thomson, R., & Murphy, J. (2017). *The Routledge Handbook of Contemporary English Pronunciation*. Abingdon: Routledge.

Kaur, J. (2017). Intelligibility in global contexts. In O. Kang, R. Thomson, & J. Murphy (Eds.), *The Routledge Handbook of Contemporary English Pronunciation* (pp. 542 -555). Abingdon: Routledge.

Kennedy, S., & Trofimovich, P. (2008). Intelligibility, comprehensibility, and accentedness of L2 speech: The role of listener experience and semantic context. *The Canadian Modern Language Review*, *64*(3), 459-489.

Levis, J. M. (2006). Pronunciation and the assessment of spoken language. In R. Hughes (ed.): *Spoken English, TESOL and applied linguistics: Challenges for theory and practice*. Palgrave Macmillan, 245-70.

Ludwig, A., & Mora. (2017). Processing time and comprehensibility judgments in non-native listeners' perception of L2 speech. *Journal of Second Language Pronunciation, 3*(2), 167-198.

Major, R. C., Fitzmaurice, S. F., Bunta, F., & Balasubramanian, C. (2002). The effects of nonnative accents on listening comprehension: Implications for ESL assessment.

TESOL Quarterly, 36, 173-190.

Matsuura, H. (2007). Intelligibility and individual learner differences in the EIL context. *System, 35*(3), 293-304.

Matsuura, H., Chiba, R., & Fujieda, M. (1999). Intelligibility and comprehensibility of

Minematsu, N. (2019, March 19). *How can speech technologies support learners to improve their skills of speaking, listen-ing, conversation and more?* CUED Speech Group Seminars, Cambridge.

Munro, M. J. (2017). Dimensions of pronunciation. In O. Kang, R. Thomson, & J. Murphy (Eds.), *The Routledge Handbook of Contemporary English Pronunciation* (pp. 413-431). Abingdon: Routledge.

Munro, M. J., & Derwing, T. M. (1995a). Foreign accent, comprehensibility and intelligibility in the speech of second language learners. *Language Learning*, *45,* 73-97.

Munro, M. J., & Derwing, T. M. (1995b). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language & Speech*, *38,* 289-306.

Munro, M. J., & Derwing, T. M. (2015). Intelligibility in research and practice: teaching priorities. In M. Reed & J. M. Levis (Eds.). *The handbook of English pronunciation* (pp. 377-396). Malden, MA: Wiley Blackwell.

Munro, M. J., Derwing, T. M., & Morton, S. L. (2006). The mutual intelligibility of L2 speech. *SSLA, 28,* 111-131.

Nelson, C. L. (1995). Intelligibility and world English in the classroom. *World Englishes, 14,* 273-279.

Nelson, C. L. (2008). Intelligibility since 1969. *World Englishes, 27,* 297-308.

Nelson, C. L. (2011). Intelligibility in world Englishes: Theory and application (ESL and

applied linguistics professional series). London: Routledge.

Nihalani (2010). Globalization and international intelligibility. In M. Saxena & T. Omoniyi (Eds.). *Contending with globalization in world Englishes.* Bristol: Multilingual Matters.

Thomson, R. (2018). Measurement of accentedness, intelligibility, and comprehensibility. In O. Kang & G. April (Eds.). *Assessment in second language pronunciation* (pp. 11-29). Abingdon: Routledge.

Osimk, R. (2009). Decoding sounds: An experimental approach to intelligibility in ELF. *Vienna English Working PaperS, 18,* 64-89.

Patrick, M. (2019, January 30). *Surveys: Overview of tools and good practice.* IT Learning Center Seminars, Oxford.

Minematsu, N. (2019, March 19). *How can speech technologies support learners to improve their skills of speaking, listen-ing, conversation and more?* CUED Speech Group Seminars, Cambridge.

Perlmutter, M. (1989). Intelligibility rating of L2 speech pre- and post-intervention. *Perceptual and Motor Skills 68*: 515–521.

Rose, H., & Galloway, N. (2019). *Global Englishes for language teaching.* Cambridge.

Saito, K., Suzukida, Y., Magne, V., Tran, M., Sun, H., & Ilkan, M. (2019). How do second language listeners perceive the comprehensibility of foreign-accented Speech? *Studies in Second Language Acquisition,* 1-17.

Saito, K., Trofimovich, P., & Isaacs, T. (2015). Using listener judgments to investigate linguistic influences on L2 comprehensibility and accentedness: A validation and generalization study. *Applied Linguistics, 38*(4), 439-462.

Smith, L. E. (1992). Spread of English and issues of intelligibility. In Kachru, B. B. (Ed.), *The other tongue: English across cultures* (2nd ed.). (pp. 75-90). Urbana

and Chicago: University of Illinois Press.

Smith, L. E., & Nelson, C. L. (1985). International intelligibility of English: Directions and resources. *World Englishes, 4,* 333-342.

Smith, L. E., & Rafiqzad, K. (1979). English for cross-cultural communication: The question of intelligibility. *TESOL Quarterly, 13(3),* 371-380.

Stibbard, R., & Lee. (2006). Evidence against the mismatched interlanguage speech intelligibility benefit hypothesis. *The Journal of the Acoustical Society of America, 120*(1), 433-442.

Takeuchi, O., & Mizumoto, A. (2014). *Gaikokugo kyoiku kenkyu handbook kaiteiban* [Handbook of foreign language education (2nd ed.)]. Tokyo: Shohakusha.

Thomson, R. (2018). Measurement of accentedness, intelligibility, and comprehensibility. In O. Kang & G. April (Eds.). *Assessment in second language pronunciation* (pp. 11-29). Abingdon: Routledge.

Trisitichoke, T., Ando, S., Inoue, Y., Saito, D., & Minematsu, N. (2018). Influence of content variations on native speakers' fluency of shadowing. *SLP, 16,* 1-3.

Tsushima, E. (2010). *Iryokei ronbun no yomikata matomekata* [The way to read and summarize medical papers]. Tokyo: Tokyotosho.

Yamato, K. (2011). L2 speech kenkyu ni okeru hatsuon no meiryousei no toriatsukai: meiryouna hyoutei no tameni [The treatment of pronunciation “intelligibility” in L2 speech research: For the purpose of intelligible evaluation]. *Gaikokugo kyouiku Media Gakkai (LET) Kansai Shibu Mesodorogii Kenkyuubukai 2011nendo Houkoku Ronshuu* [LET Kansai Chapter Methodology Workshop 2011 Annual Report], 41-49.

Appendix A. Speech Material

The highlighted sentences were excluded after the pilot study

Sentence No.	Speaker Category	Sentece	Pronunciation Scoring by NS		
			Segmental score	Rhythem score	Intonation score
1	HP Male	It is very hard to judge.	3.84	4.15	3.80
2	HP Male	He works in the ticket office.	3.86	3.75	4.00
3	HP Male	Few people live to be a hundred.	4.76	4.65	4.30
4	HP Male	Jane may earn more money by working hard.	3.24	3.90	3.30
5	HP Male	She put everything into a large bag.	3.20	2.50	2.65
6	HP Male	There are quite a few cars in the park.	3.18	3.84	3.20
7	HP Male	The full moon shone brightly that night.	3.14	3.50	3.25
8	HP Male	Just drop notices in any suggestion box.	3.10	2.80	3.45
9	HP Male	The water is boiling.	4.98	4.65	3.85
10	HP Male	They often go out in the evening.	3.54	3.50	3.10
11	HP Female	He ate four extra eggs for breakfast.	4.03	4.10	3.70
12	HP Female	It's healthier to cook without sugar.	3.92	4.35	3.10
13	HP Female	This gear is too heavy to handle.	4.03	3.95	4.30
14	HP Female	Guess the question from the answer.	3.62	3.85	3.65
15	HP Female	Too much curiosity can get you into trouble.	3.15	4.15	3.20
16	HP Female	John tried to put his coat on the rock.	3.17	3.80	3.55
17	HP Female	We are busy preparing for the shop's fashion show.	3.42	4.10	4.00
18	HP Female	They all like long hot showers.	3.95	4.55	3.75
19	HP Female	They are running fast.	4.17	4.80	4.00
20	HP Female	His name is quite common.	3.95	3.60	3.60
21	LP Male	They are digging a hole.	2.60	2.75	2.85
22	LP Male	That caused her a lot of worry.	1.86	2.90	2.90
23	LP Male	I like American food very much.	2.30	3.25	2.40
24	LP Male	She is thinner than I am.	2.58	2.65	3.30
25	LP Male	He is driving along the road.	2.38	2.85	2.65
26	LP Male	The small boy put the worm on the hook.	2.42	3.20	2.90
27	LP Male	Gwen grows green beans in her vegetable garden.	2.62	2.75	2.65
28	LP Male	Her chocolate cake tastes great.	2.22	2.70	3.11
29	LP Male	He is a good pitcher.	2.24	2.55	2.35
30	LP Male	We welcome many new students each year.	2.34	2.25	2.85
31	LP Female	That man is Nancy's father.	2.02	2.05	2.60
32	LP Female	The children are sitting on a muddy road.	2.58	3.60	3.65
33	LP Female	I walked several times up and down the street.	2.52	3.90	3.20
34	LP Female	John's brother repainted the garage door.	2.40	3.25	2.65
35	LP Female	We find ourselves involved in a difficulty.	2.65	3.50	3.15
36	LP Female	John doesn't want to fail it.	2.65	3.58	2.75
37	LP Female	They need cars for their performance.	2.60	3.21	3.40
38	LP Female	There are two birds in the cage.	2.27	3.20	3.15
39	LP Female	That noise problem grows more annoying each day.	2.50	2.95	2.55
40	LP Female	Swing your arm as high as you can.	2.55	2.75	3.30

Apeendix B. Invitation Letter

Dear [redacted]

Hi, I hope you are doing well. I finally managed to start my experiment and I would like to ask if you could take part in my study. If interested, please use the following link to access the task form.
https://oxfordeducation.eu.qualtrics.com/jfe/form/SV_0qUPTK0w2wEjM2N

In this survey, you will be asked to answer a questionnaire and listen to 32 very short recordings of Japanese native speakers reading English sentences. You will be asked to write down what you heard the speaker say and rate the passage on how easy it was to understand. Please conduct the listening task in a quiet room with earphones, headphones, or speakers. It takes about 25 minutes.

