Examining the Segmental and Suprasegmental Correlates of the IELTS Pronunciation Scale

Author: Yui Suzukida

Birkbeck, University of London

British Council ELT Master’s Dissertation Awards: Commendation
Examining the Segmental and Suprasegmental Correlates of the IELTS

Pronunciation Scale
Abstract

The current study identified the crucial segmentals and suprasegmentals for developing IELTS pronunciation syllabus and explored in depth the extent to which trained raters actually follow the IELTS rubric to distinguish different levels of candidates in order to suggest which pronunciation features test takers should selectively work on for ultimate improvement of overall test results. This study additionally investigated the practicality of improving IELTS pronunciation skills.

Forty speech samples of native Japanese speakers were first rated by seven trained judges based on the publically available IELTS pronunciation rubric and degree of comprehensibility, then the same samples were objectively coded by the researcher according to eight pronunciation measures and six problematic segmental groups for Japanese learners.

The results of the correlation analysis suggested that more errors in productions of major segmentals (/l, ɹ, ð, q, v/), syllable and word stress lower the raters’ evaluations on IELTS pronunciation marking. Confidence intervals revealed that accurate production of major segmentals (/l, ɹ, ð, q, v/), secondary segmentals (/æ, ʌ, ə/), syllables, and diphthongs are important for beginner group (Band scores 4–5), while accurate production of word stress as well as the minor segmentals (/w, ñ, h, n, p, t, k/, Contractions) are crucial for intermediate group (Band score 6) to attain advanced level (Band scores 7–8). Finally, the result of the correlation analysis yielded significantly high correlation of IELTS pronunciation rating with comprehensibility judgements, supporting the high applicability of the IELTS pronunciation skill to the success in L2 communication in real-life context.
# Table of Contents

Tables and Figures.................................................................................................................. 6

Chapter 1: Introduction ........................................................................................................... 8
  1.1 Background ...................................................................................................................... 8
  1.2 Rationale ........................................................................................................................ 9
  1.3 Aim and Structure of the Current Study ........................................................................ 10

Chapter 2: Literature Review ................................................................................................. 11
  2.1 Introduction .................................................................................................................... 11
  2.2 Teaching L2 Pronunciation ........................................................................................... 11
    2.2.1 Paradigm shift in L2 oral proficiency ........................................................................ 11
    2.2.2 Defining Comprehensibility .................................................................................... 12
    2.2.3 Role of pronunciation in comprehensibility .............................................................. 13
    2.2.4 First language influence ......................................................................................... 14
    2.2.5 Setting priority in pronunciation teaching ............................................................... 15
  2.3 Issues in Realising Pronunciation Instructions for the IELTS Pronunciation Skill .......... 17
  2.4 Summary ........................................................................................................................ 21
  2.5 The Current Study ......................................................................................................... 22

Chapter 3: Methodology ......................................................................................................... 24
  3.1 Introduction .................................................................................................................... 24
  3.2 Methodological approach ............................................................................................... 24
  3.3 Instruments .................................................................................................................... 25
    3.3.1 Participants ............................................................................................................. 25
    3.3.2 Speech data ........................................................................................................... 26
3.3.3 Error coding measures ........................................................................... 28
3.3.3.1 Phonological measures ................................................................. 29
3.3.3.2 Segmental measures ...................................................................... 29
3.3.4 Speech rating ...................................................................................... 30
3.3.5 Error coding ....................................................................................... 32
3.3.6 Data analysis procedure ...................................................................... 32

Chapter 4: Results .......................................................................................... 34
4.1 Introduction .............................................................................................. 34
4.2 Rater Consistency .................................................................................. 34
4.3 Speech Data ........................................................................................... 34
4.4 Correlation Between Pronunciation Features and the IELTS Pronunciation Rating .................................................. 35
4.4.1 Eight phonological measures .............................................................. 36
4.4.2 Six segmental measures .................................................................... 36
4.5 Level Determinants of the IELTS Pronunciation Rating ...................... 37
4.5.1 Eight phonological measures .............................................................. 38
4.5.2 Six segmental measures .................................................................... 45
4.6 Correlation between the IELTS Pronunciation Rating and Comprehensibility ................................................................. 51

Chapter 5: Discussion .................................................................................... 53
5.1 Introduction ............................................................................................. 53
5.2 First Research Question ......................................................................... 53
5.3 Second Research Question .................................................................... 58
5.4 Third Research Question ......................................................................... 61

Chapter 6: Conclusion ..................................................................................... 63
List of Tables and Figures

Tables

121 Table 1: Length of Regidence and Age for 40 Japanese Speakers ......................... 27
122 Table 2: Result of the Classifications of 40 Speech Samples .............................. 35
123 Table 3: Pearson’s Correlations Between Eight Pronunciation Measures and the
IELTS rating Results .................................................................................................. 36
125 Table 4: Result of Person’s Correlations between Six Segmental Groups and the
IELTS Pronunciation Rating ...................................................................................... 37
127 Table 5: Summary of Confidence Intervals of Eight Pronunciation Measures........ 39
128 Table 6: Summary of Confidence Interval of Six Segmental Measures ................. 46

Figures

131 Figure 1: Confidence intervals of segmental errors .............................................. 40
132 Figure 2: Confidence intervals of syllable errors .................................................. 40
133 Figure 3: Confidence intervals of word stress misplacement............................... 41
134 Figure 4: Confidence intervals of word stress absence ....................................... 42
135 Figure 5: Confidence intervals of total word stress errors ................................... 43
136 Figure 6: Confidence intervals of intonation misplacement .................................. 44
137 Figure 7: Confidence intervals of intonation absence ......................................... 44
138 Figure 8: Confidence intervals of total intonation errors ..................................... 45
139 Figure 9: Confidence intervals of major segmentals ............................................. 46
140 Figure 10: Confidence intervals of errors in assimilation problems ..................... 47
141 Figure 11: Confidence intervals of secondary segmental errors ......................... 48
142 Figure 12: Confidence intervals of errors in diphthong problems ....................... 49
Figure 13: Confidence intervals of minor segmental errors

Figure 14: Confidence intervals of other segmental errors

Figure 15: Result of Pearson’s correlation between IELTS pronunciation rating and comprehensibility rating
Chapter 1: Introduction

1.1 Background

In a long history of controversy among second language (L2) research on the development of pronunciation teaching, one of the central focuses is what model to follow for instruction (Walker, 2010). Most recently, due to the expansion of variety and use of English following increasing globalisation, the standard of pronunciation instruction has been subject to changes from pursuing established regional varieties of English such as British English Received Pronunciation to achieving mutual intelligibility for wider communication in a global context (Brinton, 2012). Despite such transition in the standards however, most language instructions and pronunciation syllabi still hold implicit emphasis on the native speaker model and teachers’ intuitive decisions based on their norm of native-like pronunciation (Murphy, 2014), indicating the pressing need for fostering an awareness and development of the legitimate instructional model instead of mastering native-like ability.

Beside the communicative success in L2 oral communication in a global context, what educators need to be aware of when teaching L2 pronunciation is the needs of individual learner. In effect, quite a lot of learners of English who pursue academic careers in English-medium universities are eager to learn English pronunciation to obtain a certificate in their English exams. This is because the certificate is used as a tangible proof of one’s sufficient L2 proficiency for meeting the language requirements set by those institutions for admission purposes. In the context of the UK, International English Language Testing System (IELTS) is a dominant and recommended test to prove one’s proficiency level when applying for UK universities (British Council, n.d.-b). Since Band 7
on 9-point scale is commonly regarded as a benchmark to be accepted by these universities, candidates of the IELTS are in need of instructions for acquiring sufficient skills in every four skill (speaking, reading, writing, and reading).

1.2 Rationale

Importantly, it should be mentioned that there is already a plethora of evidence that shows the importance of pronunciation skills in order for one to be successful in the speaking section of the English proficiency test. For instance, Iwashita, Brown, McNamara and O’Hagan (2008) speculated that pronunciation errors could be an initial determining factor for examiners to make an overall proficiency judgment based on their finding that some learners, who have high vocabulary richness and poor pronunciation skill were assigned lower scores by raters. Thus, pronunciation is a crucial skill, which could negatively impact on overall rating and should be a first challenge for learners to overcome in order to make their speech assessable by examiners. Notably, in the case of IELTS pronunciation assessment, interplay of examiners’ rating process and their use of the scales have remained underresearched (see Isaacs, Trofimovich, Yu & Chereau, 2015). Given the minor status and unsystematic organization of pronunciation instruction within the speaking syllabus in general (Breitkreutz, Derwing & Rossiter, 2001) and a pressing need from learners to develop test-taking strategies for speaking tests (Issitt, 2008), it is reasonable to examine the legitimate pronunciation instructions and develop a tailor-made pronunciation syllabus of IELTS for test-takers who seek academic success in their careers. The development of such a syllabus would be especially helpful for native speakers of Japanese (NJs), who often receive limited amount of instructions on proper pronunciation in the classroom setting and need different instructions specialised for taking the IELTS test in
In order to study in UK universities. In fact, researchers claimed that Japanese learners face difficulties in articulating English in terms of various dimensions ranging from segmentals (Nishi & Kewley-Port, 2007; Riney & Takagi, 1999; Sekiyama & Tohkura, 1993) to suprasegmental levels (Ohata, 2004; Tsujimura, 2013) due to the great difference in phonetic system from that of English.

1.3 Aim and Structure of the Current Study

Therefore, in order to move beyond the current L2 pronunciation research and cast light on the development of pronunciation syllabus for test-taking purpose, the current study aim at developing strategic pronunciation teaching for IELTS test-takers, and in-depth understanding of the constructs to be measured in IELTS pronunciation skill. Also, the present study examines the adaptability of the IELTS pronunciation skills in real-life communication by setting the degree of correlation with comprehensibility as a barometer.

This thesis is organised in five parts: the first section overviews previous literature, which illustrates the underresearched respects of pronunciation teaching and generates the research questions of this study. Then the second section addresses the detailed account of the methodological approach and research design. The subsequent section begins with the presentation of the results, followed by the discussion with respect to the three research questions. Lastly, the findings of the study are summarised with research implications and limitations.
Chapter 2: Literature Review

2.1 Introduction

This chapter discusses the L2 oral proficiency model for language teaching/learning and the underlying issue of the adoptability in applying research outcomes to the context of the IELTS pronunciation skill development. The description of the conceptual change in L2 speaking proficiency principles (i.e., comprehensibility as a feasible goal) is introduced with empirical evidence, which is followed by the overview of the key components of the highly comprehensible pronunciation in past research findings. Subsequent section raises issues in the modelling of comprehensible pronunciation teaching in the IELTS pronunciation assessment.

2.2 Teaching L2 Pronunciation

2.2.1 Paradigm shift in L2 oral proficiency. L2 pronunciation research had experienced a major paradigm shift in its pedagogical practice/goal from nativelike pronunciation (i.e., accent-free speech) to intelligible pronunciation (i.e., accented but understandable pronunciation; Derwing & Munro, 2005; Levis, 2005). According to Levis (2005), this shift involved two contradictory principles: Nativeness principle which is defined as the instructions for the eradication of a foreign accent whereas intelligible principle’s main focus is described as the instructions for the production of understandable speech with a foreign accent. Although there are handful of exceptional cases (see Derwing, 2003; Kang, 2010; Scales, Wennerstrom, Richard & Wu, 2006; Timmis, 2002; Tokumoto & Shibata, 2011), researchers have found that unattainability of accent-free speech is due to
various factors such as learners’ language background (Flege, 2003), age of acquisition (Flege, Munro, & MacKay, 1995), quality and quantity of input (Flege, Yeni-Komshian, & Liu, 1999), attitude, motivation, and aptitude toward L2 speech learning (Moyer, 1999). Therefore, foreign accented but intelligible pronunciation is the realistic goal for L2 learners hoping to achieve international communicative success (e.g., Derwing & Munro, 2009; Tutor et al., 2016a; Tutor et al., 2016b; Tutor et al., 2016c).

2.2.2 Defining comprehensibility. With the shift of paradigm, enormous amounts of research have been carried out to deconstruct the components of intelligible and unintelligible pronunciation though listener-judgment of accentuatedness and comprehensibility (for review of the various approach, see Derwing & Munro, 2005). Accentuatedness refers to the listeners’ perception of how close the speaker’s language is to the speech patterns of the target-language community (Munro & Derwing, 1999), thus best described as linguistic nativelikeness (Tutor et al., 2015c), while comprehensibility is frequently used in L2 pronunciation research as the synonym of intelligibility (Levis, 2005). Although Isaacs (2008) pointed out that the confusion in the use of intelligibility between each study is due to its manifold definition and interpretation of the term (for more discussion of the term, see Isaacs, 2008), a rather clear account has been provided by Munro & Derwing (1999). According to their explanation, intelligibility and comprehensibility both indicate the listener’s ability to understand L2 speech in the broader meaning (see Levis, 2005), yet, in a narrower sense, they are different in the way they measure the listener’s understanding: intelligibility measures the amount of understanding via orthographical transcribing whereas comprehensibility measures the degree of understanding through a scalar rating. Thus, the distinction of the two terms seems to be made by the way of operationalising the measurement (Isaacs et al., 2015). This study, therefore, defines accentuatedness as the listeners’ perceptual degree of nativeness of the L2 speech, comprehensibility as the listeners’ degrees
of understanding of L2 speech based on rating, and intelligibility as, in a broader sense, listeners’ ability to understand L2 speech. Assuming that learners can communicate successfully with an accented L2 (Derwing & Munro, 2009), researchers have been attempting to reveal the unique and shared linguistic factors that underlie comprehensibility and accentedness for developing practical L2 pronunciation pedagogy.

