Accentedness, comprehensibility and intelligibility of L2 speech:
A replication and extended study

Catherine Pease

MA by Research

University of York

Language and Linguistic Science

December 2016

Abstract

This thesis examines the relationship between the accentedness, comprehensibility and intelligibility of Second Language (L2) speech, and is based on the replication of an empirical study carried out in the 1990s by Munro and Derwing (1995a). Following the methodology of the original study, ratings were collected from native English listeners for the accentedness, comprehensibility, and intelligibility of spontaneous L2 English utterances of mother tongue Arabic speakers. These ratings were then correlated with each other to ascertain what bearing a foreign accent has on the comprehensibility and intelligibility of speech. The three listener ratings were also correlated to error counts for phonetic and phonemic errors, and non-native intonation, in an attempt to establish which error types affect an utterance most in terms of its accentedness, comprehensibility and intelligibility.

The original study resulted in two major findings: firstly, accentedness and intelligibility were found to be orthogonal; and secondly, non-native intonation was found to be more highly correlated with problems of comprehensibility than were phonetic and phonemic errors. These two findings were re-examined in the context of the present study. The results suggest that if language is treated as a complex system, whose behaviour is based on the interaction of all of its parts, as Dynamic Systems Theory (De Bot, 2008) proposes, then i) accentedness does in fact affect the intelligibility of L2 speech, and ii) both segmental and suprasegmental issues contribute to problems of comprehensibility. The main finding of this study is that whether speech is comprehensible and intelligible is based on the outcome of the interactions of its parts. It concludes that to examine individual linguistic aspects of an utterance as isolated phenomena is ineffective.

List of Contents

List of Tables

List of Figures

Acknowledgements

Many thanks go first of all to my supervisor Sam Hellmuth for her dedication and academic support, and for the mountain of time she miraculously found in her busy professional life to discuss this project at length, and guide me through it. I always came away from our meetings inspired. Thanks go also to my Thesis Advisory Panel member Tamar Keren-Portnoy for her very useful suggestions.

I am extremely grateful to Marta Szreder of the United Arab Emirates University for recording my speech samples at the UAE University in Al-Ain for me when I was unable to do so myself, and to the UAE University for providing the opportunity to record the speech samples with their students, as well as to the people – students and others - both in Al-Ain and in York, who participated in the studies. My thanks also go to Santander Universities UK for funding my trip to Accents 2016 in Lodz, Poland, to disseminate the results of the study. Thanks go also to Francis Duah for spending several hours helping me to become better acquainted with the statistical approaches used in such research. I am also very grateful to Tracey Derwing, one of the authors of the original study, for the considerable time she spent clarifying parts of the original study for me, and for the interest she showed in mine. Thanks, also, to Volker Dellwo for sharing his Praat script for calculating rhythm metrics, and of course to Paul Boersma and David Weenink, for giving free access to their acoustic analysis software. I am also grateful to my long-time linguist friend Piklu Gupta for proof-reading the final version of this thesis and making some useful comments.

Last but not least, a big thank you goes to my son for his patience whilst waiting for me to be free to spend time with him, and for making endless cups of tea.

Author’s Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References. I declare that part of the results of my thesis was presented at the Accents 2016 conference, in collaboration with Sam Hellmuth, University of York, and Marta Szreder, United Arab Emirates University.

1. INTRODUCTION

This study examines the relationship between the accentedness, comprehensibility, and intelligibility of L2 speech. It is based on the replication of a study (Munro & Derwing, 1995a) carried out in the 1990s, and addresses two research questions which are related to the findings of that study.

The first research question is based on their first major finding: are the accentedness and intelligibility of L2 speech orthogonal - ie statistically independent? The results of their study suggested this to be the case, and this finding has formed the basis in recent decades for the idea that foreign accentedness is not an obstacle to communication, although it may make communication more difficult.

The second research question is based on their second major finding: do suprasegmental aspects of L2 speech – here, non-native intonation - have a bigger impact on the comprehensibility of L2 speech than segmental issues? Their study found this to be the case, and there has been a considerable increase in the study of prosodic systems in recent decades.

The remainder of this chapter is given over to looking at what it means to have a foreign accent, in particular within a modern communicative setting. Chapter 2 gives an overview of previous studies addressing these issues. Chapter 3 describes the original study which the replication study was based on. Chapter 4 details the method of the replication study. Chapter 5 discusses the linguistic analysis of the speech samples as well as the predictions for the results, based on a cross-linguistic study. Chapter 6 presents the results. Chapter 7 provides a general discussion of the results, and addresses the two research questions again. Chapter 8 talks about conclusions which can be drawn from the study, and suggests some possible directions for future research.

The Existence of Foreign Accent

Before embarking on an examination of the effects of foreign accentedness, we need to define exactly what is it that we are referring to when we talk about foreign accent. Munro (2008) defines this in terms of both segmental and suprasegmental differences to native speech: on a segmental level it involves the insertion, omission or substitution of a phoneme, or variation at sub-phonemic level, such as VOT, vowel duration, and syllabification (cf. Ioup, 2008). On a suprasegmental level, both rhythmic and intonational differences play a role, as first shown in experiments carried out by Van Els & DeBot in their comparison of native and L2 Dutch utterances (cited in Munro, 2008). The properties described above are all properties of speech; Derwing & Munro (2005) also acknowledged the role of the listener in the perception of accentedness.

Theoretical models of L2 acquisition differ in their descriptions of the causes of foreign accentedness: in the Critical Period Hypothesis (Lenneberg, 1967), it is claimed that neural plasticity ends with puberty and that from this point on native-like accentedness is therefore not possible. The Perceptual Assimilation Model (Best, 1994) claims after (L1) phonemic categories have been learnt, non-native phonemic categories are assimilated to native categories based on their articulatory similarity. The Native Language Magnet Model (Kuhl, 1992) posits that phonetic prototypes are formed during L1 acquisition (before phonemic categories are formed), which then act as a magnet for new (L2) sounds. Flege's Speech Learning Model (1995) links the decline in the ability to assume an L2 native-like accent to age but not to a critical period, describing this as a gradual process as the L1 phonological system becomes more and more established. Brown's model (2000) is based on the idea that the internal structure of the phonemic system of the L1 leads to a rigidification of perceptual ability (as opposed to its phonetic properties). Escudero’s model of Second Language Linguistic Perception (Escudero, 2009) builds on previous models but focuses on the importance of L2 sound perception. Despite the differences in these models, they all agree that a) the L1 affects the L2 accent and b) older learners generally have stronger foreign accents.

Is it a problem?

It is only in recent years that a distinction has been made between foreign accentedness per se, and foreign accentedness which is detrimental to communication. Munro (2008, p194) comments: “Current views of foreign-accented speech hold that native pronunciation in the L2 is not only uncommon but also unnecessary”. This view, to which the findings of the study (Munro & Derwing, 1995a) replicated within the present study (which will be described in more detail in the next chapter) has contributed, has led to a series of studies on the relationship between accentedness and intelligibility, and forms the basis of the ideas published in Derwing & Munro (2015). There are two points to make here. The first is that because of the many different varieties of English that exist around the globe, and the variation in native pronunciation which this has led to, nativeness is difficult define for English. The second is that as English has become more and more an international language, used by many L2 speakers for purposes of international communication in commerce and many other fields, and many of the world's conversations in English take place between non-native speakers, the concept of what constitutes comprehensibility and intelligibility in terms of accentedness has also inevitably changed. Jenkins (2000) identified a 'Lingua Franca Core', which she describes as a scaled-down phonological core 'to focus attention on those items which are essential in terms of intelligible pronunciation' (p124) for speakers of EIL (English as an International Language). Functional load of phonemes plays a significant role in the definition of what is deemed essential. It is outside of the scope of this study, however, to address issues of comprehensibility and intelligibility in reference to non-native listeners of English: the remainder of this study examines the comprehensibility and intelligibility of L2 English with regard to native listeners only.

When tackling the question of whether foreign accentedness is a problem or not, it is useful to consider it from the perspective of Dynamic Systems Theory. This theory was first developed in the fields of physics and mathematics but is now used across many different disciplines to describe the behaviour of complex, adaptive, non-linear systems. Elman (1995) argued for the need for a new theory in linguistics because advancement in cognitive theories of how the brain works (based on there being no static, discrete, passive or context-free representations) means that the traditional view of mental processing is no longer valid. De Bot (2008, p167) describes DST as “part of the study of systems, in which systems are studied as a whole rather than with focus on their parts”. Within linguistics, DST has typically been applied to language acquisition and development rather than to language processing and studies of the comprehensibility and intelligibility of L2 speech, but parallels can be seen between this latter area of research and some of the applications in language acquisition. Lorenzo (1963, cited in De Bot, 2008), showed that small differences in initial conditions of complex systems can lead to big changes (ie there is a discrepancy between input and effect), incorporating the idea of non-linearity. By way of example, the ‘Breakthrough’ concept (Lightbrown & Spada, 1999, cited in De Bot, 2008) represents the idea of “a sudden jump in proficiency because all the disconnected information seems to fall into place” (De Bot, 2008, p173). By analogy, whether the disconnected information contained in an utterance a listener hears falls into place, makes the difference between it being intelligible or not.

2. LITERATURE REVIEW

In recent decades, a considerable amount of research has been carried out into the effects of foreign-accented speech. Amongst others, Munro & Derwing (1995a, 1995b, 1997, 2001) have carried out extensive research on the relationship between the accentedness, comprehensibility, and intelligibility of L2 speech. Their empirical work includes both cross-sectional and longitudinal studies on the speech of Mandarin speakers in L2 English. Their work has been ground-breaking in the study of accentedness: it was found that it is possible to accurately interpret heavily accented speech.

This chapter first examines the main concepts behind the subject of these studies – accentedness, comprehensibility and intelligibility - and attempts to provide a definition for them, then reviews the empirical work to date on the effect of foreign accentedness on L2 comprehensibility and intelligibility.

2.1 Defining Accentedness, Comprehensibility and Intelligibility

Before engaging further with theoretical and empirical work on these three concepts, it is necessary to define the ideas behind the terminology used. As the empirical study undertaken here is a replication of the work of Munro & Derwing (1995a), it is useful to look at their definitions first.

Accentedness is not defined in concrete terms in the original Munro/Derwing study which is replicated here, but in a study which was carried out at about the same time, they define it as “non-pathological speech that differs in some noticeable respects from native speaker norms.” (Munro & Derwing, 1995b, p289). They detail this further by saying that 'the deviations may include phone substitutions, phonetic distortions, and non-native prosodic patterns' (p302). As such, it is based on the properties of speech. The definition of accentedness has, however, undergone a subtle change in recent years. In ratings studies, accentedness and the quantifiable segmental and suprasegmental errors in L2 speech do not necessarily amount to the same thing, as studies which use accentedness ratings as a measure of accentedness introduce a listener perspective to the equation. The definition which includes the *perception* of accentedness on the part of the listener is represented here in the accentedness ratings.

Comprehensibility is defined as 'the listener's perceptions of difficulty in understanding particular utterances' (Munro & Derwing, 1995b, p291). Comprehensibility has also been defined as the 'ease of interpretation' (Munro & Derwing, 1995a, p290, quoted from Varonis & Gass, 1982).

Intelligibility has been defined as the 'extent to which an utterance is actually understood' (Munro & Derwing, 1995a, p289), and this is the definition which is adopted in both the original and the replication study, though they concede that there is no universally accepted way of assessing it.

There is some debate about the definition of comprehensibility and intelligibility, as there is a certain amount of overlap in the concepts they represent. Comprehensibility is included in the broad definition of intelligibility given by Levis (2006, p252, cited in Isaacs & Trofimovich, 2012, p477), whereby comprehensibility and intelligibility are not distinct from each other but are considered to be part of a continuum of the same concept. Smith & Nelson (1985) use the term *intelligibility* to denote word or utterance recognition, and *comprehensibility* to denote word or utterance meaning. In the distinction which Derwing and Munro make, described above, comprehensibility can be seen as the process of interpreting an utterance, and intelligibility as the result of that interpretation. Their distinction is often adopted in L2 research, and distinguishing between the recognition of an utterance and the effort needed on the part of the listener to understand an utterance certainly makes empirical studies easier by providing an obvious way of measuring intelligibility. Zielinski (2006) points out, however, that understanding the meaning of an utterance may well contribute to its intelligibility: there is evidence to suggest that listeners use contextual, lexical and syntactic knowledge when trying to identify the words of a utterance (see studies by Bard, Shillcock & Altmann (1988), as well as Cutler & Clifton (1999), cited in Zielinski, 2006, p7), and it is therefore questionable whether word or utterance recognition can in fact be measured separately from word or utterance meaning.

2.2 Review of Empirical Work

The empirical work carried out on the effects of foreign accent for either comprehensibility or intelligibility have led to vastly differing conclusions about the sources of the problems of both comprehensibility and intelligibility. As Trofimovich & Isaacs find: 'although listener perceptions are central to the construct of comprehensibility, little is known about the dimensions that underlie listeners' L2 comprehensibility judgements' (2012, p476). The contradictions in findings are diverse, both in comparative studies of the effects of accentedness versus other factors such as grammar, as well as in comparative studies involving different aspects of pronunciation, such as segmental and suprasegmental factors.

Derwing & Munro (2015) cite the following studies carried out on pronunciation as against other factors being an influence on the comprehensibility and intelligibility of L2 speech: Gynan (1985) found in studies on Spanish L2 speakers of English that listeners thought that foreign accent influenced comprehensibility more than grammatical errors, whereas Ensz (1982) found the opposite in his studies on English American L2 speakers of French. Politzer (1978) also found that lexical and grammatical errors affected comprehensibility more than foreign accentedness, and Albrechtsen et al. (1980) also found little to suggest a correlation between pronunciation problems and comprehensibility in their studies with Danish-accented English. Fayer & Krasinski (1987), on the other hand, found non-native pronunciation had a profound effect on comprehensibility.

They also cite studies comparing specific aspects of pronunciation: on a segmental level, Gimson (1970) found that for English learners, the pronunciation of consonants was more important than vowels, but Schairer (1992) found the exact opposite for learners of Spanish. Comparing segmental and supra-segmental aspects of pronunciation, both Anderson-Hsieh et al. (1992), Johansson (1978) and Palmer (1976) found that prosodic errors were more detrimental to understanding L2 speech than segmental errors, whilst Koster & Koet (1993), and Fayer & Krasinski (1987) argued the opposite.

In their attempt to break down the elements contributing to lack of comprehensibility of

L2 speech, Trofimovich & Isaacs (2012) also examined the above three categories, but added discourse as a fourth category. Their study included ratios of error counts of the following areas of phonology: segmental errors; syllable structure errors; word stress errors; vowel reduction; pitch contour; and pitch range. Their results showed a strong correlation between comprehensibility ratings and word stress error ratios, as well as vowel reduction ratios. The other areas showed moderate correlations with the exception of pitch range, which showed no correlation at all. Their study was carried out with native French learners of English, and although French is argued by some to lack word stress altogether (Dupoux & Peperkamp, 2002), and this could therefore be a problem specific to French speakers, English word stress (and rhythm) cause problems for speakers of other languages too, and Swan and Smith (2001, cited in Isaacs et al., 2012, p498) suggest that English stress patterns are a general cause of global comprehensibility issues.

Isaacs et al. (2012) suggest that word stress and rhythm problems "create a bottleneck at the phonological encoding and articulation stage of speech production". This affects MLR (mean length run) – i.e. fluency – which in turn affects comprehensibility. They also remark that segmental errors should not be discounted from affecting comprehensibility, as they did find some correlation. The significance of segmental errors depends on the functional load of a particular segmental contrast.

Recent work suggests that lexis also plays a role in the comprehensibility of L2 speech (Saito et al., 2015, 2016). They examine several lexical factors (appropriateness, fluency, variation, sophistication, abstractness, sense relations) which may influence whether communication is successful or not, and their findings show that lexical accuracy (appropriateness) and complexity (variation of sophistication) are especially relevant.

Dirven & Oakeshott-Taylor (1984, p333, cited in Pennington & Richards, 1986, p211) say: "To interfere with stress, timing, fundamental frequency [and other aspects of prosodic continuity in discourse] usually has more drastic consequences for comprehension than removing the cues of a particular [phonological] segment".

From this short literature review, it can be seen that the field is wrought with conflicting views and findings on the diverse potential effects of foreign accentedness. Saito et al.

(2015, p4) note that “while influences of individual properties of speech on comprehensibility and accentedness are relatively well understood, it is still unclear how multiple dimensions interact…”. All studies hope to shed new light on their particular topic, or at the very least provide food for thought, and this one, by studying the interactions between these individual properties of speech, is no exception.

3. BACKGROUND: THE ORIGINAL STUDY

This chapter describes the study which was replicated here: a study by Derwing and Munro (1995a) from the 1990s on the accentedness, comprehensibility and intelligibility of second language speech, based on the L2 English speech of L1 Mandarin speakers.

3.1 Introduction and Method

In this study, the spontaneous speech of ten native speakers of Mandarin speaking English was recorded (in the form of descriptions of a picture sequence), as well as two British English native speaker controls. The L2 speakers had all spent between one and six years in Canada, were described as proficient in English and had all learnt English after puberty. In total, 36 spontaneous speech samples of one utterance each – three per speaker - were extracted for use in the study.

These stimuli were presented to 18 native English listeners who were students of linguistics or ESL teaching at the University of Alberta. The listeners had to perform three tasks designed to rate the speech samples according to three dimensions:

i) intelligibility, based on orthographic transcription;
ii) comprehensibility, based on evaluation on a scale of 1-9, with 1 being 'easy to understand' and 9 being 'impossible to understand';
iii) accentedness (performed 4 days after the first two tasks), based on an evaluation on a scale of 1-9, with 1 being 'no foreign accent' and 9 being 'very strong foreign accent'.

The authors then analysed the speech input according to phonemic errors, phonetic errors, intonation, grammatical errors and utterance length. Phonemic errors were defined as deletion or insertion of a segment. Phonetic errors were defined as cases where the production of segment was recognisable but distinctly non-native. Intonation was rated on a scale of 1-9 with 1 being 'native-like' and 9 being 'not at all native-like'.

Grammatical errors were defined as morphosyntactic errors. The three ratings are then correlated with each other to ascertain what bearing a foreign accent has on the comprehensibility and intelligibility of speech. These three ratings of accentedness, comprehensibility and intelligibility were also correlated to the error counts for phonetic and phonemic errors, non-native intonation, grammatical errors and utterance length. Although they found that interrater reliability was high, because of the individual differences found amongst the listeners in the strength of correlation between the accentedness and comprehensibility ratings, they calculated Pearson correlations based on each individual listener.

3.2 Results

Table 3.1 and 3.2 below show the correlations they obtained, as described in the previous section: Table 3.1 shows the correlations between the listener ratings, whilst Table 3.2 shows correlations between the ratings and the stimulus measures (phonetic and phonemic errors, non-native intonation, grammatical errors, and utterance length).

	# of Listeners showing a Significant Correlation
Accentedness-Comprehensibility	17/18
Comprehensibility-Intelligibility	15/18
Accentedness-Intelligibility	5/18

Table 3.1: Listeners showing Significant Correlation between Ratings in Original Study

The results suggest a strong correlation between accentedness and comprehensibility, and comprehensibility and intelligibility, but also that accentedness and intelligibility are orthogonal (which forms the basis of the first research question).

Number of Significant Correlations Between Perceived Comprehensibility, Accent, and Intelligibility Scores and a Stimulus Measure (p<.05)

Stimulus Measure	Perceived Comprehensibility No.	%	Accent No.	%	Intelligibility (Words Correct) No.	%
Phonemic Errors	8	44	14	78	5	28
Phonetic Errors	2	11	13	72	0	0
Intonation	15	83	16	89	4	22
Grammatical Errors	10	56	14	78	3	17
Utterance Length	0	0	0	0	0	0

Table 3.2: Correlations Table for Ratings and Stimulus Measures of Original Study

The results suggest that comprehensibility is strongly correlated to intonation (which forms the basis of the second research question), accent is strongly correlated to phonetic and phonemic errors, as well as to intonation and grammatical errors, and that intelligibility is not strongly correlated to any of these linguistic measures.

4. METHOD

The present study replicated the original study with native Arabic speakers instead of native Mandarin speakers. This chapter details the methodology used in the replication study, and is divided into four parts. Section 4.1 describes the collection of the stimuli, whilst section 4.2 describes the collection of speech ratings. In section 4.3, the linguistic analysis of the stimuli is documented. In section 4.4 the methodology used for the statistical analysis of the speech ratings and the results of the linguistic analysis is described.

The present study follows the same format as the original study: ratings were collected from 18 native English listeners for the accentedness, comprehensibility, and intelligibility of three spontaneous L2 English utterances for each of ten L1 Emirati Arabic speaking intermediate-to-advanced L2 learners of English, as well as two native British English speakers. These ratings were then correlated with each other to ascertain what bearing a foreign accent has on the comprehensibility and intelligibility of speech. The three listener ratings are also correlated to error counts for phonetic and phonemic errors, non-native intonation, morphosyntactic errors and utterance length.

The listener ratings served as the dependent variables in the study and the linguistic analyses were used as predictors.

4.1 Stimuli Collection

Speakers

The speakers were ten native Emirati Arabic speakers with an intermediate to advanced level of English, who were all students of Linguistics at the University of the United Arab Emirates (UAEU), and two native British English speakers. All were female. A sample consent form from both the UAEU, as well as the University of York, can be found in Appendix G (the latter was also used for the listeners).

Recordings

The recordings of the native Arabic speakers were made in the phonetics lab of the University of the UAE, using an Apogee MIC 96K table-top microphone recorded directly to Audacity. The recordings of the native English speakers were made in a quiet residential room in York, using a Marantz PMD 620 with Shure SM10 headset microphone recorded directly to wav format 44.1KHz sample rate 16 bit.

Procedure

In pairs, speakers were asked to perform a map task, and the results were recorded. The idea for this task was taken from the HCRC Map Task Corpus developed at the University of Edinburgh (Anderson, 1991) and was used to elicit a spontaneous conversation for the collection of spontaneous speech samples. This method was chosen over the picture task used in the original study as it provided a more natural communicative environment.

Three utterances for each speaker were taken from these recordings, giving a total of 36 utterances. They were not all complete sentences, but those that were not began and finished at clear phrasal boundaries. The 36 stimuli can be found in Appendix A, which contains both the orthographic form and the IPA (SSBE) transcription.

4.2 Speech Ratings of Stimuli

The utterances were played to native English listeners, who were asked to perform three different rating tasks. The utterances were randomised in each case, so that each listener heard the 36 utterances in a different order.

Listeners

The listener were 18 native English speakers who all had a connection with linguistics or language teaching – they were either students of linguistics, EFL teachers, or retired EFL teachers, ranging in age from 18 to 75. 3 were male and 15 were female. All reported having no hearing problems.

Procedure

The three rating tasks are described below. In the original study, the first two tasks (in which the intelligibility and comprehensibility ratings were collected) were performed in one session, and the final task (in which the accentedness ratings were collected) was performed in a separate session which took place four days after the first one. For practical reasons, it was not possible to ask the listeners to attend two separate sessions, so all three tasks were performed on the same occasion. Because the accentedness ratings were considered to be the most instinctive, they were collected first – i.e. before the listeners had started to analyse or reflect on what they had heard. The Praat (Boersma & Weenink, 2013) MFC interface was used to collect the results.

Listening Task 1: Accentedness Ratings

In the first task, the 36 stimuli were played to the listeners, and after hearing each one they were asked to give a rating according to their perception of how strong the foreign accent was for each utterance. The stimuli for which the results were collected were preceded by two practice samples, which were not evaluated. These ratings were based on an evaluation on a Likert scale of 1-9, with 1 being 'no foreign accent' and 9 being a 'very strong foreign accent'. Each stimulus could be played as many times as necessary before proceeding to the following one.

Listening Task 2: Intelligibility Ratings

The listeners then heard the 36 stimuli again, and were asked to write down in standard English orthography exactly what they heard. Each stimulus could be played as many times as necessary before proceeding to the following one.

Listening Task 3: Comprehensibility Ratings

After each orthographic transcription, the listeners were asked to give a rating according to their perception of how easy it was to understand that utterance. These ratings were based on an evaluation on a Likert scale of 1-9, with 1 being 'easy to understand' and 9 being 'impossible to understand'.

4.3 Linguistic Analysis of Stimuli

In this section, the various types of linguistic analysis carried out by the author on the stimuli are described. These were carried out in accordance with the criteria and methodology described by Derwing and Munro (1995a).

Procedure

The 36 utterances were analysed on five counts (referred to as stimulus measures), as in the original study. They are organised in three groups, as shown in Table 4.1 below: the first group contains factors which may have some bearing on the listener ratings of accentedness, comprehensibility and intelligibility, but are not related to phonological acquisition. The second group contains factors which relate to segmental issues, and the third group is made up of factors which relate to suprasegmental issues. Examples of these can be found in the Chapter 5, where the results of the linguistic analysis are discussed.

GROUP 1: **Extra-phonetic/ phonological factors**	**GROUP 2:** **Segmentals**	**GROUP 3:** **Suprasegmentals**
Utterance Length	**Phonetic Errors**	**Intonation Rating**
Morphosyntactic Errors	**Phonemic Errors**	

Table 4.1: Stimulus Measures used in the Experiment

Segmental errors were counted manually based on auditory impression made with reference to spectrograms and waveforms in Praat for all utterances. Phonemic errors were defined in the original study as deletion or insertion of a segment. Phonetic errors were defined in the original study as cases where the production of a segment was recognisable but distinctly non-native.