In addition, due to the nature of the study a large number of participants is required and if you would like to share this information with your friends, your assistance would be greatly appreciated. I am looking for adult (18 or over) British native English speakers, Mandarin Chinese learners of English, and Japanese learners of English with no listening impairment. The following is an invitation letter that I have created.

Thank you in advance for your cooperation.

Best wishes
[redacted]

My name is [redacted]. I am a master student at the University of Oxford, supervised by [redacted]. I am looking for adult (18 or over) British native English speakers, Mandarin Chinese learners of English, and Japanese learners of English, with no listening impairment who can participate in my research. For my research project, I am examining whether first language background affects the intelligibility of non-native English speakers.

This is an online survey, and in the survey, you will be asked to listen to 32 very short recordings of Japanese learners of English. You will be asked to write down what you hear the speaker say and rate the passage on how easy it was to understand. The survey will take around 25 minutes to complete. Please conduct the survey in a quiet room with earphones, headphones, or speakers.

To participate, please click on the following link:
https://oxfordeducation.eu.qualtrics.com/jfe/form/SV_0qUPTK0w2wEjM2N

You can choose the language of instruction. Please select the language of the survey instruction from the list provided in the top right corner on the task form.

Thank you for your time and attention.

Sincerely yours,

[redacted]
MSc in Applied Linguistics and Second Language Acquisition
Department of Education | University of Oxford
[redacted]

Appendix C. Online Task Form (English)

Invitation

I would like to invite you to join a research project which I am conducting. Please read the following research information carefully before agreeing to participate.

Note: This study includes listening exercises, thus please conduct the task in a quiet room with earphones, headphones, or speakers. Some web browsers are not compatible with the audio player used in this study, such as Internet Explorer. Please check whether you can play the following sample audio file before continuing.

What is the purpose of this research?

This study aims to understand the way English speakers understand each other based on their first language background.

Why have I been invited to take part?

You have been invited because we are seeking British native English speakers, native Mandarin speakers, and native Japanese speakers (with no hearing impairment) to take part. Do I have to take part? No. You can ask questions about the research before deciding whether or not to participate. If you do agree to participate, you may withdraw from the study at any time, without giving a reason. If you decide to stop taking part, I will not use any of the data I have collected about you up to that point.

What will happen to me if I take part in the research?

You will be asked to do three tasks. Firstly, you will be asked to answer a questionnaire. The questionnaire is about your background and language learning experience. It will take about 5 minutes. Following this, you will be asked to listen to 32 very short recordings of Japanese native speakers reading English sentences. You will be asked to write down what you heard the speaker say and rate the passage on how easy it was to understand. This process will take about 20 minutes.

Are there any potential risks and benefits in taking part?

I do not know of any risks that could arise from participating in this study. There will be no direct benefit to you from taking part in this research, but by taking part in the study you will help researchers understand more about how speakers of English understand each other. This might be very helpful in a world in which English is spoken by many people, most of whom are not ‘native speakers’.

Who has reviewed this study?

This study has been reviewed by, and received ethics clearance through, the University of Oxford Central University Research Ethics Committee (Reference number: ED-CIA-19-146).

What happens to the data provided?

Your answers will be stored confidentially using a password protected computer. The researcher and supervisor will have access to the data. Responsible members of the University of Oxford may be given access to data for monitoring and/or auditing of the research. The only personal identifying data will be e-mail addresses. Your e-mail address is collected if you are willing to leave your e-mail address at the end of this form in order to be sent a summary of this research. The e-mail address will be deleted following my sending of the research summary to you. All the other data will be stored for at least three years after publication or public release of the work of the research. The data will be presented in my dissertation. Furthermore, the results of this study may be published in an academic journal or book or presented at a conference. However, your personal information will not be shown in any cases.

Contact information

If you have a concern about any aspect of this study, you may contact [redacted], the researcher or his supervisor [redacted].

Researcher: [redacted]
Email Address: [redacted]
Supervisor: [redacted]
Email Address: [redacted]

If your enquires do not receive a reply within 10 working days, or you are unhappy with the response, or wish to make a formal complaint, please contact the relevant chair of the Research Ethics Committee at the University of Oxford (ethics@socsci.ox.ac.uk) who will seek to resolve the matter in a reasonably expeditious manner.

Research Consent Form

Project title: Intelligibility in English as a Lingua Franca

Researcher: [redacted]

1. I confirm that I have read and understood the information sheet for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.

2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, and without any adverse consequences or academic penalty.

3. I understand that research data collected during the study may be looked at by designated individuals from the University of Oxford where it is relevant to my taking part in this study. I give permission for these individuals to access my data.

4. I understand that this project has been reviewed by, and received ethics clearance through, the University of Oxford Central University Research Ethics Committee.

5. I understand who will have access to personal data provided, how the data will be stored and what will happen to the data at the end of the project.

6. I understand how this research will be written up and published.

7. I understand that the recordings used on this website are for the purposes of this study alone, and I will not distribute them to anyone else.

8. I understand how to raise a concern or make a complaint.

9. I agree to take part in the research.

By clicking “Next”, you agree to participate in this study. If you do not wish to participate, please do not continue.

O Next

O I do not wish to continue

Skip To: End of Survey If Research Consent Form = I do not wish to continue

Questionnaire

Q1 How old are you?

_____

Q2 What is your gender?

O Male

O Female

O Other

Q3 What level of education have you completed?

O High School or earlier

O College

O Bachelor's Degree

O Master's Degree

O Doctorate or beyond

Q4 What do you do?

I am

O a student

O employed

O self-employed

O retired

O unemployed

O other _____

Q5

If you are a student, at what level of education are you currently?

O College

O Undergraduate

O Pursuing Masters

O Pursuing Doctorate

O Other _____

Q6

If you are employed or self-employed, what is your current job?

_____

Q7 What is your first language?

O English

O Mandarin Chinese

O Japanese

O Other _____

Skip To: Q11 If What is your first language? = English

Q8 How long have you studied or learned English?

O less than six months

O six months to a year

O a year to three years

O three years to five years

O five years to ten years

O more than ten years

Q9 If you have taken any standardized English proficiency tests, please note the score or level that you got here and give the year that you took the test. If you have taken the same test on more than one occasion, please provide the most resent score/level.

	Score/Level	Year
TOEIC		
TOEFL (iBT)		
TOEFL (ITP)		
IELTS		
CET		
STEP 英検		
Cambridge English Qualifications		

Q10 Please tick the option that best represents your current English proficiency. The criteria for judging your proficiency are shown below.

O Beginner

O Elementary

O Intermediate

O Upper Intermediate

O Advanced

O Master

Criteria for language proficiency:

Beginner

I can understand and use familiar everyday expressions and very basic phrases aimed at the satisfaction of needs of a concrete type. I can introduce myself and others and can ask and answer questions about personal details such as where I live, people I know and things I have. I can interact in a simple way provided the other person talks slowly and clearly and is prepared to help.

Elementary

I can understand sentences and frequently used expressions related to areas of most immediate relevance (e.g. very basic personal and family information, shopping, local geography, employment). I can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. I can describe in simple terms aspects of my background, immediate environment and matters in areas of immediate need.

Intermediate

I can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. I can deal with most situations likely to arise whilst travelling in an area where the language is spoken. I can produce simple connected text on topics which are familiar or of personal interest. I can describe experiences and events, dreams, hopes & ambitions and briefly give reasons and explanations for opinions and plans.

Upper Intermediate

I can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in my field of specialisation. I can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party. I can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.

Advanced

I can understand a wide range of demanding, longer texts, and recognise implicit meaning. I can express myself fluently and spontaneously without much obvious searching for expressions. I can use language flexibly and effectively for social, academic and professional purposes. I can produce clear, well-structured, detailed text on complex subjects, showing controlled use of organisational patterns, connectors and cohesive devices.

