Migrant Pronunciation:
What do Employers find Acceptable?

by

Marty Pilott

A thesis

Submitted to the Victoria University of Wellington

in fulfilment of the requirements for the degree of

Doctor of Philosophy

Victoria University of Wellington

2016

CONTENTS

LIST OF FIGURES

LIST OF TABLES

List of Models

Notes

Unlabelled references in bold (e.g. **7.2.1**) are numbers of sections within this thesis.

Pronunciation features used in this thesis, and Global Measures of pronunciation (e.g. Acceptability), as described in **3.2**, are written with a capital when necessary to distinguish them from other uses.

Statistical concepts are explained in **7.1.1**.

ABSTRACT

Many migrants from non-English speaking backgrounds need to look for unskilled and low-skilled work in a range of industries in New Zealand. A number of barriers to employment have been identified, amongst which are numerous reports of migrants' difficulties with English language. As many of these migrants speak languages which are very different from English they often have accents which native speakers find unfamiliar or hard to understand.

The aim of this thesis was to identify which features of migrant pronunciation are more or less acceptable to employers, so that ESOL professionals can assist migrants more effectively when learning English.

In my literature review, I begin by placing the need for pronunciation teaching in the context of New Zealand's rapidly increasing ethnic and language diversity, and the barriers to migrant employment. I then describe how proficiency in pronunciation is measured, and how acceptability fits in to these measurements. In the following chapter I discuss what predictions can be made about the features which are likely to cause pronunciation difficulties for current groups of learners. Finally, I review research on the main levels of pronunciation (segmental, prosody and fluency) and how these can be expected to predict Acceptability.

My research questions were:

1. Does pronunciation affect employers' assessments of the acceptability of migrants for employment?
2. If so, which pronunciation features are the strongest predictors of the assessment?
3. Are there other employer or speaker factors which affect employer ratings of acceptability?

To address these questions, I obtained speech samples from 40 male and female migrants from a range of L1s. These were rated through an online survey by employers or human resource managers (n=95) from industries employing low-skilled workers in the main centres around New Zealand. In addition, ratings of the speakers' pronunciation features were obtained from Experienced Raters to use as a baseline for analysis.

The data from the online survey was then analysed to determine which features predicted the employers' acceptability ratings. Three factors comprising a wide range of pronunciation

features, segmental and suprasegmental (the latter divided into prosodic and fluency), were found to be highly significant in the employers' ratings of acceptability, while most other factors were not. However, parts of the survey found that acceptability was significantly influenced by the employers' assessment of whether the speaker was a hard worker, and by employers' judgements about some of the speakers' pronunciation features.

In addition, judgements of acceptability differed from those of the other global measures of intelligibility, comprehensibility and accentedness, leading to a re-evaluation of what the ultimate goal of pronunciation teaching should be. While the Intelligibility Principle has been emphasised recently, it does not acknowledge the contextual nature of communication. This thesis concludes that Acceptability is a more useful goal for language learners, and that this can be achieved by including all levels of pronunciation including fluency.

This thesis proposes that ESOL pronunciation programmes, which typically focus on developing intelligibility through segmental and some prosodic features, also need to meet acceptability standards by including fluency features from the earliest stages, such as appropriate use of Pausing, Variety, and Smoothness.

1 INTRODUCTION

1.1 MOTIVATION

I arrived in New Zealand with my family in 1964. We were white, indistinguishable from other Pakeha[1] New Zealanders, and from Britain, the source of most Pakeha migrants. The reception we received, however, was unexpected. It seemed that New Zealand was kicking against its dependence on Britain and using English migrants as a target. As soon as I spoke I was identified as a "Pom[2]", and mocked accordingly. At lunchtimes the fish and chip shop near my school was full of children ordering their meals; as soon as I called out "One shilling fish and chips" I would hear a loud chorus of "cheeps, cheeps!" My 6-year-old brother's teacher corrected his pronunciation, telling him to say "Iddie" instead of "Eddie". At high school even my social studies teacher, while teaching the curriculum-required history and geography of Britain, found every opportunity to make fun of Pommie third-formers. Events which might have been trivial to an adult were less easy for a child to cope with. As late as 1972, broadcaster Tim Bickerstaff gained both laughter and outrage from his "Punch a Pom a day" campaign (Gifford, 2009). As a result, I gained some insight into the migrant's point of view, and also an awareness of language and the effect of accent. I never lost mine.

Many years later, while teaching ESOL and managing language schools, I noticed that some students could study English for years yet still have barely intelligible pronunciation. They wanted to interact with New Zealanders, apply for jobs or go on to further study, but their accents would surely hinder them. Pronunciation teaching, as I knew it, was little more than getting segmentals and word stress right, but I was not sure whether they were even learning that. What did teachers need to know in order to ensure the best outcome for their students?

From this I realised that the goal of language teaching should not just be to meet the criteria of a curriculum or an exit test, but the criteria of real life – communicating effectively with whoever the learners would be interacting with or applying to for work. What features of pronunciation would make a difference to those people? Was it enough for students to be

---

[1] White people, usually of British origin.
[2] A mildly derogatory term for English people.

intelligible to ESOL teachers? This led me to the goal of discovering which migrant pronunciation features would be acceptable to employers. There has been surprisingly little research on this topic: a recent investigation, *The Effect of Foreign Accent on Employability* (Timming, 2016), noted that very few studies have focused on discrimination on the basis of foreign accent or whether context affects acceptability (in Timming's study, hireability in the context of dealing with customers or not).

In addition, this is an age of accountability. Government departments no longer ask whether education providers are using "best practice"; they want evidence that programmes are producing the intended outcomes. Providers of English language programmes are now asked to prove that learners meet the "graduate profile" and that their subsequent work or study shows they possess the skills taught. At present there is no evidence at all that the pronunciation skills taught (or not taught) in ESOL classes up and down the country produce graduates who can communicate acceptably with employers. Anyone in the ESOL industry would find this information valuable.

My goal was approved by eminent linguist James Emil Flege, who wrote "What you want to do is to determine whether, and to what extent, potential employers' evaluation of speech samples might relate to [intelligibility, foreign accent and comprehensibility]. That, on the surface, would seem to be an interesting and, indeed, practical goal. If to be considered worthy of being hired an individual would have to be readily highly intelligible but not necessarily comprehensible (to cite one of many purely logical possibilities), then the goal of language teachers would be to focus their efforts on the former and not the latter" (2011, personal email).

Speech may have high intelligibility yet still be unacceptable to the listener for a variety of reasons (**3.3**). For everyday conversation the result may be no more than irritation, but employers are gatekeepers whose attitudes have high-stakes outcomes, and ESOL teachers and course designers need to be aware of their attitudes so that English programmes can include appropriate pronunciation content. Fraser (2000, p. 7) wrote that "Pronunciation is the aspect that most affects how the speaker is judged by others, and how they are formally assessed in other skills" and Harrison (2014, p. 255) also commented "Employers routinely form initial impressions of job applicants on the basis of their linguistic presentation, using accent to make inferences about social group membership and level of competence." Some

of the former refugees interviewed for the speech samples used in this research would have little to put on a CV, and employers are likely to use immediate clues such as how well the applicant speaks.

As yet there has been little research to provide advice on pronunciation teaching for this purpose. Isaacs (2014, p. 12) wrote "To date, only a handful of empirical studies have examined which pronunciation features are most important for intelligibility and comprehensibility." Much overseas research has been on general native speaker (NS) attitude to non-native accents (Rich, Maeda, Cargile, & Rodriguez, 2010), but some of it has been focused on the effect on employment of having a foreign accent – in Australia (Ho & Alcorso, 2004; Tombs, Rao Hill, Lee, & Lings, 2014), Canada (Hekman et al., 2010; Krahn, Derwing, Mulder, & Wilkinson, 2000; Xu, 2006), the United States (Cargile, 2000; Segrest Purkiss, Perrewé, Gillespie, Mayes, & Ferris, 2006), and Britain (Schellekens, 2001; Shields & Price, 2001). Most of this research points to the disadvantages faced by migrants on account of attitudes towards their accents, although there are occasional reports of positive stereotyping as well (Cargile, 2000) as a result of favourable attitudes towards the migrant group.

I believed, therefore, that knowing how employers rate migrant pronunciation would improve ESOL teaching and benefit learners.

1.2   Scope

This thesis focuses on migrants to New Zealand who have learned some English and are likely to seek low or unskilled employment, and the acceptability of their accents by potential employers. Underlying the justification for such a study and its conclusions is an extensive body of research on English pronunciation, what it consists of, how it is learned, why adults struggle to learn it, what specific problems are faced by speakers of different languages, and which features they and their teachers ought to focus on. This of course raises the question "Which English?" since it is always possible to find a dialect which belies any general statement about English pronunciation.

Firstly, the English referred to here is not ELF (English as a Lingua Franca) or International English, which is now seen as a newly developing dialect with its own phonological rules (Christiansen, 2014). The participants interviewed are all migrants who have settled in New Zealand and need to work here and communicate with others who live here. The focus is on ESL (English as a Second Language), EAL (English as an Additional Language), or ESOL (English for Speakers of Other Languages) as it is usually known in New Zealand. This means that ELF (Jenkins, 2000) is not relevant, although articles about the subject are used when they contain useful information.

The employers I am concerned with are any who offer low or unskilled work, such as in supermarkets, rest homes, cleaning services, and manual jobs. These are the kinds of jobs which migrants look for when they lack qualifications (as is the case with many of refugee background), or their qualifications are not recognised in New Zealand, or their English proficiency is too low to work in their areas of expertise. They may also need to find unskilled part-time work while studying.

New Zealand demographics mean that most employers are native speakers of English, with 74% of the population described as being of European (usually British) origin (Statistics New Zealand, 2015b) and many others also being native speakers. Throughout this thesis – specifically, for the purpose of rating the pronunciation of migrant speech – I have in mind the dialects of English which are most widely used and heard in New Zealand, and which are likely to be the target, openly or tacitly, of English classes, and the language accepted by

New Zealand employers. Of course, this is an idealised situation, and there is no single description of these dialects. There are books and articles about New Zealand English (Bauer & Warren, 2008; Bell, Harlow, & Starks, 2005; Bell & Kuiper, 2000; Hay, Maclagan, & Gordon, 2008; Warren, 1998) and on approaches to teaching and learning English here, a country outside of the major English standards such as General American or RP / Standard Southern British English (Romova, Smith, & Neville-Barton, 2008; J. Smith, 2012; Tan, 2009). But the vast bulk of work on aspects of English pronunciation has inevitably been done in Britain, the United States and Canada, describing the standard dialects of those countries, and this thesis could not have been written without frequent reference to that research.

The limited worldwide use of New Zealand English means that its users and learners are surrounded by examples of the major dialects, from media or other immigrants. Surveys show that native speakers of New Zealand English are aware of these dialects and may even accord them higher status than New Zealand English (Bayard, Weatherall, Gallois, & Pittam, 2001), and ESOL students too are aware of them and may choose the accent they want to use (Kang, 2010a; Pilott, 1995; Ray & Zahn, 1999; Y. Watanabe, 2008). To this extent the model of English referred to in this thesis can be described as variationist (Pennington, 1996, p. 17) in acknowledging that a single standard does not exist.

1.3 AIMS

The research questions for this thesis are as follows:

RQ1 Does pronunciation affect employers' assessments of the Acceptability of migrants for employment?

RQ2 If so, which pronunciation features are the strongest predictors of the assessment?

RQ3 Once pronunciation features are taken into account, are there other employer or speaker factors which affect employer ratings of acceptability?