2.2.3 Role of pronunciation in comprehensibility. To date, while the distinctiveness of comprehensibility and accentedness were argued (e.g., Jenkins 2000; Kang, Rubin & Pickering, 2010; Munro & Derwing, 1995), research evidence supports that pronunciation features both segmental and prosodic aspects impact on both accentedness and comprehensibility judgments. Accentedness seems to be strongly associated with segmental accuracy, temporal measures, syllable duration, stress, and pitch range (Anderson-Hsieh, Johnson & Koehler, 1992; Winters & O’Brien, 2013), while comprehensibility is linked to segmental sounds in stressed syllables (Zielinski, 2008), word stress (Field, 2005), primary stress in sentences (Hahn, 2004), tone choice (Pickering, 2001; Wennerstom, 2001) as well as lexicogrammatical accuracy (Munro & Derwing 1995; Varonis & Gass 1982).

While these studies seem to have a bias towards the negative impact of non-segmental aspects of speech, without a shadow of a doubt it is clear that some segmental difficulties should be given attention as well. One of the segmental-focused studies on comprehensibility, for example, was conducted by Munro and Derwing (2006) who selected the segmentals from Functional Load (FL) theory, which is a ranking list of segmental contrasts based on their importance in English pronunciation, developed by Brown (1991) and Catford (1987). The finding revealed that high FL (/l, f, n, s, d/) errors were rated worse than that of low FL (/ð, θ/) to greater extent on both accentedness and comprehensibility (Munro & Derwing, 2006). This indicates that not all but certain segmentals appear to be crucial for comprehensibility, underpinning the essential contribution of segmentals on
These comprehensibility studies were more extensively examined by Isaacs and Trofimovich (2012), and Trofimovich and Isaacs (2012), who employed various ranges of measures including pronunciation, fluency, lexis, grammar, and discourse structure for linguistic coding. For instance, Trofimovich and Isaacs (2012) found that comprehensibility and accentedness were considerably linked to pronunciation measures such as word stress and rhythm, but comprehensibility was also influenced by grammatical accuracy and lexical type frequency and accentedness judgments were uniquely tied to the listeners’ perceptual salience of vowels and consonants errors (Trofimovich & Isaacs, 2012). However, while a follow-up study by Tutor et al. (2015c) confirmed the impact of word stress errors on both accentedness and comprehensibility, they also revealed that segmental errors not only affect accentedness but also comprehensibility judgement. Furthermore, the result indicated that the overall rating score of comprehensibility were higher than that of accentedness, supporting the claim that one’s accented L2 speech can be highly comprehensible (Derwing & Munro, 2009; Jenkins, 2000; Munro & Derwing, 1995).

2.2.4 First language influence. Yet there is a general consensus that one’s first language (L1) can impact on L2 development and this topic has been widely discussed in the L2 phonological development literatures with multiple learning models (e.g., Darcy, Dekydtspotter, Sprouse, Glover, Kaden, McGuire & Scott, 2012; Eckman, 2004; Eckman & Iverson, 2013; Escudero & Boersma, 2004), little attention has been paid to how speaker’s L1 come into play in listener’s judgment of comprehensibility. One of the predominant approaches used in the L2 development research is Lado’s (1957) contrastive analysis hypothesis (CAH), which gives an account of L2 learners’ phonological development in terms of similarities and differences between speaker’s L1 and L2 phonologies. On one hand, differences of speaker’s L1 and L2 are emphasized to describe the pattern of L2
development (e.g., Lado, 1957), on the other hand, some researchers focus on the similarities of these two languages (Oller & Ziahosseiny, 1970; Young-Scholten, 1985), however, both lines of the research are devoted to giving the account of the difficulty L2 learners face (for reviews, see Eckman, 2004; Major, 2008). Yet still underdeveloped, this array of studies demonstrated that learners’ learning difficulties and error types in L2 pronunciation may vary depending on their L1 backgrounds (Eckman & Iverson, 2013).

Pertaining to listener-based comprehensibility studies, however, the L1 variable has not receive sufficient attention except for few studies such as Tutor et al. (2015a). They looked at how L1 differences affect the correlation between errors in linguistic features and degree of comprehensibility. Although no strong association was found with Farsi speakers, their results showed that comprehensibility was linked to pronunciation variables with Chinese speakers, lexicogrammar with Hindi speakers, both pronunciation and lexicogrammar variables with Romance speakers, indicating that L1 background certainly bears relevancy to comprehensibility. Thus, further expansion and refinement of the research is required to find which features are uniquely tied to the specific L1 group’s L2 comprehensibility.

**2.2.5 Setting priority in pronunciation teaching.** The range of studies discussed above have empirically proven that pronunciation features especially word stress, segmentals and rhythm seem to be the overlapping components of comprehensible speech and native-like speech. However, regarding the development of pronunciation syllabi for teaching, these research outcomes are sparse especially with respect to segmentals. While non-segmental features are relatively clear on the importance of certain non-segmental features over others (e.g., word stress), further clarification needs to be done on segmentals. This is because they often treated as a homogeneous measurement category albeit they are an aggregation of the individual sounds except for one exception of Munro and Derwing’s (2006) use of
Functional Load theory. Yet this study confirmed that certain segmentals should take precedence over other segmentals due to their different degree of impact on comprehensibility, fine-grained research on segmental classification is fairly limited. Beside the Functional Load theory (Brown, 1991; Catford, 1987), Jenkins (2002), from the world English paradigm, sought to find the crucial pronunciation features and its teaching priorities to attain global intelligibility between non-native speakers of English (NNEs). Her pronunciation syllabus (i.e., *Lingua Franca Core*) is based on the observation of the breakdowns of interaction between NNEs, and comprise of *core* and *non-core* which separate crucial features especially individual segmentals from the peripheral features. Fairly recently, Tutor (2014) developed an exhaustive list of segmentals and suprasegmentals to be prioritized for native Japanese learners of English. Based on the cross-linguistic analysis between English and Japanese, 25 problematic pronunciation features including specific segmentals were carefully selected and rated on importance by 120 experienced teachers (NEs and NJs). The results revealed that pronunciation needs to be taught in the following order: major segmentals (/l, ɹ, ð, q, v/), L1 effect at syllable levels (cognates, syllabification), assimilation (/si, ŋi, ti/), stress/intonation (sentence/lexical stress, intonation), secondary segmentals (/æ, ʌ, ɪ/), diphthongs (/aʊ, aɪ, oʊ, ɔɪ, eɪ/) minor segmentals (/p, t, k, w, n, ŋ, h/) and contractions, and fluency problems (Tutor, 2014). Compared to the *Lingua Franca Core* (Jenkins, 2002) practice which considers NNEs of various language backgrounds and deals with universally essential features for NNEs, the uniqueness of Tutor’s (2014) study is its focus on a single L1 group (i.e., Japanese) with its theoretical basis from L2 phonological development theory (i.e., contrastive analysis hypothesis). Nevertheless, research on prioritisation of certain segmentals is still unsatisfactory to identify universal and L1 specific segmental features to develop pronunciation syllabus for improving one’s comprehensibility.
2.3 Issues in Realising Pronunciation Instructions for the IELTS Pronunciation Skill

Although the research on constructs of comprehensibility certainly informs instructional agenda of L2 pronunciation, practitioners need to consider how such empirical evidence can be applied to the specific contexts, where learners’ success is beyond the accomplishment of L2 communication. In the academic context, L2 learners, who pursue academic careers in English-medium universities, need to obtain the skill to meet the standard measured through the English proficiency tests as a part of their application to the university (Feast, 2002).

Among UK universities, for instance, International English Language Testing System (IELTS) is the dominant and preferred high-stake English proficiency tests for assessing applicants for admission purposes. The IELTS score, which consists of the discrete marks of four disciplines (listening, writing, reading, and speaking) and overall score, is regarded as a predictor of the applicants’ proficiency levels to be successful in their courses (Hayatt & Brooks, 2009; Hayes & Read, 2004; Issitt, 2008). Upon selection, most higher education providers set the minimum requirements of the IELTS score (for more detail of the minimum Band scores for application, see British Council, n.d.-c). The IELTS assesses the test-takers pronunciation skill as one of four subcomponents of their speaking skill (i.e., fluency, grammatical accuracy, lexical richness) together with the other three skills (i.e., listening, writing, reading).

With respect to IELTS pronunciation skill development, there are at least two issues to consider for operationalising comprehensibility practice: (a) the inconsistency of the measuring construct, (b) vagueness in the rating descriptor. Firstly, the construct (conceptual model) being measured in IELTS pronunciation skill is inconsistent (Isaacs et al., 2015) thus
the appropriateness of the adaptation of comprehensibility practice to IELTS pronunciation skill is uncertain. Isaacs et al. (2015) examined the pronunciation scales consulted by accredited IELTS examiners (the scales are different from publically available version), and found that the scale includes a mixture of three constructs (intelligibility, comprehensibility, and accentedness), which seems to be a common shortcoming among L2 oral proficiency tests (Harding, 2013; Isaacs & Trofimovich, 2012). The actual criteria for the score judgement are provided by Seedhouse, Harris, Naeb, and Üstünel (2014): “Pronunciation refers to the capacity to produce comprehensible speech in fulfilling the speaking test requirements. The key indicators will be the amount of strain caused to the listener, the amount of unintelligible speech and the noticeability of L1 influence” (p. 5). According to the definitions discussed above, it can be speculated that the terms “comprehensible speech” and “strain caused to the listener” refer to comprehensibility, and “unintelligible speech” corresponds to intelligibility as a synonym of comprehensibility, since IELTS pronunciation assessment involves examiners’ subjective scalar ratings, which is the same measuring system as comprehensibility (Isaacs et al., 2015), indicating that their construct is built on a comprehensibility paradigm. However, “the noticeability of L1 influence” is associated with accentedness. Thus, from the researchers’ perspective, it is not clear to what extent the degree of accentedness comes into play in the raters’ judgement, as empirical evidence supports that a L2 speaker with an accent can be highly comprehensible (Isaacs et al., 2015).

On top of the examiner’s assessment criteria, the ambiguity also exists in key indicators of the public rubric (see Appendix A). Comprehensibility-associated indicators are observed in Band 9 (“effortless to understand”), Band 8 (“easy to understand throughout”), Band 6 (“generally be understood throughout”), Band 4 (“cause some difficulty for the listener”), and Band 2 (“unintelligible”), while accentedness-related indices do exist in Band 8 (“L1 accent has minimal effect”), and in Bands 6 and 4 (“mispronunciation”). Such
conflation may cause teachers and learners difficulty in understanding what model to follow for improving IELTS pronunciation skill (i.e., nativelike or comprehensibility model).

The second issues underlying the operation of IELTS pronunciation instruction is the vagueness of the distinction between each band due to the lack of information in pronunciation features illustrated in the public version of score indicators. The vagueness issue is two folds: vagueness in the description of pronunciation features in Bands 4, 6, 8, and 9, and vagueness in the description of Bands 3, 5, and 7. Regarding the first vagueness, with the exception of Band 2 (“often unintelligible”) and Band 1 (“no communication possible, no rateable language”), the differentiation of the expected pronunciation features to be performed in Bands 2, 4, 6, 8 and 9 is crude: “a full range” in Band 9, “a full range” in Band 8, “a range” in Band 6, “a limited range” in Band 4 respectively. This vagueness of the features indicates that further specification is needed for clear understanding of what components pronunciation features contain (i.e., segmentals and suprasegmental features) for teachers and examiners (Isaacs et al., 2015). The second vagueness is the underspecification of Bands 3, 5, and 7 compared to that of relatively detailed explanation of Bands 2, 4, 6, 8 and 9. As Isaacs et al. (2015) summarized, these three bands are briefly defined by referring to the descriptions of the band description immediately below and above. For instance, Band 7 is defined as “shows all the positive features of Band 6 and some, but not all, of the positive features of Band 8”. This vagueness of definition in between bands (Yates, Zielinski, & Pryor, 2011, p. 34), again creates greater confusion for examiners in assigning the score as well as teachers and test-takers to clarify what they need to aim for (Isaacs, et al., 2015). For Band 7, this vagueness is especially problematic for international students who seek a degree in English-medium universities since Band 7 is often used as a threshold for the admission purposes in the universities. Therefore, although the pronunciation rubric has been revised to a 9-point scale (for the detail of the history of
revision, see DeVelle, 2008) due to its coarse classification of the old 4-point based old scale
where (Brown, 2006), the indicators still lack of its detailed discrimination between bands
and components of the pronunciation features they assess.