Master

I can understand with ease virtually everything heard or read. I can summarise information from

different spoken and written sources, reconstructing arguments and accounts in a coherent presentation. I can express myself spontaneously, very fluently and precisely, differentiating finer shades of meaning even in more complex situations.

Display This Question: If What is your first language? ≠ Japanese

Q11 Have you learned or studied Japanese before?

O Yes

O No

Skip To: Q15 If Have you learned or studied Japanese before? = No

Display This Question: If What is your first language? ≠ Japanese

Q12 How long have you studied or learned Japanese?

O less than six months

O six month to a year

O a year to three years

O three years to five years

O five years to ten years

O more than ten years

Display This Question: If What is your first language? ≠ Japanese

Q13 If you have taken the Japanese Language Proficiency Test (JLPT), please note the most recent level you achieved (e.g. N1 to N5 or level 1 to level 4), as well as the year you took the test

	Level	Year
Japanese Language Proficiency Test		

Display This Question: If What is your first language? ≠ Japanese

Q14 Please tick the option that best represents your current Japanese proficiency. The criteria for judging your proficiency are shown below.

O Beginner

O Elementary

O Intermediate

O Upper Intermediate

O Advanced

O Master

Criteria for language proficiency:

*The criteria are omitted here, since it is the same criteria used for measurement of English proficiency (Q10).

Q15 Do you have English interactions with native Japanese speakers?

(e.g., daily conversation, university lecture, or business meeting).

O Never

O Once a year

O Once a month

O Once a week

O Often

Q16 On a scale from one to nine, how familiar are you with English spoken by native Japanese speakers?

I am very familiar with English spoken by native Japanese speakers = 1

I am not familiar with English spoken by native Japanese speakers at all = 9

o	o	o	o	o	o	o	o	o
1								9

Thank you for answering this questionnaire. Now, I would like you to work on the listening tasks.

Listening task instruction: Please conduct the listening task in a quiet room with earphones, headphones, or speakers. Please adjust the volume while listening to the following sample audio file. You will listen to 32 very short English sentences. You will work on two tasks per recording.

Task 1. Click the play button and listen to the recording once. You will hear a sentence after the beep sound. Then, please type the words you hear.

Note: Please describe the words you hear as best as you can, and do not worry if you think your transcribed sentences do not make sense. This is not an exam and minor spelling mistakes are not a problem.

Task 2. Listen to the same recording again and judge how easy it is to understanding the speech with a nine-point scale (1 = very easy to understand, 9 = very difficult to understand).

Note: Please do not modify your answers to task one after you have listened to the audio the second time. This is because the study aims to measure participants' immediate responses.

Sample audio file

Q17 Listen to the recording once. What did the speaker say? Please type the words you hear.

_____

Listen to the recording again. How easy or difficult was it to understand this clip?

very easy to understand = 1, very difficult to understand = 9

o o o o o o o o o

1 9

*Q18 to Q49 are omitted here, since they are the same question type as Q17.

Thank you for cooperating with my research. If you would like to receive the summary of my research, please leave your e-mail address. I will send you the summary of my study around October after the submission of my dissertation. I will use your address only for sending the results of my research. I will delete your e-mail address once I send you the report.

_____

Appendix D. Online Task Form (Japanese)

研究へのご協力のお願い

以下の研究目的や実施内容等をご理解いただき、本研究にご協力いただくようお願いいたします。　ご協力いただける場合は、次ページの同意書に署名いただいて、アンケートへの回答、聞き取り実験へと進んでください。

*本研究は聞き取り課題を含みます。そのためイヤホン、ヘッドホン、スピーカーなどを使用の上、静かな場所で行ってください。

*特定のウェブブラウザーは本研究で使用する音声再生ソフトに対応しておりません（現在 Internet Explore では再生できないと報告を受けております)。以下のサンプル音声を再生できるかどうかを続ける前にご確認ください。

研究目的

本研究は、聞き手の第一言語（母語）の違いが英語の聞き取りにどのような影響を与えるかを明らかにすることを目的として、実施いたします。

研究対象者

本研究は聴覚障害がない日本語母語話者、イギリス人英語母語話者、中国語母語話者を対象としております。

本研究への参加について

本研究への参加は任意です。参加前に何か質問がある場合はご連絡ください。また、参加に同意いただき作業途中であっても受諾を撤回することができます。その場合は、それまでに収集したデータを分析対象とせず、廃棄いたします。

研究方法

本研究では、３つの課題に取り組んでいただきます。初めに参加者の方の背景情報（年齢、性別等)、言語学習履歴に関するアンケートに答えていただきます。アンケートへの記入は、5 分ほどで完了します。　次に日本人母語話者が読み上げた 32 個ほどの文を聴いていただきます。参加者の方には聞こえた音声の書き取りと音声の理解のしやすさを評価いただきます。聞き取り課題は 20 分ほどで完了します。

研究に参加していただくことにより得られる利益と不利益

本研究に参加されることにより不利益が生じることはございません。また、本研究に参加することにより、直接的に参加者に利益が生じることはありません。しかしながら、本研究に参加されることで、多くの人々、それも大半が英語を母国語としない人々が英語でコミュニケーションを取る今日の社会において、英語を話す人々がどのようにお互いを理解するかを明らかにすることに役立ちます。

倫理審査

本研究はオックスフォード大学の倫理委員会の審査を受け、承認を得ております。(照会番号：ED-CIA-19-146)。

研究データの取り扱いについて

研究データはパスワードを設定したコンピューターで厳重に保管されます。研究データにアクセスするのは研究者および指導教官になります。また、研究の監視/監査のためオックスフォード大学の担当者がデータにアクセスする場合があります。　本研究における唯一の個人の特定につながる情報は Email アドレスになります。Email アドレスは研究対象者がこのオンラインフォームの最後で本研究結果の要旨の受け取りを希望した場合にのみ収集いたします。いただいた Email アドレスは、要旨の送付後、破棄します。　他の研究データに関しては本研究終了後から最低 3 年以上保管され、研究者の修士論文作成のために使用されます。また、研究結果は学術誌、書籍として出版、学会において発表する場合があります。いずれの場合も個人の情報は表示されません。

研究者、および問い合わせ先について

本研究に関して何か問い合わせ事項がございましたら、以下の研究者もしくは指導教官までご連絡ください。

研究者: [redacted]
Email Address: [redacted]
指導教官: [redacted]
Email Address: [redacted]

問い合わせへの返信を 10 日以内に受け取れなかった場合、もしくは返信内容に満足ができない場合はオックスフォード大学の倫理委員会(ethics@socsci.ox.ac.uk)までご連絡ください。なお、指導教官及び倫理委員会への連絡は英語のみの受付となります。

研究参加の同意書

研究名: 国際英語論からみる英語発話の明瞭性の研究

研究者: [redacted]

1. 私は研究内容を理解しました。また、私は研究内容に関して、実施者へ質問し確認することができることも理解しました。

2. 私は今回の研究への参加が無償であり、いつでも研究への参加を辞退することができることを理解しました。

3. 私は今回の研究データがオックスフォード大学の担当者によって閲覧されることがあることを理解し、オックスフォード大学の担当者が研究データにアクセスすることを許可します。

4. 私はこの研究がオックスフォード大学の倫理委員会の審査を受け、承認を得ていることを理解しました。

5. 私は個人データに誰がアクセスし、どのようにデータが保管され、処理されるかも理解しました。

6. 私はこの研究が、研究者の修士論文作成のために使用され、学術誌や書籍として出版される場合もあることを理解しました。

7. 私はこの研究の聞き取り実験で使用する音声ファイルがこの研究のためだけに使用されることを理解し、第三者には配布いたしません。

8. 私はどのようにこの研究に関する懸念、苦情を申し立てるかを理解しました。

9. 私はこの研究に参加することに同意します。

「同意する」を選択することであなたは実験に参加していただけます。参加を希望されない場合は、「アンケートの終了」を選択してください。

O 同意する
O アンケートを終了する

Skip To: End of Survey If 研究参加の同意書 = アンケートを終了する

Questionnaire

Q1 あなたの年齢をお知らせください。

_____

Q2 あなたの性別をお知らせください。

O 男性

O 女性

O その他

Q3 あなたの最終学歴（すでに完了したもの）をお知らせください。例：現在大学生の場合は多くの方が「高等学校かそれ以下」になるかと思います。

O 高等学校かそれ以下

O 専門学校

O 大学

O 大学院（修士課程）

O 大学院（博士課程）かそれ以上

Q4 あなたは現在何をされていますか。

O 学生

O 会社員・公務員

O 自営業者

O 定年退職者

O 無職

O その他 _____

Q5 学生の場合は、所属する教育機関を選択肢からお選びください。

O 専門学校

O 大学

O 大学院（修士課程）

O 大学院（博士課程）

O その他 _____

Q6 有職者の場合はどのような仕事をされているかご記入ください。

_____

Q7 あなたの母語はなんですか？

O 英語

O 中国語

O 日本語

O その他 _____

Skip To: Q11 If あなたの母語はなんですか？= 英語

Q8 これまでに英語をどれぐらい勉強しましたか？

O 6 ヵ月未満

O 6 ヵ月以上 1 年未満

O 1 年以上 3 年未満

O 3 年以上 5 年未満

O 5 年以上 10 年未満

O 10 年以上

Q9 英語に関するテストや検定等を受けたことがある方は、その点数や級等、取得年月をご記入ください。同じテストを複数回受験したことがある場合は一番最近取得した点数/級をご記入ください。