The stimuli were low-pass filtered in Praat in order to collect the **intonation ratings**. The author listened to each utterance and recorded a rating for each one according to how native or non-native the intonation of the utterance sounded, on a scale of 1-9, with 1 being 'native-like' and 9 being 'not at all native-like'.

4.4 Statistical Analysis of the Results

Pair-wise Pearson correlation coefficients were calculated as for the original study between all the speech ratings (i.e. for each pair of accentedness, comprehensibility and intelligibility). Inter-rater reliability was calculated using the intraclass correlations suggested by Shrout & Fleiss (1979). Pearson correlation coefficients were then calculated between each speech rating and each of the five linguistic analyses described in Section 3.1 (phonetic error count, phonemic error count, intonation rating, utterance length and morphosyntactic error count). This methodology mirrors that of the original study.

A more modern statistical method was then adopted for the remainder of the current study: the statistical methods used in linguistics have advanced since the time of the original study, and it has become common practice to use more complex statistical procedures rather than simple pair-wise correlations. Linear mixed models are an efficient way of dealing with multiple continuous independent variables, and are also able to take into account the influence of subject variability by calculating the random intercepts and slopes of subjects. Indeed, they are preferred by Baayen et al. (2008) over by-subject analyses as “they allow the researcher to simultaneously consider all factors that potentially contribute to the understanding of the structure of the data” (p410). The lme4 package (Bates, 2015) in the R environment (R Core Team, 2016) was used for all mixed models described in this study.

5. LINGUISTIC ANALYSIS OF THE SPEECH SAMPLES

This chapter documents the five different types of linguistic analysis, and is divided into the groups described in the Method: the first group covers the extra-phonetic/phonological factors, the second group the segmental aspects of the stimuli, and the third the suprasegmental analysis. They are documented in the subsequent sections in that order.

5.1 Extra-phonetic/phonological factors

Morphosyntactic Errors: 27 morphosyntactic errors were found. There were 9 utterances with one error, 6 utterances with 2 errors, and 2 utterances with 3 errors. Of the 30 non-native utterances, 13 were error-free. The errors are shown in Table 2, and exemplified below. All morphosyntactic errors are documented in Appendix B.

Utterance Length: Utterance length was counted in words. The mean length for the 36 utterances was 10.8 words, with a range of 6 to 19 words.

MORPHOSYNTACTIC ERROR TYPE	NUMBER OF OCCURRENCES	EXAMPLE
Incorrect Word Order: (in indirect questions (adverb misplacement	8 5) 3)	*Can you tell me where is the..* (#9) *Can you tell me please where..* (#23)
Incorrect prepositions	7	*opposite to the.. State College* (#6)
Omitted articles	6	*It's next to post office* (#13)
Superfluous articles	1	*…near the River Street…* (#16)
Omitted copula	1	*The railway station in area seven..* (#19)
Incorrect number (on noun)	1	*… in the Main Streets… (#4)*
Omitted preposition	1	*It's in area one.. the east of, er..* (#32)
Incorrect tense	1	*Where can I found the…* (#31)
Incorrrect pronoun	1	*He's near..*. (#16)

Table 5.1: Morphosyntactic Errors found in the Stimuli

5.2 Segmentals (Phonetic and Phonemic Error Analysis)

This section contains an in-depth analysis of phonetic and phonemic errors in the L2 stimuli. First, potential sources of errors in the realisation of consonants and vowels are described, based on a cross-linguistic analysis of Arabic and English. Each section concludes with a description of typical errors of that type which occurred in the stimuli. Transcriptions are given between forward slashes (ie /) in the case of both phonetic and phonemic errors, as the reference is phonemic.

The target assumed is Standard Southern British English (SSBE), or Standard American: there is not a clear target given to students in the United Arab Emirates, but these are the two varieties which are generally presented, and so nothing was counted as an error which was acceptable in either of these varieties. The SSBE variety has been used in the transcriptions. A detailed error analysis of each of the 36 stimuli is given in Appendix C (Phonetic and Phonemic Analysis of the Stimuli).

5.2.1 Consonants

Obstruents

Variations in Voice Onset Time (VOT) is one of the main differences in the realisation of **plosives (stops)** between languages. VOT measures the time between the release of a stop and phonation onset of the following vowel. There are three types: *pre-voiced* – where phonation starts during the closure phase of the stop, *unaspirated* - where phonation starts at or close to the end of the closure, and *aspirated* – where there is a delay between the release of the closure and the onset of phonation (Di Paolo et al., 2011). VOT is measured from the point of the stop release, with the result that with pre-voicing – where phonation starts before the release – a negative VOT value is given. Lisker and Abramson (1964, cited in Flege & Port, 1981) talk of *voicing lead* and *voicing lag* (of which the latter is generally divided into short-lag and long-lag), whereby voicing lead overlaps with negative VOT values, and voicing lag with positive VOT values. As well as differences in VOT, the voicing of stops can have two other effects: firstly, it can influence the duration of a preceding vowel, and secondly, in word-final position, stop closure times of voiced and unvoiced stops can vary (Flege & Port, 1981). These effects vary between languages.

Although there is a certain amount of variation between dialects, unaspirated stops are common in English for /b d g/, and aspirated stops are used for /p t k/. In other words, voiced stops generally have a short-lag VOT in English, whereas voiceless stops have a long-lag VOT (Di Paolo et al., 2011). Also, vowels in English are longer when followed by a voiced stop, as opposed to a voiceless stop. Voiceless stops at the end of syllables are longer than voiced ones (Ladefoged et al., 2011).

In Arabic, voiced stops have a continuous glottal pulsing during the stop closure (Flege &

Port, 1981) - in other words they show pre-voicing – and voiceless stops have a short lag (although there is no /p/ in the phonemic inventory of Arabic, a voiceless allophone of /b/ exists, which often occurs next to voiceless sounds (Newman, 2002)).

Khattab (2002) provides a useful table comparing VOT in English and Arabic stops, adapted from Deuchar & Clark (1995):

English Stops:

/b d g/ /p t k/

------------------------|-- (VOT)

- 0 +

lead voicing short lag long lag

Arabic Stops:

/b d g/ /p t k/

------------------------|-- (VOT)

- 0 +

lead voicing short lag long lag

It is predicted that the L2 speakers will use the VOT values from Arabic stops in their voicing of the English stops.

Fricatives and **affricates** show similar patterns to stops in English in that they influence the length of preceding vowels: vowels are longer when followed by a voiced fricative or affricate as opposed to a voiceless one (Ladefoged et al., 2011). As with stops, if they occur at the end of a syllable, voiceless fricatives/affricates are longer than voiced ones. In English, voiced fricatives and affricates generally become voiceless towards the end; often, only a small portion at the beginning of the sound is voiced, and voicing is only maintained throughout if they are followed by a voiced segment.

In contrast to English voiced fricatives, in Arabic voiced fricatives voicing is maintained throughout the segment. The two affricates which exist in Arabic (/dʒ/ and /tʃ/) are not used in many dialects of Arabic, though they do exist in Gulf Arabic (Holes, 1990).

As voiced consonants in English (including fricatives) are almost never voiced throughout, it is predicted that the voiced fricatives and affricates produced by the L2 speakers may be 'over-voiced' – i.e. voicing will continue for longer than in native speech.

Approximants/Liquids

Liquids, or approximants, are complex sounds which can function syllabically but are not considered to be true vowels (Di Paolo et al, 2011). They show a considerable amount of variation cross-linguistically.

In the two varieties of English which are taken in this study as potential targets for Emirati Arabic speakers of English, the usage of the **non-lateral** /ɹ/ varies. SSBE is a nonrhotic variety of English, in which /ɹ/ has undergone deletion in post-vocalic environments, whereas Standard American is a rhotic variety of English, and thus maintains /ɹ/ in all environments.

The phonetic realisation of the **lateral** /l/ can either be palatal or velar. At one end of the scale, it can be a 'bright' or 'clear' sound, which is palatised (anterior), and where the front of the tongue is raised, and at the other end it is velarised, or 'dark', and where the dorsum is raised or retracted. Clear /l/s have higher F2 values than dark /l/s (Di Paolo, 2011).

In English, in onset position, a palatal /l/ is more common, whereas in coda position, it is more likely to be velarised, though the extent of both of these phenomena is dialect-dependent, and the place of articulation in onset position is also influenced by the frontedness or backness of the following vowel. In Arabic, the /l/ is described as dento-alveolar (Newman, 2002), and contrasts with the emphatic (pharyngealized) /l/ (l dot) (Watson, 2002), which is only found in the word for 'God' – [/al/ ˘aah] (transcription from Newman, 2002), though the pharyngealised allophone of /l/ does occur by assimilation when near other pharyngeal consonants (Newman, 2002). It is predicted that the dental-alveolar articulation of /l/ may be adopted by the L2 speakers for English.

5.2.1.1 Phonetic Errors found in the Stimuli in the Realisation of Consonants

Examples are given below for repeated phonetic errors, with the context in which they occurred. Where possible, an example was selected from the data where the consonant in question occurred in intervocalic position (or if necessary before a syllabic sonorant), to minimise coarticulation effects.

Obstruents

The non-native speech samples showed several occurrences of lead voicing in the voiced plosives /b d g/. Examples of this in the data are the pre-voiced /b/ in the word *bus* in sample 8, the pre-voiced /d/ in the word *middle* in sample 36. Figure 5.1 shows the waveform produced in Praat showing voice onset time (VOT) in the L2 word *middle*, followed for comparison by a waveform showing an example of VOT in the segment /d/ (in an utterance containing the words ..*where Delaware..*) produced by a native speaker, in figure 5.2.

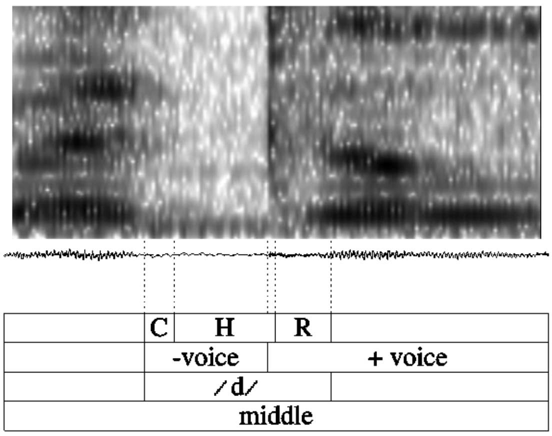

Figure 5.1: L2 /d/ in *middle* showing pre-voicing

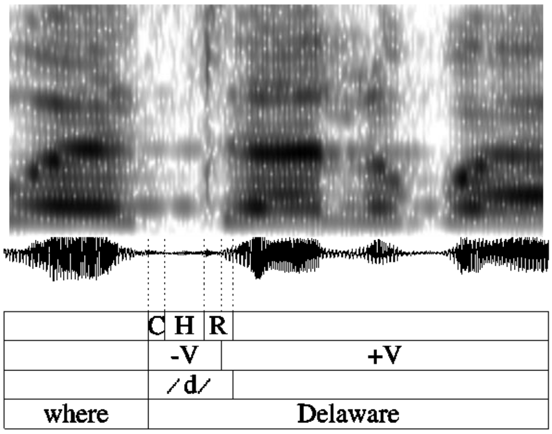

Figure 5.2: L1 /d/ in *Delaware* showing short-lag voicing

The voiceless plosives /p t k/ were often produced with short-lag VOT (in contrast to the long-lag VOT of native English speech) in pre-vocalic position, such as /k/ in the word *can* in sample 2, /t/ in the word c*ity* in sample 15, and /p/ in the word *post* in sample 32.

Figures 5.3 and 5.4 below show the VOT in /t/ produced by an L2 speaker and by one of the native speaker controls respectively.

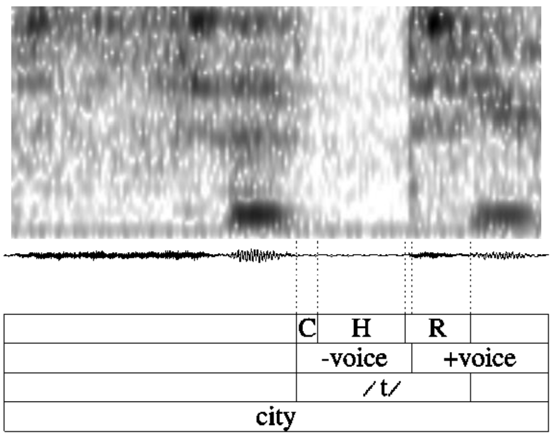

Figure 5.3: L2 /t/ in *city* showing short-lag voicing

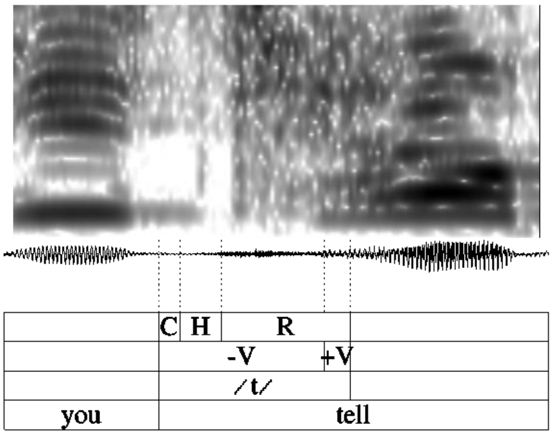

Figure 5.4: L1 /t/ in *you tell* showing long-lag voicing

Voiceless plosives in word-final position were sometimes produced with a shorter closure time in contrast to that of native speech, such as in /t/ in the word *market* in sample 7. This also occurs in pre-consonantal position, such as /k/ in the word *bookshop* in sample 17, where the burst is stronger than in native speech (which is barely noticeable in this position as the stop is unreleased), as well as in post-consonantal position, where in native speech voiceless plosives are often unaspirated, such as /k/ in the word *school* in sample 36.

As predicted, voicing in fricatives, specifically in /v ð z/, was frequently observed to be present for a greater proportion of the segment than in the native samples. Examples from the data are the extra voicing in /v/ in the word *of* in sample 25, or in /z/ in the word *is* in sample 28. In the latter example, /z/ was fully voiced, compared with an instance of /z/ in the same word in one of the native samples, where the fricative had a total length of 125 msecs, but only the first 26 msecs were voiced.

Liquids

The phoneme /ɹ/ does not exist in the phonemic inventory of Arabic, and was often rendered as a tap (/ɾ/) by the L2 speakers (e.g. in the word *library* in sample 1). This is given as a phonetic error as the tap does not exist in the target SSBE English, and the tap of Standard American is an allophonic variant of /t/ or /d/, rather than /ɹ/. Although the transcriptions given of the stimuli are in SSBE, the use of the tap in place of the rhotic /ɹ/ in post-vocalic environments in the non-native speech samples was considered a phonetic error rather than a phonemic insertion because of the influence of Standard American as a potential L2 target.

The phonetic realisation of the /l/ in Arabic and English differs in that although there exists in both languages a 'clear' and a 'dark' variant, the use of the contrast depends on its position in the syllable in English, which is not the case in Arabic, but on the place of articulation of neighbouring consonants in Arabic (which is not the case in English). As there are no pharyngeal consonants in English which might evoke L1 transfer from Arabic of the dark /l/, it was predicted that the L2 speakers would produce a clear /l/ in English rather than a dark one in most instances, in both onset and coda position, and regardless of vocalic environment. This turned out to be the case.
There are numerous examples in the data; one example is in the word *hall* in sample 4. As shown in the examples in figures 5.5 and 5.6, higher F2 values were found in the L2 speech samples as compared with those found in the native speaker samples.

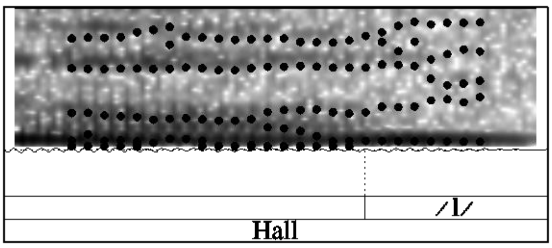

Figure 5.5: Formants of /l/ in L2 sample

The second formant of the /l/ in the word *hall* in figure 5.5 is 882 Hz.

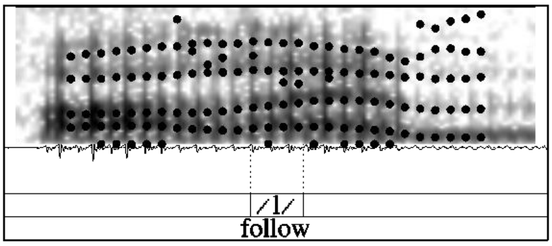

Figure 5.6: Formants of /l/ in L1 sample

The second formant of the /l/ in the word *follow* in figure 5.6 is 757 Hz.

5.2.1.2 Phonemic Errors found in the Stimuli in the Realisation of Consonants

Obstruents

/p/ was very frequently rendered as /b/ in the L2 speech samples, probably because the former does not exist as a phoneme in Arabic.

Although /ʤ/ and /tʃ/ do exist in the Gulf variety of Arabic, /ʤ/ is often realised as /ʒ/, and /tʃ/ as /ʃ/. An example from the data is /tʃ/ in the word *Manchester* in sample 17, which is rendered as /ʃ/.

Glottal stops (/ʔ/) were often inserted before vowels at the beginning of words, probably because syllables need a consonantal onset in Arabic (Watson, 2002). This error was not counted as a phonemic error, as its main effect was on the syllabification process. (It was not counted as a segmental error at all, but may have influenced the intonation ratings if the speech rhythm was changed by the effects of differing syllabification.)

Liquids

No phonemic errors were found in the speech samples relating to the realisation of liquids.

5.2.2 Vowels

Vowel quality is defined by three features: height, backness, and roundedness (Ogden, 2009). The first two can be established acoustically by the first two formants of a vowel (although these values are subject to individual variation, determined by the length of the vocal tract of the speaker). Vowel height correlates to the first formant, and backness correlates to the difference in frequency between the first and second formant (Ladefoged et al, 2011). The first formant is inversely proportional to vowel height, whereas the second formant is inversely proportional to the place of articulation. So F1 decreases as vowel height increases, and F2 decreases as the place of articulation becomes further back.

Vowel quality in English varies a great deal from one dialect to another, as does the number of vowels in any one dialect. For RP English, Roach (2004, cited in Ogden, 2009) established an inventory of 20 vowels: 12 monophthongs and 12 diphthongs (vowels of both distinct quantity as well as quality were counted separately). For General American, Ladefoged (2001) established an inventory of 22 vowels: 19 monophthongs (including rhotacised vowels) and 3 diphthongs. There are dialacts with fewer than this, but one general characteristic of all varieties of English is the crowded vowel space compared to other languages.

Though there is some dialectal variation, Arabic (including Gulf Arabic), in contrast to English, has a basic vowel inventory which is well below the mean (Newman, 2002), with just 3 short vowels and 3 long vowels, all of which are monophthongs. For some of these there exist pharyngeal allophones, which occur when the vowel follows (or precedes) a pharyngeal consonant. These allophones are more retracted than their non-pharyngeal counterparts, have a lower F2 value, and involve a constriction of the upper pharyngeal area (Watson, 2002). The process of pharyngealisation in vowels does not occur at a very great extent in the high vowels, /i:/ and /u:/.

English has been described as having a qualitative vowel system (ie each vowel is distinct because of differing formant values), whereas Arabic has a quantitative system (length, rather than changing formants, is what distinguishes vowels from each other), and this is predicted to cause problems for Arab L2 learners of English. In addition to this, it has been shown (Mitleb, 1984) that the two languages differ significantly in the duration of short and long vowels, thus vowel quantity may also be an issue in the L2 English speech samples.

The Arabic long, high vowels /i:/ and /u:/ are similar in quality to those in English, except that in English, vowels are generally diphthongal in nature (Cowell, 1964, cited in Almbark, 2012), which is not the case for Arabic: Chladkova & Hamann (2011, cited in Almbark, 2012) found that /i:/ in English showed a rising F2 contour, whilst /u:/ showed a falling F2 contour. This change in the F2 value is an indication of some degree of diphthongisation; as F2 remains constant in monophthongal vowels. Because of the reported monophthongal nature of Arabic, diphthongs are generally assumed not to exist in Arabic, thus errors in the production of the English diphthongs were predicted in the L2 speech samples. The existence of the two sounds [aj] and [aw] in Arabic does somewhat complicate this issue in terms of predicted behaviour, as they are not far removed from the English diphthongs /aɪ/ and /aʊ/.

5.2.2.1 Phonetic Errors found in the Stimuli in the Realisation of Vowels

Di Paolo and colleagues (2011) specify optimal conditions for the acoustic analysis of vowels, which include choosing an environment where the vowels are easily identifiable, i.e. do not occur next to voiced consonants. Some of the acoustic analyses described below are not in accordance with these recommendations: the reason for this is that the optimal conditions described were lacking in the data in those cases.

Unreduced vowels were considered to be phonetic errors in words which have a citation form but which is only ever used in rare instances of narrow focus, i.e. function words – for example, in the phrase *…next to the State College* (sample 35), the vowel in the word *to* was considered to be a phonetic error if it was not reduced. The schwa of unstressed syllables of content words, which can have no other form, was categorised as a phonemic error if it was replaced (usually for apparent orthographic reasons) by another vowel. An example of this is the last vowel in the word *Beethoven* (sample 11), which was pronounced as a /ɛ/ instead of /ə/.

Monophthongs: the long high vowels /i:/ and /u:/ were often found to be too monophthongal in the L2 speech samples – i.e. the typical rising and falling F2 contours of the English vowels were not present. An example of these differing formant values can be seen in /i:/ in the word *street* (samples 19 and 22), and for /u:/ the word *supermarket* (7 and 14). These are shown in Table 5.2. Interestingly, the L2 speech samples showed inverse F2 movement from those of the native samples, though the consequent fall in /i:/ and the rise in /u:/ was probably due to a change in articulatory position in preparation for the following consonant.

Monophthong	1st Formant	2nd Formant
L2 /i:/	601	2465 (falling from 2587 to 2289)
L1	545	2170 (rising from 1873 to 2330)
L2 /u:/	633	1862 (rising from 1614 to 2348)
L1	433	2063 (falling from 2098 to 1929)

Table 5.2: Monophthong Formant Values for /i:/ in street, and /u:/ in supermarket

Diphthongs: Probably because of the lack of diphthongs in Arabic, the diphthongs of the target varieties of English were often produced by the L2 speakers in a fashion which was closer to one of the monophthongs. These cases were counted as phonetic errors as the resulting sound was a variation of the target rather than a distinctly different phoneme.

The two main diphthongs (used in the contexts below in the assumed target varieties of English) which the L2 speakers had difficulties with were /eɪ/ and /əʊ/. In sample 2, /eɪ/ in the word *location* was closer to /e/. Although both components of the diphthong were clearly distinguishable, both audibly and acoustically, they were not as distinct as the two components of /eɪ/ in a similar context in one of the native samples, in the word *station* in sample 29. A similar case occurs in sample 35, where /əʊ/ in the word *row* was closer to /ɔ:/, and where the formant values were not as distinct as the two components of /əʊ/ in a similar context of one of the native controls - in the word *road* in sample 22. (The formant values for all diphthongs are shown in Table 5.3).

Diphthong	1st/2nd part	Speaker type	1st Formant	2nd Formant
/eɪ/	/e/	L2	399	1964
		L1	590	1855
	/ɪ/	L2	299	2054
		L1	543	2213
/əʊ/	/ə/	L2	696	1594
		L1	561	1035
	/ʊ/	L2	544	1659
		L1	429	1043

Table 5.3: Diphthong Formant Values for /eɪ/ in location/station and /əʊ/ in row/road

In the case of /əʊ/, the vowel produced sometimes approached a /ɒ/, though these occurrences may be based on reasons of orthography, as in the word *post* in sample 13.

5.2.2.2 Phonemic Errors found in the Stimuli in the Realisation of Vowels

Monophthongs

There were no instances where the L2 production of English monophthongs mapped onto other phonemes.

Diphthongs

Some of the diphthongs described in the previous section on phonetic errors were realised not only as less diphthongal vowels than in native speech, but completely as monophthongs, and because they mapped onto an existing monophthongal phoneme of English, they were categorised as phonemic errors.

In sample 17, /əʊ/ in the word *road* was auditorily identified as /ɔː/. In fact there was no acoustic evidence to support its perception as a monophthong: two distinct components were identified in Praat, indicating that the phonetic realisation of the sound was in fact the required diphthong. Possibly the reason for the aural identification as a monophthong was the difference in duration when compared to native speech – the L2 realisation of /əʊ/ had a duration of 58 msecs, as against the sample in the same context as used in the previous section, which was 108 msecs. Such cases were still counted as phonemic errors, despite the lack of acoustic evidence, as they were also likely to be identified as monophthongs by the listeners in the experiment, and thus potentially contribute to the ratings of accentedness, comprehensibility and intelligibility.

In sample 11, /eɪ/ was rendered as /e/, in the word *Beethoven*. There was a slight variation in the F2 values in the two barely recognisable components (raising from 2155 to 2224), though compared to the values of the native sample (which rose from 1855 to 2213), this was too slight to be identified impressionistically as a diphthong. Perhaps, also, the durational aspect of production played a role. The duration of the L2 vowel was 43 msecs, whereas the L1 vowel was 115 msecs (both occurred in unaccented syllables).

5.3 Suprasegmentals (Intonation Analysis)

This section contains an in-depth analysis of the intonational ratings for the L2 stimuli. First, potential differences in the realisation of intonation are described, based on a cross-linguistic analysis of Arabic and English. This is followed by an analysis of the ratings obtained, with both a description of overall conclusions which can be drawn from the data, as well as a detailed examination of some of the individual cases. The ratings for each of the 36 stimuli are documented in Appendix D (Intonation Ratings and ToBI Analysis of the Stimuli).