RQ1 taps into the broader issue of the lack of effective pronunciation teaching in ESOL classes (see section **3.1.1**) by reporting on how important pronunciation is for learners. If, for example, the majority of employers who responded to the Employer Survey (described in **1.4** below) found most of the migrants highly acceptable on the basis of their pronunciation, then there would be no call for additional or different pronunciation teaching. In fact, many of the migrants were given low acceptability ratings. Knowing this, teachers will then want to know what they need to focus on in order to improve learners' pronunciation. Previous studies asked raters, including employers, to judge whether groups of migrants were suitable for positions at various levels of work such as factory worker or manager.

RQ2 sought specific answers to this by asking how the employers' ratings of acceptability could be related (as far as possible) to an objective judgement of the migrants' pronunciation, including intelligibility and comprehensibility. This would reveal which features of pronunciation would predict higher acceptability ratings, and in turn guide teachers as to what they needed to teach. Three levels of pronunciation – labelled in this thesis as Segmental, Prosody and Fluency – were identified in the research, and all were found to be significant predictors of acceptability. This has implications for pronunciation teaching, since it often goes no further than segmentals, intonation and word stress.

RQ3 asked what other features would influence ratings of acceptability, as discussed in section **3.3**. The question refers solely to the context of audio communication, since the employers surveyed had no knowledge of the migrants other than how they spoke. I could therefore separate judgements based on pronunciation from those based on beliefs about

ethnicity or actual language pronunciation differences. None of these turned out to be significant predictors. I asked about other features and employer information such as type of work offered, which also had no predictive effects. Effects were found from some employer ratings of pronunciation and their interpretation of the speakers' personal qualities.

I also wanted to investigate the role of Acceptability as a Global Measure alongside Intelligibility, Comprehensibility and Accentedness (see **3.3**). The Intelligibility Principle, using Intelligibility in its broad meaning to include Comprehensibility (Isaacs, 2014, p. 8), sets as a goal for pronunciation teaching that listeners should be able to understand L2 speech. Would this then mean that an intelligible speaker would therefore be an acceptable speaker? If not, what does this say about the value of Intelligibility as a goal? And if the goal shifts to Acceptability, how can the teacher hope to know what is acceptable? These questions are answered in Chapter 8.

1.4 STRUCTURE AND SURVEY

Chapters 2-5 refer to a wide range of literature, but they do not simply offer a literature review. Their purpose is to build an understanding of the migrants arriving in New Zealand, their likely pronunciation learning needs, and how pronunciation attainment is measured.

Chapter 2 describes the huge increase in immigration to New Zealand since the 1980s, particularly from non-traditional sources such as China and the Middle East. It describes government immigration policies and concludes that there is a high likelihood that there will continue to be large numbers of migrants in need of English language tuition and assistance with pronunciation.

I also discuss the barriers which migrants often face and the difficulties they have in finding work, which is partly driven by their English proficiency. This provides evidence that providing quality pronunciation teaching would improve outcomes for migrants, particularly those of refugee background.

Chapter 3 discusses the field of pronunciation teaching and the changes that have taken place in what to teach and what the goals of teaching should be. I introduce Acceptability, the reasons why it should be viewed as a Global Measure alongside others such as Intelligibility and Comprehensibility, and why teachers should work towards it as a goal.

Chapter 4 examines the features of pronunciation and shows why English is a particularly difficult language to learn for many of the new migrants to New Zealand. Specific learning challenges for the migrants interviewed for the Employer Survey are described, including some such as the Myanmarese and Amharic speakers for whose languages very little data exists or is readily available.

Chapter 5 reviews what is known about the features of pronunciation under the three levels of Segmental, Prosody and Fluency (the latter two usually grouped as Suprasegmentals) and asks what effect each feature is likely to have on predicting Acceptability. The outcome of this thesis is that all levels are significant and that Fluency is highly important, although it is defined in many different ways and its features have a variety of descriptions.

Chapter Six describes the methodology used. A survey (in one general and two specific parts) was designed, and implemented by recording migrants, finding employers to rate

them for acceptability and other features, and using Experienced Raters to create a benchmark rating for comparison.

The statistical outcomes of the Employer Survey are described in Chapter 7. The *R* platform (R Core Team, 2015a) was used to run statistical models on the survey output, and each of the variables used was tested in the model.

I make reference throughout to the **Employer Survey**, **Experienced Raters, Base rating**, and the **3-Factor Model**. They are all explained in detail in Chapter 6 (Methodology) and the full surveys appear in **Appendix B**. Below is a brief description of these terms so that references to them can be understood.

I wanted to gather data about the acceptability to employers of the speech of migrants with a low level of English (Elementary to low Intermediate). To do this I recorded mock job interviews with 40 migrants and had 95 employers rate them (usually 5 samples each) through an online survey. There were two alternative surveys (Parts A and B) to which the Employers were randomly assigned, asking different questions, although all Employers answered the same questions about themselves and the overall Acceptability question. This, as a whole, is referred to as the **Employer Survey**.

Five experts rated the migrants' speech, and they are referred to as the **Experienced Raters** (**6.2.3**). For reasons explained in Chapter 6, part of these ratings were unusable, so I completed the **Base Rating** (**6.5**, **7.2**) myself. A factor analysis of the features used in the Base Rating produced three Factors which I named **Segmental**, **Prosody** and **Fluency**. A description of this model appears in **7.2.2**, and an explanation of how the three factors are labelled and the justification for the Fluency construct appears in **7.2.3**.

To answer **RQ1** a statistical model was used to predict the Employers' ratings of acceptability by the Base Rating, and the output is referred to as the **3-Factor Model** (**7.3**). Other variables were then introduced to this model in order to answer **RQ2** and **RQ3**.

1.5 CONCLUSIONS

There are two key results from this thesis which researchers, course designers, ESOL teachers and learners will benefit from.

The first result is the proposed status of Acceptability as a Global Measure, and as the primary pronunciation goal for ESOL teaching. All levels of pronunciation (Segmental, Prosody and Fluency) were found to be significant in employer judgements of Acceptability, which suggests that being intelligible or comprehensible is not the only pronunciation criterion for a non-native speaker to be acceptable for employment.

The second result follows on from the first. Employers' judgements of Acceptability are influenced by all levels of pronunciation, not just segmentals and prosody. Segmental and Prosody were significant factors, but Fluency was the strongest predictor throughout. The implication for teaching is that learners ought to be introduced to Fluency features right from the start. Such an approach should greatly improve their language development, because speakers perceived as more fluent get more speaking time with native speakers, and their abilities will be reinforced.

It is therefore suggested that aiming for Acceptability and teaching all levels of pronunciation from the earliest stages should greatly improve the pronunciation outcomes and therefore the communication and employment prospects for migrants in New Zealand.

2 BACKGROUND

2.1 WELCOME TO NEW ZEALAND

2.1.1 A new multi-ethnic society

Since the 1980s, New Zealand has undergone a massive ethnic transformation, the full effects of which are still unknown. From a society which was predominantly Pakeha with smaller numbers of Maori and Pacific people, the country, especially in the larger centres, has seen an increasing admixture of general migrants and refugees from many parts of the world (Marlowe, Bartley, & Hibtit, 2014, p. 5) and the rate of immigration is increasing: the 2014-5 year saw a net migration gain of 58,300 people (from actual immigration of 63,900) – the highest net gain ever recorded (Ministry of Business, Innovation and Employment, 2015) compared to 44,008 people in 2013/4 (Figure **2.1-1**).

Figure 2.1-1 Increase in immigration 2005-15 (Statistics New Zealand, 2015c)
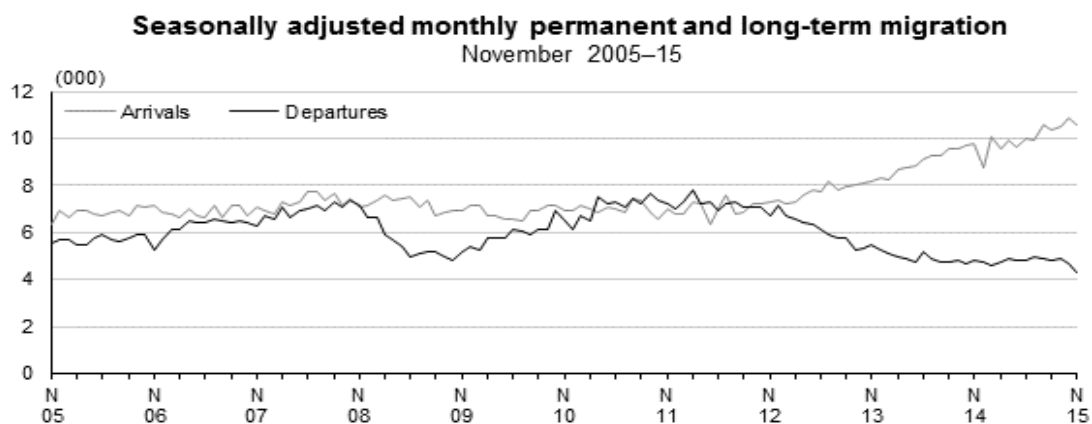
As a result, New Zealand has been described by Spoonley and Bedford (2012, pp. 11–13) as one of a small number of countries which are culturally and linguistically "superdiverse". The concept of superdiversity was introduced by Vertovec (2007) to underline the differentiation not only in the wide range of ethnic origins, languages and religions in Britain, but the "significant new conjunctions and interactions of variables" (p. 1025) such as differing immigration statuses, rights, and labour market experiences. Superdiversity indicates a "level of cultural complexity surpassing anything previously experienced" and will develop in New Zealand at a faster rate than in most nations (Royal Society of New Zealand, 2013, p. 1).

This development comes from increasingly large numbers of migrants entering New Zealand from new areas of the world, such as East Asia and the Middle East. Figure **2.1-2** shows the changes in three migrant groups since 1996, including both New Zealand born and immigrants.

Figure 2.1-2 Growth in minority ethnicities 1996-2013 (Statistics New Zealand, 2015a)
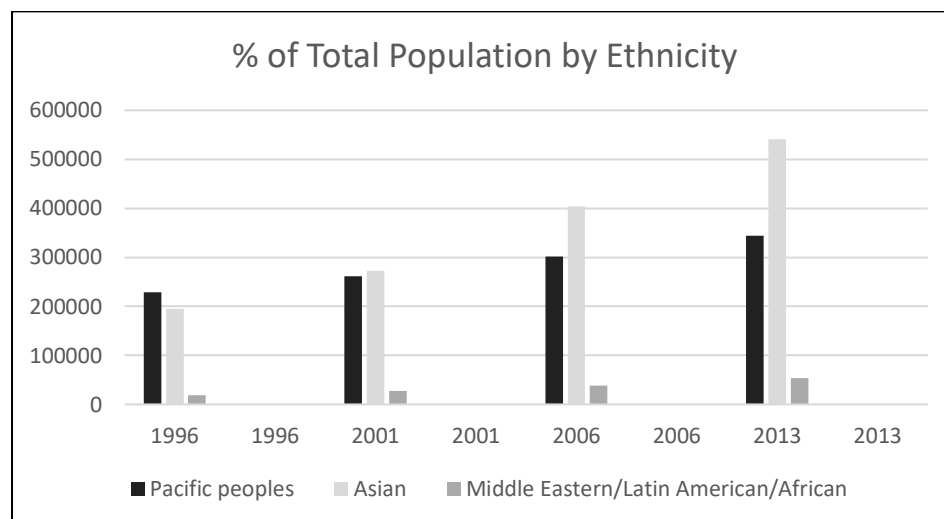
While the increase in Pacific populations from 1996 to 2013 has been slow and steady, there has been a dramatic rise in numbers from the Middle East, Latin America and Africa (MELAA) and from Asia. During this period the total New Zealand population increased by 19.03% from 3,732,000 to 4,442,100, while the number of European / Other[1] increased by less than 8%, Maori by over 20%, Pacific by 50%, Asian by nearly 178% and MELAA by nearly 188%. The Asian increase, from 194,800 to 541,300 or over one eighth of the whole population, is clearly the most substantial change (Statistics New Zealand, 2015a) and this diversity is expected to continue while the proportion of Europeans will fall (Statistics New Zealand, 2015d).