	点数/級	取得年
TOEIC		
TOEFL (iBT)		
TOEFL (ITP)		
IELTS		
CET		
STEP 英検		
Cambridge English Qualifications		

Q10 あなたの現在の英語の習熟度を最も適切に表しているものを以下の選択肢からお選びください。習熟度の判断基準は以下の表を参照ください。

O 初級

O 初中級

O 中級

O 中上級

O 上級

O 超上級

言語習熟度評価表:

初級

私は具体的な欲求を満足させるための、よく使われる日常的表現と基本的な言い回しを理解し、用いることもできる。自分や他人を紹介することができ、どこに住んでいるか、誰と知り合いか、持ち物などの個人的 情報について、質問をしたり、答えたりできる。 もし、相手がゆっくり、はっきりと話して、助け船を出してくれるなら簡単なやり取りをすることができる。

初中級

私はごく基本的な個人的情報や家族情報、買い物、近所、仕事など、直接的関係がある領域に関する、よく使われる文や表現が理解できる。簡単で日常的な範囲なら、身近で日常の事柄についての情報交換に応ずることができる。自分の背景や身の回りの状況や、直接的な必要性のある領域の事柄を簡単な言葉で説明できる。

中級

私は仕事、学校、娯楽で普段出会うような身近な話題について、標準的な話し方であれば主要点を理 解できる。その言葉が話されている地域を旅行しているときに起こりそうな、たいていの事態に対処することができる。身近で個人的にも関心のある話題について、単純な方法で結びつけられた、脈絡のある文章を作ることができる。経験、出来事、夢、希望、野心を説明し、意見や計画の理由、説明を短く述べることができる。

中上級

私は自分の専門分野の技術的な議論も含めて、抽象的かつ具体的な話題の複雑な文章の主要な内容を理解できる。私は緊張しないで母語話者とやり取りができるくらい流暢かつ自然な発話ができる。かなり広汎な範囲の話題について、明確で詳細なテクストを作ることができ、さまざまな選択肢 について長所や短所を示しながら自己の視点を説明できる

上級

私はいろいろな種類の高度な内容のかなり長い文章を理解することができ、含意を把握できる。言葉を探しているという印象を与えずに、流暢に、また自然に自己表現ができる。 社会的、学問的、職業上の目的に応じた、柔軟な、しかも効果的な言葉遣いができる。 複雑な話題について明確で、しっかりとした構成の、詳細な文章を作ることができる。その際テクストを構成する字句や接続表現、結束表現の用法をマスターしている。

超上級

私は聞いたり、読んだりしたほぼ全てのものを容易に理解することができる。いろいろな話し言葉や書き言葉から得た情報をまとめ、根拠も論点も一貫した方法で再構成できる。自然に、流暢かつ正確に自己表現ができ、非常に複雑な状況でも細かい意味の違い、区別を表現できる。

Display This Question: If あなたの母語はなんですか？≠ 日本語

Q11 あなたはこれまでに日本語を学んだり、勉強したりしたことがありますか？

O はい

O いいえ

Skip To: Q15 If あなたはこれまでに日本語を学んだり、勉強したりしたことがありますか？= いいえ

Display This Question: If あなたの母語はなんですか？≠ 日本語

Q12 これまで日本語をどれぐらい勉強しましたか？

O 6 ヵ月未満

O 6 ヵ月以上 1 年未満

O 1 年以上 3 年未満

O 3 年以上 5 年未満

O 5 年以上 10 年未満

O 10 年以上

Display This Question: If あなたの母語はなんですか？ ≠ 日本語

Q13 日本語能力試験（JLPT）を受けたことがある方は、もっとも最近取得した級と取得年月をご記入ください。

	級	取得年
日本語能力試験（JLPT）		

Display This Question: If あなたの母語はなんですか？≠ 日本語

Q14 あなたの現在の日本語の習熟度を最も適切に表しているものを以下の選択肢からお選びください。習熟度の判断基準は以下の表を参照ください。

O 初級

O 初中級

O 中級

O 中上級

O 上級

O 超上級

言語習熟度評価表:

*The criteria are omitted here, since it is the same criteria used for measurement of English proficiency (Q10).

Q15 あなたは日本人と英語で交流する機会がありますか？
(例：友人との会話、学校での英語の授業、外国人を交えた会議など)

O 一度もない
O 一年に一度
O 一カ月に一度
O 一週間に一度
O 頻繁にする

Q16 あなたが日本語母語話者が話す英語にどれだけ慣れ親しんでいるかを 9 段階評価でお知らせください。

私は日本人の話す英語にとても詳しい = 1
私は日本人の話す英語に全く詳しくない = 9

o	o	o	o	o	o	o	o	o
1								9

アンケートへの回答ありがとうございました。続いて聞き取り実験に進んでいただきます。

聞き取り実験：

聞き取り実験はイヤホン、ヘッドホン、スピーカーなどを使用の上、静かな場所で行ってください。以下のサンプル音声ファイルを再生して音量を事前に調整ください。　これからあなたには合計 32 の短い英語の文を聴いていただきます。1 文につき行っていただくタスクは 2 つあります。

タスク 1. 音声ファイルの再生ボタンをクリックして音声を一度聴いてください。ビープ音後、課題の音声が再生されます。聞こえた音声を回答欄に打ち込んでください。 *聞き取った内容が文として意味を成していないと思っても記入をお願いします。 *これはテストではございません。細かな綴りのミスは気にしないでいただいて大丈夫です。

タスク 2. 音声ファイルをもう一度再生し、音声の理解しやすさを 9 段階で評価してください（1 = たいへん理解しやすい、9 = たいへん理解しにくい)。*タスク 2 で音声ファイルを再生後、タスク 1 の回答を修正しないようにお願いします。

サンプル音声

Q17

音声を一度聴いてください。発話者は何と言っていましたか？聞こえた文を記入してください。

_____

音声をもう一度聴いてください。音声はどれぐらい理解しやすかったですか？

たいへん理解しやすい = 1

たいへん理解しにくい = 9

o o o o o o o o o

1 9

*Q18 to Q49 are omitted here, since they are the same question type as Q17.

実験にご協力いただきありがとうございました。

ご希望の方には、私の研究結果の要旨を送付させていただきます。ご希望の方は、以下の欄にご自身のメールアドレスをご記入ください。今年の 10 月頃に送付させていただきます。

ご記入いただいたメールアドレスは結果の送付のみに使用し、結果の送付後、削除いたします。

Appendix E. Online Task Form (Mandarin)

邀请

我想邀请您加入我正在研究的一个项目。在同意参加此研究项目前，请仔细阅读以下研究信息。
注意：本研究包括听力练习，因此，请在安静的房间里使用耳机、头戴式耳机或扬声器来完成此项任务。一些浏览器与本研究中使用的音频播放器是不兼容的，如 Internet Explorer。请检查您的浏览器是否可以正常播放下面的音频样本，然后再继续。
音频样本

此项研究的目的

本研究的目的为了解以英语为母语的人如何理解对方

邀请您参与此研究的原因

我们邀请您的原因是我们在寻找以英语为母语的英国人，以普通话为母语的人，以及以日语为母语的人（无听觉障碍）来参与这项研究。

关于参加此项研究

本研究为自愿参与。在决定是否参加之前，您可以问一些关于这项研究的问题。如果您同意参加，您可以在任何时候退出研究且无需给出任何理由。如果您想停止参加，请告诉我。如果您决定停止参加，我将不会使用我收集到的关于您的任何数据。

研究的内容

您将被要求做三项任务。首先，您需要完成一份问卷。调查问卷的内容是关于您的背景以及语言学习经验。完成问卷大约需要五分钟。接下来，您需要去听 32 段简短英文录音。您需要对这篇文章的容易理解程度进行评分，并写下您听到的说话者所说的话。这个过程大约需要 15 分钟。

参与此研究的风险与好处

我不知道参与这项研究可能会产生任何的风险。参加这项研究对您没有任何直接的好处，但是通过参与这项研究，您将帮助研究人员更好地了解说英语的人是如何理解对方的话。在一个多数人都说英语的世界里，这可能对大多数不以英语为母语的人们会很有帮助。

对此项研究的审核

本研究经牛津大学中央大学研究道德伦理委员会审核通过,并取得许可(参考编号: ED-CIA-19-146)。　　数据将被如何使用　您的答案将会被保密储存在使用密码保护的计算机中。研究人员和主管将有权访问数据。牛津大学可靠负责的人员可访问数据，以监测和/或审计研究。唯一可以识别个人数据的信息则是电子邮件地址。您的电子邮件地址只有在您愿意在本表格末尾处留下您的电子邮件地址以便接收本研究总结的情况下被收集。向您发送研究总结后，您的电子邮件地址将会被删除。其他所有研究数据将在研究成果发表或公开发表后至少存储三年。这些数据将在我的论文