5.3.1 Intonation: A Cross-linguistic Comparison

Intonation has three main functions in English (Wells, 2006) – which is likely to be the same for Arabic: expressing attitudes, structuring the message of an utterance, and focusing attention on a particular element of that message.

Crystal (1975, p11) says of intonation that “[it] is not a single system of contours and levels, but the product of the interaction of features from different prosodic systems – *tone*, *pitch-range*, *loudness*, *rhythmicality* and *tempo* in particular." There is as yet no theory of second language acquisition which adequately deals with prosody (though work is ongoing in this area, cf. So and Best (2010), with their extension to Best's PAM-L2 model, and Mennen (2015) with her L2 intonation learning theory 'LiLt'), but differences in language typologies relating to intonation and rhythm seem a good place to start (in Mennen's proposed model, these differences cover information on the inventory, distribution, structure, and frequency of the phonological elements of a language, as well as their phonetic implementation and intonational function).

Traditionally, intonation has been divided into three components - tone, tonicity and tonality. *Tone* is indicated by the shape of the pitch contours of an utterance, and can be used by the speaker to convey attitudinal factors. Typical ‘tunes' have been proposed for English for different types of utterance (Wells, 2006), but empirical studies of different dialects of English show that the inventory of observed pitch contours is extremely variable (Grabe et al, 2005). *Tonicity* involves the accenting of particular words in an utterance: certain words are made more prominent than others by the use of pitch accents (which signal the beginning of a pitch movement). It can be used to show the focus of the utterance. Pitch accents are marked by an increase in pitch, duration or volume, or any combination of these. *Tonality* involves the phrasing of the utterance: it divides speech into *intonation phrases.* These are sometimes subdivided into lower level phrases (see discussion below). The boundaries of these phrases often (but not always) mirror the grammatical structure of the utterance, and are marked in some languages by the use of boundary tones, and by phrase and domain-final syllable lengthening.

Samples from the stimuli were examined with respect to four areas of intonation; in the autosegmental-metrical model for intonation (Ladd, 2008), the two areas of prominence and phrasing are considered to be the cornerstones of intonation. Jun (2014) established a useful typology for classifying languages based on intonational differences, and in her typology, in addition to prominence and phrasing, she adds a third area which she calls macro-rhythm (or tonal rhythm), and which describes “the phrase-medial, global, tonal pattern of an utterance“ (2014, p521). The basic concept of this latter category – that tone also provides an utterance with a rhythm - will not be pursued further here, but the three resulting parameters which Jun uses to classify languages are useful as a practical framework with which to investigate these first three areas (tonality, tone, and tonicity), namely: i) the use (or lack) of tonal marking of heads and/or edges of prosodic phrases (for tonality), ii) the size of the inventory of pitch accents and boundary tones (for tone), and iii) the distribution of pitch accents (for tonicity). Rhythm in its conventional sense (which Jun refers to as micro-rhythm) is not traditionally included in intonational phonology (Jun, 2014) but is included here as a fourth area of examination, a) because it is often considered to be a component of intonation (cf. Crystal above), and b) it is quite probable that it plays a role as a suprasegmental aspect in the processing (and thus the comprehensibility) of L2 speech.

There are two methods for transcribing intonation – the 'British-school' style (cf. Cruttenden, 1997), which describes intonational patterns in terms of a series of pitch contours based on the nucleus of the utterance, such as a fall, a rise, and a fall-rise, and a method which was first developed for American English by Pierrehumbert (1980), and extended by Beckman and Pierrehumbert (1986), which describes intonational patterns in terms of a series of pitch accents (of which the nucleus in the British system is only one) and edge tones. The latter has been formalised in the ToBI system of transcription (Silverman et al., 1992), and it is this system which we shall use in the descriptions below. A useful conversion table between the two systems is given in Ladd (2008, p91).

5.3.1.1 Tonality

The first parameter of the language typology described above concerns the marking of heads and/or edges of phrases, and thus relates to tonality. Generally speaking, languages are either head-marking or edge-marking, but can be both (Jun, 2014). Edges can be marked in several ways, including domain-final syllable lengthening, and/or by boundary tones. Cruttenden (1997, p33) states that 'some of the reasons for final syllable lengthening may be language-specific, but it does seem nevertheless that the phenomenon itself may be an intonational universal'. Gussenhoven (2004) talks of the cumulative nature of lengthening processes in English: if a syllable with a pitch accent occurs at a prosodic boundary, both lengthening processes occur concurrently. Although no research has been published on Emirati Arabic, Hellmuth (2016) comments that most varieties of eastern Arabic are probably head-marking. Algethami (2013) notes that some similarities have been found in the use of lengthening in the prosodic head and at prosodic edges between Arabic and English, but lengthening is more common in English than in Arabic. No research has been published on lengthening in Emirati Arabic, but domain-final boundary lengthening has found to be common in some varieties of Arabic (Chalal and Hellmuth, 2014). In terms of boundary tones, although Blodgett et al. (2007) suggest that two of the four boundary tones in Emirati Arabic are complex (HL% LH%), Chalal & Hellmuth (2014) suggest that the use of complex boundary tones is very rare - though their research was based on the Egyptian and Syrian varieties of Arabic.

How a language is affected by the marking of phrasal heads and edges depends on the prosodic phrasing of a language, or language variety, in the sense that the more levels of prosodic phrasing are identified in the language, the more heads and edges there will be in the utterances of that language. Utterances are chunked in different ways in different languages. In English, it has been proposed (Beckman & Pierrehumbert, 1986) that there are three constituent levels: the prosodic word (PW), and two phrasal levels: the intermediate phrase (ip), and the intonational phrase (IP) (although the existence of the intermediate phrase has been disputed - cf. Gussenhoven (2004) – and Grabe (2001) does not use intermediate phrases in the IViE ('Intonational Variation in English') labelling system). In Arabic, prosodic phrasing appears to depend on the variety of Arabic in question. Research carried out by Chahal & Hellmuth (2014) shows that some dialects function in a manner similar to English in this respect (eg. Lebanese), whilst for others, the marking of prominence on the two lower levels is often conflated (eg. Egyptian). They propose replacing the intermediate phrase with two 'subdivisions' for Egyptian: the minor and major phonological phrases. Blodgett et al. (2007) go one step further and propose just one level of prosodic phrasing for Emirati Arabic, leaving out the intermediate phrase completely (although they do suggest that this conclusion may change with further research). Both SSBE (Grabe, 2000) and Emirati Arabic (Blodgett et al., 2007) employ compression when phrasal heads and edges occur on the same syllable - i.e. all the accents and tones are realised on the existing segmental material (as opposed to truncation, where the boundary tone is not fully realised (Grabe et al., 2000).

To summarise, in both languages, the head of a phrase can be marked by syllable lengthening, and the right-hand edge can be marked both by lengthening of the final syllable, and by boundary tones. However, when predicting the realisation of phrasing in the English utterances of the L2 speakers in this study, there are two points to make: firstly, because lengthening takes place to a greater extent in English than in Arabic, and secondly, because Arabic and English have been described differently in terms of their prosodic phrasing, we can expect to see differences when compared to the native speaker utterances if L1 transferal is taking place.

5.3.1.2 Tone

As mentioned above, this area of intonation can be correlated to the global tonal pattern of an utterance (Jun's second parameter), in the sense that it provides the speakers of a language with a range of possibilities with which to produce the tone of an utterance. Pierrehumbert (1980, cited in Ladd, 2008) established the following inventory of pitch accents and boundary tones for English (though some varieties of English do not contain all of them): seven pitch accents (H* (and its variant !H*), L*, L*-H, L-H*, H*-L, H-L* and H*-H, two phrase accents (H- and L-), and two boundary tones (H% and L%). Blodgett et al. (2007) propose the following as an inventory of Emirati Arabic: two pitch accents (H* (and its variant !H*), and LH*), and four bitonal phrase accents (LL%, LH%, HL% and HH%). As they propose only one level of prosodic phrasing for Emirati Arabic, the phrase accents in their proposed inventory are combined with the boundary tones – unlike in the English inventory, where phrase accents can occur at the end of an intermediate phrase, without a boundary tone. To summarise, the phrase accents and boundary tones are identical in the two inventories (though their use may potentially differ), but the English inventory of pitch accents is significantly larger than that of Emirati Arabic.

Intonational differences between languages are notoriously difficult to capture, not least because of the variation present within individual languages themselves – both between dialects but also within dialects and even for individual speakers (Grabe et al., 2005) – which causes great difficulties for the establishing of a 'norm'. Because of the variation in the inventories of observed pitch contours discussed above (Grabe et al., 2005), it is difficult to establish what exactly a non-native pitch contour – i.e. tone – might be. Although there may be pitch contours that have a non-native ring to a native listener, these have not yet been formalised sufficiently in acoustic terms to be used as a source of comparison in a scientific study. Because of this, instead of examining the shape of the pitch contours of the stimuli, the realisation of the contours in terms of pitch span and intervals in pitch height (between successive pitch accents, phrasal accents and boundary tones) were examined, in the search for non-native sounding jumps within the contours.

Patel et al. (2006) examined both pitch span and variability in pitch intervals in their study of British English and continental French intonation using a computational model called “prosogram“, developed by Mertens (2004). This model converts the F0 contour of an utterance into discrete tonal segments, stylising the signal into semitones, as in a musical score. Instead of using either the raw F0 contour, or the tone sequences produced by intonation transcription methods such as ToBI (cf. Beckman & Pierrehumbert, 1986), prosogram is instead based on the psychoacoustic principles, and is designed to capture intonation patterns as perceived by human listeners. In their experiment, Patel et al. found that the varieties of English and French used as input were similar in terms of the variability of pitch height, but differed when it came to variability in pitch interval size, with English showing more variability then French. Dilley’s study of intonational phonology (Dilley, 2005) suggests that pitch intervals may be important for the mental representation of intonation, and based on this work, Patel et al. comment that one possible reason for the differences between English and French is the suggested existence of three phonologically distinct pitch levels in the English intonation system, but only two in the French system.

No work has been published which examines Arabic in this context, but speech samples examined of Omani Arabic (which is close to Emirati Arabic) show F0 intervals for adjacent accented syllables which are regularly over 100-150 Hz. It was therefore predicted that similar jumps in pitch height, and resulting wide pitch spans, might occur in the English L2 speech samples.

5.3.1.3 Tonicity

The two languages in question differ quite considerably with respect to the third parameter: Wells (2006) says of tonicity that it is one of the most important functions of English intonation. In English (with the exception of certain dialects (Ladd, 2008)), a distinction is made between old and new information: given information tends to be deaccented (Ladd, 2008). Accent distribution tends to take place on a phrasal level (Hellmuth, 2007), with relative semantic weight being an important factor in its placement (Ladd, 2008), although there is some variation in pitch accent distribution both between and within English dialects (Grabe et al, 2005). Research into some varieties of Arabic, such as Hellmuth's work with Egyptian Arabic (Hellmuth, 2011), shows that accent distribution takes place at word-level, with every content word bearing an accent. On the basis of their preliminary study, Blodgett et al. (2007) suggest that this is also the case for Emirati Arabic. This fact, added to a general tendency found in L2 learners of different L1s not to deaccentuate non-prominent words (Mennen, 2015), led to the prediction that there would be a higher number of pitch accents in the English L2 speech of native Emirati Arabic speakers than in native speech.

Grabe et al. (2005) suggest that, although the way nuclear accent is marked can vary between dialects in native English, the location of the nuclear accent is less likely to vary, as this depends on focus structure. Research into Lebanese and Egyptian Arabic (Chahal & Hellmuth, 2014) shows that nuclear accent placement is generally on the final pitch accent of the utterance in broad focus utterances, as in English. Despite this concurrence between the two languages, the potential general accenting of content words in Emirati

Arabic (cf. Blodgett et al., 2007) as well as the learner effects mentioned by Mennen (2015), would have as a consequence that the nuclear accent is less prominent than in English, potentially making the utterance more difficult to process for the native listener.

5.3.1.4 Rhythm

White & Mattys (2007, p501) provide a useful description of what rhythm is: "Rhythm derives from the repetition of elements perceived as similar". In speech, these elements are (stressed) syllables. Although the distinction between syllable-timed and stress-timed languages is under debate, as well as the definition of what it actually means for a language to be stress-timed, both Arabic (McCarthy, 1979, Watson, 2002, cited in Chalal & Hellmuth, 2014) and English (Colantoni et al., 2015) have been described as stress-timed languages, which means that the intervals between stressed syllables are (perceived to be) roughly of equivalent length (Colantoni et al, 2015). Despite this similarity between Arabic and English, however, there are cross-linguistic differences in connected speech which lead to predictions regarding non-native pronunciation.

Turk et al (2013) give the following list of criteria which affect the rhythm of a language (which they call the global timing profile): contrastive phonological categories (and their frequency) such as tense and lax vowels, long and short vowels/consonants; phonotactic structure; word-prosody patterns (lexical stress and accent); phonetic correlates of phonological categories; language-specific performance principles, and speech rate/stylistic factors.

Hellmuth (2013) comments that the differences in both the syllabification and vowel reduction processes between Arabic dialects point to a difference in their rhythmicality: although Arabic has been classified as a stress-timed language, more westerly dialects are considered to be more stress-timed than easterly ones. This may affect the rhythmicality of L2 English spoken by native Emirati Arabic speakers, as their L1 is one of the less stress-timed varieties of Arabic. Vocalic syllables can contain either a full vowel or a reduced vowel. Whether a full or a reduced vowel is used is often considered to reflect the level of stress in the utterance, though some phonologists (Bolinger (1958), Halliday (1967), and Vanderslice & Ladefoged (1972), cited in Ladd, 2008) treat it as an independent phonetic or phonological phenomenon rather than a level of stress. In either interpretation, vowel reduction results from de-accentuation, and is one of the principal processes affecting rhythmicality in English. In Arabic, reduction in both the quality and duration of vowels in unstressed syllables does occur (Algethami, 2013), although there is more reduction in Western than in Eastern dialects (Ghazali et al., 2002). Algethami notes that even where vowel reduction does occur, it is not to the same extent as in English. Because of this (exacerbated by the fact that, according to Blodgett et al. (2007), in Emirati Arabic content words generally bear a pitch accent), it was predicted that de-accentuation, and thus vowel reduction, would not occur in the L2 speech samples as much as in the L1 speech samples. In native speech, these vowel reductions serve to provide the syllabic 'off-beats' of the utterance, and if these are less evident in the L2 utterances, as predicted, this might lead to a change in rhythm.

Re-syllabification is a dynamic process which often cuts across word boundaries. The Maximum Syllable Onset Principle (Selkirk, 1982) states that the onset of a syllable is maximised, in conformance with the principles of Basic Syllable Composition of a language. In other words, at junctures in speech strings where syllabification could occur at more than one place, the preference is for a heavier onset in the second syllable rather than a heavier coda in the first syllable. In English, an onset can be vocalic, or can contain up to three consonants within certain restrictions for consonant type. Codas are optional. This often leads to re-syllabification across word boundaries in English. In Arabic, all syllables require a consonantal onset, so when words beginning with a vowel occur mid-utterance, and follow a closed syllable, re-syllabification takes place. If a morphological sequence occurs which cannot be syllabified, either an epenthetic vowel is added (thus creating a new syllable) or consonant prosthesis occurs, or, in the case of closed syllables containing long vowels, these are shortened (Watson, 2002). Consonant prosthesis generally only occurs in utterance-initial position, where a minimal consonant is added – ie. a glottal stop. Following the rules for syllabification in Arabic, L1 transfer for native Arabic speakers of L2 English should theoretically not cause a problem for syllabification in English utterances – ie. where words in mid-utterance position begin with a vowel in English, and follow a closed syllable in the previous word, re-syllabification using the coda of the previous syllable occurs in both languages. It is thought, though, that less confident L2 learners might tend towards pronouncing each word individually, and might therefore show a preference for consonant prosthesis over re-syllabification. This non-native realisation would have an effect on the rhythmicality of the utterances as there would be more heavy syllables than in native speech.

Rhythm Metrics Rhythm calibrating metrics can be used as a way of measuring the rhythm of an utterance. There are a variety of measures available: White & Mattys (2007) found that those calculating %V (the proportion of the utterance which is comprised of vocalic intervals), VarcoV (the rate-normalised standard deviation of vocalic interval duration), and n-PVI-V (the measure of durational variability between successive pairs of vocalic intervals) were the most effective ones, both in terms of distinguishing between and within rhythm classes. Wiget et al. (2010) also used the CRI (Contrast Regularity Index), with the aim of measuring the regularity of sentence stress patterns, and found the results to correspond very closely to those of the nPVI-V measures. The first three metrics, then, were adopted for the present study, as recommended by Wiget et al.

Although, as mentioned above, both Arabic and English have been classified as stress-timed languages, the influence of lack of vowel reduction and incorrect syllabification may have an effect on the rhythm of the L2 English utterances. Dasher & Bolinger (1982) and Roach (1982) comment that the extent of vowel reduction in unstressed syllables renders stressed syllables more salient, thus causing durational differences in vocalic intervals. Prieto (2009, cited in Wiget et al., 2010, p1565) showed that the phonological structure of syllables affects rhythm.

5.3.2 Intonation Ratings and Analysis

The ratings on the scale of 1-9 (with 1 being 'native-like' and 9 being 'not at all native-like') for the low-pass filtered speech samples ranged from 1 to 8 for the 36 utterances, distributed for each rating as shown in Table 5.4 below. The ratings for the three utterances of each of eight of the twelve speakers were within 2 scalar units of each other, whilst those of the remaining four speakers ranged from 4 to 6 scalar units apart. The overall mean rating was 3.75. The mean rating for the six native speaker samples was 1.5, leaving a mean rating for the L2 speaker samples of 4.2.

INTONATION RATINGS	NUMBER OF OCCURRENCES
1 (native-like)	6
2	6
3	2
4	7
5	7
6	4
7	3
8	1
9 (non-native-like)	0

Table 5.4: Distribution of Intonation Ratings

In order to establish what factors may have played a role in these ratings, two examples from each utterance type (questions and statements) were taken which were similar in length and prosodic structure, but which had received very different intonational ratings.

Contrasting Statements

In the group of statements – which included grammatically simple as well as complex statements with embedded clauses, one utterance (#15) was given a rating of 8 (ie. very 'non-native-like') but had a similar prosodic (and grammatical) structure to another (#35) which received a rating of 1 (ie. 'native-like'). The only difference was that in the first two ips, the first utterance contained a low pitch accent on the final syllable, followed by a monotonal high boundary, whereas the second had no pitch accent on the final syllable, but instead a bi-tonal boundary.

The two utterances were:

It's in the middle north, er, in the first row, near to the City Hall (with a score of 8)

It's in the third row, in the third area, next to the State College (with a score of 1).

The pitch traces of each are shown in figures 5.7 and 5.8 below.

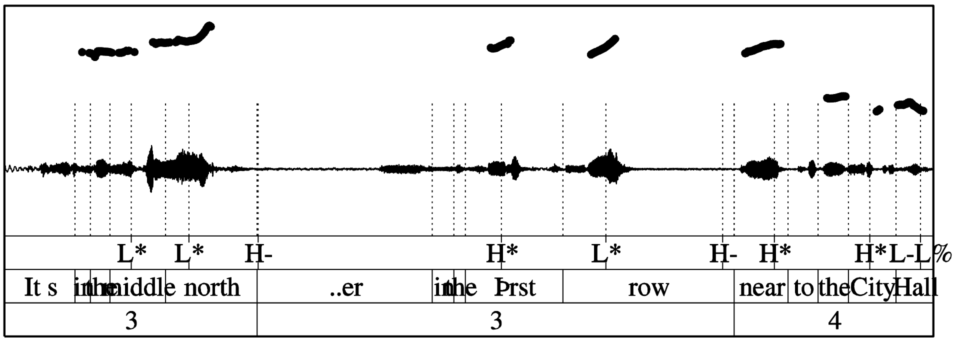

Figure 5.7: Pitch Trace for less Native-like Statement [Intonation Rating: 8]

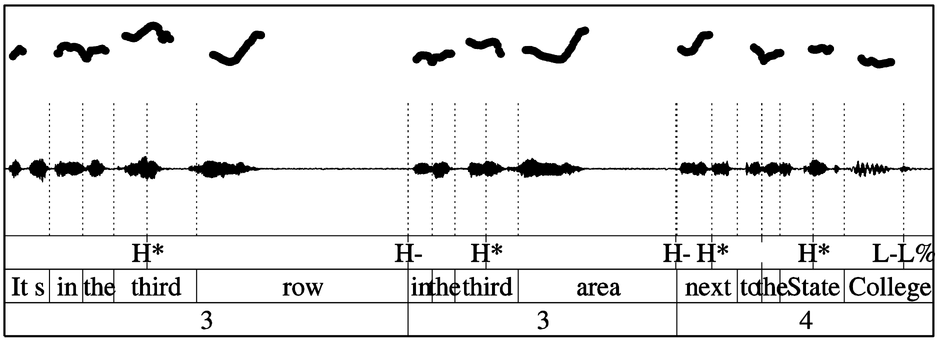

Figure 5.8: Pitch Trace for more Native-like Statement [Intonation Rating: 1]

The two utterances were examined in terms of tonality, tone, tonicity and rhythm, and the results are collated in Table 5.5 below, first for Utterance 1 (with a rating of 8), and then for Utterance 2 (with a rating of 1).

Utterance 1:

Pitch Span	Contour of Pitch Accents plus Boundary Tone (in Hz and semi-tone intervals)	Pitch Accent Distribution and Prominence	Mean intensity and mean pitch for whole utterance	Head/Edge lengthening	Rhythm Metrics
196-369 Hz (173)	1st ip: 240-285-369 (2 ST up – 5 ST up) 2nd ip: 261-286-329 (1 ST up – 3 ST up) 3rd ip: 265-243-196 (1 ST down – 4 ST down)	1st ip: 66 dB/240 Hz (*mi*-ddle) 73 dB/285 Hz (*north*) 2nd ip: 65 dB/261Hz (*first*) 71 dB/286 Hz (*row*) 3rd ip: 68 dB/265 Hz (*near*) 60 dB/243 Hz (*ci*-ty)	65 dB 213 Hz	*1st ip:* *north (head and edge)*: unlengthened *2nd ip:* *row (head and edge)*: slightly lengthened *3rd ip:* *ci-ty (head)*: unlengthened hall (edge): unlengthened	34 (%V) 65 (Varco V) 70 (nPVI)

Utterance 2:

Pitch Span	Contour of Pitch Accents plus Boundary Tone (in Hz and semi-tone intervals)	Pitch Accent Distribution and Prominence	Mean intensity and mean pitch for whole utterance	Head/Edge lengthening	Rhythm Metrics
202-363 Hz (161)	1st ip: 343-297 (3 ST down) 2nd ip: 278-307 (2 ST up) 3rd ip: 300-249-196 (3 ST down – 4 ST down)	1st ip: 67 dB/343 Hz (*third*) 2nd ip: 65 dB/278 Hz (*third*) 65 dB/307 Hz 3rd ip: (*next*) 65 dB/300 Hz (*state*) 67 dB/249 Hz	65 dB 249 Hz	*1st ip:* *third* (head): lengthened row (edge): lengthened *2nd ip:* *third* (head): lengthened *area* (edge): lengthened *3rd ip:* *state* (head): lengthened *college* (edge): lengthened	37 (%V) 47 (VarcoV) 53 (nPVI)

Table 5.5: Analysis of Tonality, Tone, Tonicity and Rhythm of Statement Samples

As can be seen in the table above, more head and edge lengthening takes place in the more native-like utterance. In terms of tone, both utterances followed the same basic pitch contour, and the pitch spans were quite similar. The pitch height intervals were also quite similar, though there was one jump of 5 semi-tones in the less native-like utterance. With regard to tonicity and pitch accent distribution, there were six pitch accents in the less native-like utterance, compared with only four in the more native-like utterance. The nuclear accent in both utterances was slightly higher in F0 than the mean of the respective utterances, but in the less native-like utterance it had a lower intensity than the mean of the utterance. The %V and VarcoV values are similar for both utterances; the metric that shows the most disparity in the two utterances is nPVI, which is probably the most relevant for the current study: because it compares the vocalic sections pair-wise, instead of comparing the vocalic and non-vocalic components proportionally over the whole utterance, the problem of the vocalic intervals 'evening themselves out' (which might occur if long vowels are shortened but there is no vowel reduction) does not apply as each interval is measured as a distinct unit. In fact, though, the nPVI value for the less native-like utterance was within the range of the results obtained in the experiments cited by Wiget et al. (2010, p1560: Ramus (2002), Grabe & Low (2002), White & Mattys (2007)) for SSBE English (63-73), whilst the more native-like utterance was outside this range.

Contrasting Questions

Of the questions, six utterances fell in the upper half of the ratings (with scores ranging from 1 to 4) – i.e. closer to native-like intonation - and six in the lower half (with scores ranging from 5 to 7) – i.e. closer to non-native-like intonation. They included both yes-no and wh-questions. The two utterances chosen for comparison which were similar in grammatical and prosodic structure but received very different ratings were number 26, which scored 7 (i.e. fairly non-native), and number 9, which scored 2.

The two utterances examined were:
Can you tell me please where is a mobile phone shop? (with a score of 7)
Can you tell me where is the supermarket please? (with a score of 2).

The pitch traces for each utterance are shown in figures 5.9 and 5.10 below.