Although direct contribution to the IELTS pronunciation syllabus has not been made,
a study by Isaacs and her colleagues’ (2015) provides insights into the two issues discussed
above. They examined the linguistic correlates, which are attended to by the accredited
IELTS examiners when they discriminate between scores in the pronunciation rating as well
as how comprehensibility is modelled in the actual rating. For their study, 80 pre-rated
speeches of the IELTS candidate from various L1 background such as Chinese, Arabic,
Tagalog, Spanish, Thai, Kannada were re-rated by eight IELTS raters in terms of seven
linguistics measures (i.e., grammatical accuracy, lexical richness, speech chunking, vowel
and consonant errors, word stress, intonation, and speech rate), comprehensibility and IELTS
pronunciation rubric. The pronunciation rating was strongly correlated with grammatical
accuracy, lexical richness and speech chunking whilst most of the pronunciation features
such as vowel and consonant errors, word stress, intonation and speech rate had a low
correlation with the rating. Notably, they also reported that comprehensibility showed the
least correlation with the IELTS pronunciation rating. These results indicated that the IELTS
pronunciation skill is not just about comprehensibility but other factors were taken into
account for the judgement. Also, pronunciation features had little impact on the rating but
were influenced by other factors such as lexicogrammar. With regard to the score
discriminating factors, all the measures contributed to the discrimination between Bands 6
and 7. In addition, speech rate and lexical richness also distinguished Bands 5 and 6, but
other measures were not strong determinants of separating Bands 5 and 6 or Bands 7 and 8.
Their findings shed light on the issue of rubric vagueness and inconsistency of the measuring
construct in IELTS pronunciation assessment by revealing linguistic factors involving the
examiners’ actual rating process and discrimination of the bands. Their investigation is the very first study, which looked at IELTS pronunciation assessment and explored the phonological features, which differentiate the pronunciation scores. Therefore, setting this study as a starting point, further details need to be researched in terms of the contribution of pronunciation features such as the weight of individual segmentals. In addition, it is worth examining how a learner’s L1 background comes into play on the pronunciation rating since research evidence on L2 pronunciation development and comprehensibility discussed above have supported the impact of the L1 on L2 production (e.g., Tutor et al., 2015b).

2.4 Summary

This section discussed the literature regarding the key concepts and findings involved in this present research. First, the paradigm shift in L2 pronunciation model was introduced by an overviewing of the key literatures. Due to the ample amount of research evidence of unfeasibility in attaining nativelike pronunciation, researchers have proposed that comprehensibility is a realistic goal for L2 learners, and explored the linguistic factors, which contribute to L2 comprehensibility. Followed by the defining of three core terms in this study (Intelligibility, comprehensibility, and accentedness), the essential constructs of comprehensibility and accentedness have been described through comparison of relevant studies. Through the examination of the research findings, it was confirmed that pronunciation features not only impact accentedness but also on comprehensibility. Further, based on the L2 pronunciation development literature, the influence of learner’s L1 on comprehensibility was presented in the latest findings of Tutor et al. (2015b). Then, the lack of the detail in the selection of segmentals for the development of a pronunciation syllabus towards comprehensibility was pointed out, followed by the introduction of available
research evidence, which attempt to describe the priorities of certain segmentals. Lastly, the
relevancy of developing the syllabus for IELTS pronunciation teaching was emphasised and
the two issues in adopting the comprehensibility research outcome to the IELTS context was
pointed out with the findings of a key study on linguistic correlates of the IELTS
pronunciation rating.

2.5 The Current Study

As we reviewed in the previous section, the scope of the pedagogical research of
pronunciation does not cover the development of instructional guidelines for testing contexts
such as the IELTS pronunciation assessment. Also, the relevant evidence of the IELTS
pronunciation rating from the language testing research side is limited to one study done by
Isaacs et al. (2015). Therefore, in order to establish a strategic guideline for the IELTS
pronunciation assessment, further research is required with an in-depth examination of the
contribution of individual segmentals and test-takers’ L1 backgrounds to the rating results.

Thus, to advance the current development of the pronunciation teaching and deepen
the latest understanding of the key constructs measured by the IELTS pronunciation ratings,
the present study set its goals to identify crucial pronunciation features and segmentals
which link to the IELTS pronunciation rating itself and discrimination of the band scores by
focusing on one L1 group (i.e., Japanese adult learners). This study also explores the
relationship between comprehensibility and IELTS pronunciation skill to examine the
practicality of the IELTS pronunciation skill for real life communication success. The
research questions are as follows:
1. Which pronunciation aspects are priorities for teaching the IELTS pronunciation skill to native Japanese learners of English in order to improving their score?

2. What pronunciation aspects most distinguish different proficiency levels of NJ learners of English in IELTS pronunciation rating?

3. Is the IELTS pronunciation skill useful for the learner’s real life communication success?
3.1 Introduction

In this chapter, the methodological approach and research design used to arrive at the answers to the research questions are introduced together with the discussion of its suitability for the current research. This chapter also covers data collection procedure and how they are analysed.

The first section gives an account of a quantitative approach employed in the current study, this is then followed by the overview of the research design, which contains the illustration and justification of the methods adopted to obtain the valid data for examining the research questions. The subsequent section involves the description of the data collection procedure including the participants, instruments and data collection step, which lead into the elucidation of data analysis.

3.2 Methodological approach

In line with the series of research on L2 prosodic features and development (e.g., Tutor et al, 2015a; Tutor et al., 2016b), the current study employs a quantitative approach to seek answers for the three research questions (cf. Chapter 2). The nature of quantitative approach is to deal numerical data and enables researchers to run descriptive analysis and inferential statistics (Lazaraton, 2005; Loewen & Plonsky, 2016). Since the current study seeks to identify the numbers of errors L2 speakers make, how many errors different proficiency groups make (i.e., mean of the errors and possible error ranges), how high and low the speech is rated based on the 9-point scale (i.e., average scores) for the analysis, quantitative approach is regarded as best suited. For data collection, in order to analyse the
relationship between assigned IELTS pronunciation score and the number and types of errors in the speech, speech rating by native judges and phonological error coding are used. These two methods are predominant combination in the empirical research (e.g., Isaacs & Trofimovich, 2012; Trofimovich & Isaacs, 2012).

Regarding listener-based speech rating, it has been extensively used for L2 speech evaluation. While rating consistency has been a focus of the debate among number of studies (e.g., Calloway, 1980; Thompson, 1991, Rossiter, 2009), relatively high degrees of reliability have been proven by rater-judgement study of segmental judgement (Cunningham-Anderson & Engstrand, 1989; Piske, MacKay, & Fledge, 2001), and comprehensibility and accentedness judgements (Isaacs & Thomson, 2013), suggesting that scale-based listener judgements hold sufficient reliability for the current study. As for an ethical issue of participant recruitment for rating, present study followed ethical research procedure set by author’s university, and obtained an ethical approval from the department ethics officer. Phonological coding or linguistic coding in broader sense, have been directed either auditory measure to count the frequency of the specific phenomenon or coding software such as speech coding software Praat (Boersma & Weenink, 2010) or lexical tutor (Cobb, n.b.) for measuring lexical aspects. While some studies such as Trofimovich and Isaacs (2012) used both auditory and analysis software, present study only employed auditory measure for error frequency counting in order to focus on errors that are detectable to teachers.

3.3 Instruments

3.3.1 Participants. The participant (henceforth, raters) for speech rating in the present study included seven teachers (M_{age} = 36 years, range = 29–56) recruited via email in
London. They speak English as their L1 and were borne and/or raised in English-speaking home environment with at least one English-speaking parent except for one rater whose parents’ L1 were Urdu. They reported that English is the major medium of the communication ($M = 96\%$, range = 90–100). In order to minimise the inconsistency and inaccuracy in evaluating the speech, the current study controlled two rater variables via language background questionnaire (see Appendix B): (a) rater’s familiarity with non-native speakers L2 (see Isaacs & Trofimovich, 2012 for detailed discussion of listener factor), and (b) raters’ linguistic and teaching experience (for the high reliability and accuracy in pronunciation rating of linguistic and teaching experience, see Tutor et al., 2015c). With respect to familiarity of Japanese-accented English, the raters reported high familiarity with Japanese-accented English ($M = 5.14$, range = 4–6) on 6-point likert scale (1 = not at all, 6 = very much) with frequent contact with NJs ($M = 5$, range = 4–6) on 6-point likert scale (1 = very infrequent, 6 = very often). Four out of seven have experience of visiting Japan (two weeks, one month, six month, and two years) and two of the four raters took short-term Japanese language courses. Thus, all the raters were regarded as consistent in their familiarities with Japanese accented English. Considering their linguistic knowledge and teaching background, all the raters responded that they have teaching experience in EFL/ESL contexts ($M = 9.9$ years, $range = 3$-25), and have extensive linguistic/phonological knowledge obtained through either enrolled in or completed their master degree in Applied Linguistics, TESOL, or TEFL courses in English-medium university. Therefore, the seven raters are regarded as relatively homogenous in their phonological and linguistic knowledge as well as their teaching experience.

### 3.3.2 Speech data
The present study used 40 speech data from Tutor’s (2011) unpublished corpus, which contains more than 200 audio data collection of Japanese learners who completed various speaking tasks in Japan and Canada. For the sake of the
investigation, careful selection had been made to include various proficiency levels of the
speakers based on their amount of L2 immersion, which is regarded as a contributing factor
to define one’s L2 developmental stage (see detailed discussion Fledge & Liu, 2001). Table
1 illustrates the distribution of the length of residence in English speaking countries, ranging
from 0 to 24 months ($M = 4.6$) with their age information (see Table 1). On the selection of
the audio, those 40 speakers’ performance were chosen specifically because the speakers
answered to the similar prompt to the one used in the IELTS independent long-turn task,
which was also chosen for the prior IELTS pronunciation study by Isaacs et al. (2015).

<table>
<thead>
<tr>
<th>Length of the stay (months)</th>
<th>$n$</th>
<th>Age</th>
<th>$n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>16</td>
<td>18–19</td>
<td>5</td>
</tr>
<tr>
<td>0–1</td>
<td>9</td>
<td>19–20</td>
<td>16</td>
</tr>
<tr>
<td>1–10</td>
<td>7</td>
<td>20–30</td>
<td>11</td>
</tr>
<tr>
<td>10–24</td>
<td>8</td>
<td>30–53</td>
<td>8</td>
</tr>
<tr>
<td>Total</td>
<td>40</td>
<td>Total</td>
<td>40</td>
</tr>
</tbody>
</table>

The IELTS speaking test takes an interview style with an examiner. In the long-term
part, which is the second part of the test, the examiner provides a test-taker a prompt sheet
and asks the test-taker to speak after one or two minutes of preparation time. The prompt
used in the task is “Describe the hardest and toughest challenge in your life” and detail of
the prompt sheet is shown in Appendix C. Being in line with previous speech judgement
studies (e.g., Iwashita et al., 2008; Derwing & Munro, 1997), first 30 seconds of the entire
speech were excised from each of the 40 speaker’s performance (approximately three
minutes each) for preparation of the ratings. Since this study’s focus is on pronunciation, not
lexicogrammar or fluency, this study decided that 30 seconds is long enough to identify
pronunciation errors due to the empirical evidence such as Munro, Derwing, and Burgess’s
(2010), which revealed that accented detecting ability of NEs was fairly reliable even with
single word. This is for avoiding raters’ fatigue, which affect the accuracy of the rating result
due to its time-consuming nature, and completing all the error coding in the limited time
allocation.