中出现。此外，本研究的结果可能将被发表在学术期刊或图书中，或被呈现在会议上。但是，您的个人信息将不在任何情况下出现。

研究人员与联系信息

如您对此项研究的任何方面存有疑问，您可以联系研究人员 [redacted]，或他的导师 [redacted]。

研究员：[redacted]
电子邮件地址：[redacted]

导师：[redacted]
电子邮件地址：[redacted]

如果您在 10 个工作日内没有收到就您所提出问题的任何回复，或者您对回复不满意，或者希望正式投诉，请联系牛津大学研究道德伦理委员会的相关主席(ethics@socsci.ox.ac.uk)，该相关主席将以合理迅速的方式解决该问题。全程交流必须使用英语。

研究同意书

方案题目：英语作为通用语言的可理解性

研究者：[redacted]

1. 我确认我已阅读并理解上述针对这项研究的信息内容。我曾有机会考虑这些信息、提出问题，并已得到满意的答复。

2. 我理解我的参与是自愿的，我可以在任何时候退出且不需提供任何理由，并且没有任何不良后果或学术惩罚。

3. 我理解在研究期间收集的研究数据可能会被牛津大学指定的个人进行查看且这些个人与我参与这项研究有关。我允许这些人访问我的数据。

4. 我了解该项目已通过牛津大学中央大学研究道德伦理委员会审核，并获得许可。

5. 我了解可查阅所提供的个人数据的人是谁、数据将被如何储存，以及在研究方案结束时数据将被如何处理。

6. 我了解这项研究将被如何写成文章并被发表。

7. 我了解如何提出疑问或进行投诉。

8. 我了解本网站所使用的录音只供本研究之用，我不会将与其他任何人分享。

9. 我同意参与这项研究。

通过点击“下一步”，您将同意参与此项研究。如果您不愿意参与此项研究，请不要继续。

O 下一步

O 我不想继续参与

Skip To: End of Survey If 研究同意书 = 我不想继续参与

Questionnaire

Q1 您的年龄？

_____

Q2 您的性别？

O 男

O 女

O 其他

Q3 您的教育程度？

O 高中或高中以前

O 专科

O 本科

O 硕士

O 博士或博士以上

Q4

您的职业？

我是

O 学生

O 雇员

O 自雇

O 退休人员

O 无业

O 其他 _____

Q5 如果你是一个学生，请问你现在是教育水平是？

O 专科

O 本科

O 硕士

O 博士

O 其他 _____

Q6 如果您是一名雇员或是自雇，那么您现在的工作是什么？

_____

Q7 您的母语是什么？

O 英语

O 汉语

O 日本语

O 其他 _____

Skip To: Q11 If 您的母语是什么？ = 英语

Q8 您已经研究或学习英语多长时间？

O 不足 6 个月

O 6 个月到 1 年

O 1 年到 3 年

O 3 到 5 年

O 5 年到 10 年

O 10 年以上

Q9 如果您参加过标准化的英语水平测试，请在这里写下您的分数与水平程度，并注明您参加测试的年份。如您曾多次参加同一项测试，请提供最近一次测试的分数/水平。

	总分/水平	年份
TOEIC		
TOEFL (iBT)		
TOEFL (ITP)		
IELTS		
CET		
STEP 英検		
Cambridge English Qualifications		

Q10 请在表示您目前的英语程度的方框内打钩。判断您的熟练程度的标准如下所示。

O 入门级

O 初级水平

O 中级水平

O 中高级水平

O 高级水平

O 精通

语言能力标准

入门级

我能理解和使用熟悉的日常用语和非常基本的短语，以满足具体类型的需求。我可以介绍或询问、回答关于自己或他人有关个人的信息，例如居住地、人际关系、所有物，对于他人进行缓慢而清晰的对话，只能以简单的方式予以回应。

初级水平

我能理解在与自己最密切相关的环境中常用的句子和表达，例如非常基本的个人和家庭信息、购物、区域地理、和就业，能与人沟通简单而例行性的工作，这类工作通常只需要简单而直接的日常信息交流。我能够用基本的词语描述自身背景，以及与自己最密切相关的环境中的事物。

中级水平

对于自己在工作、学校、休闲活动等生活中经常遇到的熟悉事物，我能够理解清晰规范的要点。我可以使用该语言处理在当地旅行时可能出现的各种状况，对于自己感兴趣或熟悉的事物也可以做出简单的相关表达。另外我能够描述经验、事件、梦想、愿望和雄心大志，并能对自己的观点或计划做出简单的解释。

中高级水平

我能够理解复杂文本的具体和抽象主题，包括对专业领域的技术讨论。我可以自然而流畅的与该语言的母语使用者进行正常互动，可以针对广泛的主题撰写思路清晰、描述具体的文字，并且可以就一个议题阐述观点，提出各种选项的利弊分析。

高级水平

我能够理解大量要求高、篇幅长的文本，并能识别隐含的含义。我能够自然而流畅地表达，而没有明显的词穷状况发生，能灵活有效地运用语言进行社交、学术和专业活动。对于复杂的主题我能写出思路清晰、架构良好、描述具体的文字，展示收放自如的文字组织、连接和衔接技巧。

精通

我能够轻易理解任何接收到的信息，并且针对不同的口语和书面来源总结信息、重建论点、并能进行流畅叙述。我能够自然、流利、准确地表达自己，即使在较为复杂的场合上也能辨别出细微的含义。

Display This Question: If 您的母语是什么？≠ 日本语

Q11 您曾经学过或研究过日语吗？

O 有过

O 没有

Skip To: Q15 If 您曾经学过或研究过日语吗？ = 没有

Display This Question: If 您的母语是什么？ ≠ 日本语

Q12 您已经研究或学习日语多长时间？

O 不足 6 个月

O 6 个月到 1 年

O 1 年到 3 年

O 3 到 5 年

O 5 年到 10 年

O 10 年以上

Display This Question: If 您的母语是什么？≠ 日本语

Q13 如果您参加过日语水平测试(JLPT)，请注明您最近时期达到的水平程度(例如 N1 至 N5，或 1 至 4 级)，以及您参加测试的年份。

	级	年份
日语水平测试（JLPT）		

Display This Question: If 您的母语是什么？≠ 日本语

Q14 请在表示您目前的日语程度的方框内打钩。判断您的熟练程度的标准如下所示。

O 入门级

O 初级水平

O 中级水平

O 中高级水平

O 高级水平

O 精通

*The criteria are omitted here, since it is the same criteria used for measurement of English proficiency (Q10).

Q15

您和以日语为母语的人进行英语交流互动吗？

(例如，日常会话、大学讲座或商务会议等)。

O 从不

O 一年一次

O 一月一次

O 一周一次

O 经常

Q16 从 1 到 9，你对日式英语的熟悉程度如何?
我非常熟悉以日语为母语的人所说的英语 = 1
我对以日语为母语的人说的英语一点也不熟悉 = 9

o o o o o o o o o
1 9

非常感谢您回答这份问卷。现在，我想请您做听力部分。
说明：请在安静的房间里使用耳机、头戴式耳机或扬声器进行听力部分。请在听下面的音频文件时注意调整音量。您将听到 20 个英文短句。每个录音将有两个任务
任务一：点击播放按钮并且只听一遍录音。在“哔”的一声之后您会听到一个句子。然后请输入您听到的词。注意:请尽可能输入您听到的词，如果您认为您记录的句子没有意义，请不要担心。此任务不是考试，小的拼写错误是没有关系的。
任务二：再听一遍录音，请用 9 分制来判断此录音的理解难易程度(1=非常容易理解，9=非常难理解)。注意：第二次听完录音后，请不要修改任务一的答案。
音频样本

Q17
听一遍录音。说话者说了什么？请输入您听到的词。
_____

再听一遍录音。这个片段的理解难易程度如何?
非常容易理解 = 1
非常难理解 = 9

o o o o o o o o o
1 9

*Q18 to Q49 are omitted here, since they are the same question type as Q17.