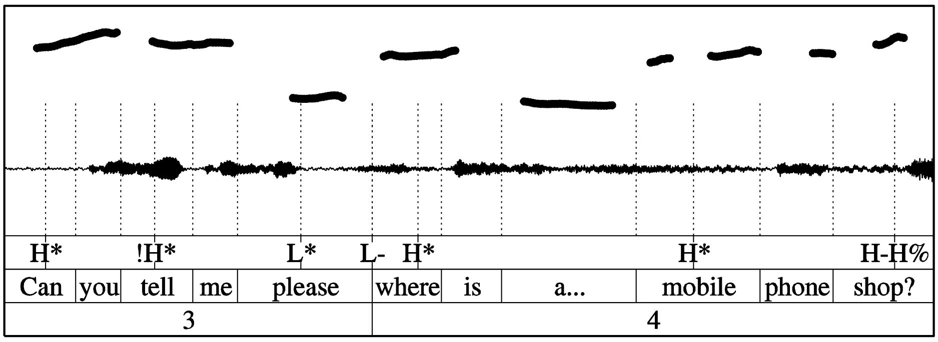

Figure 5.9: Pitch Trace for less Native-like Question [Intonation Rating: 7]

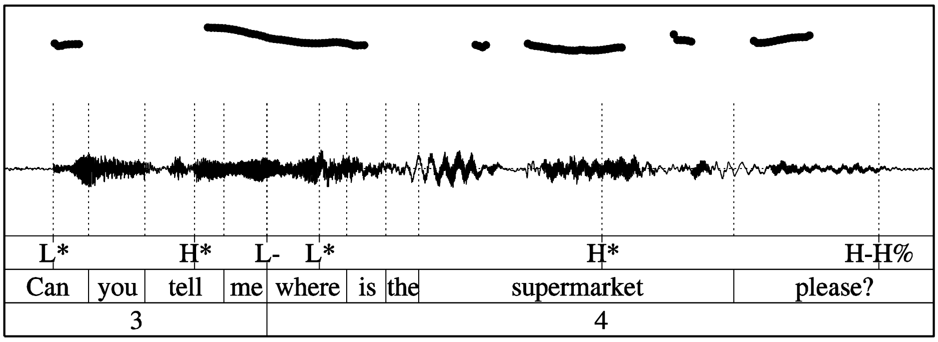

Figure 5.10: Pitch Trace for more Native-like Question [Intonation Rating: 2]

The two utterances were examined in terms of tonality, tone, tonicity and rhythm, and the results are collated in Table 5.6, first for Utterance 1 (with a rating of 7), and then for Utterance 2 (with a rating of 2).

Utterance 1:

Pitch Span	Contour of Pitch Accents plus Boundary Tone	Pitch Accent Distribution and Prominence	Mean intensity and mean pitch for whole utterance	Head/Edge lengthening	Rhythm metrics
99-327 Hz (228 Hz)	1st ip: 269-271-114 (-13ST down) 2nd ip: 223-234-288 (1ST up - 4ST up)	1st ip: 64 dB/ 269 Hz (*can*) 64 dB/271 Hz (*tell*) 2nd ip: 66 dB/223 Hz (*where*) 64 dB/234 Hz (mo-*bile*)	64 dB 205 Hz	*1st ip:* *tell* (head): unlengthened *please* (edge): lengthened 2nd ip: *mobile* (head): lengthened *shop (edge)*: unlengthened	40 (%V) 33 (VarcoV) 37 (nPVI)

Utterance 2:

Pitch Span	Contour of Pitch Accents plus Boundary Tone	Pitch Accent Distribution and Prominence	Mean intensity and mean pitch for whole utterance	Head/Edge lengthening	Rhythm metrics
183-360 Hz (177 Hz)	1st ip: 269-360-320 (5ST up 2ST down) 2nd ip: 274-247-293 (2ST down - 3ST up)	1st ip: 74 dB/269 Hz (*can*) 69 dB/ 360 Hz (*tell*) 2nd ip: 73 dB/274 Hz (*where*) 70 dB/247 Hz (super-*mar*-ket)	70 dB 272 Hz	*1st ip:* *tell* (head): unlengthened *me* (edge): unlengthened 2nd ip: super-*mar*-ket (head): lengthened *please* (edge): lengthened	29 (%V) 59 (VarcoV) 56 (nPVI)

Table 5.6: Analysis of Tonality, Tone, Tonicity and Rhythm of Question Samples

As can be seen in the table, more head and edge lengthening takes place on the nuclear accent and final boundary tone of the more native-like utterance (though not in the first ip). Both utterances followed the same basic pitch contour, but the pitch span of the less native-like utterance covered 50 Hz more than the native-like one, and also included one pitch height interval of 13 semi-tones – i.e. just over an octave, which is very rare in L1 English speech. With regard to tonicity, there was an equal distribution of pitch accents. The nuclear accent was in fact slightly more prominent in the less native-like utterance (they were both equal in intensity to the mean of the respective utterance, but the nuclear accent in the utterance with the less native-like rating had a higher FO than the mean of the utterance). In terms of rhythm, the less native-like utterance had a higher proportion of vocalic intervals, and much lower varcoV and nPVI values. Interestingly, the nPVI value for the less native-like utterance fell within the range calculated in the experiments described above by Wiget et al. (2010, p1560: Ramus (2002), Grabe & Low (2002), White & Mattys (2007)) for Castilian Spanish (30-42), which is classified as a syllable-timed language.

To sum up, utterances of the same utterance type which were similar in prosodic and grammatical structure but which received very diverse ratings were compared. The comparison shows that in the utterances which received a less native-like rating, there was overall slightly less lengthening at prosodic domain boundaries, slightly bigger jumps in pitch height between series of pitch accents, and between nuclear accents and boundary tones, a slightly higher number of pitch accents, and a rhythm which was sometimes slightly less typical of a stress-timed language. None of these phenomena were consistently found, however, even for the two utterances examined here which were judged to have non-native-like intonation. Here, though, it is important to note Mennen’s comment (2006, p1): the 'cumulative effect of continuously using slightly inappropriate intonation should not be underestimated'.

6. RESULTS

This chapter looks first at the distribution of the intelligibility scores, then the comprehensibility and accentedness ratings individually, in section 6.1. In section 6.2, the correlations between the ratings are examined. In section 6.3, the correlations between the stimulus measures and the ratings are described and interpreted, mirroring the original study.

In section 6.4, the data is re-analysed using more sophisticated statistical methods, and the new results are discussed and interpreted.

6.1 Distribution of Intelligibility Scores and Listener Ratings

Intelligibility Scores

Transcription scores were based on the percentage of words which matched the author's own transcriptions exactly. The latter were accepted as the 'correct' transcriptions based on familiarity with the recordings and the contexts in which the samples were collected. Instances in the speech samples for which any doubt existed were verified by a third party.

Transcription errors were counted for the following categories: omission, substitution, insertion and regularisation. In keeping with the criteria of the original study, omission was sub-divided into the omission of content words and omission of function words; substitution was where a word was replaced by a phonetically or semantically similar word (for example *route* instead of *road*); insertion was where a novel word was used which bore no phonological resemblance to any word in the utterance (there were no examples of this in the data); and regularisation was where the listener changed a word to make it grammatically correct (for example the pronoun *he* became *it* when referring to a place).

Overall, the speech samples were found to be extremely highly intelligible: a total of 73 errors were found in the 540 orthographic transcriptions, i.e. 98.8% of all L2 words were transcribed correctly. (There were a total of 86 errors in 648 transcriptions if the L1 utterances are included). The percentage of correct transcriptions ranged from 78%-100% for the L2 speakers, and 89%-100% for the L1 speakers. 13 of the 30 L2 utterances and 4 of the 6 L1 utterances were 100% intelligible to all listeners. For each of the utterances containing transcription errors a mean rate was calculated over all listeners, and these ranged between 93% and 99.5%.

The results for the error count distribution can be seen in Table 6.1 below.

	L2 speakers #	L2 speakers %	L1 speakers #	L1 speakers %
Omission (function word)	2	2.7	-	-
Omission (content word)	14	19	-	-
Novel Word	-	-	-	-
Substitution	55	75	13	100
Regularisation	2	2.7	-	-
Total	73		13	
Errors per Speaker	7.3		6.5	

Table 6.1: Transcription Error Distribution

As the table above shows, the vast majority of errors were substitution errors. The mean number of errors per speaker was not much different between the L2 and the L1 speakers (7.3 vs 6.5).

These results differ significantly from those of the Derwing/Munro study, where a total of 636 errors in the 540 L2 utterances were found (680 errors in the 648 transcriptions if the L1 utterances are included) – thus 89.1% of words were transcribed correctly (as opposed to 98.8%). The mean utterance length was, however, almost the same (10.8 words as compared with 10.7 of the original study), so the total word count was very similar in both studies. Figure 6.1 shows a comparison of the distribution of the intelligibility scores for L2 utterances in the present study (left) and the original (right):

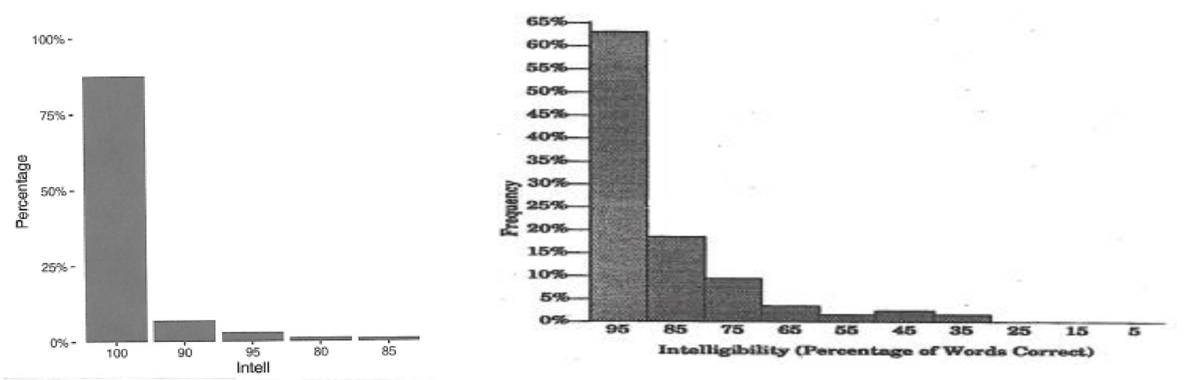

Figure 6.1: Comparison of Distribution of Intelligibility Scores

Comprehensibility Ratings

The mean of the comprehensibility ratings of the current study, based on a Likert scale of 1-9, with 1 being 'easy to understand' and 9 being 'impossible to understand', was 3.4 for all utterances. The distribution of the ratings for utterances which were 100% intelligible to all listeners (13 utterances, as opposed to 5 in the original) is shown on the left in Figure 6.2, together with that of the original study (shown on the right).

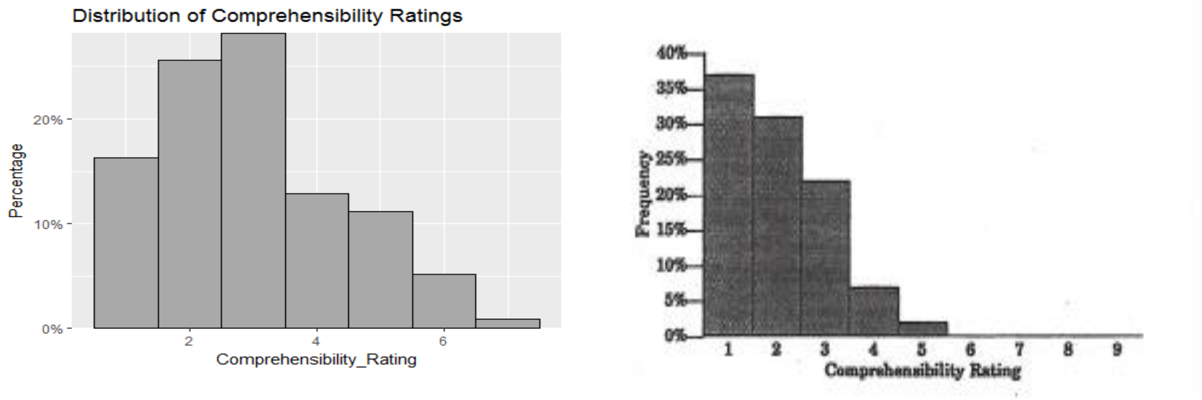

Fig. 6.2: Comparison of Distribution of Comprehensibility Ratings

The ratings of the current study showed a mean of 2.9. Where most utterances received a 1 in the original study, the most frequent rating in the current study was 3. This may have been affected by the fact that the comprehensibility rating task was performed just after the accentedness rating task in the same session, thus the listeners may have been affected by the accent ratings.

Accentedness Ratings

The mean of the accentedness ratings of the current study, based on a Likert scale of 1-9, with 1 being 'no foreign accent' and 9 being a 'very strong foreign accent', was 5.4 for all utterances.

The distribution of the accentedness ratings for utterances which were 100% intelligible to all listeners, is shown below (left), together with that of the original study (right).

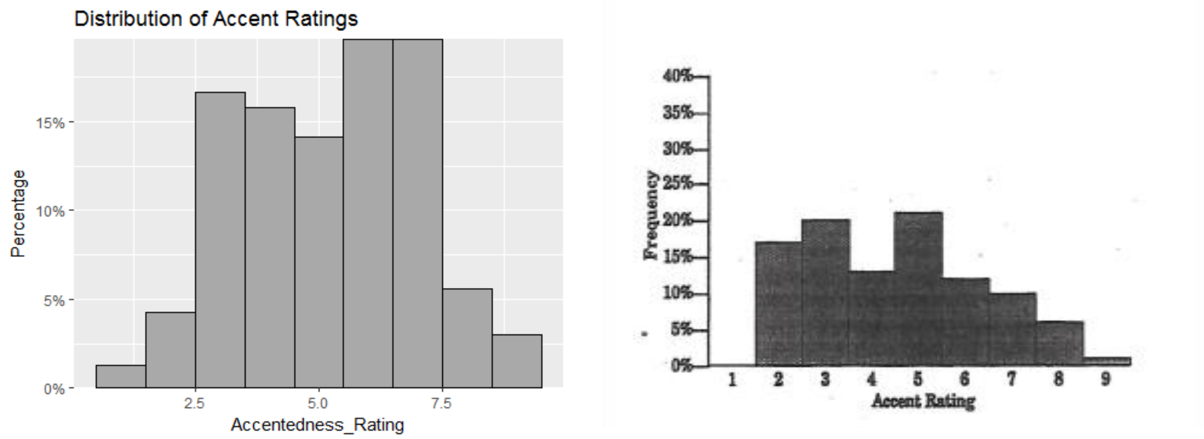

Figure 6.3: Comparison of Distribution of Accentedness Ratings

The accentedness ratings of the current study showed a mean of 5.2 for these 13 utterances. This is against a backdrop for 5.5 for those which were not 100% intelligible, thus there is not much difference between the accent ratings of those utterances which were intelligible, and those which were not, suggesting orthogonality, as in the original study.

6.2 Correlations between Accentedness, Comprehensibility, and Intelligibility

Below are the plots of the correlations of the three speech ratings for all L2 utterances for all listener ratings. These plots are produced as jitter plots in R, which are different to scatter plots in that the positions of the data points are separated out slightly in order to better show the density of each area.

Accentedness and Intelligibility

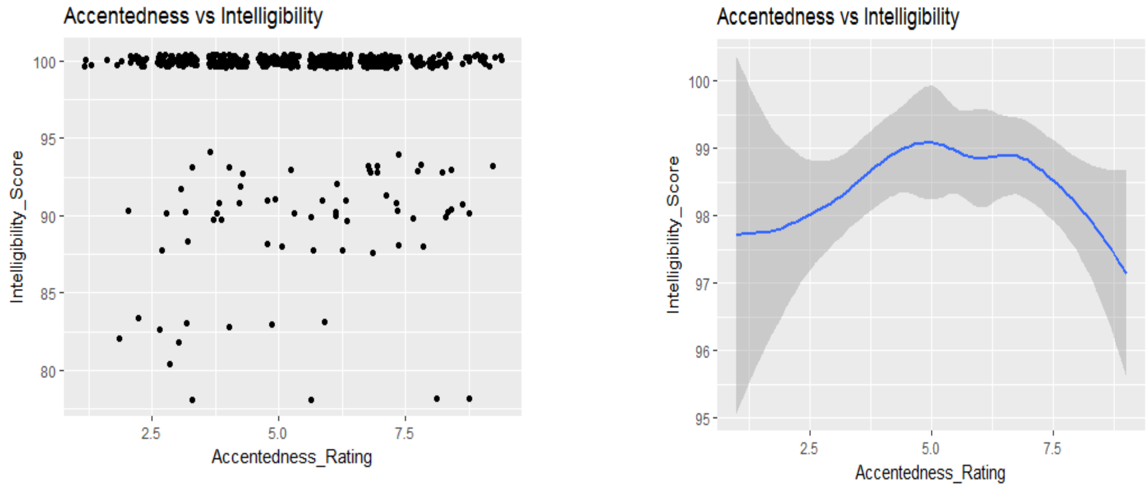

Figure 6.4 Correlation between Accentedness and Intelligibility

The results show that there is no significant correlation between accentedness and intelligibility: Pearson's r =.007; p =.853. This suggests that accentedness and intelligibility are orthogonal, as the first research question asked.

Comprehensibility and intelligibility

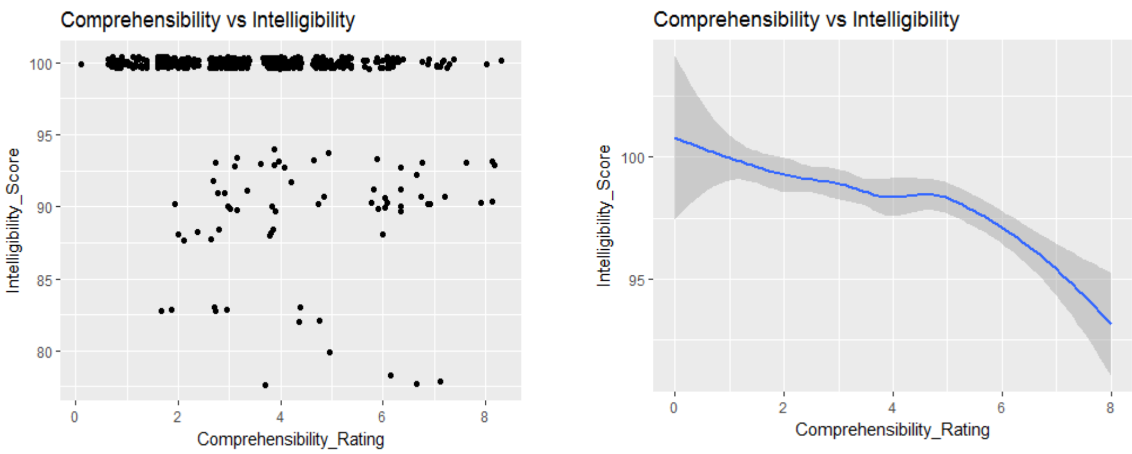

Figure 6.5: Correlation between Comprehensibility and Intelligibility

The results show that there is a significant negative correlation between comprehensibility and intelligibility: Pearson‘s r =-.257; p <.0001. This suggests that comprehensibility and intelligibility are not orthogonal.

Accentedness and Comprehensibility

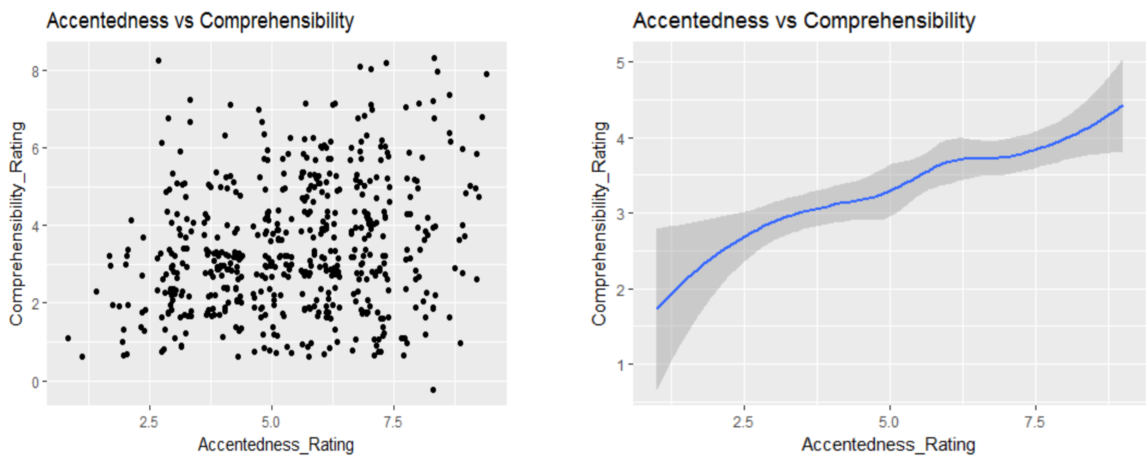

Figure 6.6: Correlation between Accentedness and Comprehensibility

The results show that there is a significant correlation between accentedness and comprehensibility: Pearson‘s r =.248; p <.0001. This suggests that accentedness and comprehensibility are not orthogonal.

6.3 Correlations between Accentedness, Comprehensibility, Intelligibility, and the Stimulus Measures

Pair-wise Correlations

Before running Pearson Correlations between the whole set of data for each stimulus measure and the whole set of data for each of accentedness, comprehensibility and intelligibility, intraclass correlations (cf. Shrout et al, 1979) were carried out in order to check inter-rater reliability for the ratings. It was found that although both the accentedness and comprehensibility ratings showed a high level of inter-rater reliability, there were significant differences in the strength of correlations *between* the accentedness and comprehensibility ratings of individual listeners (with Pearson's r values ranging from .03 to .68), as was the case in the original study (which showed a range of r=.41 to r=.82). Pearson correlation coefficients were thus calculated between all ratings, and the results are shown for both studies in the table below. Although the results for the overall correlations in the original study mirror to a large extent the individual ones shown in the table below, for the current study, this shows a different story from the correlations calculated for the across-listener ratings. Table 6.2 below gives a summary of the number of individual listeners for whom there was a significant correlation.

	Current Study	Original Study
Accentedness-Comprehensibility	10/18	17/18
Comprehensibility-Intelligibility	6/18	15/18
Accentedness-Intelligibility	1/18	5/18

Table 6.2 Number of Listeners showing Significant Correlation between Ratings

When compared to the findings of the original study, the current study also suggests the orthogonality of accentedness and intelligibility, but the correlations between both accentedness and comprehensibility, and comprehensibility and intelligibility, are much weaker than in the original study.

Pair-wise Correlations between Speech Ratings and Linguistic Analyses

Likewise, Pearson correlation coefficients were also calculated for the ratings and the stimulus measures for each individual listener. These are shown in Table 6.3 below for the current study, followed by Table 6.4 from the original study for comparison (the latter table was already shown in Chapter 3).

Stimulus Measure	Perceived Comprehensibility #	%	Perceived Accent #	%	Intelligibility (Words Correct) #	%
Phonemic Errors	1	5.5%	1	5.5%	2	11%
Phonetic Errors	1	5.5%	10	55.5%	0	0%
Intonation	0	0%	0	0%	1	5.5%
Grammatical Errors	1	5.5%	1	5.5%	1	5.5%
Utterance Length	0	0%	1	5.5%	0	0%

Table 6.3: Correlations Table for Ratings/Stimulus Measures of Current Study

Number of Significant Correlations Between Perceived Comprehensibility, Accent, and Intelligibility Scores and a Stimulus Measure (p<.05)

Stimulus Measure	Perceived Comprehensibility No.	%	Accent No.	%	Intelligibility (Words Correct) No.	%
Phonemic Errors	8	44	14	78	5	28
Phonetic Errors	2	11	13	72	0	0
Intonation	15	83	16	89	4	22
Grammatical Errors	10	56	14	78	3	17
Utterance Length	0	0	0	0	0	0

Table 6.4: Correlations Table for Ratings/Stimulus Measures of Original Study

6.4 A Comparison of the Results of the Original and the Replication Study

As shown Table 6.3 above, there were very few listeners for whom there was a significant correlation between the speech ratings and any of the stimulus measures in the current study, with the exception of the correlation between the number of phonetic errors and the accentedness rating, for which just over half the listeners showed a significant correlation. In contrast, in addition to the correlation between the phonetic error count and the accentedness ratings, the original study also showed a significant correlation for at least 70% of the listeners (14 or above of the 18 listeners) between accentedness and the number of both phonemic and grammatical errors, as well as the intonation rating – i.e. accentedness is correlated to all the linguistic measures other than utterance length. It also demonstrated a correlation for 83% of the listeners between intonation and comprehensibility (the second research question of this dissertation).

The major difference between the two studies is the surprising lack of correlations in the present study, either between accentedness, comprehensibility and intelligibility, or between any of the ratings and the linguistic measures examined in the stimuli (other than accentedness and the number of phonetic errors). A new statistical approach was adopted to explore the data further.

6.5 Modelling Interactions of Linguistic Factors

As shown in the last section, according to the results of the Pearson correlations run in the experiment, the concepts of comprehensibility and intelligibility (what makes L2 speech difficult, or impossible, to understand) have not been 'explained' by any of the linguistic factors which were examined – in other words, none of the linguistic error counts (or ratings) were significantly correlated to the comprehensibility rating or the intelligibility score for more than a minority of the listeners.

One important point to make at this stage is that this was also the case for intelligibility in the original study: the results from the Pearson correlations showed which linguistic measures were *not* significantly correlated to intelligibility, but not which *were*: the highest number of listeners who showed a significant correlation between intelligibility and any of the linguistic measures examined was for phonemic errors, which showed a correlation with intelligibility for just 28% of the listeners in the original study, and 11% in the present study.