This is also reflected in the number of languages spoken in New Zealand – now 160 (Royal Society of New Zealand, 2013, p. 1). Figure **2.1-3** below shows the number of people with English as an additional language from a range of geographic origins in three arrival points: NZ born, arrived before 1996, and arrived 1996-2006. This figure highlights the recent dramatic rise in speakers from the Middle East and other regions of Asia.

---

[1] "Other" is a Statistics NZ category referring to very small groups not included elsewhere, who describe themselves as "New Zealander" or "Inuit" (Statistics New Zealand, 2007).

Figure 2.1-3 New Zealanders with English as an additional language (Earle, 2009, p. 7)
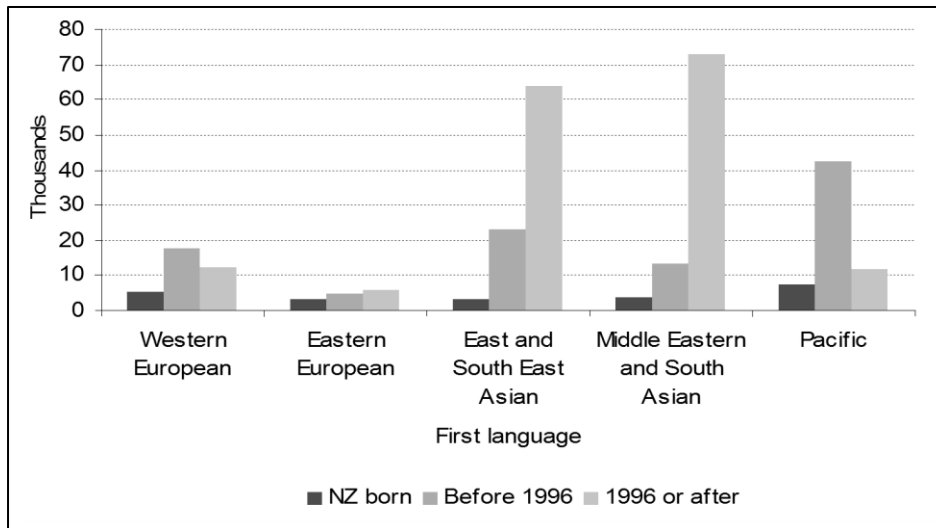
2.1.2   Immigration policies

This change in the origins of immigrants to New Zealand reflects changing government poli-cies. Immigrants may now enter New Zealand under a number of categories (Immigration New Zealand, 2015a) and many must show a sufficiently high score in IELTS (International English Language Testing System) or equivalent score if they cannot provide other evidence of English competence such as schooling in English. Coming from an English-speaking coun-try is not mentioned as sufficient evidence.

What is relevant for any policy of English language teaching, and for awareness of the Eng-lish language ability of migrants applying for work, is that for the foreseeable future there will be a wide range of non-native English speakers in New Zealand. In 2006, they accounted for 17 percent of all New Zealanders aged 25 to 65, totalling 362,000 people[1] (Earle, 2009). New Zealand employers will continue to deal with a great number of migrant job applicants speaking with a variety of familiar and unfamiliar accents.

Policies for refugees

Many of the migrants whose voices were recorded for the Employer Survey undertaken for this thesis were of refugee background[2]. Recent policies of refugee acceptance suggest that

---

[1] 4% of this group were students, including international students. In comparison, 2% of the remaining 83% of New Zealanders with English as their first language were students.
[2] It is considered more appropriate to use terms such as "refugee background" or "former refugee" to describe refugees once they have settled in their destination country (McBrien, 2014, p. 2, footnote).

they will continue to arrive in similar numbers, although the decision by the New Zealand government to accept more refugees in response to public pressure (Vance, 2015) has increased the number somewhat, and the 2015 European refugee crisis may encourage governments to take more.

In the decade from 2004 to 2014, New Zealand accepted a total of 7065 refugees under a quota (Immigration New Zealand, 2015d). Of all these refugees, only those from Sri Lanka have a high likelihood of prior exposure to English. Refugees are housed throughout the main centres of New Zealand, currently excluding Christchurch because of the disruptions caused by the 2011 earthquake. In addition to these, former refugees often arrange for family members to join them under Refugee Family Support (Immigration New Zealand, 2015c), and while the origin of new refugees is likely to change from year to year these family members are brought in by established migrant groups. 1,385 residence visas were granted under the Refugee Family Support Category from 2007 to 2014, the largest groups being from Afghanistan, Ethiopia, Somalia, Iran, Vietnam, Myanmar and Iraq (Immigration New Zealand, 2015c) – again, non-English speaking countries. Furthermore, people not included in the quota may claim asylum or refugee status, and 884 of these were approved in the calendar years from 2004 to 2014. These may come from countries not involved in the quota, including Fiji, India, China, Iran and Malaysia (Immigration New Zealand, 2015b).

Over a decade, this is a total of over 9,000 refugees who have needed a wide range of support from different agencies which assist with trauma counselling, housing and adapting to New Zealand society. A report from NZ Immigration suggested that, as a conservative estimate, 1,250 migrants of refugee background are now resettled in this country every year (Dunstan, Shorland, & Dibley, 2004, p. 44). It is clear that English language education will be needed for many of these people for the foreseeable future so that they can be integrated into our society and be able to find employment.

Post-arrival policies

Government policy in New Zealand and other countries has tended to focus on getting migrants employed at the earliest opportunity, regardless of their readiness in terms of settlement, health or education, and this can be damaging. A former immigration minister, Aussie Malcolm, saw no value in offering migrants opportunities to learn English and therefore deliberately provided no funding for it (Watts, White, & Trlin, 2002). Mortensen (2008) argued

that government strategies for former refugees were too narrowly focused on employment rather than responding to their overall needs, so that they were additionally disadvantaged. Such policies set up poverty traps which can condemn migrants to low-paid work because their formal English skills are less likely to improve much while they are working.

An investigation into migrants' experiences of a bond imposed in the 1990s concluded that "an inability to communicate in the host country language … can result in migrants experiencing difficulties meeting their physical, economic and social needs in New Zealand, as well as having notable cost implications for the New Zealand economy" (Forsyte Research, 1998, p. 23). Fletcher (1999, p. 66) also saw English language proficiency as critical to good labour market integration, especially for skilled migrants.

Governments have generally recognised that language support is needed for new migrants, but this has often been as part of broader educational policies rather than designed for the purpose, and employment has generally been given a higher priority than health or education, regardless of official policy (McBrien, 2014, p. 22). Training Opportunities, for example, was a government-funded get-to-work programme which was often used to offer ESOL as there were few other forms of funding available. The providers were tied to outcomes such as percentage of students going into long-term full-time work. Results from the TEC (Tertiary Education Commission, the government tertiary funding agency) demonstrated the difficulty migrants have in finding work. In 2010, 47% of all Training Opportunities students moved into full-time employment, while the figure for ESOL learners was less than 32% (Service Centre, Tertiary Investment Directorate, personal email). This was less than the minimum target set for each provider by the TEC.

The result is that opportunities to learn English have usually been inadequate. While a minimum of five years is needed to reach an "academic understanding" of English (Cummins, 1999, p. 4; McBrien, 2014, p. v) the Student Allowance (a non-refundable benefit for students) is limited to 200 weeks (Ministry of Social Development, 2014) which means a substantial portion of this allocation can be used up studying English. However, if migrants remain on Jobseeker Support (the unemployment benefit), they may be pulled out of class by Work and Income staff requiring them to attend seminars or accept a job (McBrien, 2014, p. 34 and personal observation). While employment is an important goal, it should not be seen as the only measure of successful settlement (Marlowe et al., 2014, p. 4).

The ESOL teaching profession works within the constraints imposed by migration policy, funding policy (via the TEC), academic policy (via NZQA, the New Zealand Qualifications Authority) and the economic and policy constraints of the language providers. Nevertheless, teachers aim to enable learners of English to meet their goals, which include academic success and employment, by developing their English language skills. The recently developed NZCEL (New Zealand Certificate in English Language), a national qualification taught throughout the country, sets graduate performance outcomes in terms of language performance skills. For example, the NZCEL Level 3 Qualification Details (NZQA, 2015) include the requirement that "Graduates of this qualification will have the English language skills to:

"… speak with some coherence, fluency and spontaneity to express personal ideas and sustain interaction in familiar contexts."

Such an outcome requires competence in pronunciation, but teachers are often poorly trained for pronunciation teaching (**3.1.1**) and it is often not clear what competences they need to be teaching. Chapters 3-5 examine that question in detail.

2.2   BARRIERS TO MIGRANTS

2.2.1   Reluctant hosts

An increase in migration is not always popular. In a survey of 1100 New Zealanders conducted in 2003, Spoonley, Gendall and Trlin (2007) found conflicting views: a majority opposed the number of immigrants and believed they put a strain on resources, but there was greater support from those who had contact with migrants. Some of the attitudes found were also based on mistaken beliefs about the numbers of different groups. The report found that "Overall, the level of expressed racism in New Zealand is very high and, it seems, universal in terms of its occurrence although it might only be infrequent" (p.15).

There has also been little change in the visibility of other languages. New Zealand was described fifteen years ago by Kuiper and Bell (2000, p. 14) as "overwhelmingly monoglot" and five years ago by Macalister as still "firmly monolingual" (2010, p. 72). Macalister investigated signs in the main street of Picton, a small town in the South Island which is the arrival and departure point for the Cook Strait Ferry (which links the two main islands of New Zealand) and hence is host to numerous tourists. He found that "languages other than English remain largely inaudible" (p.69), with virtually no signs in any language other than English.

Maori greetings are now often heard on RNZ (the national radio service), but other changes appear to be driven by business rather than government or community: the main language visibility is shop signs in Chinese along Queen St in Auckland for the benefit of language students, and Chinese signage and interpreters at Auckland Airport (Scoop Media, 2014).

Migrants have also reported many difficulties in acquiring English, including lack of opportunities to speak with native speakers of English and the inexperience of New Zealanders in speaking with people from other backgrounds (Watts, White, & Trlin, 2001). Skilled migrants have described "widespread negative attitudes towards people from different cultures and discrimination" (Watts & Trlin, 2000, p. 92), and one research interviewee commented: "Kiwis don't like Asian migrants. To learn English I need to have contact with New Zealanders, people living in New Zealand, but I know they don't like me and so it is really hard to approach [them]" (Forsyte Research, 1998, p. 64).

It is therefore worth asking whether the famous friendliness of New Zealander extends to supporting migrants or offering them jobs.