非常感谢您配合我的研究。如果您想获取我的研究总结，请留下您的电子邮件地址。我将在 10 月份左右论文提交后将我的研究总结发给您。我将只使用您的邮箱来发送我的研究结果。在我将报告发给您后，我就会删除您的电子邮件地址。

Appendix F. Results of Pilot Study

The highlighted sentences were excluded after the pilot study

Sentence No.	Speaker Category	Sentece	The results of pilot study	
			Intelligibility scores (*M*)	comprehensibility ratings (*M*)
1	HP Male	It is very hard to judge.	0.98	2
2	HP Male	He works in the ticket office.	0.98	2.6
3	HP Male	Few people live to be a hundred.	0.72	5.3
4	HP Male	Jane may earn more money by working hard.	0.73	7
5	HP Male	She put everything into a large bag.	0.76	3.7
6	HP Male	There are quite a few cars in the park.	0.90	5
7	HP Male	The full moon shone brightly that night.	0.56	6.4
8	HP Male	Just drop notices in any suggestion box.	0.79	7.6
9	HP Male	The water is boiling.	1.00	1.3
10	HP Male	They often go out in the evening.	1.00	2.1
11	HP Female	He ate four extra eggs for breakfast.	0.97	2.8
12	HP Female	It's healthier to cook without sugar.	0.84	5.9
13	HP Female	This gear is too heavy to handle.	0.83	6.9
14	HP Female	Guess the question from the answer.	0.87	4.3
15	HP Female	Too much curiosity can get you into trouble.	0.96	3.8
16	HP Female	John tried to put his coat on the rock.	0.78	6.3
17	HP Female	We are busy preparing for the shop's fashion show.	0.89	4
18	HP Female	They all like long hot showers.	0.83	4.9
19	HP Female	They are running fast.	1.00	1.8
20	HP Female	His name is quite common.	1.00	1.7
21	LP Male	They are digging a hole.	0.88	4.1
22	LP Male	That caused her a lot of worry.	0.64	7.4
23	LP Male	I like American food very much.	0.88	3.2
24	LP Male	She is thinner than I am.	0.55	7.3
25	LP Male	He is driving along the road.	0.78	4.4
26	LP Male	The small boy put the worm on the hook.	0.88	4.9
27	LP Male	Gwen grows green beans in her vegetable garden.	0.79	5.5
28	LP Male	Her chocolate cake tastes great.	0.82	4.6
29	LP Male	He is a good pitcher.	0.86	3.8
30	LP Male	We welcome many new students each year.	1.00	2.6
31	LP Female	That man is Nancy's father.	0.73	7.1
32	LP Female	The children are sitting on a muddy road.	0.56	7.4
33	LP Female	I walked several times up and down the street.	0.82	5.6
34	LP Female	John's brother repainted the garage door.	0.68	6.7
35	LP Female	We find ourselves involved in a difficulty.	0.74	6.4
36	LP Female	John doesn't want to fail it.	0.82	6.8
37	LP Female	They need cars for their performance.	0.70	6.3
38	LP Female	There are two birds in the cage.	0.80	5.3
39	LP Female	That noise problem grows more annoying each day.	0.90	3.1
40	LP Female	Swing your arm as high as you can.	0.95	3.9

Appendix G. Confirmation of Ethical Approval

Dear [redacted]

Title: Intelligibility in English as a Lingua Franca
Ref: ED-CIA-19-146

The above application has been considered on behalf of the Departmental Research Ethics Committee (DREC) in accordance with the procedures laid down by the University for ethical approval of all research involving human participants.

I am pleased to inform you that, on the basis of the information provided to DREC, the proposed research has been judged as meeting appropriate ethical standards, and accordingly, approval has been granted. (Please note that the stated start date on your CUREC form is earlier than this response from DREC. You should not begin any research until this approval is received.)

Should there be any subsequent changes to the project which raise ethical issues not covered in the original application you should submit details to research.office@education.ox.ac.uk for consideration.

Good luck with your research study!

Yours sincerely,

Laura Molway
Member of DREC

Laura Molway
PGCE Curriculum Tutor & D.Phil Candidate in Education
University of Oxford, Department of Education
Tel: 01865 611008
Join our mailing list

New Article: Changing mindsets in the modern foreign languages classroom: an intervention combining intelligence theories and reading strategies.

Are you a teacher in the first 3 years of your career? Come along to the Early Career Teacher Conference on Saturday 11th May!

Please note that Thursdays are my research days and I may not be in my office



Appendix H. Output of SPSS for The Relationships between L1 Differences and Intelligibility

H.1. Results of The Normality Test of Intelligibly

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	L1	Statistic	df	Sig.	Statistic	df	Sig.
Intelligibility	English	.182	25	.032	.912	25	.034
	Mandarin	.232	22	.003	.840	22	.002
	Japanese	.101	20	.200*	.964	20	.634

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

H.2. Results of The Brown-Forsythe Test of Intelligibly

Robust Tests of Equality of Means

Intelligibility

	Statistic[a]	df1	df2	Sig.
Brown-Forsythe	27.328	2	31.183	.000

a. Asymptotically F distributed.

H.3. Results of The Kruskal-Wallis Test of Intelligibility

Ranks

	L1	N	Mean Rank
Intelligibility	Japanese	20	36.45
	Mandarin	22	14.89
	English	25	48.86
	Total	67	

Test Statistics[a,b]

	Intelligibility
Kruskal-Wallis H	36.052
df	2
Asymp. Sig.	.000

a. Kruskal Wallis Test

b. Grouping Variable: L1

H.4. Results of The Normality Test of Intelligibly (HP and LP)

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	L1	Statistic	df	Sig.	Statistic	df	Sig.
Intelligibility (HP)	Japanese	.114	20	.200*	.976	20	.874
	Mandarin	.254	22	.001	.760	22	.000
	English	.221	25	.003	.893	25	.013
Intelligibility (LP)	Japanese	.153	20	.200*	.941	20	.245
	Mandarin	.224	22	.005	.855	22	.004
	English	.138	25	.200*	.922	25	.056

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Appendix I. Output of SPSS for The Relationships between L1 Differences and Comprehensibility

I.1. Results of The Normality Test of Comprehensibility

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	L1	Statistic	df	Sig.	Statistic	df	Sig.
Comprehensibility	Japanese	.141	20	.200*	.944	20	.283
	Mandarin	.133	22	.200*	.934	22	.148
	English	.194	25	.016	.940	25	.145

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

I.2. Results of The Kruskal-Wallis Test of Comprehensibility

Ranks

	L1	N	Mean Rank
Comprehensibility	Japanese	20	23.03
	Mandarin	22	43.36
	English	25	34.54
	Total	67	

Test Statistics[a,b]

	Comprehensibility
Kruskal-Wallis H	11.452
df	2
Asymp. Sig.	.003

a. Kruskal Wallis Test

b. Grouping Variable: L1

I.3. Results of The Normality Test of Comprehensibility (HP and LP)

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	L1	Statistic	df	Sig.	Statistic	df	Sig.
Comprehensibility (HP)	Japanese	.141	20	.200*	.944	20	.283
	Mandarin	.128	22	.200*	.925	22	.099
	English	.194	25	.016	.940	25	.145
Comprehensibility (LP)	Japanese	.126	20	.200*	.951	20	.378
	Mandarin	.094	22	.200*	.976	22	.842
	English	.114	25	.200*	.976	25	.796

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Appendix J. Output of SPSS for The Relationships between Familiarity with Japanese Accented English and Intelligibility

J.1. Results of The Normality Test

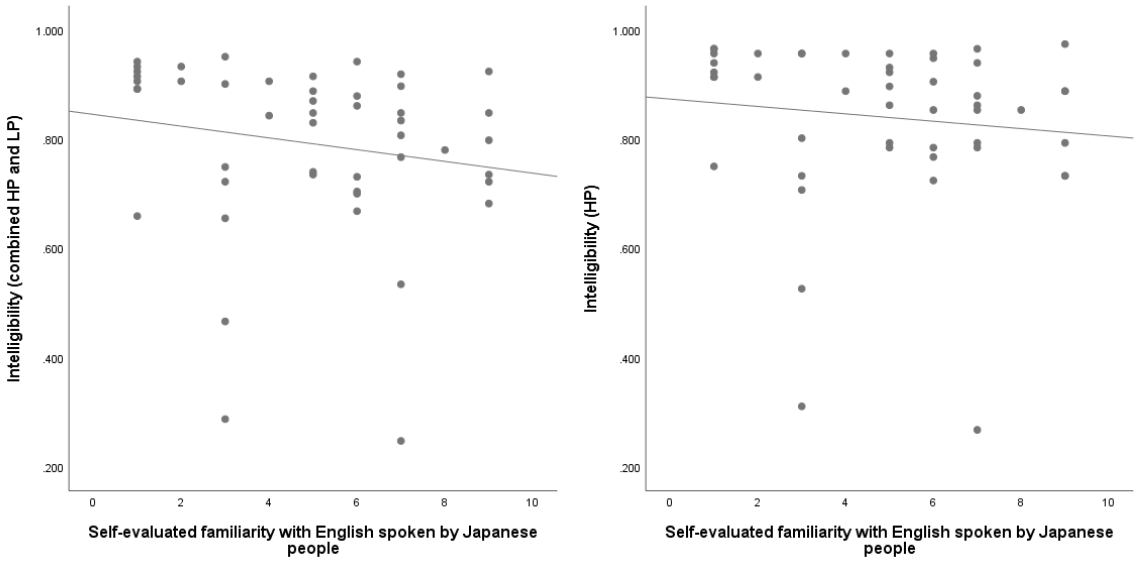
Tests of Normality

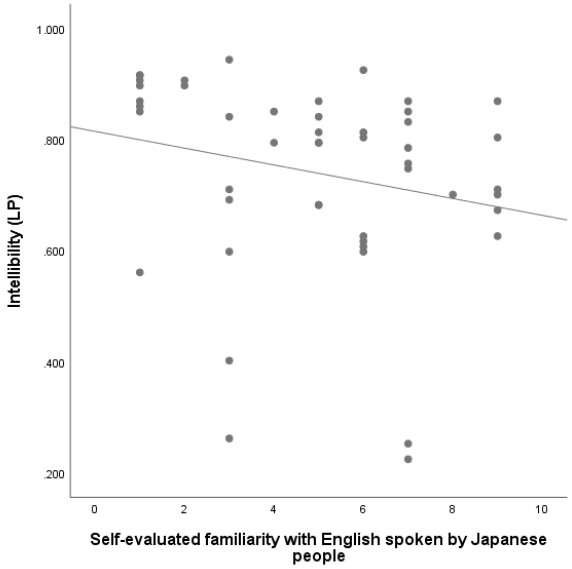
	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Familiarity with English spoken by Japanese	.123	48	.067	.924	48	.004
Frequency of interaction with Japanese people	.214	48	.000	.843	48	.000
Intelligibility (combied HP and LP)	.158	48	.004	.809	48	.000
Intelligibility (HP)	.185	48	.000	.743	48	.000
Intelligibility (LP)	.164	48	.002	.842	48	.000
Comprehensibility (combied HP and LP)	.090	48	.200*	.975	48	.375
Comprehensibility (HP)	.114	48	.153	.962	48	.122
Comprehensibility (LP)	.052	48	.200*	.994	48	.997