This lack of explanatory factors amongst the linguistic measures examined for both the comprehensibility and intelligibility of L2 speech in the current study, and intelligibility of L2 speech in the original study, indicates that there is a missing link. To look for this missing link, it is time to take a closer look at the utterances that were not 100% intelligible to all listeners.

Lexis

It was mentioned in Chapter 2 that lexis plays a role in the comprehensibility (and intelligibility) of L2 speech (Saito et al., 2015, 2016). On examination of the utterances in this study which were less than 100% intelligible, it was in fact found that 56 of the 73 L2 transcription errors involved words that were either proper nouns (and therefore generally devoid of meaning and hence much less predictable from context than other words) or were unusual names, or were words classified (by the author) as inappropriate lexical choices in the given context (such as the word *row* in the context of giving directions).

Lexis was therefore included as a new linguistic measure. Lexical items of the types described above were marked as ‘unpredictable’ in the data, and an ‘unpredictability score’ was calculated based on the number of syllables in the word, relating to the likelihood of the existence of a minimal pair or similar-sounding word, which might lead to a transcription error. One-syllable words were given an unpredictability score of 1, and for every extra syllable, the score was halved, so two-syllable words scored 0.5, and three-syllable words scored 0.25.

Stress

Another new factor which was included as a linguistic measure was *stress*, as this error type had not been accounted for either in the phonetic/phonemic error counts, or in the intonation rating, as stress errors are not obvious when using low-pass filtering. The *stress* error count includes both lexical and compound stress errors. An example of a lexical stress error found in the data was the accenting of *mobile* [məʊ’baɪl] in the phrase *mobile phone shop* on the second syllable, whilst an example of compound stress was supermarket: [su:pə’mɑ:kɪt].

When the utterances which scored below 100% intelligibility were examined in detail, it was observed that cases of lexical unpredictability did not consistently cause a problem for intelligibility: closer scrutiny of the data shows a clear relationship between the intelligibility of these lexical items and pronunciation errors of the phonetic/phonemic/stress type. For example, the word *row* (which was classified as an inappropriate lexical item in this context) was used in 4 of the 30 L2 utterances. The instance of *row* with the most transcription errors (6) was the only instance which contained a phonemic error.

To investigate this further, a more complex analysis is required than that which is offered by calculating the effect of isolated phenomena. Because of the very low numbers of listeners for whom there was a significant Pearson correlation between the isolated linguistic phenomena and both the comprehensibility or intelligibility of the L2 speech samples, mixed models were designed to examine how possible contributing factors work in combination with each other. The two measures which were unconnected to pronunciation were lexical predictability, and morphosyntactic errors. Models were developed to examine whether, in combination with measures concerning pronunciation (phonetic and phonemic errors, errors of stress, and non-native intonation – henceforth referred to as pronunciation factors), a significant result was obtained.

6.5.1 Effects of Interactive Linguistic Factors on Comprehensibility

Linear mixed models were designed to investigate the effects of pronunciation errors (*phonetic*, *phonemic intonation*, *stress*) in combination with lexical and morphosyntactic unpredictability/errors (*lexis* and *grammar*) on *comprehensibility* with random effects of *listener* and *speaker* by integrating interactions between the fixed factors (ie between lexical and morphosyntactic unpredictability/errors, and the different types of pronunciation error). The workings of the models are shown in Appendix F.

The best-fit model shows a main effect of lexis and grammar, together with an interaction of lexis*phonetic and lexis*phonemic, as well as an interaction of grammar*phonemic, grammar*inton, and grammar*stress ($\chi 2=26.327;df=3;p=8.147e\text{-}06^{***}$). Figure 6.7 shows the strength of the interactions.

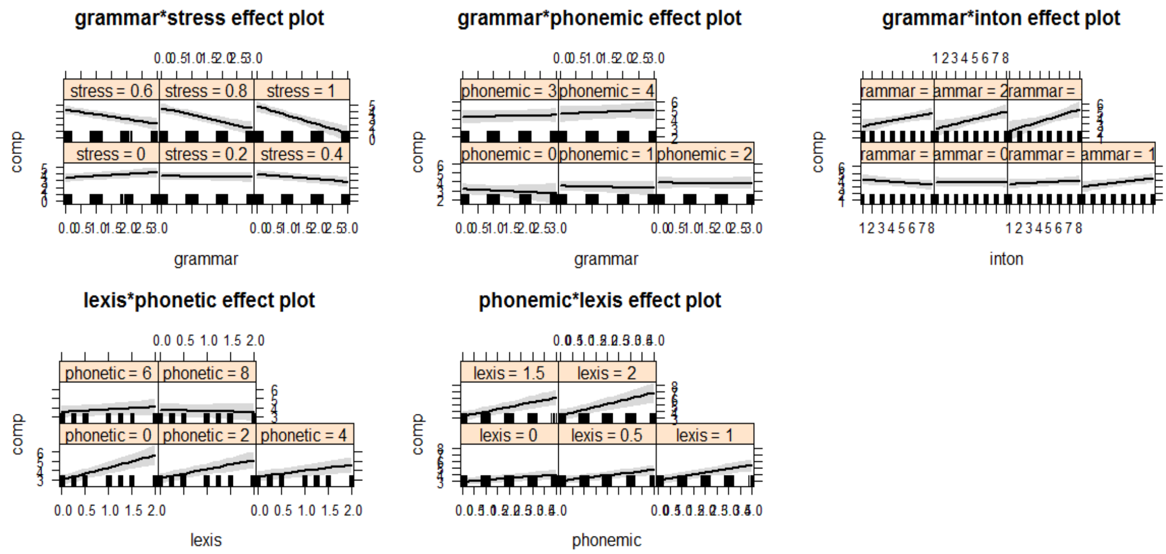

Figure 6.7: Effects of Grammar and Lexis on Comprehensibility, with Interactions with Pronunciation Factors

In other words, when interactions are allowed in the data between the pronunciation factors and the lexical and morphosyntactic properties of L2 speech described above, it is shown that pronunciation factors do have a significant effect on comprehensibility: phonetic and phonemic errors have a significant effect on comprehensibility when they occur with unpredictable lexical items, and phonemic errors, non-native-like intonation and errors of stress have a significant effect on comprehensibility when they occur with morphosyntactic errors.

6.5.2 Effects of Interactive Linguistic Factors on Intelligibility

Linear mixed models were designed to investigate the effects of pronunciation errors (*phonetic*, *phonemic intonation*, *stress*) in combination with lexical and morphosyntactic unpredictability (*lexis* and *grammar*) on *intelligibility* with random effects of *listener* and *speaker* by integrating interactions between the fixed factors (ie between lexical and morphosyntactic unpredictability/errors, and the different types of pronunciation error). The workings of the models are shown in Appendix E.

The best-fit model shows a main effect of lexis and grammar, together with an interaction of lexis*phonetic, lexis*phonemic and lexis*inton, as well as an interaction of grammar*phonemic, grammar*inton, and grammar*stress ($\chi 2$=19.369;df=4;p=.0006648***).

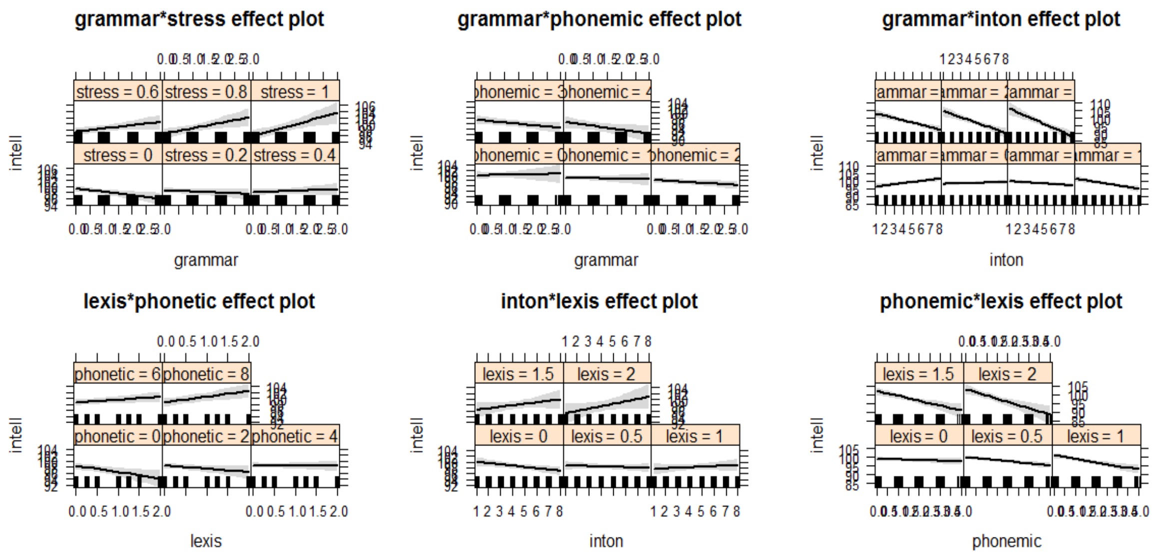

Figure 6.8: Effects of Grammar and Lexis on Intelligibility, with Interactions with Pronunciation Factors

In other words, when interactions are allowed in the data between the pronunciation factors and the lexical and morphosyntactic properties of L2 speech described above, it is shown that pronunciation factors do have a significant effect on intelligibility: phonetic and phonemic errors, and non-native-like intonation have a significant effect on intelligibility when they occur with unpredictable lexical items, and phonemic errors, non-native-like intonation and errors of stress have a significant effect on intelligibility when they occur with morphosyntactic errors.

In the figures above, however, the interactions for *comprehensibility* (fig. 6.7) between *grammar* and *stress*, and the interactions for *intelligibility* (fig. 6.8) between *grammar* and *stress*, and between *lexis*, and *phonetic* and *intonation* appear to show a negative effect (the reader is reminded that the scores for intelligibility and comprehensibility are inverse: a higher intelligibility score means that an utterance is more intelligible, whereas a higher comprehensibility rating means that an utterance is less comprehensible). There is no guarantee in a mixed model analysis of any cause-and-effect relationship between variables, and it is highly improbable that, for example, the effect on comprehensibility of morphosyntactic errors is reduced *because* they occur in conjunction with errors of stress. It is therefore likely that these negative interactive effects are simple cases of co-occurrence.

These negative interactive effects show a general weakness in the statistical analysis of the data, as some of the errors examined are local errors – ie. local to a particular part of the utterance, such as lexis, phonemic and phonetic errors, and errors of stress – and others may be global or local, such as morphosyntactic errors and the intonation rating. The statistical analysis carried out is, however, at an utterance-global level regardless of error type. A more sophisticated statistical analysis would be needed to take this into consideration. The listener, for example, is faced with a different situation if a phonemic error occurs in an unpredictable lexical item than if a phonemic error and an unpredictable lexical item occur in the same utterance but in differing locations.

These two best-fit models for comprehensibility and intelligibility suggest (even if the negative interactive effects are manually removed) that one type of information deficit on its own (presumably within a certain limit) does not have a significant effect on the intelligibility of an utterance. However, in combination with other forms of information deficit (in this case, certain pronunciation factors), both the comprehensibility and the intelligibility of L2 speech will be reduced.

6.5.3 Re-examining the effect of Accentedness on Intelligibility and Comprehensibility

Intelligibility

As shown in Table 6.2, there were very few listeners in both the original and the current study for whom there was a significant correlation between accentedness and intelligibility, leading to the idea that accentedness and intelligibility are orthogonal.

Linear mixed models were designed to investigate the effects of *accentedness* in combination with lexical and morphosyntactic unpredictability/errors (*lexis* and *grammar*) on *intelligibility* with random effects of *listener* and *speaker* by integrating interactions between the fixed factors (ie between lexical and morphosyntactic unpredictability/errors, and accentedness). The workings of the models are shown in Appendix E.

The best-fit model shows a main effect of lexis and grammar, together with an interaction of grammar*accent ($\chi 2=6.5644; df=1; p=.0104^*$).

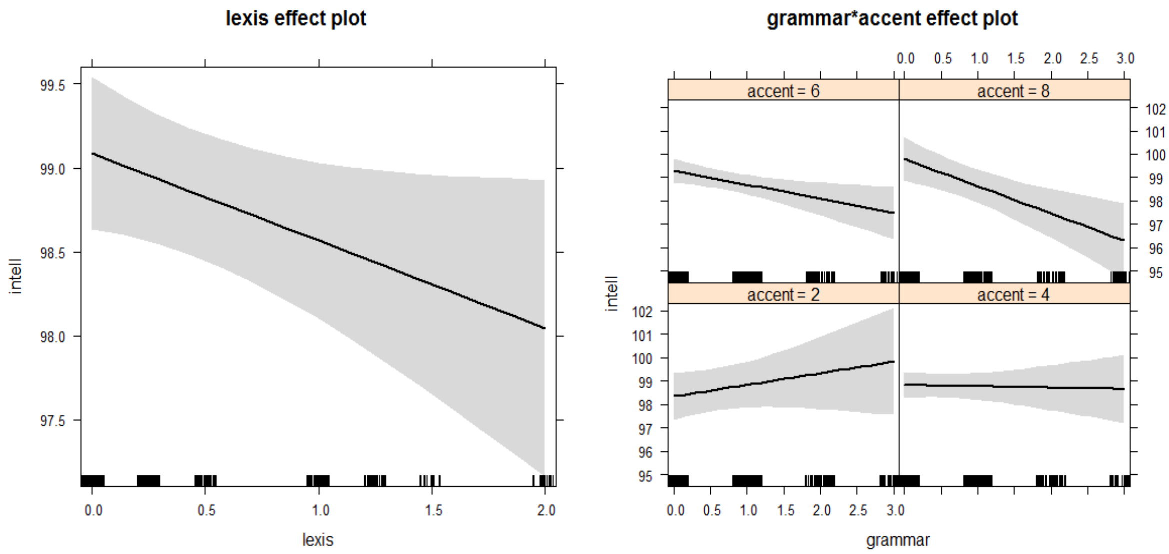

Figure 6.9: Effects of Grammar and Lexis on Intelligibility, with Interactions with Accentedness

Comprehensibility

Linear mixed models were designed to investigate the effects of *accentedness* in combination with lexical and morphosyntactic unpredictability/errors (*lexis* and *grammar*) on *comprehensibility* with random effects of *listener* and *speaker* by integrating interactions between the fixed factors (ie between lexical and morphosyntactic unpredictability/errors, and accentedness). The workings of the models are shown in Appendix F.

The best-fit model shows a main effect of lexis and grammar, together with an interaction of grammar*accent (χ2=.12.445;df=1;p=.0004191***).

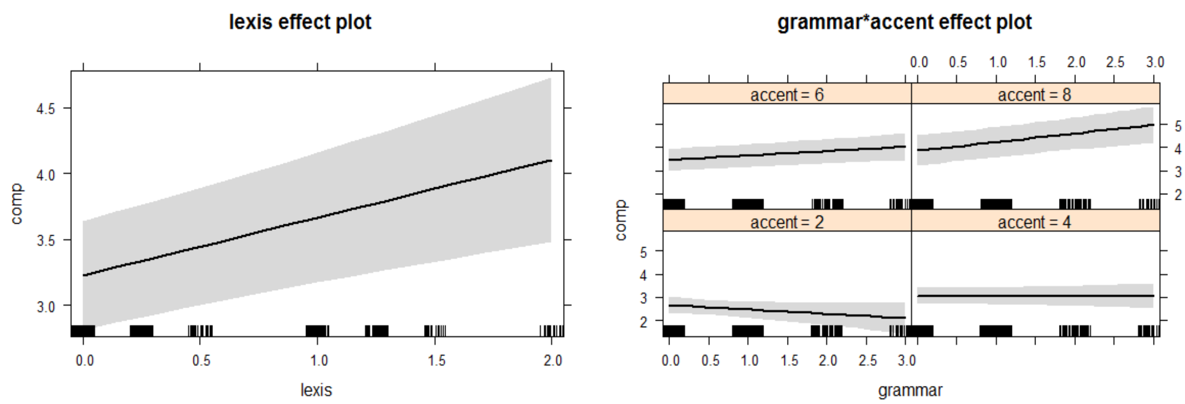

Figure 6.10: Effects of Grammar and Lexis on Comprehensibility, with Interactions with Accentedness

6.5.4 Comparing Accentedness as a Listener Property and Pronunciation as a Speech Property

Analysis showed that the quantifiable pronunciation factors are a better predictor of intelligibility than the accentedness ratings are: The best-fit model showed a main effect of lexis and grammar, together with an interaction of lexis*phonetic, lexis*phonemic and lexis*inton, as well as an interaction of grammar*phonemic, grammar*inton, and grammar*stress ($\chi2=69.636$;df=13;p=9.367e-10***). The full model comparisons are shown in Appendix E.

(Even if the negative interactive effects are manually removed from the best-fit 'pronunciation' models for comprehensibility and intelligibility, on the grounds that they are non-sensical, the new 'manual' best-fit pronunciation models still show that pronunciation factors are a better predictor for both comprehensibility and intelligibility than the accentedness ratings are. The 'manual best-fit' pronunciation model for comprehensibility shows a significant effect of grammar and lexis with interactions of grammar*phonemic, grammar*inton, lexis*phonetic and lexis*phonemic ($\chi2=1276.7$; df=11; p=< 2.2e-16***) - i.e. without the negative grammar*stress interaction. The 'manual best-fit' pronunciation model for intelligibility shows a significant effect of grammar, lexis and phonetic, with interactions of grammar*phonemic, grammar*inton, and lexis*phonemic ($\chi2=26.412$;df=10;p=.003224**) – i.e. without the negative interactions of grammar*stress, lexis*phonetic, and lexis*inton, phonetic has a main effect.)

7. DISCUSSION

The two research questions asked at the beginning of this study were:

1. Are the accentedness and intelligibility of L2 speech orthogonal?
2. Do the suprasegmental aspects of L2 speech have a bigger impact on the comprehensibility of L2 speech than do segmental issues?

To return to the first research question, it was shown that accentedness does in fact affect the intelligibility of L2 speech in certain circumstances – here, in the presence of grammatical errors. In other words, accentedness and intelligibility are not orthogonal.

This suggests that there is a minimum amount of information which a listener needs in order to understand an utterance. This information can come from different sources. We already know that listeners use contextual, lexical, syntactic information when identifying words in the speech string (see studies by Bard, Shillcock & Altmann (1988), and Cutler & Clifton (1999), cited in Zielinski, 2006, p7). Phonetic and phonological information is another source available to the listener. If the information threshold required for the listener to process an utterance successfully is not reached, the utterance will become unintelligible. In the current study it was demonstrated that when morphosyntactic errors occur, other sources of information are needed to make up the deficit: it would seem that foreign accentedness deprives the listener of the necessary information to make up this deficit.

This is not just an issue in L2 speech: the coarticulatory freedom native speakers allow themselves is much reduced in contexts where little other information is available. By way of example, in the L1 utterance *Is there a supermarket around here anywhere?*, the word *there* was transcribed perfectly by all listeners, despite it being almost absent in phonetic terms: the /z/ of the preceding word (*is)* was slightly dentalised, and the vowel which followed it (in the word *a*) was slightly r-coloured). In other words, a broad transcription would give /ɪz ə su:pəmɑ:kɪt…./. It is unlikely that this realisation would have been recognised as the word *there* outside of the context. This suggests that phonetic and phonological information is a contributing factor in a combination of factors which lead to the successful processing of the speech signal.

At this point, it is helpful to return to the definition of accentedness, and the usefulness of accentedness ratings. There have been a number of studies which show the effects of the listener's experience, and acknowledge diverse influences on accentedness ratings. Beinhoff (2014) found that L2 proficiency levels as well as L1 background significantly influence accentedness ratings. Huang et al. (2016) said that raters self-reported that their accent familiarity affected their evaluations of accentedness, and might have made them more lenient toward speakers with familiar accents (although systematic differences were not found in the study). In a study on the effects on accentedness ratings of lexical properties, Porretta et al. (2016, p2449) comment that: "an L2 talker's ability to approximate typical native speaker values on acoustic measures is only part of what affects the strength of perceived foreign accentedness".

Whether accentedness is defined as a property of speech – contained, in other words, within quantifiable errors of pronunciation (for example those examined in this study, and were referred to as *pronunciation* factors) - or as a property of the listener, pertaining to the *perception* of accentedness, and measured by the impressionistic accentedness ratings collected, this study has shown that when interacting with other linguistic components of speech, neither accentedness as a listener property nor pronunciation as a speech property are orthogonal to intelligibility. One more important point should be made here: the perception of accentedness by the listener is not necessarily the most relevant to the L2 learner who has the goal of being intelligible, as quantifiable errors of pronunciation (in the form of the segmental and suprasegmental error counts) were shown to predict intelligibility better than the perceived accentedness of the accent rating.

Returning to the second research question: on the basis of the Pearson correlations, very few of the listeners showed a significant correlation between either segmental or suprasegmental factors and comprehensibility. When these factors were examined in interaction with other aspects of L2 speech – morphosyntactic errors and lexical unpredictability - however, the results show that both segmental and suprasegmental aspects of L2 speech are predictors of comprehensibility. In interaction with morphosyntactic errors, intonation had a significant effect on comprehensibility. In other words, an utterance which was grammatically incorrect and had non-native-like intonation was harder to understand than one which was just grammatically incorrect. In this sense, the results of the two studies are similar. In the original study, however, intonation played a far greater role than segmental factors – there was only a significant correlation for 44% and 11% of the listeners between comprehensibility, and phonemic and phonetic errors respectively, compared with a correlation for 83% of the listeners with intonation. In contrast, the present study showed a significant effect of phonemic errors in interaction with both grammar and lexis, and a significant effect of phonetic errors in interaction with lexis.

It is possible that the difference in the L1s of the L2 English speaker groups of the two studies may have had a bearing on their intonation in English, and thus the gravity of the intonational errors, with the consequence that intonation did not affect comprehensibility any more than the segmental errors did. According to Jun's typology, English is classified as a head-prominent stress language with a medium tonal rhythm). Emirati Arabic is not classified, but of the varieties of Arabic which have been classified, Lebanese Arabic is accorded the same classification as English, and Egyptian Arabic is also classified as a head-prominent stress language but with a strong tonal rhythm (It is, however, worth noting that Jun bases her head/edge-marking classifications on what occurs in low-level phrases in the prosodic hierarchy (for example Accentual Phrases)). Mandarin is also classified as a head-prominent language, but with a weak tonal rhythm.

In order to judge the severity of the intonational errors encountered in the L2 samples with regards to their effects on comprehensibility, the filtered stimuli were listened to again, with the aim of trying to identify the utterance type (for example, whether the utterance in question was a question or a statement) and internal prosodic structure. The idea behind this test was that if an utterance can be parsed from its intonation alone (on the basis of the information contained in the low-pass filtered version), the lower intonation ratings may be based on problems which are less relevant to the actual comprehensibility of the utterance. Of the 36 utterances, both the utterance type and the internal structure of 24 were recognised. For 10 of the utterances, the utterance type but not the internal structure was recognised, and the remaining 2 were completely unparsable. This suggests that the gravity of the intonational 'errors' – i.e. the factors causing a less native-like rating to be given – were perhaps not as severe as those that occurred in the L2 English of the Mandarin speakers (although a comparison of the data of the two studies would be necessary to establish if this is in fact the case).

8. CONCLUSION AND FUTURE DIRECTIONS

The study found that accentedness and intelligibility are not orthogonal when interactions are allowed in the data, but that accent plays a role as one thread of many threads in the interwoven fabric of linguistic (and meta-linguistic) information that is available to the listener. The study also found that the subjective input of the listener perspective to the ratings of accentedness is not the most useful if the goal of the L2 speaker is to be intelligible, as speech properties such as quantifiable phonetic and phonemic errors counts are a better predictor of intelligibility.

The study also found that although intonation plays a role in the comprehensibility of L2 speech, it is not the main factor, but is one contributing factor of several, both segmental and suprasegmental in nature. Analysis of the data suggested that this may be affected by the properties of the L1 of the speaker, and that the similarities between Arabic and English in the realm of intonation may have as a consequence that the intonation of L1 Arabic speakers of L2 English does not present as great a problem for comprehensibility as does that of L1 Mandarin speakers of L2 English. It would interesting to compare the two data sets directly.

It serves us well to consider these results in conjunction with empirical studies in word segmentation. In their hierarchy of speech segmentation cues, Mattys et al. (2005) suggest that lexical, semantic and syntactic information is of foremost importance in speech segmentation, followed by segmental cues, and then, in English, cues provided by stress. They find that in 'rich interpretative conditions', sublexical cues contribute little to the speech segmentation process. However, they conclude (p492) that

> "In ambiguous parsing conditions (caused by mild noise, e.g.), lexical and postlexical evidence could be sufficiently impoverished to make the activity difference between sublexically aligned and misaligned candidates decisive. Similarly, the hierarchical approach gives stress-initial candidates a measurable

> boost primarily when both lexical and segmental sources of information fail to provide an unequivocal segmentation solution—for example, ambiguous embeddedness, novel words, and severely impoverished input."

In other words, if lexical and post-lexical cues fail (for reasons of ambiguous embeddedness, or use of novel words), then a listener turns to segmental cues for the word segmentation process. If the segmental cues fail (because of severely impoverished input, such as when the speech takes place in a noisy room), the listener uses cues of stress. When adapted to the L2 context of this study, incorrect morphosyntactic structures and inappropriate words constitute poor lexical and post-lexical cues, and pronunciation errors can result in severely impoverished input. (It is important to remember, though, that word recognition – i.e. the aim in the intelligibility task - takes the processing of speech one step further than word segmentation: words may be correctly segmented but incorrectly transcribed, i.e. the breakdown in the process occurs later.)