2.2.2   Ethnic stereotyping

Ethnic preferences

A persistent theme in research over the past two decades has been discrimination on the basis of ethnic origin. At the turn of this century a survey published in the National Business Review (Hunt, 1999) found that nearly half of those questioned thought there were too many Pacific people in New Zealand and 40% thought there were too many Asians, yet a majority were happy with the number of British, Australian and South African migrants or would be willing to see more of them. A strong distinction is often noted between attitudes towards immigrants from Europe (including Russia), South Africa and North America (often abbreviated to ESANA) as opposed to those from South Asia, the Pacific and "other" regions. ESANA immigrants are preferred and have much better labour outcomes (Butcher, Spoonley, & Trlin, 2006, p. 18).

Migrants also report that ESANA accents are accepted by New Zealand employers but Asian, Chinese or Indian are seen as foreign (McIntyre, 2014, p. 190). Some media reports describe outright prejudice ranging from frequent personal abuse to institutional discrimination, especially against non-European migrants (Devereux, 2004). Ward and Masgoret's research (2007) found that recruitment agencies responded more favourably to native born candidates than to migrants with equivalent qualifications and work experience.

Examples from specific groups

Various researchers have investigated attitudes towards different ethnic groups settling in New Zealand, and discrimination and barriers to employment have been found for Taiwanese immigrants (Boyer, 1996, p. 66), Koreans (Meares, Spoonley, Peace, & Ho, 2010b), Chinese (Meares, Spoonley, Peace, & Ho, 2010a) and others. New Zealanders are generally unable to distinguish Indian and Sri Lankan migrants, whose experiences have been similar. Lewin, Meares, Cain, Spoonley, Peace and Ho (2011) note that "Overall, Indian immigrants are competent English language communicators" and the skilled Indian migrants reported few language-based problems. However, older educated Indian women with fluent English found their form of the language was not accepted in New Zealand (Pio, 2005). According to one respondent, "As soon as they hear an Indian accent, it's like the door gets shut in our

face. And a typical Indian name means it's even more difficult to get called for an inter-view . . . leave alone being short listed."

Basnayake (1999) reported that nearly half of a sample of highly qualified Sri Lankan mi-grants seeking work had not even been interviewed, while at the other extreme one had been interviewed 200 times. Many had had difficulty finding work and there was often an instant negative response to their accent, but the migrants themselves saw the main prob-lems as lack of local experience or qualifications, and employers not understanding people from other countries (ibid, p.27). It is possible that, as native or fluent speakers of English, these migrants underestimate the effect on New Zealanders of the difference and unfamili-arity of their accents. Oliver (2000, p. 26) noted that Sri Lankans and Iraqi migrants generally had good to excellent fluency but were often told their job applications had been declined because of their "poor English". The main barrier was the attitude of employers towards their accents.

Another ethnic group to face discrimination is Koreans. Large numbers started to arrive in the 1980s and by the 1990s were the fastest-growing ethnic group in New Zealand (Chang, Morris, & Vokes, 2006, p. 6). Although most had arrived as skilled migrants they struggled to find work or had to take employment well below their skill level, partly because of language difficulties, but also – they felt – because of "anti-Asian feeling, discrimination and racism" (ibid. p.22) even towards those who had grown up here. Meares et al. (2010b) found in 2009 that 75% of Koreans interviewed reported some kind of discrimination, and 92% re-ported problems finding work because of their English. Around 50% of all Korean migrants rated their English as poor (ibid. p 34).

A study of Chinese migrants (Henderson, Trlin, & Watts, 2001) found that after 18 months residence, only 36.1% of the participants had found employment – even self-employment. Employment agencies reacted negatively to Chinese accents or pretended that the adver-tised position had been taken.

2.2.3   Impact on refugees

Prejudice and reluctance to employ migrants impacts more heavily on those of refugee background, who are likely to be less prepared for living in New Zealand and to have few or no qualifications. They are also likely to speak languages which are very different from English, and many (such as Myanmarese) need considerable help with their pronunciation. "The issue around language and accent in getting a job is particularly salient for refugees, who may speak no English at all and have particularly strong accents" (Butcher et al., 2006, p. 25). Dunstan, Shorland and Dibley (2004, p. 243) also noted: "Of the 109 established refugees [in New Zealand five years or more] who had looked for work in New Zealand ... the most common difficulties mentioned were English language problems (30 responses) and lack of work experience (22 responses)."

According to McBrien (2014), employment is a key element in successful resettlement for most refugees, and there are refugee-background adults who are able and willing to begin work in the first months of resettlement. The biggest challenges most of these job-seekers face are insufficient English language, skills, and lack of New Zealand work experience (p.6). In a study of youth from migrant and refugee backgrounds, 75% of the sample said that the main barriers to settlement and social inclusion were language difficulties including accent, issues with New Zealand accents and jargon (Sobrun-Maharaj, Abramson, Hoque, & Rossen, 2008, p. 42).

A survey of refugee background migrants by the Department of Labour found that after 6 months residence, 50% of all former refugees said that they could speak or write English well, but only 10% of quota or family support refugees could do so. Even after two years residence, only 28% of quota and 46% of family support migrants said they could speak English well or very well (Dunstan et al., 2004, p. 189), and after five years only 38% of all former refugees were in paid work (ibid. p.225). These figures demonstrate the importance of providing free English classes to former refugees, as now happens under the widely offered Intensive Literacy and Numeracy and the Refugee English Fund (Tertiary Education Commission, 2015).

2.3 LANGUAGE AND EMPLOYMENT

2.3.1 Accent discrimination

Much of the published material describing barriers to migrant employment refers to English language in general terms without specifying what aspects of language cause the problem. As Smith and Brown (2011) suggest, "very little attention has been paid to the role that pronunciation plays in workplace communication" (p.31). There is, on the other hand, plenty of evidence of antagonism towards unfamiliar accents. A language institute director summed up the situation in this way: "The biggest barrier faced by immigrants and international students seeking employment is their accent, and a lack of Kiwi experience" (Tan, 2009). The same article quotes the director of a language school on New Zealanders' attitude to different speech: "New Zealand is not open to foreign languages, or even to those who speak English differently to themselves."

Although Singer and Eder (1989) found no effect of accent in job selection (**2.3.3**), discrimination on the basis of accent (and what it represents in the mind of the hearer) was described by Henderson (2004) as a "widespread phenomenon" supported by a considerable body of research. "Analyses of the employment rates of immigrants in New Zealand have found limited skills transferability and a penalty for being from a non-English speaking background" (Henderson, 2004, p. 15). Butcher et al. (2006) described this as an "accent ceiling" for senior positions. In one case a senior Work and Income employee said that an applicant was not suitable for a job in McDonald's because of a lack of "New Zealand English" – yet she had IELTS 7 with a Band 8 in speaking, which is near native-speaker performance (Henderson et al., 2001, p. 115).

Li and Campbell (2009) wrote that the main barriers to employment were insufficient language skills, lack of New Zealand work experience, accent, skin colour, prejudice and discrimination, lack of cultural understanding, and lack of helpful support from recruitment and government agencies. "A number of the respondents commented that their accents were viewed as a barrier, in some cases, stating that a better accent would contribute to career advancement" (p.382). Younger people from refugee backgrounds face the same problems, as they find that language and lack of networks are barriers to work experience and

employment. Poor English language skills and a lack of social connections are seen as barriers to gaining work experience, and even those who speak good English believe having an accent is a disadvantage (R. O'Connor, 2014, p. 15). According to Oliver (2000, p. 31), late-teenage children of Muslims had Muslim names but New Zealand accents and were able to find work more easily than their parents, suggesting that accent rather than discrimination was the problem.

The dangers of seeking a "better accent" have been highlighted by criticisms of the accent-reduction industry (Derwing & Munro, 2009b; Thomson, 2014); a research-based approach is therefore needed to advise ESOL teachers on what needs to be done to make the speech of migrants more acceptable to employers.

2.3.2   Learning English

It has often been pointed out that communication is a two-way process, and comprehension is as much the listener's responsibility as the speaker's (Zielinski, 2006, p. 20). An article by Field, *Intelligibility and the Listener,* "emphasizes the perceptions of listeners rather than the productions of speakers" (2005, p. 400). Smith and Brown (2011, p. 31) point out that "Most often, when a person says of another that he/she has 'bad pronunciation' … the assumption is that the speaker needs to change the way he/she speaks, to more closely align his/her pronunciation with the local norm." This is especially true when the speaker is believed to be a non-native speaker. Ideally, listeners should be trained to accommodate to unfamiliar accents; but training all New Zealand employers to understand a range of accents is far less practicable than changing the pronunciation of migrants, although increasing familiarity (**3.3.2**) may improve the situation in the longer term.

Objections to some foreign pronunciation can be justified. My experience is that many teachers lack pronunciation teaching skills and students can still have very poor pronunciation after years of study. In addition, migrants may be fluent in English yet their accents are not well understood (Oliver, 2000). For example. Hogan (2013) reported that Filipino nurses were having difficulties in communicating because of their accent. McBrien (2014, p. 45) also commented: "One refugee director I interviewed felt that more needed to be done in terms of English pronunciation. He said that even after years of study, too many resettled

refugees lacked courage to speak in public because they were uncertain of pronunciation." Hunter (2007) reported the reasons given by one employer: "We've got to put them out to a client, and if we put someone whose English isn't up to scratch out there, then it reflects badly on us" (see further research on Accentedness in **3.2.3**).

It can therefore be difficult to clarify whether employers are reacting to poor or merely unfamiliar pronunciation, or acting on ethnic prejudice. For example, Earle (2009) reported that people whose first language was a South-East or East Asian language had much lower wages, total income and household income, but the cause is unknown. Daldy, Poot, Roskruge and van Dyke (2011) found discrimination against migrants in the workplace, but also that this diminished with length of residence, which suggests some integration effects such as improved language, cultural knowledge or work experience.

A migrant participating in the survey by Meares, Spoonley, Peace and Ho (2010a, p. 57) acknowledged this concern: "When I speak English with a strong accent many bosses are afraid that their customers may have difficulty understanding my English. It is understandable. So, good oral English is very important in finding a job."

2.3.3   Employer attitudes

Intolerance of "others" can be caused by fear of a threat to one's way of life or to resources, prejudice, or by simple stereotyping. Some of the earliest New Zealand research was done by Huang and Singer (1984), who found clear stereotypical attitudes towards Maori, Samoan, NZ Chinese and English immigrants. The first formal attempt in New Zealand to examine the effect of ethnicity bias in job selection was by Singer (1988), who found that respondents to video interviews with Chinese, Maori and European New Zealanders showed a preference for the Chinese, supporting previous results on ethnic stereotyping.

To examine ethnicity effects, Singer and Eder (1989) used four NZ ethnic groups (Maori, Chinese, Dutch, Pakeha[1]) to test the effect of ethnicity, accent and job status in simulated applications for jobs of high and low status. As well as rating the applicants for position ("Selection"), respondents were asked which dimensions they considered most important in

---

[1] In the study, meaning of British origin. "Pakeha" now usually means any fair-skinned New Zealander of European origin (Deverson & Kennedy, 2005).

making the decision ("Importance"), the key factors for research being accent and ethnicity. Respondents reported that they placed moderate importance on accent and little importance on ethnicity for the semantic differential questions, yet this was the reverse of the actual statistical results of their decisions, suggesting prejudicial attitudes towards certain ethnicities.

Singer and Eder (1989) found no significant interactions between job status and applicant accent for the Selection question, and suggested that contrary findings in previous United States and Canadian research by Hopper and Williams (1973) and Kalin and Rayko (1978) (both described in more detail in **3.3.3**) came about because they had confounded accent with ethnicity when reporting a significant effect for accent. Accent was not found to be significant for the Selection decision but there was clear ethnic stereotyping of the Maori and Chinese applicants. More recently, an Auckland survey by Work and Income on barriers against migrants seeking work concluded "The main barrier was not English competence but degrees of employer discrimination and stereotyping, together with inadequacies in the ESOL courses available" (Oliver, 2000, p. 26).