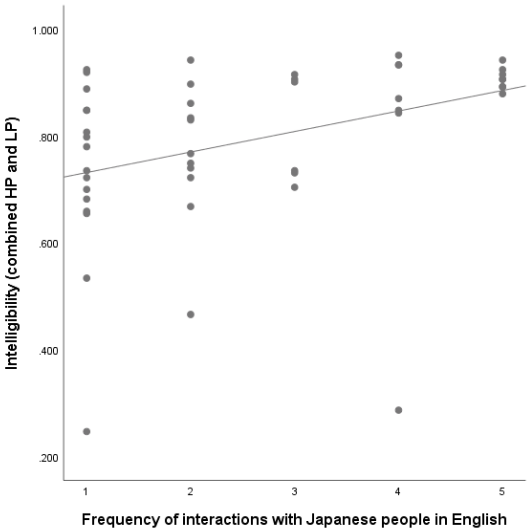
*. This is a lower bound of the true significance.

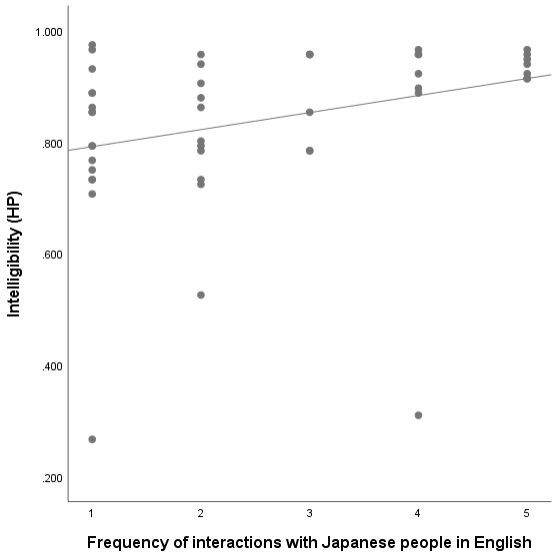
a. Lilliefors Significance Correction

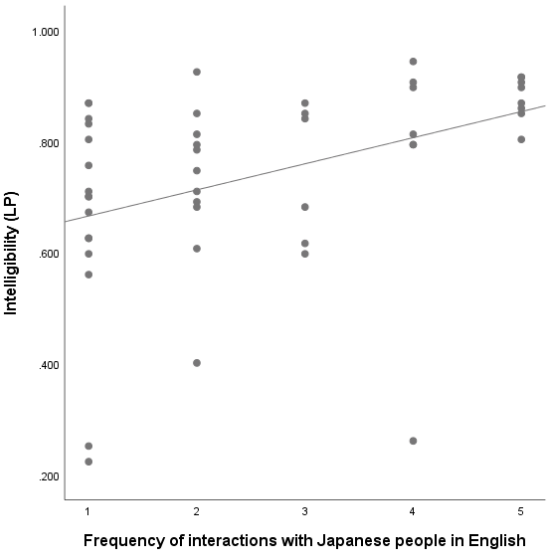
J.2. Scatter Plots (Familiarity with Japanese Accented English vs. Intelligibility)



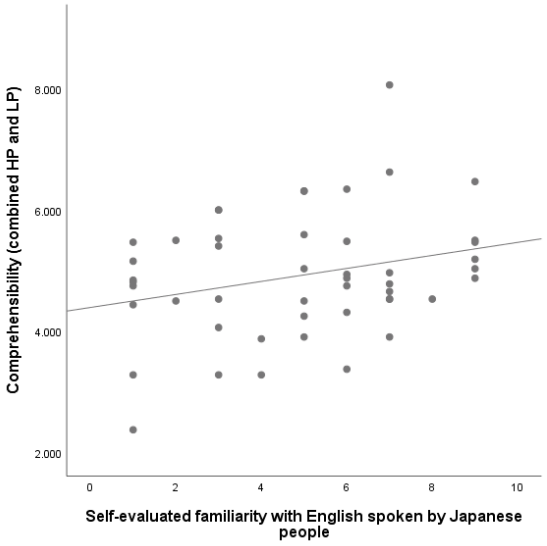

Intellibility (LP)
1.000
.800
.600
.400
.200
0
2
4
6
8
10
Self-evaluated familiarity with English spoken by Japanese people

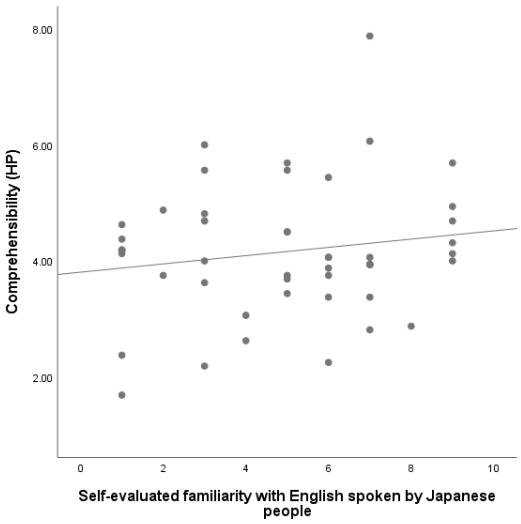

Intelligibility (combined HP and LP)
1.000
.800
.600
.400
.200
1
2
3
4
5
Frequency of interactions with Japanese people in English

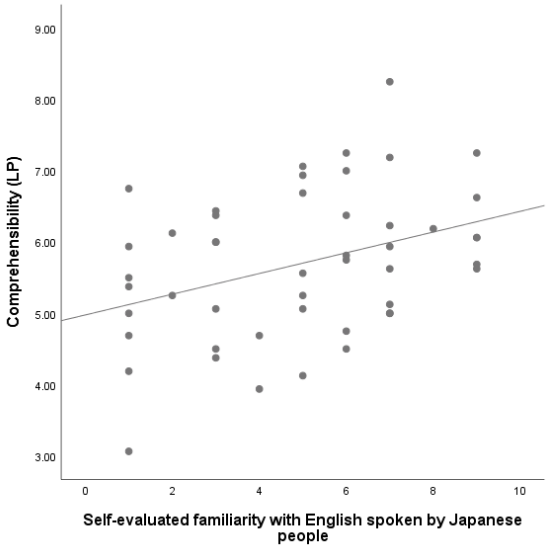

Intelligibility (HP)
1.000
.800
.600
.400
.200
1
2
3
4
5
Frequency of interactions with Japanese people in English

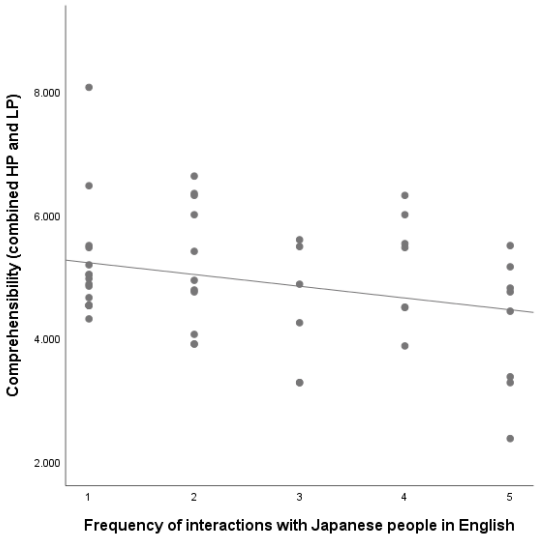

Intelligibility (LP)
1.000
.800
.600
.400
.200
1
2
3
4
5
Frequency of interactions with Japanese people in English

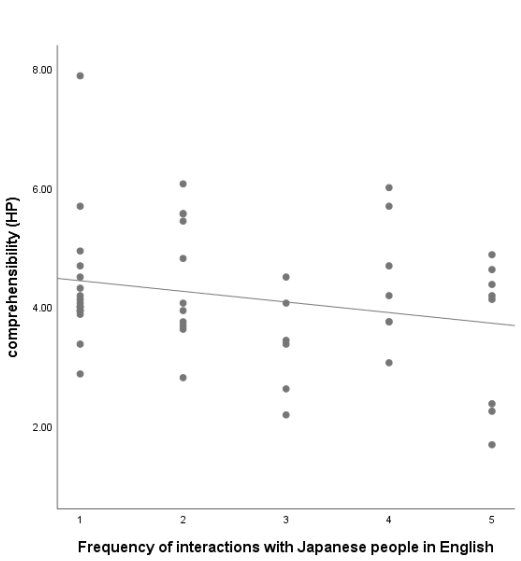
J.3. Scatter Plots (Familiarity with Japanese Accented English vs. Comprehensibility)