When examining the comprehensibility and intelligibility of speech, there is a need to consider all factors which may contribute information to making speech comprehensible and intelligible, as no component of speech functions in isolation, but in conjunction with all others. In their study, Munro & Derwing (1995a) conclude that even heavily accented speech is sometimes perfectly intelligible. The current study suggests that it is possible to understand heavily accented speech *if and only if* there is enough other information available in the utterance and the context in which it is uttered. If this is not the case, communication may break down.

One point should be made here with regard to the amount of data used in such studies. In the same way that there is no guarantee in this analysis of a cause-and-effect relationship between the negative interactive effects reported in the Results chapter for comprehensibility and intelligibility (which were disregarded as they were considered non-sensical), this is also the case for the positive interactive effects (which were retained as they were considered to make sense). There is no way of knowing that these latter 'effects' are not also cases of mere co-occurrence, and when using mixed models this possibility will never be eliminated. The more samples are analysed, the further the likelihood of drawing conclusions based chance co-occurrences is reduced. A data set of 30 (L2) utterances from 10 (non-native) speakers such as the one used in this experiment is not large enough to serve as a basis from which to draw conclusions, and the negative interactive effects obtained here provide evidence for the case that pleas for increased data sets not be ignored. A suggestion for further research, then, is to re-run the experiment with a much larger data set. Another point is that the analysis of the data was carried out solely by the author, and, as such, needs verification.

Possibly the main finding of this study is that looking for effects in isolation, or studying specific linguistic phenomena in isolation, is ineffective. There is a need to look at the whole picture. The findings for both research questions support this observation, and this is in line with the growing popularity of Dynamic Systems Theory in linguistics (De Bot (2008), discussed in Chapter 1, which describes language as a complex, nonlinear, and dynamic system. In de Bot's research, DST was applied to second language acquisition, but there are strong arguments for extending its application to language processing studies such as the comprehensibility and intelligibility of (L2) speech, and indeed its application would help to explain the interactions between components described as unclear by Saito et al. (2015), and thus provides a direction for further research in this area. Bar-Yam (1997, p8) comments: “Since interactions between parts of a complex system are essential to understanding its behaviour, looking at parts by themselves is not sufficient. It is necessary to look at parts in the context of the whole.” Language is one such system.

APPENDIX A: THE STIMULI (WITH SPEAKER ID) AND THEIR IPA (SSBE) TRANSCRIPTIONS

1: You can find it near the City Library. (01A)

/ju: kɑ:n faɪnd ɪt nɪə ðə sɪtɪ laɪbɹəɹi/

2: Can you tell me the location of the railway station? (05B)

/kan ju: tɛl mi: ðə leʊkeɪʃn əv ðə ɹeɪlweɪj steɪʃn/

3: It's on a street called, well, Main Street, which is this street, so just follow Main Street straight up. (EN-01B)

/ɪts ɒn ə stɹi:t kɔ:ld wɛl meɪn stɹi:t wɪtʃ ɪz ðɪs stɹi:t səʊ dʒʊt fɒləʊ meɪn stɹi:t stɹeɪt ʌp/

4: It is in the Main Streets and er, of, back of, er, City Hall. (02B)

/ɪt ɪs ɪn ðə meɪn stɹi:ts and ɜ: əv bak əv ɜ: sɪtɪ hɔ:l/

5: You'll find it on the second row and the fifth avenue, er, in front of the bus station. (08A)

/jɔ:l faɪnd ɪt ɒn ðə sɛkənd ɹeʊw ənd ðə fɪθ avɪnju: ɜ: ɪn fɹʌnt əv ðə bʌs steɪʃn/

6: It's, er, opposite to the, er, State College. (08B)

/ɪts ɜ: ɪn opəzɪt tu: ðə ɜ: steɪt kɒlɪdʒ/

7: Could you please, er, tell me where is the supermarket? (02B)

/kʊd ju: pli:z ɜ: tɛl mi: wɛ:ɹ ɪz ðə su:pəmɑ:kɪt/

8: It's in front of the bus station. (02A)

/ɪts ɪn fɹʌnt əv ðə bʌs steɪʃn/

9: Can you tell me where is the supermarket please? (07B)

/kan ju: tɛl mi: wɛ:ɹ ɪz ðə su:pəmɑ:kɪt pli:z/

10: Ah, I think it's probably going to take you about 5, 10 minutes. (EN-01A)

/a: aɪ θɪnk ɪts pɹɒbəbli: gəʊɪŋ tə təɪk ju: əbaʊt faɪv tɛn mɪnɪts/

11: Can you show me the way to the Beethoven Museum? (01A)

/kan ju: ʃəʊ mi: ðə weɪ tə ðə beɪthəʊvən mju:zi:əm/

12: It's actually on the third avenue. (08A)

/ɪts aktu:əli: ɒn ðə θɜ:d avɪnju:/

13: It's next to post office, in area one and two. (04A)

/ɪts nɛkst tə pəʊst ɒfɪs ɪn ɛ:ɹiə wɒn and tu:/

14: Er, is there a supermarket round here anywhere? (EN-01B)

/ɜ: ɪz ðɛ: ə su:pəmɑ:kɪt ɹaʊnd hɪə ɛniwɛ:/

15: It's in the middle north, er, in the first row, near to the City Hall. (07A)

/ɪts ɪn ðə mɪdl nɔ:θ ɜ: ɪn ðə fɜ:st ɹəʊw nɪə tə ðə sɪtɪ hɔ:l/

16: He's near the River Street and, er, near bus station. (02B)

/hi:z nɪə ðə ɹɪvə stɹi:t and ɜ: nɪə bʌs steɪʃn/

17: What is the road to the Manchester Bookshop? (08B)

/wɒt ɪz ðə ɹəʊd tə ðə mantʃɛstə bukʃɑp/

18: It's in the third row, next to the Country Hospital. (05B)

/ɪts ɪn ðə θɜ:d ɹəʊw nɛkst tə ðə kʌntɹi: hɒspɪtəl/

19: The railway station in area seven, er, at, er, Grove Street, er, and it's, er, near of Elementary School. (04A)

/ðə ɹeɪlɰeɪj steɪʃn ɪn ɛ:ɹiə sɛvən ɜ: at ɜ: gɹəʊv stɹi:t ɜ: ənd ɪts ɜ: nɪə əv ɛlɪmɛntəɹi sku:l/

20: It's in area four, er, next to State College. (07B)

/ɪts ɪn ɛ:ɹiə fɔ: ɜ: nɛkst tə steɪt kɒlɪdʒ/

21: Er hello, excuse me, can you, erm, take me to the bowling alley? (02A)

/ɜ: hɛləʊ ɛkskju:z mi: kan ju: ɜ:m teɪk mi: tə ðə bəʊlɪŋ ali:/

22: You get to another road on your left, and the railway station's just on the corner of that street. (EN-01A)

/ju: gɛt tu: ənʌðə ɹəʊd ɒn jɔ: lɛft and ðə ɹeɪlweɪj steɪʃnz dʒʌst ɒn ðə kɔ:nəɹ əv ðat stɹi:t/

23: Can you tell me please where is the Liverpool Hotel? (04B)

/kan ju: tɛl mi: pli:z wɛ:ɹ ɪz ðə lɪvəpu:l həʊtɛl/

24: It's next to the bus station. (08B)

/ɪts nɛkst tə ðə bʌs steɪʃn/

25: It's next to the river, er, on the left of Elementary School. (01A)

/ɪts nɛkst tə ðə ɹɪvə ɒn ðə lɛft əv ɛlɪmɛntəɹi sku:l/

26: Can you tell me please where is, er, mobile phone shop? (07A)

/kan ju: tɛl mi: pli:z wɛ:ɹ ɪz ɜ: məʊbaɪl fəʊn ʃɒp/

27: It's, er, in the right of, erm, City Park and in, in the left of City High School. (02A)

/ɪts ɜ: ɪn ðə ɹaɪt əv ɜ:m sɪtɪ pa:k and ɪn ɪn ðə lɛft əv sɪtɪ haɪ sku:l/

28: Can you tell me please where is the Delaware Cafe? (04A)

/kan ju: tɛl mi: pli:z wɛ:ɹ ɪz ðə dɛləwɛ: kafeɪ/

29: You'll go past the City Court on your right, you'll have the fire station on your left.. (EN-01B)

/jɔ:l gəʊ past ðə sɪtɪ kɔ:t ɒn jɔ: ɹaɪt jɔ:l hav ðə faɪə steɪʃn ɒn jɔ: lɛft/

30: It's in area three and four, it's in, next to the bus station. (04B)

/ɪts ɪn ɛ:ɹiə θɹi: ənd fɔ: ɪts ɪn nɛkst tə ðə bʌs steɪʃn/

31: Excuse me, where can I found the bowling alley? (08A)

/ɛkskju:z mi: wɛ: kan aɪ faʊnd ðə bəʊlɪŋ ali:/

32: It's in area one, er, the east of, er, post office. (07B)

/ɪts ɪn ɛ:ɹiə wɒn ɜ: ðə i:st əv ɜ: pəʊst ɒfɪs/

33: Could you tell me where Delaware Cafe is? (EN-01A)

/kʊd ju: tɛl mi: wɛ: ə dɛləwɛ: kafeɪ ɪz/

34: It's in the area number six, er, near to the City Middle School. (07A)

/ɪts ɪn ɛ:ɹiə nʊmbə sɪks ɜ: nɪə tə ðə sɪtɪ mɪdl sku:l/

35: It's in the third row, in the third area, next to the State College. (05B)

/ɪts ɪn ðə θɜ:d ɹəʊw ɪn ðə θɜ:d ɛ:ɹiə nɛkst tə ðə steɪt kɒlɪdʒ/

36: I think it's behind City Middle School. (04B)

/aɪ θɪnk ɪts bɪhaɪnd sɪtɪ mɪdl sku:l/

APPENDIX B: MORPHOSYNTACTIC ERROR ANALYSIS OF THE STIMULI, WITH SPEAKER ID

1: You can find it near the City Library. (01A)

2: Can you tell me the location of the railway station? (05B)

3: It's on a street called, well, Main Street, which is this street, so just follow Main Street straight up. (EN-01B)

4: It is in the Main Streets and er, of, back of, er, City Hall. (02B)

ERRORS: plural instead of singular noun (streets)

article omitted (City Hall)

5: You'll find it on the second row and the fifth avenue, er, in front of the bus station. (08A)

6: It's, er, opposite to the, er, State College. (08B)

ERRORS: preposition inserted (opposite to)

7: Could you please, er, tell me where is the supermarket? (02B)

ERRORS: direct question used in embedded clause (where is)

8: It's in front of the bus station. (02A)

9: Can you tell me where is the supermarket please? (07B)

ERRORS: direct question used in embedded clause (where is)

10: Ah, I think it's probably going to take you about 5, 10 minutes. (EN-01A)

11: Can you show me the way to the Beethoven Museum? (01A)

12: It's actually on the third avenue. (08A)

13: It's next to post office, in area one and two. (04A)

ERRORS: article omitted (post office)

14: Er, is there a supermarket round here anywhere? (EN-01B)

15: It's in the middle north, er, in the first row, near to the City Hall. (07A)

ERRORS: preposition inserted (near to)

16: He's near the River Street and, er, near bus station. (02B)

ERRORS: incorrect pronoun (he)

article inserted (the)

article omitted (bus station)

17: What is the road to the Manchester Bookshop? (08B)

18: It's in the third row, next to the Country Hospital. (05B)

19: The railway station in area seven, er, at, er, Grove Street, er, and it's, er, near of

Elementary School. (04A)

ERRORS: copula omitted (station.. in)

incorrect preposition (at)

preposition inserted (near of)

20: It's in area four, er, next to State College. (07B)

ERRORS: article omitted (State College)

21: Er hello, excuse me, can you, erm, take me to the bowling alley? (02A)

22: You get to another road on your left, and the railway station's just on the corner of that street. (EN-01A)

23: Can you tell me please where is the Liverpool Hotel? (04B)

ERRORS: direct question used in embedded clause (where is)

24: It's next to the bus station. (08B)

25: It's next to the river, er, on the left of Elementary School. (01A)

ERRORS: article omitted (Elementary School)

26: Can you tell me please where is, er, mobile phone shop? (07A)

ERRORS: direct question used in embedded clause (where is)

27: It's, er, in the right of, erm, City Park and in, in the left of City High School. (02A)

ERRORS: incorrect prepositions (1st and 2nd in)

28: Can you tell me please where is the Delaware Cafe? (04A)

ERRORS: direct question used in embedded clause (where is)

29: You'll go past the City Court on your right, you'll have the fire station on your left.. (EN-01B)

30: It's in area three and four, it's in, next to the bus station. (04B)

31: Excuse me, where can I found the bowling alley? (08A)

ERRORS: incorrect tense (found)

32: It's in area one, er, the east of, er, post office. (07B)

ERRORS: preposition omitted (east)

article omitted (post office)

33: Could you tell me where Delaware Cafe is? (EN-01A)

34: It's in the area number six, er, near to the City Middle School. (07A)

ERRORS: preposition inserted (near to)

35: It's in the third row, in the third area, next to the State College. (05B)

36: I think it's behind City Middle School. (04B)

APPENDIX C: PHONETIC & PHONEMIC ERROR ANALYSIS OF THE STIMULI

Each stimulus is given, followed by its phonetic transcription. The transcription of the stimuli is given in SSBE, though realisations in Standard American are accepted (thus the use of [ɾ] instead of [ɹ] in pre-consonantal position is considered to be a phonetic rather than a phonemic error). For each error found, the orthographic form of the word containing the error is given in italics, followed by the phonetic and phonemic transcriptions of the errors, followed by a description of the problem.

1: You can find it near the City Library.

/ju: kɑ:n faɪnd ɪt nɪə ðə sɪtɪ laɪbɹəɹi/

PHONETIC: *you:* [u:] no F2 fall (starts too rounded)

it: [t] short-lag VOT

library: [ɾ] instead of [ɹ]; [i:] no F2 rise (starts too close)

PHONEMIC: *can:* /t/ inserted at end

library: /a/ not /ə/

2: Can you tell me the location of the railway station?

/kan ju: tɛl mi: ðə leʊkeɪʃn əv ðə ɹeɪlweɪj steɪʃn/

PHONETIC: *can:* [k] short-lag VOT

location: [eɪ] more like [e]

railway: [ɹ] back of tongue too raised

PHONEMIC:-

3: It's on a street called, well, Main Street, which is this street, so just follow Main Street straight up.

/ɪts ɒn ə stɹi:t kɔ:ld wɛl meɪn stɹi:t wɪtʃ ɪz ðɪs stɹi:t səʊ dʒʊt fɒləʊ meɪn stɹi:t stɹeɪt ʌp/

PHONETIC:-

PHONEMIC:-

4: It is in the Main Streets and er, of, back of, er, City Hall.

/ɪt ɪs ɪn ðə meɪn stɹi:ts and ɜ: əv bak əv ɜ: sɪtɪ hɔ:l/

PHONETIC: *It:* [t] short-lag VOT

street: [r] instead of [ɹ], [iː] no F2 rise (starts too close)

PHONEMIC:-

5: You'll find it on the second row and the fifth avenue, er, in front of the bus station.

/jɔːl faɪnd ɪt ɒn ðə sɛkənd ɹeʊw ənd ðə fɪθ avɪnjuː ɜː ɪn fɹʌnt əv ðə bʌs steɪʃn/

PHONETIC:

PHONEMIC: *second:* /ɒ/ instead of /ə/

the (2nd): /ð/ deleted

6: It's, er, opposite to the State College.

/ɪts ɜː ɪn opəzɪt tuː ðə steɪt kɒlɪdʒ/

PHONETIC: *college:* [k] short-lag VOT, [l] too clear

PHONEMIC: *opposite:* /b/ not /p/, /ɒ/ not /ə/

state: /ɛ/ inserted at the beginning

7: Could you please, er, tell me where is the supermarket?

/kʊd juː pliːz ɜː tɛl miː wɛːɹ ɪz ðə suːpəmɑːkɪt/

PHONETIC: *tell:* [l] too clear

where: [r] instead of [ɹ]

supermarket: [r] instead of [ɹ] (both instances), [t] too short a closure time

PHONEMIC:-

8: It's in front of the bus station.

/ɪts ɪn fɹʌnt əv ðə bʌs steɪʃn/

PHONETIC: *bus:* [b] pre-voiced

PHONEMIC: *front:* /ɒ/ not /ʌ/

9: Can you tell me where is the supermarket please?

/kan juː tɛl miː wɛːɹ ɪz ðə suːpəmɑːkɪt pliːz/

PHONETIC: *can:* [k] short-lag VOT

PHONEMIC:-

10: Ah, I think it's probably going to take you about 5, 10 minutes.

/a: aɪ θɪnk ɪts pɹɒbəbli: gəʊɪŋ tə təɪk ju: əbaʊt faɪv tɛn mɪnɪts/

PHONETIC:-

PHONEMIC:-

11: Can you show me the way to the Beethoven Museum?

/kan ju: ʃəʊ mi: ðə weɪ tə ðə beɪthəʊvən mju:zi:əm/

PHONETIC: *to:* [u:] instead of [ə] (not reduced)

PHONEMIC: *Beethoven:* /e/ instead of /eɪ/, /ɛ/ instead of /ə/

12: It's actually on the third avenue.

/ɪts aktu:əli: ɒn ðə θɜ:d avɪnju:/

PHONETIC:-

PHONEMIC:-

13: It's next to post office, in area one and two.

/ɪts nɛkst tə pəʊst ɒfɪs ɪn ɛ:ɹiə wɒn and tu:/

PHONETIC: *area:* [r] instead of [ɹ]

two: [u:] no F2 fall (starts too rounded)

PHONEMIC: *It's:* /ɛ/ instead of /ɪ/

post: /ɒ/ instead of /eʊ/

14: Er, is there a supermarket round here anywhere?

/ɜ: ɪz ðɛ: ə su:pəmɑ:kɪt ɹaʊnd hɪə ɛniwɛ:/

PHONETIC:-

PHONEMIC:-

15: It's in the middle north, er, in the first row, near to the City Hall.

/ɪts ɪn ðə mɪdl nɔ:θ ɜ: ɪn ðə fɜ:st ɹəʊw nɪə tə ðə sɪtɪ hɔ:l/

PHONETIC: *middle:* [d] pre-voiced, [l] too clear

north: [r] instead of [ɹ]

first: [r] instead of [ɹ]

row: [r] instead of [ɹ]

near: [r] instead of [ɹ]

city: [t] short-lag VOT

PHONEMIC: *the (1st):* /d/ not /ð/

first: /ɛ/ instead of /ɜ:/

middle: /ɛ/ inserted before syllabic [l]

16: He's near the River Street and, er, near bus station.

/hi:z nɪə ðə ɹɪvə stɹi:t and ɜ: nɪə bʌs steɪʃn/

PHONETIC: *near:* [i:] no F2 rise (starts too close), [r] instead of [ɹ]

station: [eɪ] more like [e]

PHONEMIC: *bus station:* vocoid between words, /ʌ/ inserted before syllabic [n]

17: What is the road to the Manchester Bookshop?

/wɒt ɪz ðə ɹəʊd tə ðə mantʃɛstə bukʃɑp/

PHONETIC: *road:* [r] instead of [ɹ]

is: [z] overvoiced

Manchester: [r] instead of [ɹ]

bookshop: [k] short closure time/aspirated burst

shop: [p] short-lag VOT

PHONEMIC: *road:* /ɔ:/ instead of /əʊ/

Manchester: /ʃ/ instead of /tʃ/

18: It's in the third row, next to the Country Hospital.

/ɪts ɪn ðə θɜ:d ɹəʊw nɛkst tə ðə kʌntɹi: hɒspɪtəl/

PHONETIC:- *to:* [u:] instead of [ə] (not reduced)

PHONEMIC: *row:* /l/ instead of /ɹ/

19: The railway station in area seven, er, at, er, Grove Street, er, and it's, er, near of Elementary School.

/ðə ɹeɪlɰeɪj steɪʃn ɪn ɛ:ɹiə sɛvən ɜ: at ɜ: gɹəʊv stɹi:t ɜ: ənd ɪts ɜ: nɪə əv ɛlɪmɛntəɹi sku:l/

PHONETIC: *area:* [r] instead of [ɹ]

at: [t] short-lag VOT

Grove: [r] instead of [ɹ]

street: [r] instead of [ɹ], [i:] no F2 rise (starts too close) *near:* [r] instead of [ɹ]

elementary: [r] instead of [ɹ], [i:] no F2 rise (starts too close)

school: [u:] no F2 fall (starts too rounded)

PHONEMIC: *railway:* /ɔ:/ instead of /eɪ/

station: /ʌ/ inserted before syllabic /n/

seven: /ɛ/ instead of /ə/ (or syllabic /n/)

Grove Street: vocoid between words

20: It's in area four, er, next to State College.

/ɪts ɪn ɛ:ɹiə fɔ: ɜ: nɛkst tə steɪt kɒlɪdʒ/

PHONETIC: *to:* [u:] instead of [ə] (not reduced)

college: [k] short-lag VOT

PHONEMIC:-

21: Er hello, excuse me, can you, erm, take me to the bowling alley?

/ɜ: hɛləʊ ɛkskju:z mi: kan ju: ɜ:m teɪk mi: tə ðə bəʊlɪŋ ali:/

PHONETIC: *excuse:* [u:] no F2 fall (starts too rounded)

PHONEMIC: *excuse:* /ɛ/ deleted

22: You get to another road on your left, and the railway station's just on the corner of that street.

/ju: gɛt tu: ənʌðə ɹəʊd ɒn jɔ: lɛft and ðə ɹeɪlweɪj steɪʃnz dʒʌst ɒn ðə kɔ:nəɹ əv ðat stɹi:t/

PHONETIC:-

PHONEMIC:-

23: Can you tell me please where is the Liverpool Hotel?

/kan ju: tɛl mi: pli:z wɛ:ɹ ɪz ðə lɪvəpu:l həʊtɛl/

PHONETIC: *can:* [k] short-lag VOT

you: [u:] no F2 fall (starts too rounded)

please: [l] too clear, [i:] no F2 rise (starts too close), [z] overvoiced

where: [r] instead of [ɹ]

is: [z] overvoiced

Liverpool: [u:] no F2 fall (starts too rounded), [l] too clear

PHONEMIC:-

24: It's next to the bus station.

/ɪts nɛkst tə ðə bʌs steɪʃn/

PHONETIC:- *to:* [u:] instead of [ə] (not reduced)

PHONEMIC: *bus station:* vocoid between words, /ʌ/ inserted before syllabic /n/

25: It's next to the river, er, on the left of Elementary School.

/ɪts nɛkst tə ðə ɹɪvə ɒn ðə lɛft əv ɛlɪmɛntəɹi sku:l/

PHONETIC: *to:* [t] short-lag VOT, [u:] instead of [ə] (not reduced)

left: [l] too clear, [t] short-lag VOT

of: [v] overvoiced

elementary: [l] too clear, [r] instead of [ɹ]

school: [u:] no F2 fall (starts too rounded), [l] too clear

PHONEMIC: *It's:* /t/ deleted

river: /ɛ/ instead of /ə/

26: Can you tell me please where is, er, mobile phone shop?

/kan ju: tɛl mi: pli:z wɛ:ɹ ɪz ɜ: məʊbaɪl fəʊn ʃɒp/

PHONETIC: *please:* [i:] no F2 rise (starts too close)

where: [r] instead of [ɹ]

PHONEMIC: *mobile:* /əʊ/ deleted

phone: /ɒ/ instead of /eʊ/

shop: /b/ instead of /p/

27: It's, er, in the right of, erm, City Park and in, in the left of City High School.

/ɪts ɜ: ɪn ðə ɹaɪt əv ɜ:m sɪtɪ pa:k and ɪn ɪn ðə lɛft əv sɪtɪ haɪ sku:l/

PHONETIC: *right:* [r] instead of [ɹ]

city: [t] short-lag VOT

park: [r] instead of [ɹ]

PHONEMIC: *park:* /b/ instead of /p/

the (2nd): /d/ instead of /ð/

left: /ɪ/ instead of /ɛ/

28: Can you tell me please where is the Delaware Cafe?

/kan ju: tɛl mi: pli:z wɛ:ɹ ɪz ðə dɛləwɛ: kafeɪ/

PHONETIC: *can:* [k] short-lag VOT

you: [u:] no F2 fall (starts too rounded)

please: [i:] no F2 rise (starts too close)

where: [r] instead of [ɹ]

Delaware: [r] instead of [ɹ]

PHONEMIC: please: /b/ instead of /p/

29: You'll go past the City Court on your right, you'll have the fire station on your left..

/jɔ:l gəʊ past ðə sɪtɪ kɔ:t ɒn jɔ: ɹaɪt jɔ:l hav ðə faɪə steɪʃn ɒn jɔ: lɛft/

PHONETIC:-

PHONEMIC:-

30: It's in area three and four, it's in, next to the bus station.

/ɪts ɪn ɛ:ɹiə θɹi: ənd fɔ: ɪts ɪn nɛkst tə ðə bʌs steɪʃn/

PHONETIC: *area:* [iə] starts too close

three: [r] instead of [ɹ]

to: [u:] instead of [ə] (not reduced)

PHONEMIC: *It's:* /ɛ/ instead of /ɪ/, /t/ deleted

the: [ð] deleted

31: Excuse me, where can I found the bowling alley?