**3.3.3 Error coding measures.** For the error coding of the 40 tracks of speech data,
eight pronunciation measures and six segmental measures were developed through two
studies (i.e., Trofimovich & Isaacs, 2012; Tutor, 2014). The eight phonological measures
contain segmental error, syllable error, misplacement of word stress, word stress absence,
misplacement of intonation, intonation absence. In order to include all the possible
phonological factors, which involve the IELTS pronunciation rating, the eight pronunciation
measures for current study were carefully selected from the research by Trofimovich and
Isaacs (2012) due to their in-depth coverage of phonological features. Regarding the six
problematic English segmental groups for NJs, the measures were adopted from Tutor’s
(2014) priority list of problematic segmentals and suprasegmentals for NJs. This list was
particularly chosen because (a) the list offers grouped individual English segmentals, which
enable the current study to investigate the individual segmental errors in efficient way, (b)
the list represents the pronunciation problems unique to Japanese learner of English. All the
segment groups were derived from the contrastive analysis between Japanese and English as
well as the perception of native English and Japanese teachers of English, which enable the
current study to develop a educational guideline specific for Japanese learners of English,
and to examine universality of the results by comparing them with the prior linguistic
correlates study of IELTS pronunciation by Isaacs et al. (2015), whose L1 background of
speech samples were varied.
3.3.1 Phonological measures. Phonological measures are ranging from the segmental level (individual vowels and consonants) to suprasegmental level (syllables, words, and phrases). The number of errors in each measure was counted based on the auditory evaluations by the researcher and two additional coders. The details of the measures are as follows:

- Segmental errors (total): the errors caused by phonemic substitutions (e.g., *life* articulated as *rife*).
- Syllable structure errors: the errors caused by vowel or consonant insertion (e.g., *it* spoken as *ito*) and deletion (e.g., *year* articulated without the initial /y/).
- Word stress misplacement: the errors caused by the misplacement of primary stress in polysyllabic words (e.g., *CHAL-lenge* spoken as *Chal-LEBGE*).
- Word stress absence: the errors caused by the absence of primary stress in polysyllabic words (e.g., *CHAL-lenge* spoken as *chal-lenge*).
- Word stress error (total): the total errors of word stress misplacement and absence.
- Intonation misplacement: the errors in the adequate choice of pitch in the place where certain type of tone is expected (e.g., falling tone at the end of statement spoken in raising tone).
- Intonation absence: the errors of failing to produce any pitch movements (e.g., falling tone at the end of statement spoken with no tones).
- Intonation error (total): the total of errors in intonation misplacement and absence.

3.3.2 Segmental measures. Segmental measures comprise six groups of segmentals (individual vowels and consonants) adapted from Tutor’s (2014) problematic segmentals and suprasegmental list (for the detailed segmental list see Appendix D). Through objective
analysis by the researcher and two additional coders, phonemes, which caused the segmental
errors, were identified and counted the numbers according to the group they belong to. The
details of the measures are as follows:

- **Major segmentals (IPA, l, θ, v, ð):** the errors in articulating two approximants (/j, l/),
  and three fricatives (/θ, v, ð/) in words.
- **Assimilation problems (IPA, si, ti):** the errors in articulating allophonic rules (/ʃi, si,
  ti/) in words.
- **Secondary segmentals (IPA, æ, ʌ):** the errors in articulating a fricative (/f/), and two
  vowels (/æ, ʌ/) in words.
- **Diphthong problems (IPA, aʊ, aɪ, oʊ, ɔɪ, ɛɪ):** the errors in articulating of diphthong
  (/aʊ, aɪ, oʊ, ɔɪ, ɛɪ/) in words.
- **Minor segmentals (IPA, ʃi, si, ti, Contraction):** the errors in articulating one
  approximants (/w/), two nasals (/n, η/), one fricatives (/θ/), plosives (/p, t, k /),
  allophonic rules (/ʃi, si, ti/) in words and contractions (e.g., won’t, can’t).
- **Other segmentals:** the errors in articulating the phonemes not included in the
  group above.

**3.3.4 Speech rating.** Seven raters judged the 40 speeches of Japanese learners of
English based on the IELTS pronunciation rubric and compressibility. The rating sessions
were conducted individually in a quiet room (each rater participated in rating sessions on
different day and time) using a rating software Praat which shows two box-shaped 9-point
likert scales (1–9) for comprehensibility and the IELTS pronunciation judgement
respectively, raters then clicked the numbers to make the judgements (see Appendix E). The
raters listened to the audios through a pair of earphones connected to a personal laptop computer.

On each rating, the raters were asked to listen to a whole audio (i.e., 30 seconds) before they make the judgements on the two scales. The software played the speech only once at randomised order, and the next speech was played right after the raters made the judgements of the previous speech. All the ratings were recorded by clicking the numbers on the two scales. Prior to the actual rating sessions, the language background questionnaire and sufficient training were given to the raters. In the first phase, the researcher asked the raters to complete the language background questionnaire shown in the computer screen in order to obtain the information about the languages they speak and their degree of the familiarity with Japanese and Japanese accented English. In the second stage, rater training session was directed by the researcher to ensure that the raters fully adequately understood the IELTS pronunciation rubric (9-point scale) and comprehensibility. In the training session, the physical copies of the IELTS pronunciation rubric (9-point scale) and adequate time were provided with the raters in order to help them fully understand the rubric. They were allowed to check the rubric anytime during the rating process if necessary. The researcher then explained typical pronunciation errors made by Japanese learners of English (e.g., syllabification problem caused by the frequent vowel insertion) to raise the raters’ awareness to the pronunciation errors. Then, based on the definition used in the previous research of comprehensibility (e.g., Trofimovich & Isaacs, 2012), the concept of comprehensibility (how effortless to understand L2 speech) was introduced to the raters with a 9-point scale (1 = hard to understand, 9 = easy to understand). For the comprehensibility rating, they were asked to make judgement intuitively and use the scales flexibly. The raters listened to three sample audios and complete practice judgement’s with the rating software and discussed the result with the researcher to confirm the raters’ accurate understanding of the two scales.
3.3.5 Error coding. Firstly, eight pronunciation measures and six segmental measures were coded by the researcher, who is a native speaker of Japanese and familiar with Japanese-accented English due to the teaching experience to Japanese learners of English. Subsequently, in order to ensure the correctness of the error coding by the researcher and pursue a fine-grained objective measurement, the result of the first error coding was re-coded by two Japanese coders. Those coders were given sufficient instructions about the error detection before the actual codings (e.g., check the segmental errors as if they were correcting their students’ pronunciation based on their own ideal pronunciation model). Two coders are experienced Japanese teachers of English (seven and five years) and they are also knowledgeable about English phonology and phonetics with master degrees in TESOL and Applied linguistics. For the coding, Japanese teachers’ subjective judgements were employed because native Japanese teachers of English are assumed to be able to detect the influence of Japanese language easily in the L2 performance compared to the native English teachers who do not speak Japanese as their L1. The results of the additional coding did not show major difference from the first coding result, suggesting that three coders’ judgements were internally consistent.

3.3.6 Data analysis procedure. In line with the prior relevant literature, which examined L2 speech from various perspectives (e.g., Tutor et al., 2015c), the quantitative methods were carefully selected for the data analysis based on the research questions set in this study. First, each speakers’ mean IELTS pronunciation scores were calculated through the averaging of seven raters’ rating results in order to estimate each speaker’s IELTS pronunciation scores. For answering the first research question, forty mean scores were then examined through Pearson’s correlation to see the correlation with the coding results of eight pronunciation measures and six problematic segmental groups. The second method, which was applied to the second research question, involved in dividing of 40 rated speech samples
into three proficiency populations (i.e., beginner, intermediate, advanced). Then three groups of confidence intervals were computed to identify the differences in possible ranges of phonological and segmental error between three groups (Loewen & Plonsky, 2016). Finally, in order to obtain the answers of the third research question, each speaker’s comprehensibility scores assigned by seven raters were averaged, and compared with the average scores of the IELTS pronunciation via Pearson’s correlation.
Chapter 4: Results

4.1 Introduction

This chapter presents an analysis of the data collected through the native judges’ rating session and the error coding result, and restates the research questions addressed in Chapter 2. Based on the quantitative results obtained via the collected data from the rating session and error coding, Pearson’s correlation and confidence intervals are calculated to identify the relationship between (a) the IELTS rating and eight pronunciation measures, (b) the IELTS rating and six segmental measures, and (c) the IELTS rating and that of comprehensibility. Each result is presented in response to the research questions.

4.2 Rater Consistency

An inter-rater reliability among seven NE raters was calculated through the averaging of the values yielded through Pearson’s correlation. The result supported a strong correlation in their ratings ($r = .633$, $p > .1$), suggesting seven raters’ judgements were consistent.

4.3 Speech Data

The rating results of 40 speech samples were manually classified into three proficiency levels (i.e., beginner, intermediate, and advanced) in order to run a statistical analysis for answering second research question (see Chapter 2). The result of the grouping is presented in Table 2 below. As shown in the table, seven NJ speakers’ mean scores were Band 4 and another seven speakers’ averaged scores were Band 5, fifteen of the speakers
had a mean score of Band 6, eight speakers had a mean score of Band 7, and three speakers had a mean score of Band 8. Owing to the small sample sizes assigned to Bands 4, 5, 7, and 8 compared to that of 6, Bands 4 and 5 were conflated into beginner set ($n = 14$), Bands 7 and 8 were grouped together to make up advanced set ($n = 15$) whereas Band 6 remained as it was and made up intermediate set by itself ($n = 11$).

Table 2

<table>
<thead>
<tr>
<th>Means of IELTS scores (Band score 0–9)</th>
<th>Number of speeches</th>
<th>Classification</th>
<th>Number of speeches in the proficiency groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>7</td>
<td>Beginner</td>
<td>14</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>15</td>
<td>Intermediate</td>
<td>15</td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>Advanced</td>
<td>11</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4.4 Correlation Between Pronunciation Features and the IELTS Pronunciation Rating

The first research question investigated the pronunciation features (i.e., segmental, syllables, word stress, and intonation) of NJ learners of English, which were most crucial in the IELTS pronunciation rating. In order to examine the strength of the association between the mean rating results of 40 speech data and coded results of eight pronunciation features (i.e., which features most affect the rating result), Pearson’s correlations were computed. Also, in order to obtain in-depth insight on segmentals for instructional purpose, the correlation between the rating result and the error means of six problematic segmental groups were examined through the correlation.

4.4.1 Eight pronunciation measures. As is presented in Table 3, the correlation
between eight pronunciation measures and the IELTS rating results were calculated. The IELTS rating result showed significantly strong negative correlation with segmental errors \( (r = -.590) \) and syllable errors \( (r = -.472) \). Similarly, the IELTS rating result was moderately and negatively correlated with word stress absence \( (r = -.294) \) and total number of word stress errors \( (r = -.304) \). The remaining four measures did not show any significance in the correlation with the IELTS rating result: misplacement in word stress category) and misplacement, absence, and total in intonation category. Especially, all the intonation measurements showed positive correlation values compared to the rest of the measurements, which all showed the negative correlation values.

Table 3

<table>
<thead>
<tr>
<th>Pronunciation measures</th>
<th>( r )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Segmental errors</td>
<td>-.590***</td>
</tr>
<tr>
<td>Syllable errors</td>
<td>-.472**</td>
</tr>
<tr>
<td>Word stress misplacement</td>
<td>-.101</td>
</tr>
<tr>
<td>Word stress absence</td>
<td>-.294*</td>
</tr>
<tr>
<td>Word stress error total</td>
<td>-.304*</td>
</tr>
<tr>
<td>Intonation misplacement</td>
<td>.026</td>
</tr>
<tr>
<td>Intonation absence</td>
<td>.164</td>
</tr>
<tr>
<td>Intonation error total</td>
<td>.066</td>
</tr>
</tbody>
</table>

Note. \( p^{*} > .1, p^{**} > .05, p^{***} > .01 \), two-tailed

4.4.2 Six segmental measures. Pearson’s correlation values of six problematic segmental groups were computed (see Table 4). The IELTS rating result showed very strong negative correlation with the major segmental group \( (r = -.473) \), and moderate negative correlation with the minor segmental group \( (r = -.275) \). There was no significance in the correlation between the IELTS rating and the rest of the segmental groups (i.e., assimilation problems, secondary segmentals, diphthong problems, other segmentals).
In summary, these results suggest that pronunciation evaluation via the IELTS pronunciation rubric considers segmental (especially major segmentals and secondary segmentals) and syllable errors as crucial degrading factors. Yet not severely, word stress errors, particularly the absence of the stress, impact on the negative evaluation of the pronunciation scores. On the contrary, intonation errors do not lower the IELTS scores (i.e., raters seem to take a lax approach regarding any type of intonation errors). A summary of the combined results of the correlation of all the measures with the IELTS rating according to the significance is given in Appendix F.

<table>
<thead>
<tr>
<th>Segmental error groups</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Major segmentals</td>
<td>-.473***</td>
</tr>
<tr>
<td>2. Assimilation problems</td>
<td>-.121</td>
</tr>
<tr>
<td>3. Secondary segmentals</td>
<td>-.228</td>
</tr>
<tr>
<td>4. Diphthong problems</td>
<td>-.217</td>
</tr>
<tr>
<td>5. Minor segmentals</td>
<td>-.275*</td>
</tr>
<tr>
<td>6 Other segmentals</td>
<td>-.099</td>
</tr>
</tbody>
</table>

Note. p*>.1, p**>.05, p***>.01, two-tailed

4.5 Level Determinants of the IELTS Pronunciation Rating

The second research question was engaged in identifying pronunciation aspects (segmentals, syllables, word stress, and intonation), which most distinguish proficiency levels of NJ learners of English in the IELTS pronunciation rating. In order to arrive at the answer of this question, eight pronunciation features and six problematic English segmental groups for NJ learners were examined according to the proficiency groups (beginner, intermediate, advanced) determined by the IELTS pronunciation rating. Quantitative approach was employed for the analysis of this research question (i.e., confidence intervals).
By using confidence intervals, the specified ranges of the errors each proficiency groups make at 95% probability were compared.