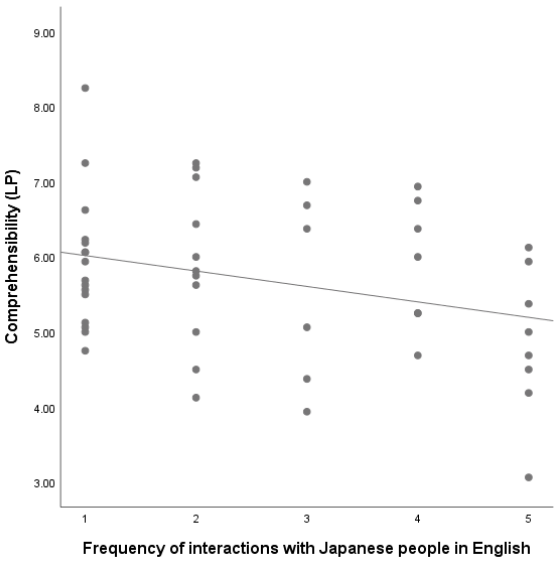












Appendix K. Output of SPSS for The Relationships between English Proficiency and Two Speech Understanding (Intelligibility and Comprehensibility)

K.1. Results of The Normality Test

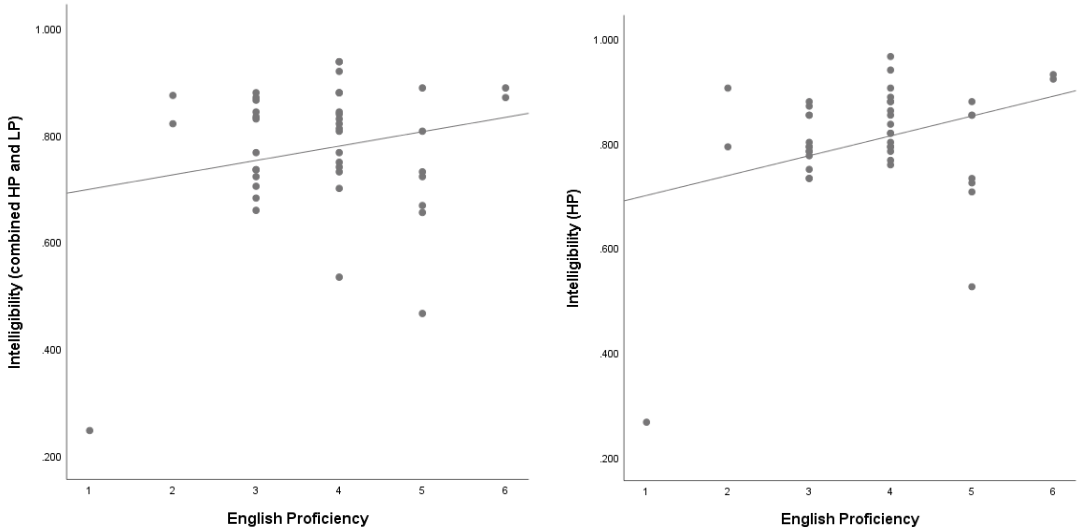
Tests of Normality

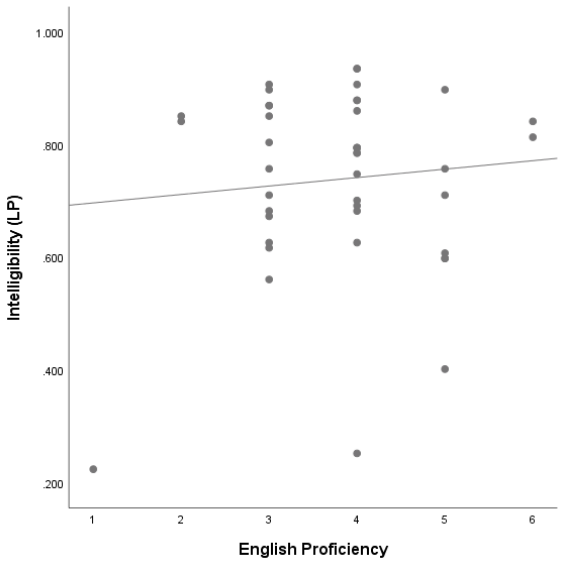
	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
English Proficiency	.203	42	.000	.914	42	.004
Intelligibility (combined HP and LP)	.149	42	.020	.838	42	.000
Intelligibility (HP)	.169	42	.004	.759	42	.000
Intellgiiblity (LP)	.136	42	.049	.866	42	.000
Comprehensibility (combined HP and LP)	.066	42	.200*	.985	42	.840
Comprehensibility (HP)	.114	42	.193	.950	42	.063
Comprehensibility (LP)	.080	42	.200*	.980	42	.656

*. This is a lower bound of the true significance.

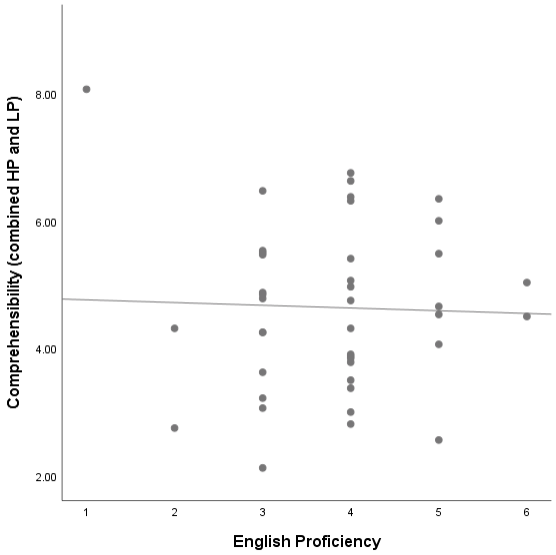
a. Lilliefors Significance Correction

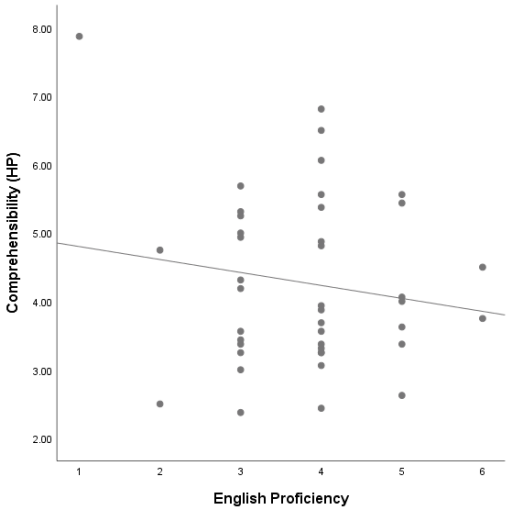
K.2. Scatter Plots (English Proficiency vs. Intelligibility)

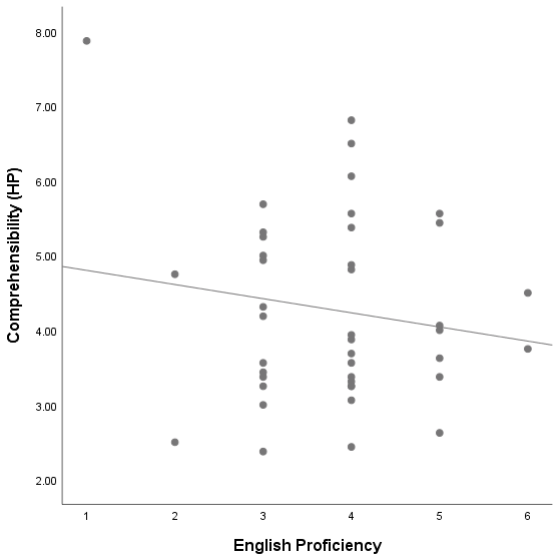




K.3. Scatter Plots (English Proficiency vs. Comprehensibility)







Appendix L. Dissertation Title Change Form

UNIVERSITY OF OXFORD
DEPARTMENT OF EDUCATION

Notification of Change of Title Form

All MSc Students

Please submit this form **electronically** to the Higher Degrees Office. Please do not submit it later than the 1st of August, or it will not be accepted.

NAME OF STUDENT:	[redacted]
CANDIDATE NUMBER:	1029481
NAME OF SUPERVISOR:	[redacted]
COURSE:	ALSLA
PREVIOUS DISSERTATION TITLE: Intelligibility in English as a Lingua Franca	
PROPOSED NEW TITLE: Intelligibility and Comprehensibility of Japanese Speakers of English from a Global Englishes Perspective	
SUPERVISOR APPROVAL Please check this box to confirm you have discussed this title with your supervisor and received his/her approval for it: ☒	