/ɛkskju:z mi: wɛ: kan aɪ faʊnd ðə bəʊlɪŋ ali:/

PHONETIC: *found:* [d] pre-voiced

bowling: [b] pre-voiced

PHONEMIC: *alley:* /eɪ/ instead of /i:/

32: It's in area one, er, the east of, er, post office.

/ɪts ɪn ɛ:ɹiə wɒn ɜ: ðə i:st əv ɜ: pəʊst ɒfɪs/

PHONETIC: *area:* [ɹ] produced as approximation *post:* [p] short-lag VOT

PHONEMIC: *east:* /ɪ/ instead of /i:/

post: /ɒ/ instead of /eʊ/

33: Could you tell me where Delaware Cafe is?

/kʊd ju: tɛl mi: wɛ: ðə dɛləwɛ: kafeɪ ɪz/

PHONETIC:-

PHONEMIC:-

34: It's in the area number six, er, near to the City Middle School.

/ɪts ɪn ɛ:ɹiə nʊmbə sɪks ɜ: nɪə tə ðə sɪtɪ mɪdl sku:l/

PHONETIC: *area:* [r] instead of [ɹ], /iə/ starts too close

six: [k] short-lag VOT

near: [ɪ] starts too close, [r] instead of [ɹ]

to: [u:] instead of [ə] (not reduced)

middle: [d] pre-voiced

school: [u:] no F2 fall (starts too rounded), [l] too clear

PHONEMIC: *the (1st):* /d/ instead of /ð/

middle: /ɛ/ inserted before syllabic [l]

35: It's in the third row, in the third area, next to the State College.

/ɪts ɪn ðə θɜ:d ɹəʊw ɪn ðə θɜ:d ɛ:ɹiə nɛkst tə ðə steɪt kɒlɪdʒ/

PHONETIC: *third:* [d] pre-voiced (both instances)

row: /əʊ/ starts too close

to: [t] short-lag VOT, [u:] instead of [ə] (not reduced)

state: [eɪ] more like [e]

PHONEMIC:-

36: I think it's behind City Middle School.

/aɪ θɪnk ɪts bɪhaɪnd sɪtɪ mɪdl sku:l/

PHONETIC: *It's:* [t] short-lag VOT

behind: [b] pre-voiced

city: [t] short-lag VOT

middle: [d] pre-voiced

school: [k] short closure/strong burst, [u:] no F2 fall (starts too rounded)

PHONEMIC: *middle:* /ɛ/ inserted before syllabic [l]

APPENDIX D: INTONATION RATINGS AND ToBI ANALYSIS OF THE STIMULI

The Speaker ID is given after the utterance number. The intonation rating is given in square brackets, at the end of each utterance. ToBI annotations are used in (2) to show intonation.

(1) INTONATION BY ORDER OF RATINGS (9-1):

15 (07A): It's in the middle north, er, in the first row, near to the City Hall. [8]

5 (08A): You'll find it on the second row and the fifth avenue, er, in front of the bus station. [7]

26 (07A): Can you tell me please where is, er, mobile phone shop? [7]

28 (04A): Can you tell me please where is the Delaware Cafe? [7]

30 (04B): It's in area three and four, it's in, next to the bus station. [6]

4 (02B): It is in the Main Streets and er, of, back of, er, City Hall. [6]

7 (02B): Could you please, er, tell me where is the supermarket? [6]

13 (04A): It's next to post office, in area one and two. [6]

1 (01A): You can find it near the City Library. [5]

11 (01A): Can you show me the way to the Beethoven Museum? [5]

18 (05B): It's in the third row, next to the Country Hospital. [5]

21 (02A): Er hello, excuse me, can you, erm, take me to the bowling alley? [5]

23 (04B): Can you tell me please where is the Liverpool Hotel? [5]

27 (02A): It's, er, in the right of, erm, City Park and in, in the left of City High School. [5]

8 (02A): It's in front of the bus station. [5]

25 (01A): It's next to the river, er, on the left of Elementary School. [4]

2 (05B): Can you tell me the location of the railway station? [4]

12 (08A): It's actually on the third avenue. [4]

34 (07A): It's in the area number six, er, near to the City Middle School. [4]

32 (07B): It's in area one, er, the east of, er, post office. [4]

16 (02B): He's near the River Street and, er, near bus station. [4]

36 (04B): I think it's behind City Middle School. [4]

6 (08B): It's, er, opposite to the, er, State College. [3]

10 (EN-01A): Ah, I think it's probably going to take you about 5, 10 minutes. [3]

3 (EN-01B): It's on a street called, well, Main Street, which is this street, so just follow Main Street straight up. [2]

9 (07B): Can you tell me where is the supermarket please? [2]

31 (08A): Excuse me, where can I found the bowling alley? [2]

17 (08B): What is the road to the Manchester Bookshop? [2]

19 (04A): The railway station in area seven, er, at, er, Grove Street, er, and it's, er, near of Elementary School. [2]

20 (07B): It's in area four, er, next to State College. [2]

14 (EN-01B): Er, is there a supermarket round here anywhere? [1]

22 (EN-01A): You get to another road on your left, and the railway station's just on the corner of that street. [1]

24 (08B): It's next to the bus station. [1]

29 (EN-01B): You'll go past the City Court on your right, you'll have the fire station on your left..[1]

33 (EN-01A): Could you tell me where Delaware Cafe is? [1]

35 (05B): It's in the third row, in the third area, next to the State College. [1]

(2) INTONATIONAL ANALYSIS BY TYPE

GROUP 1 – STATEMENTS CONSISTING OF ONE PHRASE

H* H* L* L* L-L%

1 (01A): You can find it near the Ci-ty Li-bra-ry. [5]

H* L*L- H* L-L%

8 (02A): It's in front of the bus sta-tion. [5]

H* H* L* L-L%

12 (08A): It's ac-tua-lly on the third a-ve-nue. [4]

H* H* !H* L-L%

36 (04B): I think it's be-hind Ci-ty Mi-ddle School. [4]

H* H* H- H* L-L%

6 (08B): It's, er, o-ppo-site to the, er, State Co-llege. [3]

H* !H* L-L%

24 (08B): It's next to the bus sta-tion. **[1]**

GROUP 2 – STATEMENTS CONSISTING OF MORE THAN ONE PHRASE

L* L* H-H% H* L* H-H% H* H* L-L%

15 (07A): It's in the mi-ddle north, er, in the first row, near to the Ci-ty Hall. [8]

L* H- H* L* H-H% L* H-H% H* H* L-L%

5 (08A): You'll find it on the se-cond row and the fifth a-ve-nue, er, in front of the bus sta-tion. [7]

H* H* L* H-H% H* H* L-L%

30 (04B): It's in a-rea three and four, it's in, next to the bus sta-tion. [6]

H* L- H* L-H% L* L- H* H* L-L%

4 (02B): It is in the Main Streets and er, of, back of, er, Ci-ty Hall. [6]

H* H* L* H-H% L* H* L* L-L%

13 (04A): It's next to post office, in area one and two. [6]

H* L-H% H* H- H* H* L-L%

18 (05B): It's in the third row, next to the Coun-try Hos-pi-tal. [5]

H* L* L-H% H* H- L* L-L%

25 (01A): It's next to the ri- ver, er, on the left of E-le-men-ta-ry School. [4]

L* H-H% H* H* !H* H* L-L%

34 (07A): It's in the a-rea number six, er, near to the Ci-ty Mi-ddle School. [4]

L* H- L* H-H% H* H* !H* L-L%

32 (07B): It's in a-rea one, er, the east of, er, post o- ffice. [4]

H* H* H* L* L-H% L* L- H* L* L-L%

16 (02B): He's near the Ri-ver Street and, er, near bus sta-tion. [4]

H* L-H% H* L-H% H* L-H% H* H- H* H* L-L%

19 (04A): The rail-way sta-tion in a-rea seven, er, at, er, Grove Street, er, and it's, er, near of E-le-men-tary School. [2]

H* L-H% H* H* L-L%

20 (07B): It's in a-rea four, er, next to State Co-llege. [2]

H* L-H% H* L-H% H* H* L-L%

35 (05B): It's in the third row, in the third a-rea, next to the State Co-llege. [1]

H* H* L-H% H* !H* L-L%

22 (EN-01A): You get to a-no-ther road on your left, and the rail-way sta-tion's just on the corner of that street. [1]

GROUP 3 – INCOMPLETE AND MISCELLANEOUS STATEMENTS

H* L- H* L-H% H* H* L* H-H%

27 (02A): It's, er, in the right of, erm, Ci-ty Park and in, in the left of Ci-ty High School... [5]

L* H* !H* H-H% L* L-H%

29 (EN-01B): You'll go past the City Court on your right, you'll have the fire sta-tion on your left..[1]

H*H- H* !H* !H* !H* L-L%

10 (EN-01A): Ah, I think it's pro-ba-bly go-ing to take you about five, ten mi-nutes. [3]

H* L- H* !H* H- H*L- L* L* H-H%

3 (EN-01B): It's on a street called, well, Main Street, which is this street,… so just follow Main Street straight up.[2]

GROUP 4 - QUESTIONS

H* !H* L*L- H* L* H-H%

26 (07A): Can you tell me please where is, er, mo-bile phone shop? [7]

H* !H* H* L*L- L* H-H%

28 (04A): Can you tell me please where is the De-la-ware Ca- fe? [7]

H* H* H* L* H-H%

7 (02B): Could you please, er, tell me where is the su-per-mar-ket? [6]

H* L* L*H- L*L- L* L* L-H%

11 (01A): Can you show me the way to the Beet-ho-ven Mu-seum? [5]

H*L-H% H* H- L* H* L* H-H%

21 (02A): Er he-llo, ex-cuse me, can you, erm, take me to the bow-ling a-lley? [5]

H* L* L* L* L* H-H%

23 (04B): Can you tell me please where is the Li-ver-pool Ho-tel? [5]

H* H* L*-L- L* L- L* L* L-H%

2 (05B): Can you tell me the lo-ca-tion of the rail-way sta-tion? [4]

```
                 L*        H*                       H*      L*        H-H%
9 (07B):        Can you tell me where is the su-per-mar-ket please? [2]

                     H*        H*                          L*         H* H-H%
31 (08A):       Ex-cuse me, where can I found the bow-ling a- lley?  [2]

                  H*            !H*          L*                L*     H-H%
17 (08B):       What is the road to the Man-ches-ter Book-shop?  [2]

                                H*                          L*          L-H%
14 (EN-01B):  Er, is there a su-per-mar-ket round here an-y-where? [1]

                             H*                L*             H*    L-L%
33 (EN-01A):  Could you tell me where De-la-ware Ca-fe is?  [1]
```

APPENDIX E: MIXED MODELS ANALYSIS FOR INTELLIGIBILITY

All the models described below investigated the varying fixed effects (*phonetic* errors, *phonemic* errors, *lexis*, *grammatical* errors, *stress* errors and non-native-like *intonation*) on *intelligibility* as dependent variable. Random effects were *listener* and *speaker* for all models, and the random slopes defined for *listener* were only four of the six fixed factors (phonetic, phonemic, lexis and grammar), because of the non-convergence of the models including more than four slopes.

Model 0: contained all fixed factors, but NO interactions:

```
IntellAllFixedNOINTER.model <-
lmer(intell~lexis+phonetic+stress+phonemic+inton+grammar
+(1+lexis+phonetic+phonemic+grammar|listener)+(1|speaker),data=Data)
```

(1) INTERACTIONS FOR LEXIS (the following models include interactions between lexis and the pronunciation factors, but no interactions for grammar)

MODEL NAME & DESCRIPTION	SYNTAX
lexM1 (all interactions for lex)	lmer(intell~lexis*phonetic+lexis*phonemic+ lexis*stress+lexis*inton +grammar +(1+phonetic+phonemic+lexis+grammar|listener) +(1|speaker), data=Data)
lexM2 (all lex interactions except phonemic)	lmer(intell~lexis*phonetic+lexis*stress+lexis*inton +phonemic +grammar +(1+lexis+phonetic+phonemic+grammar|listener) +(1|speaker), data=Data)
lexM3 (all lex interactions except phonetic)	lmer(intell~lexis*phonemic+lexis*stress+lexis*inton +phonetic+grammar +(1+lexis+phonetic+phonemic+grammar|listener) +(1|speaker), data=Data)
lexM4 (all lex interactions except inton)	lmer(intell~lexis*phonemic+lexis*stress+lexis*phonetic +inton+grammar +(1+lexis+phonetic+phonemic+grammar|listener) +(1|speaker), data=Data)
lexM5 (all lex interactions except stress)	lmer(intell~lexis*phonemic+lexis*phonetic +stress+grammar +(1+lexis+phonetic+phonemic+grammar|listener) +(1|speaker),

	data=Data)
lexM5b (all lex interactions except stress (completely omitted))	lmer(intell~lexis*phonemic+lexis*phonetic +grammar +(1+lexis+phonetic+phonemic+grammar|listener) +(1|speaker), data=Data)

RESULTS FOR LEX MODELS COMPARISON:

Comparison	Anova	Results	Best-fit Model
without interactions /with all interactions	Model-0/lexM1	χ2=23.87; df=4; p=8.483e-05 ***	lexM1
with all interactions except phonemic (stand-alone) /with all interactions	lexM2/lexM1	χ2=12.129; df=1; p=.0004963 ***	lexM1
with all interactions except phonetic (stand-alone) /with all interactions	lexM3/lexM1	χ2=19.204; df=1; p=1.175e-05 ***	lexM1
with all interactions except inton (stand-alone) /with all interactions	lexM4/lexM1	χ2=5.1816; df=1; p=.02283 *	lexM1
with all interactions except stress (stand-alone) /with all interactions	lexM5/lexM1	χ2=.7609; df=1; p=.383	lexM5
with all interactions, no stress /with all interactions except stress (stand-alone)	lexM5b/lexM5	χ2=.9837; df=1; p=.3213	lexM5b

(2) INTERACTIONS FOR GRAMMAR (the following models include interactions between grammar and the pronunciation factors (but none for lexis)

MODEL NAME & DESCRIPTION	SYNTAX
gramM1 (all interactions for gram)	lmer(intell~grammar*phonetic+grammar*phonemic+grammar*stress +grammar*inton +lexis +(1+phonetic+phonemic+lexis+grammar|listener) +(1|speaker), data=Data)
gramM2 (all gram interactions except phonemic)	lmer(intell~grammar*phonetic+grammar*stress+grammar*inton +phonemic+lexis +(1+lexis+phonetic+phonemic+grammar|listener) +(1|speaker), data=Data)
gramM3 (all gram interactions except inton)	lmer(intell~grammar*phonemic+grammar*phonetic+grammar*stress +inton+lexis +(1+lexis+phonetic+phonemic+grammar|listener) +(1|speaker), data=Data)
gramM4 (all gram interactions except stress)	lmer(intell~grammar*phonemic+grammar*inton+grammar*phonetic +stress+lexis +(1+lexis+phonetic+phonemic+grammar|listener) +(1|speaker), data=Data)
gramM5	lmer(

(all gram interactions except phonetic (stand-alone))	intell~grammar*phonemic+grammar*phonetic +grammar*stress +phonetic+lexis +(1+lexis+phonetic+phonemic+grammar|listener) +(1|speaker), data=Data)
gramM5b (all gram interactions except phonetic (completely omitted))	lmer(intell~grammar*phonemic+grammar*phonetic +grammar*stress +lexis +(1+lexis+phonetic+phonemic+grammar|listener) +(1|speaker), data=Data)

RESULTS FOR GRAM MODEL COMPARISON:

Comparison	anova	Results	Best-fit Model
without interactions /with all interactions	Model-0/gramM1	χ2=46.023; df=4; p=2.436e-09 ***	gramM1
with all interactions except phonemic (stand-alone) /with all interactions	gramM2/gramM1	χ2=9.792; df=1; p=.001753 **	gramM1
with all interactions except inton (stand-alone) /with all interactions	gramM3/gramM1	χ2=26.243; df=1; p=3.011e-07 ***	gramM1
with all interactions except stress (stand-alone) /with all interactions	gramM4/gramM1	χ2=31.889; df=1; p=1.633e-08 ***	gramM1
with all interactions, except phonetic (stand-alone) /with all interactions	gramM5/gramM1	χ2=.1326; df=1; p=.7158	gramM5
with all interactions, no phonetic /with all interactions, except	gramM5b/gramM5	χ2=2.8329; df=1;	gramM5b

phonetic (stand-alone)		p=.09235	

(3) INTERACTIONS FOR LEXIS AND GRAMMAR TOGETHER (the following models include interactions between grammar and the pronunciation factors AND between lexis and pronunciation factors)

MODEL NAME & DESCRIPTION	Syntax
gramANDlexBestFitM (best fit lexis and grammar models combined)	lmer(intell~grammar*stress+grammar*phonemic+grammar*inton +lexis*phonetic+lexis*phonemic+lexis*inton +(1+lexis+phonetic+phonemic+grammar|listener) +(1|speaker), data=Data)
ManualgramANDlexBestFitM (manual best fit lexis and grammar models combined (negative interctions taken out))	lmer(intell~grammar*phonemic+grammar*inton +lexis*phonemic +phonetic +(1+lexis+grammar+phonetic+phonemic|listener) +(1|speaker), Data=Data)

RESULTS FOR GRAM/LEX MODEL:

Comparison	Anova	Results	Best-fit model
With all significant lex interactions /with all significant gram and lex interactions	lexM5b/ gramANDlexBestFitM	χ2=40.302; df=4; p=3.748e-08 ***	gramANDlexBestFitM
with all significant gram interactions plus stand-alone phonemic /with all significant gram and lex interactions	gramM5b/ gramANDlexBestFitM	χ2=19.369; df=4; p=.0006648 ***	gramANDlexBestFitM

RESULT:- BEST FIT FOR INTELLIGIBILITY IS ALL INTERACTIONS btw LEXIS and PRONUNCIATION FACTORS except stress, AND ALL INTERACTIONS btw GRAMMAR and PRONUNCIATION FACTORS except phonetic

(4) COMPARISON WITH ACCENTEDNESS

First find best-fit Accentedness Model then compare with best-fit Pronunciation Factor Model

AccModel-0: contained all fixed factors, but NO interactions:

IntellAllFixedAccNOINTER.model <- lmer(intell~lexis+accent+grammar +(1+lexis+accent+grammar|listener)+(1|speaker),data=Data)

MODEL NAME & DESCR	SYNTAX
AccModel-0 (no interactions)	lmer(intell~lexis+accent+grammar +(1+lexis+accent+grammar|listener) +(1|speaker), data=Data)
lexM1-acc (accent/lexis interaction, stand-alone grammar)	lmer(intell~lexis*accent +grammar +(1+lexis+accent+grammar|listener) +(1|speaker), data=Data)
gramM1-acc (accent/gram interaction, stand-alone lexis)	lmer(intell~grammar*accent +lexis +(1+lexis+accent+grammar|listener) +(1|speaker), data=Data)

RESULTS:

Comparison	Anova	Results	Best-fit model
No interactions/ lex interactions, stand-alone grammar	AccModel-0/ lexM1-acc	$\chi 2$=.212; df=1; p=.6452	AccModel-0
No interactions/ gram interactions, stand-alone lexis	AccModel-0/ gramM1-acc	$\chi 2$=6.5644; df=1; p=.0104 *	gramM1-acc (renamed acc-intell-best-fit)
Accentedness vs pronunciation factors	acc-intell-best-fit / pron-intell-best-fit	$\chi 2$=69.636; df=13; p=9.367e-10 ***	pron-intell-best-fit
Accentedness vs pronunciation factors (manual best-fit)	acc-intell-best-fit / pron-intell-manual-best-fit	$\chi 2$=26.412; df=10; p=.003224**	pron-intell-manual-best-fit

RESULT:- BEST FIT FOR INTELLIGIBILITY IS PRONUNCIATION FACTORS, NOT ACCENTEDNESS

APPENDIX F: MIXED MODELS ANALYSIS FOR COMPREHENSIBILITY

All the models described below investigated the varying fixed effects (*phonetic* errors, *phonemic* errors, *lexis*, *grammatical* errors, *stress* errors and non-native-like *intonation*) on *comprehensibility* as dependent variable. Random effects were *listener* and *speaker* for all models, and the random slopes defined for *listener* were only four of the six fixed factors (phonetic, phonemic, lexis and grammar), because of the non-convergence of the models including more than four slopes.

Model 0: contained all fixed factors, but NO interactions:

```
CompAllFixedNOINTER.model <-
lmer(comp~lexis+phonetic+phonemic+stress+inton+grammar
+(1+lexis+phonetic+phonemic+grammar|listener)+(1|speaker),data=Data)
```

(1) INTERACTIONS FOR LEXIS (the following models include interactions between lexis and the pronunciation factors, but no interactions for grammar)

MODEL NAME & DESCRIPTION	SYNTAX
lexM1 (all interactions for lex)	lmer(comp~lexis*phonetic+lexis*phonemic+lexis*stress+lexis*inton +grammar +(1+phonetic+phonemic+lexis+grammar|listener) +(1|speaker), data=Data)
lexM2 (all lex interactions except phonemic)	lmer(comp~lexis*phonetic+lexis*stress+lexis*inton+phonemic +grammar +(1+lexis+phonetic+phonemic+grammar|listener) +(1|speaker), data=Data)
lexM3 (all lex interactions except phonetic)	lmer(comp~lexis*phonemic+lexis*stress+lexis*inton +phonetic+grammar +(1+lexis+phonetic+phonemic+grammar|listener) +(1|speaker), data=Data)
lexM4 (all lex interactions except inton)	lmer(comp~lexis*phonemic+lexis*stress+lexis*phonetic +inton+grammar +(1+lexis+phonetic+phonemic+grammar|listener) +(1|speaker), data=Data)
lexM4b (all lex interactions except i inton(completely omitted))	lmer(comp~lexis*phonemic+ lexis*phonetic +lexis*stress +grammar +(1+lexis+phonetic+phonemic+grammar|listener) +(1|speaker), data=Data)
lexM5 (all lex interactions except stress)	lmer(comp~lexis*phonemic+lexis*phonetic +stress+grammar +(1+lexis+phonetic+phonemic+grammar|listener) +(1|speaker), data=Data)
lexM5b (all lex interactions except stress (completely omitted))	lmer(comp~lexis*phonemic+lexis*phonetic +grammar +(1+lexis+phonetic+phonemic+grammar|listener) +(1|speaker), data=Data)

RESULTS FOR LEX MODELS COMPARISON:

Comparison	Anova	Results	Best-fit Model
without interactions /with all interactions	Model-0/lexM1	χ2=44.495; df=4; p=5.064e-09 ***	lexM1
with all interactions except phonemic (stand-alone) /with all interactions	lexM2/lexM1	χ2=11.324; df=1; p=.000765 ***	lexM1
with all interactions except phonetic (stand-alone) /with all interactions	lexM3/lexM1	χ2=22.467; df=1; p=2.138e-06 ***	lexM1
with all interactions except inton (stand-alone) /with all interactions	lexM4/lexM1	χ2=0.3039; df=1; p=.5814	lexM4
with all interactions, no inton /with all interactions except inton (stand-alone)	lexM4b/lexM4	χ2=.0281; df=1; p=.8669	lexM4b
with all interactions except inton and stress (stress as stand-alone) /with all interactions, no inton	lexM5/lexM4b	χ2=1.7911; df=1; p=.1808	lexM5
with all interactions, no inton or stress /with all interactions except inton and stress (stress as stand-alone)	lexM5b/lexM5	χ2=1.2935; df=3; p=.7307	lexM5b

(2) **INTERACTIONS FOR GRAMMAR** (the following models include interactions between grammar and the pronunciation factors (but none for lexis)

MODEL NAME & DESCRIPTION	SYNTAX
gramM1 (all interactions for gram)	lmer(comp~grammar*phonetic+grammar*phonemic+grammar*stress +grammar*inton +lexis +(1+phonetic+phonemic+lexis+grammar|listener) +(1|speaker), data=Data)
gramM2 (all gram interactions except phonemic)	lmer(comp~grammar*phonetic+grammar*stress+grammar*inton +phonemic+lexis +(1+lexis+phonetic+phonemic+grammar|listener) +(1|speaker), data=Data)
gramM3 (all gram interactions except stress)	lmer(comp~grammar*phonemic+grammar*phonetic+grammar*inton +stress+lexis +(1+lexis+phonetic+phonemic+grammar|listener) +(1|speaker), data=Data)
gramM4 (all gram interactions except inton)	lmer(comp~grammar*phonemic+grammar*stress+grammar*phonetic +inton+lexis +(1+lexis+phonetic+phonemic+grammar|listener) +(1|speaker), data=Data)
gramM5 (all gram interactions except phonetic (stand-alone))	lmer(comp~grammar*phonemic+grammar*phonetic +grammar*stress +phonetic+lexis +(1+lexis+phonetic+phonemic+grammar|listener) +(1|speaker), data=Data)
gramM5b (all gram interactions except phonetic (completely omitted))	lmer(comp~grammar*phonemic+grammar*phonetic +grammar*stress +lexis +(1+lexis+phonetic+phonemic+grammar|listener) +(1|speaker), data=Data)

RESULTS FOR GRAM MODEL COMPARISON:

Comparison	Anova	Results	Best-fit Model
without interactions /with all interactions	Model-0/gramM1	χ2=63.755; df=4; p=4.706e-13 ***	gramM1
with all interactions except phonemic (stand-alone) /with all interactions	gramM2/gramM1	χ2=5.6163; df=1; p=.01779 *	gramM1
with all interactions except stress (stand-alone) /with all interactions	gramM3/gramM1	χ2=46.14; df=1; p=1.101e-11 ***	gramM1
with all interactions except inton (stand-alone) /with all interactions	gramM4/gramM1	χ2=9.5538 df=1; p=.001995 **	gramM1
with all interactions, except phonetic (stand-alone) /with all interactions	gramM5/gramM1	χ2=0.6635; df=1; p=.4153	gramM5
with all interactions, no phonetic /with all interactions, except phonetic (stand-alone)	gramM5b/gramM5	χ2=.0563; df=1; p=.8125	gramM5b

(3) **INTERACTIONS FOR LEXIS AND GRAMMAR TOGETHER** (the following models include interactions between grammar and the pronunciation factors AND between lexis and pronunciation factors)

MODEL NAME & DESCRIPTION	Syntax
gramANDlexBestFitM (best fit lexis and grammar models combined)	lmer(comp~grammar*stress+grammar*phonemic+grammar*inton +lexis*phonetic+lexis*phonemic +(1+lexis+phonetic+phonemic+grammar|listener) +(1|speaker), data=Data)
ManualgramANDlexBestFitM (manual best fit lexis and grammar models combined (without negative interaction))	lmer(comp~grammar*phonemic+grammar*inton +lexis*phonetic+lexis*phonemic +(1+lexis+phonetic+phonemic+grammar|listener) +(1|speaker), data=Data)

RESULTS FOR GRAM/LEX MODEL:

Comparison	Anova	Results	Best-Fit Model
With lex best-fit interactions /with lex and gram best-fit interactions	lexM5b /gramANDlexBestFitM	χ2=47.952; df=5; p=3.633e-09 ***	gramANDlexBestFitM
with gram best-fit interactions /with lex and gram best-fit interactions	gramM5b /gramANDlexBestFitM	χ2=26.327; df=3; p=8.147e-06 ***	gramANDlexBestFitM

RESULT:- BEST FIT FOR COMPREHENSIBILITY IS ALL INTERACTIONS btw LEXIS and PRONUNCIATION FACTORS except stress and inton, AND ALL INTERACTIONS btw GRAMMAR and PRONUNCIATION FACTORS except phonetic

(4) COMPARISON WITH ACCENTEDNESS

First find best-fit accentedness Model then compare with best-fit pronunciation factor Model

AccModel-0: contained all fixed factors, but NO interactions:
CompAllFixedAccNOINTER.model <- lmer(comp~lexis+accent+grammar +(1+lexis+accent+grammar|listener)+(1|speaker),data=Data)

MODEL NAME & DESCR	**SYNTAX**
AccModel-0 (no interactions)	lmer(comp~lexis+accent+grammar +(1+lexis+accent+grammar|listener) +(1|speaker), data=Data)
lexM1-acc (accent/lexis interaction, stand-alone grammar)	lmer(comp~lexis*accent +grammar +(1+lexis+accent+grammar|listener) +(1|speaker), data=Data)
gramM1-acc (accent/gram interaction, stand-alone lexis)	lmer(comp~grammar*accent +lexis +(1+lexis+accent+grammar|listener) +(1|speaker), data=Data)

RESULTS:

Comparison	Anova	Results	Best-fit model
No interactions/ lex interactions, stand-alone grammar	AccModel-0/ lexM1-acc	χ2=.0933; df=1; p=.7601	AccModel-0
No interactions/ gram interactions, stand-alone lexis	AccModel-0/ gramM1-acc	χ2=.12.445; df=1 p=.0004191 ***	gramM1-acc (renamed acc-comp-best-fit)
Accentedness vs pronunciation factors	acc-comp-best-fit / pron-comp-best-fit	χ2=103.83; df=18; p=4.371e-14 ***	pron-comp-best-fit
Accentedness vs pronunciation factors (manual best-fit)	acc-comp-best-fit / pron-comp-manual-best-fit	χ2=1276.7; df=11; p=< 2.2e-16***	pron-comp-manual-best-fit

RESULT:- BEST FIT FOR COMPREHENSIBILITY IS PRONUNCIATION FACTORS, NOT ACCENTEDNESS

APPENDIX G: CONSENT FORMS FOR SPEAKER AND LISTENER PARTICIPANTS

1) University of York INFORMATION SHEET & CONSENT FORM:

UNIVERSITY of York

DEPARTMENT OF
LANGUAGE AND
LINGUISTIC SCIENCE
Heslington, York, YO10 5DD, UK
Email cp1034@york.ac.uk

INFORMATION SHEET

PLEASE KEEP THIS INFORMATION SHEET AND A SIGNED COPY OF THE CONSENT FORM FOR YOUR RECORDS

You are invited to take part in a research study. Before you decide whether to participate it is important for you to understand why the research is being done and what it will involve. Please take the time to read the following information carefully. If there is anything you do not understand, or if you want more information, please ask the researcher.

Title of study: The Effects of Proficiency on the Comprehensibility of Second Language Speech

Researcher: Catherine Pease

What is the research about?

The research project tries to establish what actual properties of accented running speech influence a listener's perception of accent.

What is carrying out the research?

Catherine Pease (M.A. Student at the University of York)

Who can participate?

You can take part if you are either a native Arabic speaker who can speak English at an intermediate level or above, or if you are a native speaker of English.

What does the study involve?

The study involves two parts. Speaker participants will be asked to read speech samples and to take part in a short map game with another participant. Audio recordings will be taken during one 40-minute recording session. Listener participants will be asked to give ratings of the speech samples and to write down the sentences they hear. They will also be asked to perform a word association task. This will take place during two 20-minute sessions. In both cases the author will be collecting the data.

Do I have to take part?

You do not have to take part in the study. If you do decide to take part you will be given this information sheet to keep and will be asked to sign two copies of the consent form (one copy is for you to keep). If you decide to take part you will still be free to withdraw without giving a reason, even during the session itself. If you withdraw from the study, we will destroy your data and will not use it in any way.

What are the possible risks of taking part?

There are no risks to taking part.

Are there any benefits to participating?

Should they wish, participants who are native speakers of Arabic (speaking English as a second language) will be provided with feedback on their pronunciation based on the analysis of their speech samples, with the aim of providing goals which they can work towards.

What will happen to the data I provide?

The data you provide will be used alongside the data of other participants as part of a study on second language speech. Your data will be stored securely in the University of York, Department of Language and Linguistic Science, and may be used in future research.

What about confidentiality?

Your identity will be kept strictly confidential. No real names will be used in any presentations or publications, or in my dissertation.

Will I know the results?

Once my dissertation has been published, this will be made available to any participant who requests it.

*This study has been reviewed and approved by the Departmental Ethics Committee of the Department of Language and Linguistic Science at the University of York. If you have any questions regarding this, you can contact the chair of the L&LS Ethics Committee, Traci Walker (**email:** traci.walker@york.ac.uk; **Tel:** (01904) 323611).*

If you have further questions regarding this study, please feel free to contact:

Catherine Pease
Department of Language and Linguistic Science
University of York, Heslington, York, YO10 5DD
email: cp1034@york.ac.uk

Supervisors name and details:
Dr Sam Hellmuth,
Department of Language and Linguistic Science
University of York, Heslington, York, YO10 5DD
tel: (01904) 322657
email: sam.hellmuth@york.ac.uk

The Effects of Proficiency on the Comprehensibility of Second Language Speech

Lead researcher: Catherine Pease

iv) Consent form

This form is for you to state whether or not you agree to take part in the study. Please read and answer every question. If there is anything you do not understand, or if you want more information, please ask the researcher.

Have you read and understood the information sheet about the study?	Yes ☐	No ☐
Have you had an opportunity to ask questions about the study and have these been answered satisfactorily?	Yes ☐	No ☐
Do you understand that the information you provide will be held in confidence by the research team, and your name or identifying information about you will not be mentioned in any publication?	Yes ☐	No ☐
Do you understand that you may withdraw from the study at any time before the end of the data collection session without giving any reason, and that in such a case all your data will be destroyed?	Yes ☐	No ☐
Do you understand that the information you provide may be kept after the duration of the current project, to be used in future research on language?	Yes ☐	No ☐
Do you agree to take part in the study?	Yes ☐	No ☐
Do you agree to excerpts from your audio recordings being used in presentations or in teaching by the researcher, without disclosing your	Yes ☐	No ☐

real name?
(You may take part in the study without agreeing to this).

Do you agree to the researcher keeping your contact details after the end of the current project, in order that s/he may contact you in the future about possible participation in other studies? Yes ☐ No ☐
(You may take part in the study without agreeing to this).

Your name (in BLOCK letters):
____
Your signature:
____

Researcher's name: Catherine Pease

2) United Arab Emirate University INFORMATION SHEET & CONSENT FORM:

Social Sciences Research Ethics Committee

- Consent to Participate in a Research Study-

Please read carefully before signing the Consent Form!

Effects of Language Proficiency on the Comprehensibility of Speech

You will be asked to provide or deny consent after reading this form.

Topic of the research, the researcher(s) and the location

You have been invited to take part in a study to investigate the effects of foreign accented speech.
This study will be conducted by Catherine Pease from the Department of Language and Linguistic Science, University of York, UK.

The study will take place at the United Arab Emirates University in Al-Ain, in the Phonetics Laboratory of the Department of Linguistics.

Participation in this study will take approximately 30 minutes – 5 minutes for set-up and explanation, 20 minutes for the experiment, and 5 minutes for a discussion with the researcher.

Compensation (if applicable)

You will be offered a feedback report for use in your study of English for your time. Should you withdraw from the study, you will still be offered this report for your time.

Benefit of the research

The benefit of this study to the participant will be in the increase in awareness of the areas of language addressed, as well as customised report on ways to improve their proficiency in English.

Procedure/setting

Speech recordings to be carried out in a laboratory.

About the Experiment

The experiment involves recording the speech of participants in controlled tasks.

Safety Information

There is no risk, either physical or psychological, to the participant.

Confidentiality and Privacy Information

All personal information will be treated confidentially. If the results of the experiment are published, the y shall not in any way be traceable to specific individuals.

Right to Withdraw

Participants may withdraw at any stage in the process without being penalized.

Informed Consent

v) I confirm that I have read and understood the above information sheet and have had the opportunity to ask questions.
vi) I understand that my participation is voluntary and that I am free to withdraw.
vii) I understand that my data will be kept confidential and if published, the data will not be identifiable as mine.

I agree to take part in this study:

(Name and signature of participant)	(Date)
(Name and signature of person taking consent)	(Date)
(Name and signature of witness (if participant unable to read/write)	(Date)
(Name and signature of parent/guardian/next of kin (when participant unable to give consent due to age or incapacity)	(Date)

List of Abbreviations

Contrast Regularity Index (CRI)

Dynamic Systems Theory (DST)

English as a Foreign Language (EFL)

English as an International Language (EIL)

English as a Second Language (ESL)

Fundamental Frequency F0)

1st Formant (F1)

2nd Formant (F2)

General American (GA)

Human Communication Research Centre (HCRC)

intermediate phrase (ip)

Intonational Phrase (IP)

International Phonetic Alphabet (IPA)

Intonational Variation in English (IViE)

First Language (L1)

Second Language (L2)

Second Language Intonation Learning Theory (LiLT)

Multiple Forced Choice (MFC)

Mean Length Run (MLR)

Durational Variability of Successive Pairs of Vocalic Intervals (nPVI-V)

Second Language Perceptual Assimilation Model (PAM-L2)

Prosodic Word (PW)

Received Pronunciation (RP)

Standard Southern British English (SSBE)

Tones and Break Indices (ToBI)

United Arab Emirates University (UAEU)

Rate-Normalised Standard Deviation of Vocalic Interval Duration (VarcoV)

Percentage of Vowel (%V)

Voice Onset Time (VOT)

References

Algethami, G. (2013). An Acoustic and Perceptual Investigation of English Speech Rhythm among L2 Saudi Speakers. Unpublished PhD Thesis.

Almbark, R. (2012). The Perception and Production of SSBE Vowels by Syrian Arabic Learners: The Foreign Language Model. Unpublished PhD Thesis.

Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., & Sotillo, C. (1991). The HCRC Map Task Corpus. *Language and speech*, *34*(4), 351-366.

Anderson-Hsieh, Koehler, Johnson (1992). The relationship between native speaker judgements of non-native pronunciation and deviance in syllables, prosody and syllable structure. *Language Learning Journal 42 (4): 529-555.*

Baayen, D.H., Davidson, D.J., Bates, D.M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language 59, 390–412.*

Bates, D., Maechler,M., Bolker, B., Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67(1), 1-48.*

Beckman, M.E., Pierrehumbert, J. (1986). Intonational Structure in English and Japanese. Phonology Yearbook 3: 255-310.

Beinhoff, B. (2014). Perceiving Intelligibility and Accentedness in Non-Native Speech: A Look at Proficiency Levels. *Proceedings of the International Symposium on the Acquisition of Second Language Speech Concordia Working Papers in Applied Linguistics, 5, 2014.*

Best, C. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In C. Goodman & H. Nusbaum (Eds.), The Development of speech perception: The transition from speech sounds to spoken words, Cambridge, MA: MIT Press, 167-224.

Boersma, P. & Weenink, D. (2013). Praat: doing phonetics by computer (Version 6.0.19) [http://www-.praat.org]

Blodgett, A., Owens, J., Rockwood, T. (2007). An Initial Account of the Intonation of Emirati Arabic. *Proceedings of the 16th ICPhS, Saarbrücken, Germany, 1137-1140*.

Brown, C. (2000). The interrelation between speech perception and phonological acquisition from infant to adult. In J. Archibald (ed.), Second language acquisition and linguistic theory, Oxford: Blackwell, 4-63.

Chalal, R., Hellmuth, S. (2014). The Intonation of Lebanese and Egyptian Arabic (in Prosodic Typology II, ed. by Sun Ah Jun).

Colantoni, L., Steele, J., Escudero, P. (2015) Second Language Speech: Theory and Practice. Cambridge University Press.

Cruttenden, A. (1997). Intonation. Cambridge University Press.

Crystal, D. (1975). Prosodic Features and linguistic theory, in *The English Tone of Voice*, Edward Arnold.

De Bot, K. (2008). Second Language Development as a Dynamic Process. *The Modern Language Journal, Vol. 92, No. 2, pp. 166-178*

Derwing, T., Munro, M. (1997). Accent, Intelligibility and Comprehensibility: Evidence from Four L1s. *Studies in Second Language Acquisition, 19/1: 1-16.*

Derwing, T.M., Munro, M.J. (2005). Second Language Accent and Pronunciation Teaching: A Research-Based Approach. *Tesol quarterly, Volume 39, Issue 3, 379–397*

Derwing, T.M., Munro, M.J. (2015). Pronunciation Fundamentals: Evidence-based Perspectives for L2 Teaching and Research. John Benjamins Publishing Company.

Di Paolo, M., Yaeger-Dror, M. (Eds) (2011). Sociophonetics. A Student's Guide. Routledge.

Dilley, L.C. (2005) The Phonetics and Phonology of Tonal Systems. Unpublished PhD Thesis.

Dupoux, E. & Peperkamp, S. (2002). A typological study of stress 'deafness'. In

Gussenhoven, C. & Warner, N. (eds.) *Laboratory Phonology 7*. Berlin: Mouton de Gruyter, 203-240.

Elman, J.L. (1995). Language as a dynamical system. In Port, R.F., van Gleder,T. (eds.) Explorations in the Dynamics of Cognition: Mind in Motion. MIT Press.

Escudero, P. (2009). Linguistic Perception of 'similar' L2 sounds. In Boersma, P. & Hamann, S. (Eds.). *Phonology in Perception (pp151-190).* Berlin, Mouton de Gruyter.

Flege, J. (1995). Second language speech learning theory, findings and problems. In Strange, W. (ed.) Speech perception and linguistic experience: Issues in cross-language research, Baltimore, MD: York Press, 233-277.

Flege, J. E., & Port, R. (1981). Cross-language phonetic interference: Arabic to English. *Language and speech*, *24*(2), 125-146.

Ghazali, S., Hamdi R., Barkat, M. (2002). Speech Rhythm Variation in Arabic Dialects. *Speech Prosody, 2002, 331-334.*

Ghazali, S., Bouchhioua (2003). The Learning of English Prosodic Structures by Speakers of Tunisian Arabic: Word Stress and Weak Forms. *ICPhS-15*, 961-964.

Grabe, E., Post, B., Nolan, F., Farrar, K. (2000). Pitch Accent Variation in four varieties of British English. *Journal of Phonetics, 28, 161-185.*

Grabe, E. (2001). The IViE Labelling Guide. (Version 3.0) http://www.phon.ox.ac.uk/IViE/guide.html.

Grabe, E., Kochanski, G., Coleman, J. (2005). The Intonation of Native Accent Varieties in the British Isles: Potential for Miscommunication? In Dziubalska-Kolaczyk, K.,

Przedlacka, J. (Eds.) English pronunciation models: a changing scene. Linguistic Insights Series, Peter Lang.

Gussenhoven, C. (2004) The Phonology of Tone and Intonation. Cambridge University Press.

Hardcastle, W. and Hewlett, N. (2006). *Coarticulation: Theory, Data, and Techniques*. Cambridge University Press.

Hayes, B. (2009) Introductory Phonology. Wiley & Blackwell.

Hellmuth, S. (2006). Intonational Pitch Accent Distribution in Egyptian Arabic. Unpublished PhD Thesis, SOAS.

Hellmuth, S. (2007). The relationship between prosodic structure and pitch accent distribution: evidence from Egyptian Arabic. *The Linguistic Review, 24*, 289-314.

Hellmuth, S. (2011). Accent Distribution Effects in L2 English: Causes and Implications

Hellmuth, S. (2013). Phonology in Owens, J. (ed.) The Oxford Handbook of Arabic Linguistics. Oxford University Press.

Holes, C. (1990). Gulf Arabic. Routledge.

Huang, B., Alegre, A., Eisenberg, A. (2016). A Cross-Linguistic Investigation of the Effect of Raters' Accent Familiarity on Speaking Assessment. *Language Assessment Quarterly 2016*, Vol. 13, NO. 1, 25–41.

Ioup, G. (2008). Exploring the role of age in the acquisition of second language phonology. In Hansen Edwards, J.G., Zampini, M. L. (eds.), Phonology and Second Language Acquisition. John Benjamins Publishing Company.

Isaacs, T. and Trofimovich, P. (2012). Deconstructing Comprehensibility: Identifying the Linguistic Influences on Listeners' L2 Comprehensibility Ratings. *Studies in Second*

Language Acquisition, Volume 34 Issue 3, 475-505
Jelinek, E. (2002). Agreement, Clitics and Focus in Egyptian Arabic. In Ouhalla, J. & Shlonsky, U. (Eds.), Themes in Arabic and Hebrew Syntax, pp.71-105. Springer Verlag.
Jun, S-A. (2014). Prosodic typology: by prominence type, word prosody, and macro-rhythm. In S-A. Jun (Ed.), Prosodic Typology Volume II. (pp. 520-539). Oxford: Oxford University Press.
Khattab, G. (2002). VOT production in English and Arabic bilingual and monolingual children. *Amsterdam Studies in the Theory and History of Linguistic Science Series 4*, 1-38.
Kuhl, P. (1992). Psychoacoustics and speech perception: Internal standards, perceptual anchors, and prototypes. In L. Werner & E. Rubel (eds.), Developmental psychoacoustics, Washington, DC: American Psychological Association, 293-332.
Ladd, R. (1993). On the theoretical status of "the baseline" in modelling intonation. *Language and Speech* 36: 435-451
Ladd, R. (2008). Intonational Phonology. Cambridge University Press.
Ladefoged, P. (2001). Vowels and consonants: An introduction to the sounds of languages. Oxford: Blackwells.
Ladefoged, P., Johnson, K. (2011). A course in Phonetics. 6th Ed. Wadsworth/Cengage Learning.
Lenneberg, E. (1967). Biological foundations of language. New York: John Wiley & Sons.
Mattys, S.L., White, L., Melhorn, J.F. (2005). Integration of Multiple Speech Segmentation Cues: A Hierarchical Framework. *Journal of Experimental Psychology. Vol. 134, No. 4, 477–500.*
Mennen, I. (2006). Phonetic and phonological influences in non-native intonation: an overview for language teachers. QMUC Speech Science Research Centre Working Paper.
Mennen, I. (2015). Beyond Segments: Towards a L2 Intonation Learning Theory. *In Prosody and Language in Contact (pp171-188). Springer Verlag.*
Mertens, P. (2004). The Prosogram: Semi-Automatic Transcription of Prosody based on a Tonal Perception Model. *Proceedings of Speech Prosody 2004, Nara, Japan, pp 23-26.*
Mitleb, F. M. (1984). Vowel length contrast in Arabic and English: A spectrographic test. *Journal of Phonetics.* 12: 229-35

Munro, M. J. (2008). Foreign Accent and Speech Intelligibility. In Hansen Edwards, J.G., Zampini, M. L. (eds.), Phonology and Second Language Acquisition. John Benjamins Publishing Company.

Munro, M. J., & Derwing, T. M. (1995a). Foreign Accent, Comprehensibility, and Intelligibility in the Speech of Second Language Learners. *Language Learning 45 p73-97*

Munro, M. J., & Derwing, T. M. (1995b). Processing Time, Accent and Comprehensibility in the Perception of Native and Foreign-Accented Speech. *Language and Speech, 38*(3), 289-306.

Munro, M. J., & Derwing, T. M. (2001). Modelling Perceptions of the accentedness and comprehensibilty of L2 speech. *Studies in Second Language Acquisition, 23, 451-468.*

Newman, D. (2002). The phonetic status of Arabic within the world's languages: the uniqueness of the lughat al Daad. *Antwerp Papers in Linguistics, 100. p65-75.*

Ogden, R. (2009). An introduction to English Phonetics. Edinburgh University Press.

Patel, A.D., Iversen, J.R., Rosenberg, J.C. (2006). Comparing the rhythm and melody of speech and music: The case of British English and French. *Journal of the Acoustical Society of America. 119, 3034.*

Pennington, M.C., & Richards, J.C. (1986). Pronunciation revisited. *TESOL Quarterly Vol. 20, No.2.*

Pierrehumbert, J. (1980). The Phonology and Phonetics of English Intonation. PhD thesis, MIT. Published 1988 by IULC, Bloomington, IND.

Porretta, V., Kyröläinen, A-J., Tucker, B.V. (2015). Perceived foreign accentedness: Acoustic distances and lexical properties. *Atten Percept Psychophys (2015) 77: 2438–2451.*

Post, B., & Delais-Roussarie, E. (2006). Transcribing intonational variation at different levels of analysis. *Proceedings of Speech Prosody 2006, Dresden*.

R Core Team, (2016). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria.* URL https://www.R-project.org/.

Saito, K., Trofimovich, P., & Isaacs, T. (2015). Using listener judgements to investigate linguistic influences on L2 comprehensibility and accentedness: A validation and generalisation study. *Applied Linguistics 2015: 1-25*.

Saito, K., Webb, S., Trofimovich, P., & Isaacs, T. (2016). Lexical correlates of comprehensibility versus accentedness in second language speech. *Bilingualism: Language and Cognition*, *19*, 597-609.

Selkirk, E. (1982). The Syllable. In Goldsmith, J.A. (Ed., 1999), Phonological Theory. The Essential Readings. Oxford: Blackwell Publishers.

Shrout, P.E., & Fleiss, J.L. (1979). Intraclass Correlations: uses in Assessing Rater Reliability. *Psychological Bulletin, Vol. 86, No. 2, 420-428.*

Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., Hirschberg, J. (1992). ToBI: A Standard for Labeling English Prosody. *ICSLP 92.*

Smith, L. E., & Nelson, C. (1985). International intelligibility of English: directions and resources. *World Englishes, 4 (3), 333-342.*

So, C. K., Best, C. T. (2010), Cross-language Perception of Non-native Tonal Contrasts: Effects of Native Phonological and Phonetic Influences. *Language and Speech, 2010, 53(2), 273–293.*

Thelen, E. & Smith, L.B. (2007). Dynamic Systems Theories. *Handbook of Child Pyschology. 1:6.*

Turk, A., Shattuck-Hufnagel, S. (2013) What is speech rhythm? A commentary on Arvaniti and Rodriguez, Krivokapic, and Goswami and Leong. *Laboratory Phonology 2013; 4(1): 93-118.*

Watson, J. (2002). The Phonology and Morphology of Arabic. Oxford University Press.

Wells, J.C. (2006). English Intonation: An Introduction. Cambridge University Press.

White, L., Mattys, S. (2007). Calibrating rhythm: First and second language studies. *Journal of Phonetics, 35, 501-522*

Wichmann, A. (2000). Intonation in text and discourse: Beginnings, middles and ends. Longman.

Wiget, L., White, L., Schuppler, B., Grenon, I., Rauch, O., Mattys, S. (2010). How stable are acoustic metrics of contrastive speech rhythm? *The Journal of the Acoustical Society of America*, *127*(3), 1559-1569.

Winter, B. (2013). Linear models and linear mixed effects models in R with linguistic applications. arXiv:1308.5499. [http://arxiv.org/pdf/1308.5499.pdf]

Whelan, R. (2008). Effective Analysis of Reaction Time Data. *The Psychological Record, 58, 475-482*

White,L. (2014) Communicative function and prosodic form in speech timing. *Speech* *communication, Vol. 63-64, 38-54*

Yaneer Bar-Yam. (1997). Dynamics of complex systems. Perseus Books. Reading, Massachusetts.

Zielinski, B. W. (2006b). Reduced intelligibility in L2 speakers of English. Unpublished PhD Thesis, La Trobe University, Bundoora.