2.3.4   Positive Attitudes

Despite these negative observations, the transformation described in **2.1.1** has been continuing, and it would be unfair to describe New Zealand as a society which is consistently antagonistic to migrants or other ethnicities. Pakeha New Zealanders have long prided themselves on having "the best race relations in the world" (M. King, 2003, p. 468), even when most Maori did not agree.

Employers who routinely hire migrants have shown that attitudes can change with familiarity: for example, North's survey of "top 500" employers in Auckland and Wellington revealed generally positive and welcoming attitudes (North, 1999). However, it should be noted that the positive responses in this and North's later (2007) report may come about because the employers had already made the selection of skilled migrants and were now committed to them. Outside of Auckland, a survey by Watanabe (2008) found that listeners were positive towards ESL accents, some as favourably as native local accents, although the

assessors were students who had contact with non-native speakers and therefore may have been more favourably inclined towards them than the general population.

According to Ward and Masgoret (2008), research shows that immigrants are frequently seen internationally as a threat to both resources and values, yet they discovered from over 2,000 interviews that 89% of New Zealanders endorsed the various features of multiculturalism, ranking NZ highest out of 17 Western countries. Inevitably the survey identified that "Anglo" countries were still at the top of the interviewees' ethnic group preferences, yet 75% of participants thought it important to accept a variety of cultures and half were happy with the number of migrants. The same authors also wrote they had found that "New Zealanders have generally favourable attitudes towards immigrants and that these attitudes improve as immigrant numbers increase. Furthermore, this appears to be at least partially driven by increased intercultural contact" (Ward, Masgoret, & Vauclair, 2011). This was also demonstrated by Johnston, Gendall, Trlin and Spoonley (2010), who found that negative attitudes were less pronounced in areas such as Auckland where there was more contact with migrants.

2.4 Attitudes to migrants in other countries

Attitudes towards migrants in other countries, and stereotyping on the part of employers (such as unfavourable attitudes towards Hispanic accents in the USA) are described in **3.2.3**. Overall, in English-speaking countries similar to New Zealand, migrants face the same sort of stereotypes and barriers to employment. Insufficient language skills are often cited as a barrier to employment, and employers discriminate against applicants with foreign names (E. Chen & Ward, 2013).

O'Rourke and Sinnott (2006) used information on attitudes to immigrants drawn from people in 24 countries including New Zealand and found overall similarities. On a 1-5 range from *increased a lot* to *reduced a lot*, the mean opinion results in all countries surveyed were that immigration should be reduced, but the range was 3.071 to 4.042. New Zealand was the sixth least xenophobic at 3.742 – still favouring fewer migrants. O'Rourke and Sinnott concluded that anti-immigrant sentiment was predicted in these countries by having unskilled work, greater age (amongst retirees), trade protectionist attitudes, and degree of patriotism (although the causality here was uncertain). Unemployment was not a predictor. Thus, while the degree of anti-immigration attitude in New Zealand is less, the causes are the same.

McBrien's (2014) comparison of the situation of former refugees in New Zealand and the United States found that those in the latter country faced more pressure to work before they had attained an adequate level of language and often had poor access to language classes. Freedom of speech laws made overt racism and abuse, even from teachers, far more common. In Australia, Islam and Parasnis (2014) found that migrants were better paid than non-migrants, but attributed this to the highly selective immigration policy. There is funding for workplace language tuition, but according to Fraser (2000, p. 35) "Employers are yet to be convinced of the benefits to their industries of improving workers' pronunciation" as there is a tendency to lump all language deficits into "literacy" which migrants may not need.

Schellekens (2001) in Britain found a gap between the level of ESOL provision and the employers' requirements, and commented that "There appeared to be little awareness of the difference in the teachers' and employers' perceptions" (p.15).

Migrant teachers in Canada are far from immune to prejudice. Schmidt (2010) reported hostility towards certain non-native pronunciation patterns and discrimination on the basis of accent in the school system. Xu (2006), also in Canada, interviewed Chinese women who reported difficulties because of their English, although some were hired regardless if the employer's need was great enough.

2.5 DISCUSSION

Overall, there is ample evidence to show that the experience of many migrants to New Zealand – as in other countries – has been disappointing. The reasons are not necessarily straightforward: employers have genuine reasons for not taking on people with poor language skills or who cannot be understood by customers. But it is also clear that many migrants who have adequate language skills for the jobs they apply for lose out in favour of native speakers, and there are negative or even hostile attitudes from some people.

Media reports have often commented on a lack of support for migrants or opportunities for them to socialise (McAvinue, 2012; Migone, 2012). Many migrants are still often employed in low-skilled jobs such as cleaning, driving, and construction, despite many holding qualifications or having held higher-paid or higher-skilled jobs in their countries of origin or even in refugee camps (Changemakers Refugee Forum, 2012).

There is obviously no "quick fix" to prejudicial attitudes. This thesis focuses on a single factor: the effect of migrants' pronunciation on their employability. The work of Derwing and Munro from 1995 onwards has demonstrated that Comprehensibility, Intelligibility and Accentedness are separate Global Measures (Munro & Derwing, 1999).The evidence from this thesis, as described in Chapter 8, is that Acceptability is a further Global Measure which affects the outcomes for migrants seeking employment. The implication is that teachers need to take Acceptability into account when assessing their students' pronunciation needs and how to assist them.

3 LANGUAGE LEARNING AND GLOBAL MEASURES

3.1 OVERVIEW

This chapter discusses the relevance of pronunciation teaching in the context of recent TESOL history, from its neglect in the post audio-visual era to its gradual restoration in the late 20th century, and from views that teaching pronunciation was pointless to current attitudes that it can and should be taught. The goals of pronunciation teaching are then examined, outlining the shift in emphasis from becoming like a native speaker to becoming an intelligible speaker.

I then review evidence for the Global Measures of pronunciation which are most commonly found in the literature, Accentedness, Comprehensibility and Intelligibility: how they can be predicted, how they are measured, how they are related, the subjectivity of these measures, and their implications for language learners and teachers.

Finally, I propose the hypothesis that Acceptability is a Global Measure which emerges both from the literature and from the results of the Employer Survey as reported in Chapters 7 and 8 and is correlated with but distinct from the three measures referred to above. Unlike the other measures, it is contextual, so (for example) the standard required for a job stacking shelves will not be the same as that for customer interaction. This is explained further in section **8.3.2**.

3.1.1 Learning pronunciation

It is well known that children are likely to pick up a new accent without difficulty while adults are less likely to do so. An explanation was first formulated by Lenneberg (1967). Based on his theoretical underpinning that language is biologically determined, Lenneberg suggested a "critical period for language acquisition" (p.179). He proposed that "[the critical period's] termination seems to be related to a loss of adaptability and inability for reorganisation in the brain" associated with puberty, when the brain "behaves as if it had become set in its ways" (p.158). Selinker concluded that this resulted in "fossilisation" in which learners retain certain "linguistic items, rules and subsystems … no matter what the age of the

learner or amount of explanation or instruction he receives in the TL [target language]" (1972, p. 215).

The concept of a critical period is not merely theoretical. If the brain is unable to reorganise itself after a certain age, and phonological structures become fixed, then there is little point in teaching pronunciation to anyone past that age. However, both the period and the degree to which it is critical were subsequently challenged. Selinker and Lamindella (1981, p. 201) later suggested that "age and maturational factors in general should be regarded not as an absolute, but as another set of variables which operate in complex interaction with social, cognitive and affective factors." Long (1990) supported Lenneberg's view of maturational constraints, but as a process of decline from the age of 6.

There was a lack of emphasis on pronunciation teaching in the 1980s which was encouraged by theories claiming that adults could not change their pronunciation for reasons such as those given above, by suggestions that teaching pronunciation had little benefit, and an inability to fit it into a communicative framework. Purcell and Suter (1980) concluded that the variables which predict pronunciation accuracy in adult learning are those over which the teacher has little influence (namely, L1, ability to mimic accents, length of residence, and concern for accuracy) so formal instruction was of "no importance" (p.285); and Krashen claimed that acquisition, the subconscious learning of rules, "does not seem to require or even profit from overt teaching" (1979, p. 40). It appeared that pronunciation teaching in the classroom was a pointless exercise, although teachers did not necessarily agree.

However, Schmidt (1990) made use of research in psychology to highlight the role of conscious awareness and refute the primacy of subliminal language learning; and further research (Couper, 2003; Guion & Pederson, 2007; Kissling, 2014; J. Smith & Beckmann, 2010) demonstrated the importance and effectiveness of attention, explicit teaching and self-monitoring.

Learning the pronunciation of a new language is related to perception of the L2 phonological features, or as Aslin (2014, p. 14) put it "listeners and speakers maintain a close coupling between their perception of phonetic categories and the articulatory gestures[1] required to

---

[1] "A structured movement of the vocal organs which is directed towards a specific articuatory goal" such as closing the lips to pronounce [m] (Matthews, 1997, p. 146).

produce these categories." Research into infant sound discrimination (Werker, Gilbert, Humphrey, & Tees, 1981) was cited by Munro and Derwing (2011) as marking an "extraordinary milestone" because it demonstrated that infants have an ability, much reduced in adults, to differentiate and categorise new phonological input. Werker and Tees (1984) also demonstrated that adults can in fact discriminate non-native categories, but not during natural language processing.

Early work by Flege (1980) showed that adults could develop a "fairly stable interlanguage phonetic system which admits the possibility of phonetic strategies by individual speakers" (p.117) under which they might not learn to speak like native speakers, but could make systematic progress towards the target language. This led to Flege's Speech Learning Model (1997) which predicts that, with extensive experience, adults can form new phonetic categories, offering the possibility of both discrimination and production of these categories.

Understanding of adults' abilities was refined further by Flege's Equivalence Classification (1987), and the Perceptual Assimilation Model (Best, McRoberts, & Sithole, 1988) which both provided evidence that adults have difficulty establishing a new phonetic category for an L2 phone with an L1 equivalent, but can do so with new categories – those with no L1 equivalent. For example, French learners would find it hard to create a new category for English /u:/, which is similar to French /u/ (so they continue to use the French pronunciation), but could learn to pronounce English /aʊ/, which is sufficiently distant from all French vowels to be perceived as a new category. Best et al. (op. cit.) demonstrated that different Zulu clicks can be distinguished by English speakers because these contrasts do not exist in English.

Selinker's original formulation of Interlanguage (IL) theory (1972) involved the concept of fossilisation, in which certain "linguistic phenomena" of the IL become a permanent feature of the speaker's IL. This was challenged by Acton (1984) and Derwing and Munro (1997) who claimed success in programmes for assisting learners whose pronunciation had been resistant to change; and literature reviews by Pennington (1998) and Piske, MacKay and Flege (2001) illustrated the range of studies on the general success of pronunciation instruction.

Articles began to appear at this time commenting on the lack of pronunciation teaching and the need for it (Morley, 1996, 1999) and evidence began to accumulate on the success of specific interventions in improving adult L2 pronunciation. There is clear evidence that adult

learning involves "reattunement of perceptual phonetic processes and the perceptual reorganisation of phonological categories" (Strange, 1992, p. 197); that is, the L1 phonological structure built up in childhood has to be broken down in order to perceive, accept and produce categories which were previously combined or ignored. However, this can be done: improvements in adult production of segmentals after instruction have been demonstrated by Flege (1989, 1995), Couper (2006), Munro and Derwing (2008) and Saito (2011). De Bot and Mailfert (1982) and Nagamine (2002) showed that intonation can be improved, and success on both segmental and prosodic levels was reported by Couper (2003). As a result of these and other studies, fossilisation is now one of those concepts "debunked" as pronunciation myths (Derwing & Munro, 2014b).