For the calculation of confidence intervals of 95%, all the rated speeches were sorted into the three proficient groups according to the mean IELTS pronunciation scores they were awarded (i.e., beginner for Bands 4 and 5, Intermediate for Band 6, advanced for Bands 7 and 8). Then, on the basis of eight pronunciation measures and six segmental groups, mean error and standard deviation of the coded error values (i.e., number of segmental, syllables, word stress, intonation errors, and the number of segmental errors in group 1 to 6) of the three proficiency groups were calculated to compute confidence intervals. The calculated means, standard deviations, and confidence intervals of eight pronunciation measures and six problematic segmental groups are presented in Table 5 and Table 6 respectively. Also, to visually illustrate the relationship between three proficiency levels, the error distribution figures are provided in the subsequent sections.

4.5.1 Eight phonological measures. Regarding the impact of segmental errors on scoring distinction, significantly frequent error mean of 4.36 among beginners (n = 14, CI = 1.23, SD = 2.34), less frequent error mean of 1.29 among intermediate group (n = 15, CI = 0.73, SD = 1.44) and 1.67 among advanced group (n = 11, CI = 1.14, SD = 1.94) were obtained. These results suggest the negative impact of segmental errors to distinguish beginner from intermediate and advanced levels, while intermediate and advanced levels were less distinguishable by segmental errors. A clear statistical significance of segmental error effect on distinguishing beginner from the rest of the two proficiency levels could be observed in confidence intervals at 95% probability in Figure 1: The lower limit of beginner group [3.13–5.58] is not overlapped with the upper limit of intermediate [0.56–2.01] and that of advanced level [0.52–2.81]. Therefore, segmental errors best distinguish beginner from intermediate and advanced groups.
### Summary of the Confidence Interval of Eight Pronunciation Measures vs. Three Proficiency Level

<table>
<thead>
<tr>
<th>Pronunciation error measures</th>
<th>Proficiency level</th>
<th>n&lt;sup&gt;a&lt;/sup&gt;</th>
<th>M</th>
<th>SD</th>
<th>CI</th>
<th>LL</th>
<th>UL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Segmental</td>
<td>Beginner</td>
<td>14</td>
<td>4.36</td>
<td>2.34</td>
<td>1.23</td>
<td>3.13</td>
<td>5.58</td>
</tr>
<tr>
<td></td>
<td>Intermediate</td>
<td>15</td>
<td>1.29</td>
<td>1.44</td>
<td>0.73</td>
<td>0.56</td>
<td>2.01</td>
</tr>
<tr>
<td></td>
<td>Advanced</td>
<td>11</td>
<td>1.67</td>
<td>1.94</td>
<td>1.14</td>
<td>0.52</td>
<td>2.81</td>
</tr>
<tr>
<td>Syllable</td>
<td>Beginner</td>
<td>14</td>
<td>3.54</td>
<td>3.38</td>
<td>1.77</td>
<td>1.77</td>
<td>5.31</td>
</tr>
<tr>
<td></td>
<td>Intermediate</td>
<td>15</td>
<td>1.62</td>
<td>1.66</td>
<td>0.84</td>
<td>0.78</td>
<td>2.46</td>
</tr>
<tr>
<td></td>
<td>Advanced</td>
<td>11</td>
<td>0.82</td>
<td>1.25</td>
<td>0.74</td>
<td>0.08</td>
<td>1.56</td>
</tr>
<tr>
<td>Word stress (misplacement)</td>
<td>Beginner</td>
<td>14</td>
<td>0.92</td>
<td>1.19</td>
<td>0.62</td>
<td>0.30</td>
<td>1.55</td>
</tr>
<tr>
<td></td>
<td>Intermediate</td>
<td>15</td>
<td>1.77</td>
<td>2.01</td>
<td>1.02</td>
<td>0.75</td>
<td>2.78</td>
</tr>
<tr>
<td></td>
<td>Advanced</td>
<td>11</td>
<td>0.36</td>
<td>0.67</td>
<td>0.40</td>
<td>-0.03</td>
<td>0.76</td>
</tr>
<tr>
<td>Word stress (absence)</td>
<td>Beginner</td>
<td>14</td>
<td>2.57</td>
<td>2.17</td>
<td>1.14</td>
<td>1.43</td>
<td>3.71</td>
</tr>
<tr>
<td></td>
<td>Intermediate</td>
<td>15</td>
<td>2.15</td>
<td>2.03</td>
<td>1.03</td>
<td>1.12</td>
<td>3.18</td>
</tr>
<tr>
<td></td>
<td>Advanced</td>
<td>11</td>
<td>1.36</td>
<td>2.62</td>
<td>1.55</td>
<td>-0.18</td>
<td>2.91</td>
</tr>
<tr>
<td>Word stress (total)</td>
<td>Beginner</td>
<td>14</td>
<td>3.43</td>
<td>2.06</td>
<td>1.08</td>
<td>2.35</td>
<td>4.51</td>
</tr>
<tr>
<td></td>
<td>Intermediate</td>
<td>15</td>
<td>3.87</td>
<td>2.64</td>
<td>1.34</td>
<td>2.53</td>
<td>5.20</td>
</tr>
<tr>
<td></td>
<td>Advanced</td>
<td>11</td>
<td>1.73</td>
<td>0.67</td>
<td>0.40</td>
<td>1.33</td>
<td>2.13</td>
</tr>
<tr>
<td>Intonation (misplacement)</td>
<td>Beginner</td>
<td>14</td>
<td>0.23</td>
<td>0.60</td>
<td>0.31</td>
<td>-0.08</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td>Intermediate</td>
<td>15</td>
<td>0.46</td>
<td>0.78</td>
<td>0.39</td>
<td>0.07</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>Advanced</td>
<td>11</td>
<td>0.27</td>
<td>0.65</td>
<td>0.38</td>
<td>-0.11</td>
<td>0.65</td>
</tr>
<tr>
<td>Intonation (absence)</td>
<td>Beginner</td>
<td>14</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Intermediate</td>
<td>15</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Advanced</td>
<td>11</td>
<td>0.09</td>
<td>0.30</td>
<td>0.18</td>
<td>-0.09</td>
<td>0.27</td>
</tr>
<tr>
<td>Intonation (total)</td>
<td>Beginner</td>
<td>14</td>
<td>0.23</td>
<td>0.60</td>
<td>0.31</td>
<td>-0.08</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td>Intermediate</td>
<td>15</td>
<td>0.46</td>
<td>0.78</td>
<td>0.39</td>
<td>0.07</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>Advanced</td>
<td>11</td>
<td>0.36</td>
<td>0.67</td>
<td>0.40</td>
<td>-0.03</td>
<td>0.76</td>
</tr>
</tbody>
</table>

*Note.* CI = confidence interval; LL = lower limit; UL = upper limit.

n<sup>a</sup> = number of speech categorised as given proficiency group.

The contrasts of syllable effect on the rating showed distinctive result on the beginner group with 3.54 mean error score (n = 14, CI = 1.77, SD = 3.38), while 1.62 error mean of intermediate groups (n = 15, CI = 0.84, SD = 1.66) and 0.82 error mean of advanced group (n = 11, CI = 0.74, SD = 1.25) are considered as relatively low. The results of 95 %
Confidence intervals support this distinction of beginner level from intermediate and advanced levels by presenting a relatively large range of errors of beginners [1.77–5.31] with slight overlapping with intermediate group [0.78–2.46] and no overlapping with advanced group [0.08–1.56] (see Figure 2). As for intermediate, its confidence intervals overlapped with the lower limit of the beginner group’s and the upper limit of the advanced group’s confidence intervals, suggesting that the syllable error does not clearly distinguish intermediate and advanced but it does between beginners and advanced.

Figure 1. Confidence intervals of segmental errors.

Figure 2. Confidence intervals of syllable errors.
In terms of word stress misplacement, relatively distinct difference can be seen between highest error mean of 1.77 of intermediate \((n = 15, \ CI = 1.02, \ SD = 2.01)\) and 0.36 of advanced levels \((n = 11, \ CI = 0.40, \ SD = 0.67)\). Beginners \((n = 14, \ CI = 0.62, \ SD = 1.19)\) showed in-between error mean of 0.92, which shows no distinction between beginner and intermediate levels. The result of confidence intervals (see Figure 3) yielded nearly significant difference between intermediate level \((0.75–2.78)\) and advanced level \((-0.03–0.76)\) with slight overlapping, while beginner group \((0.30–1.55)\) is overlapping to a large extent with intermediate, suggesting word stress misplacement weakly contributes to the discrimination between intermediate and advanced groups.

![Figure 3. Confidence intervals of word stress misplacement.](image)

With respect to the absence of the word stress, a gradual drop in the number of error mean values was seen throughout the three proficiency levels, with 2.57 for beginner \((n = 14, \ CI = 1.14, \ SD = 2.17)\), 2.15 for intermediate \((n = 15, \ CI = 1.03, \ SD = 2.03)\) and 1.36 for advanced groups \((n = 11, \ CI = 1.55, \ SD = 2.62)\). However, analysis of 95% confidence intervals demonstrated that absence of the word stress has no significance as score determinant between three proficiency levels, where all of the confidence intervals are
overlapping: beginner [1.43–3.71], intermediate [1.12–3.18], and advanced levels [-0.18–2.91], respectively (see Figure 4). In summary, word stress absence does not act as a crucial level determinant factor.

The overall word stress error (i.e., misplacement and absence of the stress in words) revealed the mild contribution of word stress to distinguish the proficiency levels. The calculated error means yielded 3.43 for beginner group \((n = 14, CI = 1.08, SD = 2.06)\), 3.87 for intermediate group \((n = 15, CI = 1.34, SD = 2.64)\), and 1.73 for advanced group \((n = 11, CI = 0.40, SD = 0.67)\), demonstrating clear distinction between intermediate and advanced groups. This result was also supported by 95% confidence intervals illustrated in Figure 5, which show no overlapping between intermediate level [2.53–5.20] and advanced level [1.33–2.13]. Major overlapping is seen between beginner level [2.35–4.51] and intermediate levels [2.53–5.20], which indicate that beginner and intermediate levels are not discriminated by word stress errors. Thus, word stress error in total is regarded as a strong determinant to discriminate advance from intermediate groups but it is not the case between intermediate and beginner groups.
The means of intonation misplacement between the three proficiency groups did not show any significant distinctions: 0.23 for beginner \((n = 14, \text{CI} = 0.31, SD = 0.60)\), 0.46 for intermediate \((n = 15, \text{CI} = 0.39, SD = 0.78)\) and 0.27 advanced \((n = 11, \text{CI} = 0.38, SD = 0.65)\). As is shown in Figure 6, 95% confidence intervals of three proficiency groups are overlapping each other and intermediate group \([-0.07–0.85]\) is higher than that of beginner level \([-0.08–0.54]\) and advanced level \([-0.11–0.65]\). These results suggest that intonation misplacement error does not discriminate the differentiate proficiency levels on the IELTS pronunciation rating.

Regarding the intonation absence errors, the error means were zero with beginner \((n = 14, \text{CI} = 0, SD = 0)\) and intermediate levels \((n = 15, \text{CI} = 0, SD = 0)\), whereas 0.09 of advanced level \((n = 11, \text{CI} = 0.18, SD = 0.30)\) showed slightly higher mean. As is illustrated in Figure 5, 95% confidence intervals show zero range of beginner and intermediate groups whilst advanced group \([-0.09–0.27]\) shows rather wider range and overlap with both beginner and intermediate groups. These results show that even the speeches of advanced group contained more intonation absence than that of beginner and intermediate, their speeches were rated higher. Thus, intonation absence does not impact on the IELTS.
The comparison of the error means of total intonation showed relatively high error mean of 0.46 with intermediate group ($n = 15$, CI = 0.39, $SD = 0.78$) compared to that of 0.23 with beginner ($n = 14$, CI = 0.31, $SD = 0.60$) and 0.36 with advanced ($n = 11$, CI = 0.40, $SD = 0.67$). With 95% confidence intervals (see Figure 8), three proficiency groups’ ranges are largely overlapped with each other: beginner [-0.08–0.54], intermediate [0.07–
0.85], and advanced levels [-0.03–0.76]. This indicates that intonation error has no impact on discriminating three groups.

![Figure 8](image.png)

*Figure 8. Confidence intervals of total intonation errors*

**4.5.2 Six segmental measures.** The results of confidence intervals at 95% level in six segmental groups are illustrated in Table 6. Regarding major segmentals, beginner group showed the highest error means of 2.57 (n = 14, CI = 0.84, SD = 1.60) compared to the other two groups: 1.00 of intermediate group (n = 15, CI = 0.51, SD = 1.00), and 1.27 of advanced group (n = 11, CI = 0.80, SD = 1.35). The comparisons of 95% confidence intervals of three proficiency levels are illustrated in Figure 9. As is shown in Figure 9, the impact of errors of the first segmental groups proved significant in terms of discriminating beginners from intermediate level since there is no overlap between the beginner group [1.73–3.41] and that of intermediate [0.49–1.51]. On the contrary, the errors of the major segmentals do not distinguish intermediate and advanced because their confidence intervals are almost the same: intermediate level [0.49–1.51] and advanced level [0.48–2.07]. In summary, first group segmental is a strong determinant to separate beginner from intermediate and advanced groups.
Table 6

Summary of Confidence Intervals of Six Segmental Measures

<table>
<thead>
<tr>
<th>Segmental measures</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>M</td>
<td>SD</td>
<td>CI</td>
<td>LL</td>
<td>UL</td>
</tr>
<tr>
<td>Major segmentals</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beginner</td>
<td>14</td>
<td>2.57</td>
<td>1.60</td>
<td>0.84</td>
<td>1.73</td>
<td>3.41</td>
</tr>
<tr>
<td>Intermediate</td>
<td>15</td>
<td>1.00</td>
<td>1.00</td>
<td>0.51</td>
<td>0.49</td>
<td>1.51</td>
</tr>
<tr>
<td>Advanced</td>
<td>11</td>
<td>1.27</td>
<td>1.35</td>
<td>0.80</td>
<td>0.48</td>
<td>2.07</td>
</tr>
<tr>
<td>Assimilation problem</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beginner</td>
<td>14</td>
<td>0.21</td>
<td>0.43</td>
<td>0.22</td>
<td>-0.01</td>
<td>0.44</td>
</tr>
<tr>
<td>Intermediate</td>
<td>15</td>
<td>0.20</td>
<td>0.41</td>
<td>0.21</td>
<td>-0.01</td>
<td>0.41</td>
</tr>
<tr>
<td>Advanced</td>
<td>11</td>
<td>0.09</td>
<td>0.30</td>
<td>0.18</td>
<td>-0.09</td>
<td>0.27</td>
</tr>
<tr>
<td>Secondary segmentals</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beginner</td>
<td>14</td>
<td>0.14</td>
<td>0.53</td>
<td>0.28</td>
<td>-0.14</td>
<td>0.42</td>
</tr>
<tr>
<td>Intermediate</td>
<td>15</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Advanced</td>
<td>11</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Diphthong problems</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beginner</td>
<td>14</td>
<td>0.21</td>
<td>0.43</td>
<td>0.22</td>
<td>-0.01</td>
<td>0.44</td>
</tr>
<tr>
<td>Intermediate</td>
<td>15</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Advanced</td>
<td>11</td>
<td>0.09</td>
<td>0.30</td>
<td>0.18</td>
<td>-0.09</td>
<td>0.27</td>
</tr>
<tr>
<td>Minor segmentals</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beginner</td>
<td>14</td>
<td>0.57</td>
<td>1.13</td>
<td>0.59</td>
<td>-0.02</td>
<td>1.16</td>
</tr>
<tr>
<td>Intermediate</td>
<td>15</td>
<td>0.47</td>
<td>1.13</td>
<td>0.57</td>
<td>-0.10</td>
<td>1.04</td>
</tr>
<tr>
<td>Advanced</td>
<td>11</td>
<td>0.09</td>
<td>0.30</td>
<td>0.18</td>
<td>-0.09</td>
<td>0.27</td>
</tr>
<tr>
<td>Other segmentals</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beginner</td>
<td>14</td>
<td>0.43</td>
<td>0.94</td>
<td>0.49</td>
<td>-0.06</td>
<td>0.92</td>
</tr>
<tr>
<td>Intermediate</td>
<td>15</td>
<td>0.27</td>
<td>0.59</td>
<td>0.30</td>
<td>-0.03</td>
<td>0.57</td>
</tr>
<tr>
<td>Advanced</td>
<td>11</td>
<td>0.09</td>
<td>0.30</td>
<td>0.18</td>
<td>-0.09</td>
<td>0.27</td>
</tr>
</tbody>
</table>

Note. CI = confidence interval; LL = lower limit; UL = upper limit.

1036 $n^a$ = number of speech categorised as given proficiency group.

Figure 9. Confidence intervals of errors in major segmentals
With respect to the group comparison of assimilation problems, the error means of three proficiency groups revealed no significant difference between beginner and intermediate groups: 0.21 for beginners ($n = 14$, CI = 0.22, $SD = 0.43$), and 0.20 for intermediate group ($n = 15$, CI = 0.21, $SD = 0.41$), whereas slight difference is observed between 0.20 of intermediate and 0.09 of advanced groups ($n = 11$, CI = 0.18, $SD = 0.30$).

Similarly, 95% confidence intervals yielded no significance between beginner and intermediate due to their major overlap: beginner level [-0.01–0.44], intermediate level [-0.01–0.41], while not major but some overlap is observed between intermediate level [-0.01–0.41] and advanced level [-0.09–0.27]. Thus, these results indicate that the assimilation problem group was not crucial factor to differentiate three groups.

Figure 10. Confidence intervals of errors in assimilation problem

Considering the secondary segmental group, the error means of intermediate and advanced groups showed zero values compared to that of 0.14 for beginner ($n = 14$, CI = 0.28, $SD = 0.53$). This indicates that beginner group is distinguishable from intermediate and advanced groups. As is illustrated in Figure 11, 95% confidence intervals yield the complete overlap between intermediate and advanced, while beginner [-0.14–0.42] is overlapping with
intermediate and advanced group is nearly significant. Overall, the secondary segmental group errors mildly impacted on the differentiation between beginner and intermediate groups but not between intermediate and advanced groups.

Figure 11. Confidence intervals of secondary segmentals

Diphthong problems showed nearly significant relationship between beginner and intermediate group comparison. As is illustrated in Table 6, the error mean of 0.21 for beginner group \((n = 14, CI = 0.22, SD = 0.43)\) is higher than the zero value of intermediate group \((n = 15, CI = 0, SD = 0)\). The error mean of 0.09 for advanced group \((n = 11, CI = 0.18, SD = 0.30)\) is slightly higher than intermediate group but lower than beginner group. Confidence intervals of 95\% (see Figure 12) revealed a partial overlapping between beginner \([-0.01–0.44]\) and intermediate groups \([0]\), which makes the distinction between beginner and intermediate nearly significant. However, advanced level \([-0.09–0.27]\) and beginner level have a major overlap. Overall, these results suggest that diphthong problems might impact on discriminating beginner from intermediate groups but not on separation of intermediate from advanced groups.
Figure 12. Confidence intervals of errors in diphthong problems

Regarding the minor segmental group, the error means were 0.57 for beginner level \( (n = 14, \text{CI} = 0.59, SD = 1.13) \), 0.47 for intermediate level \( (n = 15, \text{CI} = 0.57, SD = 1.13) \), and 0.09 for advanced level \( (n = 11, \text{CI} = 0.18, SD = 0.30) \). These results show relatively lower error mean of advanced group compared to that of beginner and intermediate groups. As is illustrated in Figure 13, 95% confidence intervals revealed that no significance in the comparison between beginner group \([-0.02–1.16]\) and intermediate group \([-0.10–1.04]\) due to their major overlapping. However, the comparison of confidence intervals between intermediate and advanced groups \([-0.09–0.27]\) almost reaches significance with slight overlapping. Therefore, although it is not significant, the error of minor segmental group appears to mildly contribute to the discrimination between intermediate and advanced group. Lastly, other segmentals showed gradual decline of the error mean from beginner to advanced groups: 0.43 for beginner \( (n = 14, \text{CI} = 0.49, SD = 0.94) \), 0.27 for intermediate \( (n = 15, \text{CI} = 0.30, SD = 0.59) \), and 0.09 for advanced levels \( (n = 11, \text{CI} = 0.18, SD = 0.30) \). Confidence intervals of 95% of three proficiency groups are illustrated in Figure 14 and show essential overlapping: beginner \([-0.06–0.92]\), intermediate \([-0.03–0.57]\), advanced levels \([-0.09–0.27]\) respectively. Thus, no significance is observed in the impact of other
segmental for level distinction.

Figure 13. Confidence intervals of errors in minor segmentals

Figure 14. Confidence intervals of other segmentals

Taken together, beginner group is differentiated clearly by major segmentals and
weakly by secondary segmentals and diphthongs while intermediate group is mildly
discriminated from advanced group by minor segmentals. Nonetheless, other segmental
groups are not strong determinant of separating three proficiency levels.
4.6 Correlation between the IELTS pronunciation Rating and Comprehensibility

Judgement

The third research question was set to identify the relationship between IELTS pronunciation rating and that of comprehensibility rating. This question was answered through a quantitative method, i.e., a statistical calculation. In order to see the correlation between IELTS pronunciation rating result and that of comprehensibility, the Pearson’s correlation was used to examine the strength of the relationship between two variables (-1 ≤ r ≥ 1). As a first step, mean IELTS score and mean comprehensibility score of each audio were calculated. Then, the averaged rating results of each of the 40 recordings were inserted to run person correlation coefficient. The correlation between the IELTS rating and comprehensibility rating is yielded 0.958, suggesting significant positive correlation between two rating patterns.

Figure 15. Result of Pearson’s correlation between the IELTS pronunciation rating and comprehensibility rating

According to Evans and Over (1996), the value falls between 0.80 and 1.0, the correlation between measured variable is regarded as very strong. Figure 15 illustrates
positive correlation between the results of IELTS Pronunciation rating and
Comprehensibility rating, confirming the strong link between two rating results.
Chapter 5: Discussion

5.1 Introduction

This chapter presents an extensive analysis of the crucial findings obtained through the answers of the research questions. The analysis involves the discussions of the findings in relation to the past relevant studies, which are followed by the brief re-statement of research questions with the answers. The result of the first research question is discussed in the first section with regard to the universal and L1 specific key features of IELTS pronunciation rating as well as the uniqueness of IELTS pronunciation assessment from a general L2 pronunciation syllabus. The second section emphasizes on the impact of L1 on pronunciation error patterns, which result in the difference of pronunciation features to be attended to by raters. The third section argues the importance of the further advancement in the research in IELTS rating and pronunciation pedagogy.

5.2 First Research Question

The first research question considered the most crucial English pronunciation features among segmentals and suprasegmentals (i.e., segmentals, syllable, word stress and intonation) for Japanese IELTS candidates in the IELTS pronunciation assessment. This question relates to the prioritisation within English pronunciation teaching and the relationship between IELTS pronunciation measures and the two contrastive L2 speech principles (i.e., comprehensibility and accentedness) regarding a specific L1 group (i.e., Japanese). A Pearson’s correlation was computed to see the strength of the correlation between errors of eight pronunciation measures and the IELTS pronunciation rating result of
40 NJ speeches. The correlation analysis indicated that segmental errors severely affected
the rating compared to that of suprasegmental errors. In fact, segmental errors and syllable
errors showed the strongest/significant negative correlation with pronunciation rating.

However, except for the mildly negative correlation between absence of word stress and
rating result, the negative correlation of word stress misplacement, overall word stress errors,
and all the intonation error variables (misplacement, absence and overall) on the rating were
less significant. Regarding Pearson’s correlation between the IELTS rating result and coded
errors of six problematic segmental groups, a significantly strong negative correlation was
found in errors of major segmentals. Furthermore, errors in minor segmentals showed a
mildly strong correlation. The correlation values of other segmentals were not strong enough
to reach significance. Taken together, the IELTS pronunciation rating was greatly affected
by major segmental and syllable errors and mildly affected by word stress errors especially
the absence of the expected stresses rather than misplacement. Other pronunciation features
including assimilation problems, secondary segmentals, diphthong problems are regarded as
minor factors but they still gave a negative effect on the rating. On the contrary, all the
intonation errors showed positive correlations, indicating that Intonation errors were not
considered as crucial factors in pronunciation ratings. Therefore, the following pedagogical
suggestion for IELTS pronunciation teaching and learning can be made:

1. Firstly, teachers and learners need to focus on improving the production of major
   segmentals (/ɹ, l, v, θ, ð/) as well as syllabification to make sure not to insert
   unnecessary sounds to correct articulations or drop the sounds supposed to be
   articulated.
2. Secondly, teachers and learners should concentrate on the practicing of minor segmentals (/w, η, h, n, p, t, k/, Contractions) with particular emphasis on the placing of word stress when produce words and sentences.