Pronunciation teaching was for a long time restricted to the segmental level (Pennington & Richards, 1986) and even today many ESL teachers are inadequately trained to teach pronunciation at other levels (Murphy, 2014b). When Derwing and Rossiter (2002, p. 154) surveyed adult immigrants' perceptions of their pronunciation learning problems, 79% of the problems reported were segmental, which is probably more of a reflection of the emphasis placed on segmental errors in teaching and correction than of the actual causes of their problems. The first long-term classroom-based study, by Derwing, Munro and Wiebe (1998), compared emphasis on segmental as opposed to suprasegmental instruction and concluded that, while both levels are important, the latter was more effective.

In the last three decades, then, enough evidence has been generated for teachers to be confident that adults can learn, if imperfectly, the pronunciation of an additional language; that "fossilisation" is no more than a set of long-term habits which can be changed; and that explicit teaching and noticing on the part of the student can improve learning.

Despite this, there is still a great deal of uncertainty about pronunciation teaching and its goals. After nearly half a century of comments about the neglect of this area, reviews still describe it as the "Cinderella" of language teaching (Seyedabadi, Fatemi, & Pishghadam, 2015) and teacher training is still highly inadequate (Murphy, 2014b).

The first step for teaching to improve is for teachers to understand what learners need. The following section therefore examines pronunciation learning in the context of learner goals.

3.1.2   Learner goals and attitudes

In addition to demonstrating that adults are capable of improving their pronunciation, researchers have asked whether all adults have the same ultimate pronunciation goals. The move towards learner autonomy (Cotterall, 2008) and the acknowledgement that learners set their own goals throw into question the belief that the failures of adults meeting the Nativeness (or any other) standard indicates an inability to achieve their goals.

Willingness to Communicate

Gaining competence requires practice and a willingness to talk, but this is often lacking in the classroom, and of course where the education system itself emphasises formal accuracy and does not allow for conversation or development of fluency, it is not surprising that learners struggle to use speaking skills. Lewis and Gravatt (1999) cited a number of reasons for unwillingness to talk in a New Zealand ESL class, including shyness, feeling a lack of competence, and the habit of mentally translating into L1. They found Chinese students were particularly sensitive about making mistakes, and Lucas (1984) found the same difficulty with Japanese students. Madill (2016) described three reasons why local EFL students in Korea are reluctant to speak: the belief that they were not as good as other students, lack of preparation time, and fear of failure. I have found that fear of losing face has a significant effect on willingness to communicate for many East Asian students.

Unlike other aspects of language learning, speaking aloud is a kind of performance, and not everyone is keen to perform. As Abercrombie (1991, p. 92) pointed out, learners may consider it "affected" to pronounce sounds like native speakers of the L2, and Miller (2003), in a study of language competence in schools, wrote "in a context of second (or third) language acquisition, speaking audibly and without anxiety is an enormous challenge for most students" (p. 141). Status relationships in a school or even university class can affect willingness to develop one's pronunciation (Gatbonton, Trofimovich, & Magid, 2005; Lefkowitz & Hedgcock, 2002).

Ludwig (1982, p. 279) suggested that "most students, for a variety of reasons, are threatened by a foreign language" – especially by the fear of appearing foolish in front of others. Guiora, Beit-Hallahmi, Brannon, Dull, and Scovel (1972) found that a small drink of alcohol before the class improved learning. This suggests that, for some learners at least, their lack

of progress was due to anxiety or inhibition. It is clearly important for ESOL teachers to create an atmosphere in which the students feels that it is safe to experiment and take risks. MacIntyre, Dörnyei, Clément, and Noels (1998), in a detailed study of willingness to communicate, concluded that engendering in students a "willingness to seek out communication opportunities and the willingness actually to communicate in them" was the ultimate goal of language teaching, as opposed to producing graduates who could pass tests or speak in a certain way.

Accent and Identity

Accent is a crucial part of the identity of a speaker of any language, often revealing where we are from, our education and social standing. Smit (2002) commented that "pronunciation [is] that aspect of a language which is closest to its speakers' feelings of identity" (p.102); the extent to which a learner adopts the L2 accent, and the type of accent they adopt, is likely to relate to how much they identify with the culture associated with that language. I have encountered many Francophones who speak English fluently with little concession to English phonology; it would be worth asking whether this is a choice rather than an inability to learn.

In addition, the ability to speak like a native is not necessarily a permanent feature of an L2 speaker's accent. Sicola (2014, p. 210) suggested that everyone uses different accents for different purposes and audiences, so picking up a new accent is just another style of speaking. Gnevsheva (2015), for example, found that Germans and Koreans living in New Zealand had the ability to alter their style of speech, becoming more or less L2 accented depending on the topic and audience. From interviews with well-educated and proficient Korean bilingual graduate students, Choi (2016) found that they consciously distanced themselves from "native speakerism" and looked down on those who tried to sound native by inappropriate use of slang expressions. These examples suggest that L2 accent can be under the control of the speaker.

Many learners are well aware of the value and significance of the various accents they hear and make choices about how they speak. Miller (2003, p. 170) commented: "Some minority speakers understand very well that how you sound affects how you are seen and heard, where you can go, whom you can speak or be with, how you are treated, and how you can influence events and those around you." Thomas (2014) described a wide range of accent

choices by migrant ethnic groups resulting from social and political influences. There is a need to investigate accent choices made by New Zealand migrants, because the low absolute numbers of most ethnic groups within any New Zealand community make it less easy for them to be physically or socially isolated from the majority Pakeha community, especially when seeking employment, and there may be more pressure here than in larger countries to accommodate to the majority.

Another reason for lack of progress is motivation. According to Ushioda (2008), motivation requires "involving learners in making informed choices and decisions about their learning and in setting their own goals and learning targets, and thus fostering feelings of personal responsibility." While employment and encouragement can provide extrinsic motivation, it is likely that intrinsically motivated learners will be more successful (Griffiths, 2008, p. 21).

The above evidence strongly suggests that ESOL programme designers and teachers should develop their content and assessments to work towards pronunciation outcomes which will meet learner goals while reducing their anxiety about speaking.

The following sections discuss measures of pronunciation which have been used to evaluate those outcomes, and their relevance to teaching.

3.2   THE STANDARD GLOBAL MEASURES

Linguists writing on the pronunciation aims for language teachers (e.g. "Setting pedagogical priorities" in Derwing & Munro 2005, p. 385) make frequent use of the terms *Accentedness*, *Intelligibility* and *Comprehensibility*. It is generally assumed, rather than spelled out, what class of concept they are. Some writers do not classify them at all (e.g. Leather, 1999) while elsewhere the terms are described variously as "aspects of foreign-accented speech" (Derwing & Munro, 2005, p. 385), dimensions (Munro, Derwing, & Morton, 2006, p. 112), perceptual dimensions (Derwing & Munro, 2014b, p. 41), "factors which affect native speaker perception of non-native English" (Matsuura, Chiba, & Fujieda, 1999, p. 49), and judgements divided into constructs (Kang, 2010b, p. 302). Derwing and Munro (2014a, p. 12) explain that they have "operationalised constructs pertaining to L2 speech in terms of listeners' perceptions" while Murphy, also using "construct", refers to "measures of intelligibility" (2014a, p. 261). Fayer and Krasinski (1987) refer to intelligibility as a "very global subjective judgement of comprehension by the listener."

In this thesis I use the term **Global Measure** to emphasise that certain constructs such as Intelligibility, Comprehensibility, Accentedness and Acceptability are *measures* because they can be used to rank a speaker's proficiency, and they are *global* because they are the listener's overall judgement of the speaker's pronunciation. The first three can be measured in isolation, while Acceptability must refer to a context such as a job. Global Measures are useful for research and teaching because they can be used to make judgements on the standard of an ELL's performance in broadly perceived and comparable measures of pronunciation.

The relative importance of these measures, and even the meaning of terms such as Intelligibility, have changed during the past four decades, as a result of which earlier descriptions may need interpretation or may not be easily comparable with later research. Two crucial examples are the shift away from the Global Measure of Nativeness, described in **3.2.1** to **3.2.2**, and the analysis of Intelligibility, Comprehensibility and Accentedness starting with Munro and Derwing (1995a), which is described in **3.2.3** to **3.2.5**.

3.2.1 Nativeness

The accepted goal of instruction in pronunciation was, until recently, Nativeness, explained by Griffen (1991, p. 182, writing about second language learning and speech pathology) as requiring that "the student (or patient) should learn to speak the language as naturally as possible, free of any indication that the speaker is not a clinically normal native". This appears to be a clearly-defined, desirable and measurable goal, and it persists in teaching around the world. Choi (2016), for example, claimed that the adoption of communicative language teaching in South Korea has caused the continuing emphasis on native-speaker pronunciation and recruitment of teachers from Inner Circle countries (Kachru, 1985).

The prescriptive mode of teaching towards a single set of standards ignored the constructive nature of learning (Selinker & Lamendella, 1981). The goal for pronunciation, until comparatively recently, was that the learner should speak with the "right" accent, such as RP (Leather, 1999, p. 35), an approach which, as Klein (1998) pointed out, treats the target language as a "clearly fixed entity" which is correctly described in texts (p.535) and which all learners miss to varying degrees. Such a target is attractive because "it is the natural perspective of language teachers and researchers and provides a clear yardstick" (p.536). Klein pointed out that this involves two fallacies: the "target deviation perspective" which unhelpfully describes a learner in terms of their errors or failure to achieve the goal, and the "real language hoax" because "real language is a normative fiction" (p.541). Major (1997) questioned whether the concept of "nativelike" was even possible, given that native speakers may have varying views of what the concept means (p.148).

In some cases, nativelike performance can be a disadvantage. Dalton and Seidlhofer suggested that there are "penalties" if a speaker appears to claim membership of a community they do not belong to (1994, p. 7); and Ryan (1983, p. 157) proposed that "a certain amount of nonstandardness (e.g. a language learner's accent) can sometimes attenuate the impact of another aspect of nonstandardness (e.g. grammatical or sociolinguistic errors." Other critics of Nativeness include Derwing and Munro, who have repeatedly attacked practitioners of "accent reduction" on the grounds that Accentedness (a measure of Nativeness) is not necessarily the problem so "reducing" accent cannot be the solution.

Researchers and many teachers, therefore, no longer see nativelike pronunciation as achievable, desirable or even measurable (Derwing & Munro, 2005, p. 384; Levis, 2005) and favour

the goal of Intelligibility (J. Field, 2005). Levis (2005) compared the two approaches ("shifting paradigms") noting that the influence of Nativeness had diminished; and the shift towards the "Intelligibility Principle" has since been supported by research into infant language ability (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992). However, Levis (2005, p. 371) noted that current teaching materials were still largely based on Nativeness.

Dalton and Seidlhofer (1994, p. 6) pointed out that there is a difference between a model and a goal: teachers may present certain accents, including their own, as models, without insisting that students should have to copy them. Romova, Smith and Neville-Barton (2008) found that the pronunciation of four EAP students studying in New Zealand had improved over three years but not in the direction of the local English accent. They commented that "the students' own goals (whether conscious or otherwise) need to be known in order to determine if they have achieved them" (p.12).

In terms of learner outcomes, the difference can be seen when comparing earlier research with a current "guide to interference". As noted in **3.1.1**, French speaking adults are unlikely to produce a fully English-sounding /u:/ because the L1 and L2 phonemes are similar. Swan and Smith (2001, p. 53), on the other hand, stated that English /u:/ "should … be perceived and articulated without serious difficulty" by French learners – because the L1 and L2 phonemes are similar. This is not a contradiction. The former were asking whether French adult learners of English were likely to attain native speaker quality with similar vowels, and the answer was no; the latter asked whether learners are likely to be intelligible and comprehensible when using similar vowels, and the answer is yes.

If the weight of opinion has shifted towards the Intelligibility Principle (using the "broad" definition encompassing Comprehensibility), further questions need to be asked. How is Intelligibility best determined? Intelligibility to whom? What effect does this have on teaching? And is it really a sufficient goal? These questions are addressed in the following sections.

3.2.2   Alternatives to Nativeness

Prior to Munro and Derwing (1995a), the three global measures of Accentedness, Comprehensibility and Intelligibility were treated as much the same thing. Morley's (1991) "Speech Intelligibility Index" (p.502) had at its lowest level " Speech is basically unintelligible … Accent precludes functional oral communication" and culminates in "Speech is 'near-native' … Accent is virtually nonexistent". Thus Intelligibility was described in terms of Nativeness and accent loss.

Many other studies had referred to either Intelligibility or Comprehensibility, but Munro and Derwing's (1995a) paper was the first to separate them by using an objective test (transcription) for Intelligibility and separate subjective responses for Accentedness and Comprehensibility. They also demonstrated, in the paper cited above and subsequent work (e.g. Derwing & Munro, 1997), that the three are Global Measures of different aspects of speech performance and should not be confused. They also proved that "presence of a strong foreign accent does not necessarily result in reduced intelligibility or comprehensibility" (1995a, p. 90).

This gave ESOL teachers a definable goal. Nativeness / Accentedness was merely a feature of an adult language learner's speech but not in itself a problem, while Comprehensibility and Intelligibility were both desirable and measurable. The following sections examine these Global Measures in more detail.

Section **3.2.3** onwards includes references to the Employer Survey (including Parts A and B), Base Rating and 3-Factor Model. These were introduced in Chapter 1 and are explained in detail in Chapter 6.

3.2.3   Accentedness

Accent

The term *accent* has a number of meanings. Firstly, accent can mean "a type of perceived prominence heard on a spoken word or syllable" (Crystal, 1999, p. 2). It is also used to refer to a way of speaking in a geographic area as in "she has a Welsh accent", or to mean degree

of accent (Accentedness), a measure defined by Derwing (2010, p. 29) as "a judgment of how much one's speech differs phonologically from the local variety."

Even within the concept of phonological difference, *accent* has been defined in a variety of ways, each of which captures a different aspect of how it is used. A Glossary of Applied Linguistics (Davies, 2005) defines accent as "features of the speech signal that identify individuals as belonging to certain groups, which may be geographical or social class-based" (p.4). However, a foreign accent may identify a broad language group (such as Indian subcontinent) but not necessarily the exact geography (which could be any of India, Pakistan, Sri Lanka, Myanmar or Fiji). In addition, listeners may not be able to identify which group the individual belongs to other than that they do not belong to their speech group ("a foreign accent"). Results from the Employer Survey used for this thesis show that the Employers had little ability to determine origin on the basis of accent alone (**7.4.7**).

Accent is a measure of global pronunciation rather than of specific features. For Moyer (2013) it is "a set of dynamic and suprasegmental habits that convey linguistic meaning along with social and situational affiliation" although it is not only suprasegmental.

In this thesis I define *accent* as **a set of auditory speech features which identify the speaker's geographic, social or language origin or their difference from local speakers** (e.g., a Japanese / foreign accent), and **Accentedness** to mean the measure of difference. Because of the importance of these two concepts, I do not use the term *accent* to mean prominence, which is always described in terms of stress, as in word stress or nuclear stress. The meaning of *accent* in quotations and references may have to be inferred by the reader. I use **pronunciation** when discussing the use of one or more of the specific **features** which make up an accent, such as articulation of segmentals or placement of word stress.

Attitudes to accent

Even though judgements of Accentedness can be reliably distinguished from Comprehensibility and Intelligibility, they are influenced by affective factors and stereotyping. For example, in the United States, where there is a considerable Hispanic population, there are strong stereotypes of Spanish-accented speakers (De la Zerda & Hopper, 1979; Hosoda, Nguyen, & Stone-Romero, 2012; Rey, 1977; Segrest Purkiss et al., 2006). Gluszek and Hansen (2016)

found that negative references to accent in US print media were more common than positive ones, and most were in the context of education and employment, with much discussion on accent reduction.

Rubin (2012), reviewing research into Reverse Linguistic Stereotyping, described perceiving speech as a "constructivist process" and explained: "If I am told that I am about to hear a speaker of Nigerian English, I am likely to perceive at least traces of Nigerian accent in that speaker's oral performance" (p.12). In an experiment by Hay, Nolan, and Drager (2006), New Zealand listeners, aware that the speaker they were listening to was using NZ English, nevertheless heard different vowels depending on whether the answer sheet had "New Zealander" or "Australian" written on it. Such constructivism may have influenced listener judgements of pronunciation features in the Employer Survey as described in Chapter 6.

Since Osgood (1964) described semantic differential techniques it has been shown that listeners will make judgements on social traits such as friendliness, honesty, reliability and intelligence based on accent alone (Kalin & Rayko, 1978). Attitudes towards accented speakers in the workplace are illustrated by a video produced by the NZ Office of Ethnic Communities (2012), in which a smart Chinese job applicant speaks with a strong Chinese accent and then with a much less noticeable accent, demonstrating that his accent is not related to his work ability, intelligibility or English proficiency. An experiment in the USA, using speakers with Korean and various European accents, showed that NSs perceive information as less truthful when spoken by accented speakers – even when told that the speakers were reading out statements by the researcher (Lev-Ari & Keysar, 2010).

Such stereotyping has an impact in the workplace, where customers' evaluations of service performance can be negatively impacted by foreign accent (Rao Hill & Tombs, 2011). This may in turn discourage employers from taking on employees with foreign accents, particularly strong accents. This was investigated by Hosoda and Stone-Romero (2010), who had job interview scripts read out by three ELLs: one each of standard US English, Japanese and French. College students rated the Japanese-accented applicant as less likely to be hired than the other two for a low status job with high communication demands. However, the raters were aware of the speakers' ethnicity, so attitude to accent could not be separated from ethnic stereotyping.

English speakers are used to hearing a variety of accents, some of which can be difficult to comprehend. British and American newscasts occasionally subtitle speakers from each other's lesser known dialects. Nevertheless, the degree to which listeners believe they should accept responsibility for comprehension varies between Inner Circle dialects and others. Gluszek and Dovidio (2010, p. 228) explain this as a "group membership" of native speakers which is not shared with the Expanding or Outer Circles. According to Lippi-Green, "Most of the time, we will agree to carry our share of the burden" but this depends on the status of the accents. "In extreme cases, we feel completely justified in rejecting the communicative burden, and the person in front of us" (2012, p. 73).

The fact that these stereotypes exist does not mean that every native speaker has them or acts upon them, nor that non-native speakers are always discriminated against. Lima (2011) found that student participants responded to the actual accent of foreign speakers rather than beliefs about their origins, and Eistenstein (1983) commented that "the extent to which experimental measures translate into actual behavior has often been called into question." The employers' ratings of migrants' speech in the Employer Survey were not significantly affected by what the employers believed the speaker's L1 to be. Accentedness is not always a disadvantage, but it is more likely that discrimination will be solved by changes in social attitudes rather than "reducing" learners' accents.

Accentedness

Accentedness is a Global Measure of pronunciation. It has been described as "a type of noise in the speech signal that can interfere with the message. A heavy accent causes the listener to strain to decipher the basic meaning" (Major, 1987, p. 155). This strongly emphasises the negative effects and suggests that Accentedness and Comprehensibility are identical, but later work by Munro and Derwing (1995a) is widely cited as demonstrating that they are not. Major described native speaker attitude towards Accentedness in this way: "Global foreign accent may be defined as overall pronunciation proficiency in a second language, or how native-like the accent is. Foreign accent ratings are necessarily based on native speaker judgments of authenticity (Nativeness)" (1987, p. 157). In this thesis, Accentedness is defined as **a Global Measure of how a native speaker rates an accent in terms of degree of difference from their own or local accents**.

Lippi-Green describes Accentedness as "the carryover of native language phonology and intonation into a target language" (2009, pp. 164–5) but Beebe (1984) and Major (2008) show that simple transfer is far too simplistic an explanation for the variation between native and foreign-accented speech; universal factors such as markedness (e.g. of syllable structure) are also at work.

Accentedness in itself is not a barrier to communication. In a frequently cited article whose findings have often been replicated (in this thesis also – see **7.6.2**), Munro and Derwing (1995a) demonstrated that Accentedness, Comprehensibility and Intelligibility are "related but partially independent dimensions" (p.90). In their investigation, a number of speech samples were rated by listeners as heavily accented, yet they were not necessarily rated as difficult to understand or lacking in Intelligibility. A subsequent investigation (Munro & Derwing, 1995b) found that Comprehensibility, but not Accentedness, affected listener processing time.

This was tested in more depth by Valles (2015), who asked native speakers to respond to simple commands given by accented speakers and measured the accuracy and speed (latency) of their responses. He found that only a heavy accent was correlated with a significant drop in accuracy of listener response. Interestingly, the latency response to one near NS was significantly faster than that for the NS (possibly because of slightly slower speech rate), and the response times to medium and heavy accents were not significantly different to those for a NS.

It could be expected that ratings of Accentedness will be dependent on the listener's own accent, but Kang (2010b) found that the strongest predictors of Accentedness were pitch range and word stress. Saito, Trofimovich and Isaacs (2015) found that predictions of Accentedness differed according to the speakers' proficiency: segmental and prosodic variables were important predictors at all levels, but grammatical complexity became important for intermediate to high-level speakers. These results suggest the operation of across-the-board predictors.

Accentedness is not immune to effects of NS listener interaction. Kennedy and Trofimovich (2008) were surprised to discover that semantic context of L2 speech (based on the listen-

ers' real-world expectations and semantic possibility of the utterance) affected not only ratings of Comprehensibility but also Accentedness, and suggested that the listeners' degree of understanding was therefore implicated in their Accentedness judgements.

According to Munro (2011, p. 9), Accentedness is the least relevant of the global measures because listeners are used to hearing a range of accents and can easily adjust to them. It is therefore not enough to demonstrate that a speaker has a strong accent to conclude that they must have a pronunciation problem. Unfortunately, this does not prevent discrimination against people with strong or "undesirable" accents (Lippi-Green, 2009, 2012; Sato, 1991). Nor does it discourage reputable institutions from treating Accentedness and Intelligibility as the same thing. Munro and Derwing (1995a) noted that the Test of Spoken English, used in North America, equated Accentedness and lack of Intelligibility; and Trofimovich and Isaacs (2012) criticised the Cambridge ESOL Common Scale for Speaking and the Common European Framework of Reference Scale of Phonological Control for similar reasons.

If accent can be understood as a normal aspect of every person's speech, and Accentedness is no more than a measure of difference in the sense that clothing can be uniform, similar or unusual, it is clear that Accentedness is a poor measure of pronunciation development. The Global Measures of Intelligibility, Comprehensibility and Acceptability, discussed in the following sections, are far more relevant.