3. As a final phase, secondary segmentals (/f, æ, ʌ/), diphthong problems (/aʊ, aɪ, ʊ, ɔɪ/), assimilation problems (/fɪ, sɪ, tɪ/) and other segmentals need to be introduced and practiced to further improve comprehensible articulation of English. Then the misplacements of word stress and the intonation errors (absence and misplacement) should be corrected.

The high correlation of segmental and syllable errors with IELTS rating in the current study showed a large disagreement with the findings in Isaacs et al. (2015), which lexicogrammatical factors showed highest correlation with the IELTS pronunciation rating. The cause of this difference can be due to two factors: (a) range of the linguistic measurements, (b) L1 effect. For the first factor, Isaacs et al. (2015) used a much wider range of measurement than the measurements in the current study. In fact, while the Isaacs et al. (2015) included grammatical accuracy, lexical richness speech chunking, and speech rate, the measurements employed for this study were limited to pronunciation features (i.e., segmentals, syllables, word stress, and intonation). Therefore, the lack of lexicogrammatical aspect in the current study’s measurements might have caused the high concentration of the rater’s attention to the each of the pronunciation features such as segmental errors. However, a second factor to account for an anomaly in result is also possible as this study controlled the L1 background of the speech sample while this was not the case in Isaacs et al. (2015). Based on the evidence of the effect of learner’s L1 on the L2 production being proven by the growing amount of literature in the area of L2 pronunciation development and
comprehensibility (Tutor et al., 2015b), the result obtained in this research may have reflected the L2 pronunciation errors unique to Japanese learners of English, whereas the result of Isaacs et al. (2015), which contain the speech samples of multiple L1 backgrounds, seems to be regarded as more universal features of L2 pronunciation errors among learners from various L1 backgrounds.

Another comparison of the result can be made with Tutor’s (2014) propriety ranking of pronunciation features to predict the difference in the crucial features in IELTS pronunciation instruction and general English pronunciation instruction. This is because the ranking list in Tutor’s (2014) was based on the NE and NJ teachers’ perception of importance in their general English lessons to Japanese learners but not the IELTS test-taking classes. The findings in this study were partially consistent with the ranking presented in Tutor (2014) regarding the highest importance of major segmentals and syllabification problems (and cognates). This indicates the raters in this study and the teachers who participated in Tutor’s (2014) study both perceived least comfort in understanding Japanese learners’ speeches filled with major segmental and syllable errors. One of the possible accounts for the result of the weightiness of segmental and syllable errors is the difference in the phonetic systems of Japanese and English. For major segmentals (l, 1, v, 0, ɹ), four (/v, ɹ, l, ɹ/) out of five belong to English specific segmentals (/æ, f, v, q, ɹ, w, l, ɹ/). Therefore, these sounds would be the most difficult sounds for Japanese speakers to articulate accurately. In terms of syllable production, consonant clusters are found to be particularly difficult for Japanese learners. Due to the mora-timed nature of Japanese language, which place equal stress on each syllable (Ohata, 2004), Japanese speakers tend to insert unnecessary vowels after consonants in English words. In addition, due to the existence of numerous English loanwords, which are articulated in the Japanese sound system, Japanese learners are likely to conflate these loanwords with correct English pronunciations (see
Riney & Anderson-Hsieh, 1993), indicating that use of loan words created incorrect syllabification in the audio samples examined in this study. The relatively crucial importance of major segmentals found in this study is also supported by the functional load theory (Brown, 1991; Catford, 1987; Kang & Moran, 2014; Munro & Derwing, 2006). Munro and Derwing (2006) revealed that segmentals of high functional load negatively and greatly impact on both L2 comprehensibility and accentedness. Since major segmentals include two high functional load segmentals (/ɪ, ɫ/), the errors appeared to be fatal to L2 speech perception and might have induced the raters’ negative judgments.

On the contrary to the partial consistency with Tutor’s (2014) study discussed above, the results of the relative importance of word stress, minor segmental and intonation for IELTS pronunciation rating did not fully support the priority ranking of Tutor (2014) based on teachers’ perception. Word stress error was more crucial than assimilation problems in IELTS pronunciation rating whereas assimilation is regarded as more vital than word stress in Tutor’s (2014) guideline. The relatively heavy weight absence of word stress on negative judgments might be derived from the raters’ salient perception of monotonous sounds in the speech samples. Japanese learners’ word articulations are due to the differences in the pronunciation system between English and Japanese. Compared to the vowel-focused English stress realization (i.e., longer and louder vowel of stressed syllable), stress is realized through higher pitch syllables (e.g., Gimson, 1989; Tsujimura, 2013; Vance, 1987). Minor segmentals (/w, η, h, n, p, t, k /, Contraction) lead to lower rating consequence than secondary segmentals, diphthongs, and assimilation sounds in this study compared to their worst rank among the segmental groups in Tutor’s (2014) implication. This result might be caused by the difference in perception between seven raters in this study and the NE/NJ teachers in Tutor’s (2014) study: seven raters were more sensitive to the fricatives (/h/), plosives (/p, t, k/), nasals (/η, n/) and approximal (/w/) errors. Intonation features appeared to
be the lowest priority in the current study while they need to be taught before secondary segmentals, diphthongs and minor segmentals are covered in Tutor’s (2014) implications.

Seven native judges in the current study might be tolerant towards any types of intonation errors in their rating due to their own flexible use of intonations. As Levis (1999) pointed out in his investigation of NE natural speech sample in a corpus, native speakers seem not to always follow the intonation rules described in textbooks such as raising tones for “yes/no” questions. Whereas, NE/NJ teachers in Tutor’s (2014) study might have strictly followed the intonation rules set in teaching materials and thus made intonation outrank secondary segmentals, diphthongs and minor segmentals. Overall, the results of the current study demonstrated different priority patterns from that of Tutor’s (2014), suggesting that the raters perception based on IELTS pronunciation rubric and NE/NJ teachers’ perceived pronunciation issues of native Japanese learners of English are different. Thus, when it comes to pronunciation instruction for IELTS, different approaches need to be taken it is suggested for teachers to adopt the priorities developed in the current study.

5.3 Second Research Question

The second research question examined English pronunciation aspects that most discriminate the proficiency levels of J learners in the IELTS pronunciation rating. This question is concerned with the salience of particular errors among three proficiency groups (beginner, intermediate, and advanced learners). Confidence interval of 95 % was employed for arriving at the answer to this research question. Forty rated speech samples were assigned to each proficiency group (beginner, intermediate, and advanced) according to the mean scores obtained through the averaging of the seven raters’ rating results. Then, errors of the speeches in each group were calculated to yield confidence intervals, which are the ranges of
the errors each proficiency group makes at 95% probability, were specified. Subsequently, these values were compared according to eight pronunciation features and six problematic segmental groups. The results of the group comparison regarding eight pronunciation features revealed that beginners were most distinguished by the segmental and syllable errors while other features did not show any significance in separating beginner from intermediate and advanced levels. Concerning intermediate learners, although individual word stress errors (i.e., absence and misplacement) were not significant enough to note, total errors of word stress impacted on discriminating them from advanced learners. In terms of the between-group comparison of six segmental groups, beginners were clearly discriminated by major segmental errors, and yet not significant, mildly separated by secondary segmental, and diphthong problem errors. In addition, however, it was not significant enough, minor segmentals weakly discriminated advanced from intermediate groups. Thus, the following IELTS pronunciation guideline for beginners and intermediate test-takers and teachers can be made:

1. In order to improve beginners to intermediate/advanced levels, major segmentals (/ɹ, l, v, θ, ð/) and syllables are the first features that need to be focused on among all the segmental groups. Then, secondary segmentals, and diphthongs should be instructed.

2. For the intermediate learners, word stress errors (misplacement, absence of the stress) need to be the propriety for the instruction. Subsequently, learners should work on the accurate production of minor segmentals.

Another noteworthy finding is, in terms of pronunciation features, the results were
not fully consistent with the prior research of Isaacs et al. (2015) on IELTS pronunciation rating scales in separations of beginner, intermediate, and advanced. In their study, eight accredited IELTS examiners rated 80 candidates’ speeches of various L1 backgrounds offered by Cambridge English. The ratings were conducted based on comprehensibility, segmental, prosodic, fluency, and lexicogrammatical measures and official examiner’s version of the IELTS speaking rubric (confidential). Isaacs et al. (2015) found that Bands 5 and 6 (beginner and intermediate in the current study) were best discriminated by speech rate and lexical richness, and the combined Bands 7 and 8 (equivalent to advanced in this current study) were differentiated from the combined Bands 5 and 6 by all the eight speech measures (i.e., comprehensibility, vowel and consonant errors, word stress, intonation, speech chunking, speech rate, lexical richness, and grammatical accuracy and sentence structure). Concerning the word stress errors, the current study and Isaacs et al. (2015) both found that word stress was responsible for distinguishing intermediate from advanced. Vowels and consonants (equivalent to segmentals and syllables), however, were major determinants for distinguishing intermediate from advanced groups (i.e., Band 6 from 7), while the present study revealed that they chiefly impacted on discriminating beginners from intermediate groups except for minor segmentals’ contribution to the separation between intermediate and advanced groups. Furthermore, intonation errors bore no relation to the score discrimination in the current study albeit they appeared to be degrading factors for advanced group in Isaacs et al. (2015).

The speculations on the little agreement of the findings with Isaacs et al.’s (2015) prior study described above involve two possible accounts. Firstly, the linguistic variables employed in the two studies were different. While their study considered the variables based on all the IELTS speaking criteria (fluency, lexis, grammar and pronunciation), the current study asked raters to only focus on pronunciation variables. The limited scope of the current
study may have lead raters’ attention to the accuracy of sound articulations and resulted in finding the substantial contribution of segmentals for rating, whereas the raters of Isaacs et al. (2015), who reached judgments by considering various features of the speech and resulted in finding low importance of segmentals for the score judgement. Secondly, it is possible that the result of the pronunciation rating is attributed to the nature of L1 (i.e., Japanese). As a body of literature of L2 pronunciation acquisition has been proved that the L1 has a great influence on the quality of L2 production and one’s comprehensibility (e.g., Tutor et al., 2015a; Tutor et al., 2015b). Thus, building on Isaacs et al.’s (2015) results, accuracy in word stress is likely to be a universally essential feature for advanced learners of various L1 backgrounds (except for Farci speakers), while segmental features are more important for Japanese learners and Chinese speakers.

5.4 Third Research Question

The last research question dealt with the benefit of the IELTS pronunciation skill for attaining successful L2 communication. Since this study is in line with the researchers’ belief that comprehensibility is the key facilitator of communication, comprehensibility rating was used to measure the indicator of success in L2 communication. Thus, this question investigated the correlation between IELTS pronunciation rating and the comprehensibility rating. The Pearson’s correlation showed an extremely high correlation between IELTS pronunciation and comprehensibility rating, indicating that IELTS pronunciation skill is beneficial for L2 learners to improve their comprehensibility, which leads to success in L2 communication. This result does not support the findings of the prior study directed by Isaacs et al. (2015), which showed a weak correlation between IELTS rating and comprehensibility compared to other linguistics measures they used. The cause of
the difference in the degree of correlation between the two types of judgement can be
speculated in twofold: rater factor, and measurement factor. Firstly, the raters of the present
study are different from that of Isaacs et al. (2015) in the way they were trained, which might
cause a difference in judgement patterns. Isaacs et al.’s (2015) study concerned accredited
IELTS examiners’ judgement which reflected the criteria of the assessment rubric offered
only to examiners, while the current study examined the judgement results produced by
native speakers of English who are not official IELTS examiners and trained though the
publicly available IELTS rubric, which is not as same as the one IELTS examiners consult.
Secondly, as was discussed in the earlier section, the linguistic measurements used in Isaacs
et al. (2015) and the present study are different. While Isaacs et al. (2015) employed
linguistic measures spanning from pronunciation to lexicogrammar, the present study only
used pronunciation measures. These differences in the measurement range might have
affected the raters’ degree of attentiveness to each linguistic measures and lead to differences
in the relationship between IELTS pronunciation rating and comprehensibility.

Despite the difference in the rubric and rating condition (using of only the
phonological aspects), at least this study proved that measuring the construct of public
IELTS pronunciation rubric is highly likely interpreted as a similar construct as
comprehensibility by the judges in this study. This means that pronunciation instruction
based on IELTS pronunciation rubric would help test-takers to improve their
comprehensibility (i.e., skill for L2 communicative success). However, further research is
certainly required with professional IELTS examiners to truly re-examine the relationship
between degree of comprehensibility and IELTS pronunciation scoring.
Chapter 6: Conclusion

6.1 Research Findings and Implications

The current study was designed to examine segmental and suprasegmental correlates and level determinants of the IELTS pronunciation rating. In addition, the study explored the proximity of the rating to comprehensibility judgements to see the practicality of the pronunciation skill for the IELTS speaking test.

The findings suggested that pronunciation errors, which were highly specific to Native speakers of Japanese (NJs), were linked to the raters’ judgements on the IELTS pronunciation scale. A negative correlation with the ratings was saliently observed in the production of English-specific segmentals (e.g., /ɹ, l, v, ð/) and syllabification by inserting vowels after consonants, indicating that errors derived from first L1 influence (i.e., absence of certain English sounds in Japanese phonetics, and over-application of Japanese articulation of consonant and vowel combinations to English consonant clusters). These features also distinguished beginner level from intermediate and advanced level of speakers.

Likewise, score-affecting errors caused by the difference in stress system between English and Japanese (see Tsujimura, 2013) could also be seen in the word stress errors (misplacement and absence) which showed a moderately negative correlation with pronunciation rating. In effect, failing to place stress was a common error among all proficiency levels, whilst errors in displacement of word stress were especially less frequent among advanced speakers.

The results of the current study confirmed that learner’s L1 background impacts on the types of errors and rating results. This finding would inform L2 pronunciation instructors about the importance of being aware of the L1 specific difficulties learners face not only for
improving one’s L2 comprehensibly, but also for improving IELTS pronunciation scores. In the case of the pronunciation syllabi developed in this study, they are certainly beneficial for Japanese IELTS test-takers.

6.2 Research Limitations

When considering the methodological respect of the current study, several limitations need to be addressed for future research. First, the scope to capture the pronunciation errors of the sample speech was limited to pronunciation features in the current study, while the study by Isaacs et al.’s (2015), which the current study builds on, has a much wider scope ranging from lexicogrammatical to pronunciation measures. This narrow scope prevented the study from fully capturing the complex relationship between test-takers’ errors and IELTS pronunciation teaching, especially the possible influence caused by lexicogrammatical errors. Due to the relatively short length of each speech sample, this study was not able to collect sufficient amount of intonation errors to analyse its impact on the rating. Therefore, intonation errors need to be examined with much longer length of speech samples to adequately capture the all the types of intonation errors for analysis.

Secondly, unlike Isaacs et al. (2015) and Tutor et al. (2015a) which have different L1 backgrounds in their speech samples, the current study only focused on a specific L1 background group. While the concentration on one specific language background enabled the study to generate L1 specific pronunciation syllabi, it limited the investigation into universally problematic segmentals and suprasegmentals, which can be measured though the comparisons of the result between different L1 background groups.

Thirdly, the current study did not employ a qualitative approach to capture raters’ thinking processes partially due to its sheer focus on the rating results, and partially due to
the limited time allocation of the research. Thus, employing think-aloud protocol or a post-rating interview may have been a more accurate method in which to supplement the account of the results of rater judgements. Such qualitative approach is commonly employed in related studies as a *mix-method approach* (Creswell & Clark, 2007) where qualitative data is used to complement quantitative data (Isaacs et al., 2015).

Lastly, the native judges hired in present study were not accredited IELTS judges but trained by the researcher based on publicly available IELTS rubric. Thus, admittedly, the rating results obtained in the present study would not be exactly the same as the results as professional IELTS examiners produced.
Reference List


Kang, O., Rubin, D., & Pickering, L. (2010). Suprasegmental measures of accentedness and


Appendix A

Publicly available IELTS pronunciation rubric (British Council, n.d.-a)
# SPEAKING: Band Descriptors (public version)

<table>
<thead>
<tr>
<th>Band</th>
<th>Fluency and coherence</th>
<th>Lexical resource</th>
<th>Grammatical range and accuracy</th>
<th>Pronunciation</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>speaks fluently with only rare repetition or self-correction; any hesitation is content-related rather than to find words or grammar; speaks coherently with fully appropriate cohesive features; develops topics fully and appropriately.</td>
<td>uses vocabulary with full flexibility and precision in all topics; uses idiomatic language naturally and accurately.</td>
<td>uses a full range of structures naturally and appropriately; produces consistently accurate structures apart from slight characteristic of native speaker speech.</td>
<td>uses a full range of pronunciation features with precision and subtlety; sustains flexible use of features throughout; is effortless to understand.</td>
</tr>
<tr>
<td>8</td>
<td>speaks fluently with only occasional repetition or self-correction; hesitation is usually content-related and only rarely to search for language; develops topics coherently and appropriately.</td>
<td>uses a wide vocabulary resource readily and flexibility to convey precise meaning; uses less common and idiomatic vocabulary skillfully, with occasional inaccuracies; uses paraphrase effectively as required.</td>
<td>uses a wide range of structures flexibly; produces a majority of error-free sentences with only very occasionally inappropriates or baden-sounding systematic errors.</td>
<td>uses a wide range of pronunciation features; sustains flexible use of features, with occasional lapses.</td>
</tr>
<tr>
<td>7</td>
<td>speaks at length without noticeable effort or loss of coherence; may demonstrate language-related hesitation at times, or some repetition and/or self-correction; uses a range of connectives and discourse markers with some flexibility.</td>
<td>uses vocabulary resource flexibly to discuss a variety of topics; uses some less common and idiomatic vocabulary and demonstrates some awareness of style and collocation, with some inappropriate choices; uses paraphrase effectively.</td>
<td>uses a range of complex structures with some flexibility; frequently produces error-free sentences, though some grammatical mistakes persist.</td>
<td>shows all the positive features of Band 6 and some, but not all, of the positive features of Band 5.</td>
</tr>
<tr>
<td>6</td>
<td>is willing to speak at length, though may lose coherence at times due to occasional repetition, self-correction or hesitation; uses a range of connectives and discourse markers but not always appropriately.</td>
<td>has a wide enough vocabulary to discuss topics at length and make meaning clear in spite of inappropriates generally paraphrases successfully.</td>
<td>uses a mix of simple and complex structures, but with limited flexibility; may make frequent mistakes with complex structures though these rarely cause comprehension problems.</td>
<td>uses a range of pronunciation features with mixed control; shows some effective use of features but this is not sustained; can generally be understood throughout, though mispronunciation of individual words or sounds reduces clarity at times.</td>
</tr>
<tr>
<td>5</td>
<td>usually maintains flow of speech but uses repetition, self-correction and/or slow speech to keep going; may over-use certain connectives and discourse markers; produces simple speech fluently, but more complex communication causes fluency problems.</td>
<td>manages to talk about familiar and unfamiliar topics but with repetition and/or slow speech to keep going; attempts to use paraphrase but with mixed success.</td>
<td>produces basic sentence forms with reasonable accuracy; uses a limited range of more complex structures, but these usually contain errors and may cause some comprehension problems.</td>
<td>shows all the positive features of Band 4 and some, but not all, of the positive features of Band 6.</td>
</tr>
<tr>
<td>4</td>
<td>cannot respond without noticeable pauses and may speak slowly, with frequent repetition and self-correction; limited basic sentences but with repetitive use of simple connectives and some breakdowns in coherence.</td>
<td>is able to talk about familiar topics but can only convey basic meaning on unfamiliar topics and makes frequent errors in word choice; rarely attempts paraphrase.</td>
<td>produces basic sentence forms and some correct simple sentences but subordinate structures are rare; errors are frequent and may lead to misunderstanding.</td>
<td>uses a limited range of pronunciation features; attempts to control features but apes are frequent; mispronunciations are frequent and cause some difficulty for the listener.</td>
</tr>
<tr>
<td>3</td>
<td>speaks with long pauses; has limited ability to link simple sentences; gives only simple responses and is frequently unable to convey basic message.</td>
<td>uses simple vocabulary to convey personal information; has insufficient vocabulary for less familiar topics.</td>
<td>attempts basic sentence forms but with limited success, or relies on apparent meaning of sentences; makes numerous errors even in memorised expressions.</td>
<td>shows some of the features of Band 2 and some, but not all, of the positive features of Band 4.</td>
</tr>
<tr>
<td>2</td>
<td>pauses lengthily before most words; little communication possible.</td>
<td>only produces isolated words or memorised utterances.</td>
<td>cannot produce basic sentence forms.</td>
<td>Speech is often unintelligible.</td>
</tr>
<tr>
<td>1</td>
<td>no communication possible.</td>
<td>no usable language.</td>
<td>uses a limited range of pronunciation features; attempts to control features but apes are frequent; mispronunciations are frequent and cause some difficulty for the listener.</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>does not attend.</td>
<td>no usable language.</td>
<td>uses a limited range of pronunciation features; attempts to control features but apes are frequent; mispronunciations are frequent and cause some difficulty for the listener.</td>
<td></td>
</tr>
</tbody>
</table>
Appendix B

Language background questionnaire for raters (adopted from Tutor et al., 2016a)

Name ____________________

Age ____________

Language Background

• Your birthplace ______________

• Second language proficiency (e.g., beginner, intermediate, advanced, French, Spanish)
  1. ____________
  2. ____________
  3. ____________

• Your parents’ first language background
  (Father: ____________, Mother: ____________)

• Language spoken at home (generally)
  (English ____%: _____, ____ %)

• Have you ever had any hearing problems? (yes, no)

• Have you ever taken any linguistics classes (especially phonetics/phonology)? If yes, what kinds of classes?
  ____________, ____________, ____________, ____________, ____________,
  ____________, ____________, ____________, ____________, ____________,

• How long and where have you taught English?
  ____________, ____________, ____________, ____________, ____________,
  ____________, ____________, ____________, ____________, ____________,

• Have you ever visited Japan (yes, no)

• If yes, how long did you stay? ________________

• Have you ever taken any Japanese course (yes, no)

• If yes, how long did you study Japanese? ________________

How much are you familiar with Japanese-accented English?

( 1 2 3 4 5 6 )
Not at all Neutral Very much

How often do you have contact with native Japanese speakers?

( 1 2 3 4 5 6 )
Very infrequent Infrequent Neutral Often Very often
Appendix C

A prompt sheet used for speech elicitation for 40 speech samples in Tutor’s (2011) unpublished corpus

Describe the hardest and toughest challenge in your life.

Discussion points
- When? How old and where were you?
- Why did you encounter this challenge?
- Why was it so challenging?
- Did anybody (e.g., friends, parents) help you?
## Appendix D

Ranking list of problematic segmental groups for NJs adopted from Tutor (2014)

<table>
<thead>
<tr>
<th>Segmental Measures</th>
<th>Segmentals</th>
<th>Priorities</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Major segmentals</td>
<td>/s/, /l/, /ʃ/, /θ/, /v/, /ð/</td>
<td>↑</td>
</tr>
<tr>
<td>2. Assimilation problems</td>
<td>/ti/, /si/, /ti/</td>
<td>More important</td>
</tr>
<tr>
<td>3. Secondary segmentals</td>
<td>/ɨ/, /æ/, /ʌ/</td>
<td>↓</td>
</tr>
<tr>
<td>4. Diphthong problems</td>
<td>/au/, /au/, /ou/, /ɔu/, /ʌu/</td>
<td>Less important</td>
</tr>
</tbody>
</table>
| 5. Minor segmentals      | /w/, Contraction (e.g., won’t, can’t), /ŋ/,
|                          | /h/, /n/, /p/, /t/, /k/        |            |
| 6. Other segmentals      | Other segmentals                |            |

*Note. The measures are placed in order of importance (i.e., 1 is most and 9 is least important).*
Appendix E

A screenshot of software Praat for rating

Comprehensibility (1: hard to understand-9: easy to understand)

IELTS Pronunciation (1: poor pronunciation-9: good pronunciation)
Summary of Pearson’s correlation between all the measures and the IELTS pronunciation rating

<table>
<thead>
<tr>
<th>Phonological and segmental measures</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Segmentals</td>
<td>-.590***</td>
</tr>
<tr>
<td>Major segmentals</td>
<td>-.473***</td>
</tr>
<tr>
<td>Syllables</td>
<td>-.472***</td>
</tr>
<tr>
<td>Word stress total</td>
<td>-.304*</td>
</tr>
<tr>
<td>Word stress absence</td>
<td>-.294*</td>
</tr>
<tr>
<td>Minor segmentals</td>
<td>-.275*</td>
</tr>
<tr>
<td>Secondary segmentals</td>
<td>-.228</td>
</tr>
<tr>
<td>Diphthong problems</td>
<td>-.217</td>
</tr>
<tr>
<td>Assimilation problems</td>
<td>-.121</td>
</tr>
<tr>
<td>Word stress misplacement</td>
<td>-.101</td>
</tr>
<tr>
<td>Other segmentals</td>
<td>-.099</td>
</tr>
<tr>
<td>Intonation absence</td>
<td>.164</td>
</tr>
<tr>
<td>Intonation total</td>
<td>.066</td>
</tr>
<tr>
<td>Intonation misplacement</td>
<td>.026</td>
</tr>
</tbody>
</table>

Note. $p^* < .1$, $p^{**} < .05$, $p^{***} < .01$, two-tailed