3.2.4   Intelligibility

Intelligibility is **a Global Measure of how much of the speech is actually understood by interlocutors** (Munro, 2011, p. 9). This is the definition used in this thesis. A measurement of Intelligibility should be supported by some performance on the part of the listeners which demonstrates objectively how much they have understood - usually by asking listeners to transcribe or identify what they hear. Munro, Derwing and Morton (2006) showed that assessments of Intelligibility are valid across a range of different listener groups.

Varonis and Gass (1982) found correlations between judgements of Intelligibility and listeners' ratings of Comprehensibility, and Munro and Derwing (1995a) also compared the listeners' Intelligibility ratings with Comprehensibility scores, using a 9-point scale. They found

that "most listeners showed a correlation between Intelligibility and perceived comprehension." Gallego (1990) had ESOL specialists describe and categorise the reported communication breakdowns of NNS teaching assistants (TAs) and found that the major causes were unintelligibility (80.9%), then lack of Comprehensibility in terms of the meaning of the word in context (15.3%), and finally interpretability – understanding the speaker's intended meaning (3.7%). The factors involved were overall pronunciation (64%), stress (35.8%) (not further defined), and intonation (4.5%). The Employer Survey for this thesis found a 0.8 correlation between Employer ratings of speakers' Comprehensibility and Intelligibility (Table **7.6-1**), and also that Comprehensibility formed a "floor": the Intelligibility rating could usually be any level, but no lower than the Comprehensibility score (for comments on the type of test used for Intelligibility, see Limitations).

However, Intelligibility is not an absolute. Moyer (2013, p. 151) commented that "Intelligibility is relative; we have seen that it varies according to listener expectations and attitudes … as well as contextual factors, making it an amorphous goal." Wolff (1964, p. 443) reported on two neighbouring communities speaking linguistically similar dialects; members of one claimed the others were intelligible, but members of the other, who were seeking greater autonomy, denied it.

Derwing and Munro measured Intelligibility by having native speakers transcribe what they hear (2014b) yet this does not meet the criterion of listener authenticity, because the act of formal transcription removes the listener from their cultural relationship with the speaker. Smith and Nelson (1985), summarising 163 articles and books on Intelligibility to date, concluded that Intelligibility is "interactional between speaker and listener" and depends on listener expectations: people are more likely to find a speaker intelligible if they expect to do so (p.333). The extreme positions are taken by Rajagopalan (2010), who sees Intelligibility as an imperialist trick, but has nothing to offer in its place; and Golombek and Jordan (2005), who argue for the "decentering of the primacy of intelligibility as a skill" (p.529), claiming that "intelligibility is negotiated and contested along with identity in each interaction" (p.529). Both ignore the unequal contest of a NNS in a largely monolingual country versus a NS, or versus an employer who is not interested in taking on an employee who wants to negotiate and contest Intelligibility with their customers. The response of Derwing and Munro

to the apparent subjectivity of Intelligibility is to point out that there is a "high degree of reliability across groups of listeners, such that some shared sense of what constitutes intelligible versus unintelligible L2 speech is possible" (2005, p. 381).

Despite the broad agreement as to the importance of Intelligibility, there has not been much evidence presented as to what is likely to make speech more or less intelligible. It could be argued that Intelligibility is relative to local accent and therefore it is difficult to make generalisations, but this has not prevented research into the predictors of Comprehensibility and Accentedness. Field (2005, p. 399) commented that "arguably the most pressing issue in L2 pronunciation research today is the quest to identify the factors that most contribute to speaker Intelligibility" while Derwing and Munro remarked that "it is widely accepted that suprasegmentals are very important to Intelligibility, but as yet few studies support this belief" (2005, p. 386) and six years later they repeated the call for "more studies investigating the relative importance of various accent features to Intelligibility" (2011, p. 487). Trofimovich and Isaacs (2012, p. 906) reported that, apart from generalisations about correlations with other measures and interference from stereotypes, "little research has explored which linguistic features of speech are most crucial for Intelligibility and which, while noticeable or irritating, merely contribute to the perception of an accent." If teachers are expected to help learners meet the goal of Intelligibility, the least they should expect from research is an explanation of what is needed to meet that goal.

One reason for the lack of data is the added step of requiring listeners to demonstrate, through some performance, how well they have actually understood rather than relying on their report of comprehension. There can also be confusion because of the changing meaning of the terms, e.g. in Fayer and Krasinski (1995) as mentioned in **3.2.5**. Another reason may be that much research has focused on brief stretches of language, but according to Derwing and Munro (1997, p. 14) testing Intelligibility at the discourse level requires recordings of a reasonable length. Their study reported that neither speech rate, number of words, nor utterance duration (measured in seconds) were predictors of Intelligibility.

One of the rare examples of Intelligibility testing was by Suenobu, Kanzaki and Yamane (1992) who asked speakers of American English to transcribe examples of Japanese English. They found that added or missing segmentals and incorrect word stress were both predictors of Intelligibility. Benrabah's evidence for the crucial role of word stress (1997) was also

obtained through a transcription exercise (**4.3.2**). Tiffen (1992) found that the Intelligibility of Nigerian English was most affected by rhythm and stress, then vowel segments, and finally phoneme elision.

Another study (Tajima, Port, & Dalby, 1997) tested the effect of segment duration on Intelligibility. One example of Chinese-accented speech was artificially modified to match NS segment duration, which it was noted, would "simultaneously affect higher-order temporal features, such as syllable shape and global rhythmic properties" (p.6). The response accuracy of 36 US NS improved 15-25% after the modification. Reversing the modification to NS speech reduced its Intelligibility by 5-15%. This suggests that "syllable patterning, speaking rate, and stress-related durational contrast" (p.20) may affect Intelligibility.

While there is a lack of information about which features of speech predict Intelligibility, there is useful research on which kinds of teaching create better Intelligibility ratings, which is another form of evidence towards the same goal. Focusing on segmentals, Munro and Derwing (2006) showed that the functional load principle (**4.4.2**) could be used to identify priorities in improving Intelligibility. On the prosodic level, Derwing, Munro, and Wiebe (1997, 1998) showed that general pronunciation instruction including prosodic features could improve Intelligibility. Hahn (2004) demonstrated the importance of primary stress, particularly for given versus new information.

The question has also been raised of "intelligible to whom?" Milroy and Milroy (1999, p. 138) commented on the problem of language testers who are unaware of the features of local dialects and so might mark them as incorrect. The Oxford Placement Test, for example, expects a candidate to distinguish /eə/ and /ɪə/, a feat which many younger native speakers in New Zealand cannot achieve given the conflation of these sounds in NZ English. In a classroom, the criterion may well be "intelligible to the teacher", but as Kennedy and Trofimovich (2008, p. 481) point out, teacher familiarity with accents may result in students being rated too highly.

From all this we have some direct evidence that Intelligibility of various speakers can be affected by vowel and consonant deletion, segment duration, word stress and rhythm; and evidence from teaching that functional load and prosodic features including primary stress – as well as familiarity – can affect Intelligibility ratings. Clearly a range of features are important.

3.2.5   Comprehensibility

Comprehensibility is **the listener's experience of how difficult the speech is to understand** (Munro, 2011, p. 9). This is the definition used in this thesis. According to Saito, Trofimovich and Isaacs (2015, p. 18), "Comprehensibility captures the extent to which L2 speakers have reached a certain threshold of phonological, lexical, and grammatical ability needed for their conversational partners to successfully understand them." Comprehensibility is measured by asking listeners to report on the difficulty they experience in understanding a spoken text, and therefore relies on the listener's subjective judgement.

Evidence for the distinctiveness of Comprehensibility and Accentedness was found by measuring different types of instruction. Derwing, Munro and Wiebe (1998) used a controlled classroom-based study to compare the effect of segmental instruction versus suprasegmental instruction using a "global" approach which focused on larger units incorporating stress, intonation, and rhythm. While segmental instruction did a better job of improving ratings of Accentedness for read-out sentences, the suprasegmental approach improved ratings of Comprehensibility and fluency, but not Accentedness, in the narratives. Subsequent studies have validated this result (Derwing & Munro, 2011). Predictors of Accentedness and Comprehensibility are therefore not the same.

Other researchers have found that judgements of Comprehensibility are strongly predicted by grammatical accuracy (Fayer & Krasinski, 1987; Varonis & Gass, 1982) although this was not found by Jun and Li (2010). According to a study of French speakers of English by Trofimovich and Isaacs (2012), Accentedness was associated with word stress and rhythm by novice raters, but with segmental errors, syllable errors, and sounding nativelike by experienced teachers. Comprehensibility, on the other hand, was determined by word stress, type frequency, and grammatical accuracy for novice raters, while the teachers associated it with ease of understanding (which is self-evident), grammar, and vocabulary. A study (Saito et al., 2015) with 120 Japanese speakers of English had similar results while further research by Crowther, Trofimovich, Saito and Isaacs (2015) found that "lexicogrammar" predicted nearly half of the shared variance in their speakers' comprehensibility ratings (p.832).

Kang and Pickering (2011), in contrast, measured listener ratings of Comprehensibility against suprasegmental features and found that 50% of the variance could be accounted for by prosodic measures. The most significant of these were temporal features and tone choices, followed by boundary markers. This study reinforced the importance of fluency and prosodic features for pronunciation. Warren, Elgort and Crabbe (2009) found "significant correlation of comprehensibility ratings with each of sentence prosody, word stress, consonant pronunciation, and vowel pronunciation" (p.95). Wennerstrom (1998) showed intonation to be highly important in ratings of Comprehensibility, although Kang (2010b) found pitch to be most salient for predicting judgements of Accentedness.

The effects of heavy accent described in **3.2.3** show that accent does have some effect on Comprehensibility. Gill (1994) found that participants listening to a teacher with a standard accent were able to recall more information, because fewer cognitive resources are needed to process the input. Presumably the effect of familiarity (**3.3.2**) reduces the amount of attention required to process input, so familiar accents become more comprehensible.

Comprehensibility is therefore not an immovable standard. As the Asian population in New Zealand has nearly trebled in twenty years (Statistics New Zealand, 2015a) and Chinese visitors per year via Auckland now exceed 150,000 (Scoop Media, 2014), it is reasonable to expect that Chinese accents are becoming more familiar and therefore Chinese migrants and visitors will face less difficulty being understood.

Research has shown that the predictors of Comprehensibility include:

- Segmental features: Consonant and vowel pronunciation
- Prosodic features: Word stress, tone choices, boundary markers, intonation, pitch range
- Fluency features: Temporal measures
- Discourse features: Extensive use of communication strategies
- Familiarity with topic, person and language
- Grammatical accuracy
- Lexical features: Type frequency

In other words, just about any feature of pronunciation, or indeed of language, has been observed to affect Comprehensibility to some extent.

Furthermore, Comprehensibility, like Accentedness, is not an objective measure, as it is a judgement made by the listener: "there is no useful way to assess Accentedness and Comprehensibility, except through listener responses of some sort" (Munro, 2008, p. 200). Lindemann (2002) found that Korean ELLs could achieve a task successfully with American NSs, except in cases of some of the NSs who had negative attitudes to Koreans and behaved towards them as if their speech was hard to follow: it seems the NSs took "very little responsibility for the success of the interaction". Lindemann (2010) also showed that listeners' expectations can influence how they perceive the speaker's pronunciation, so judgements of incomprehensibility "cannot be taken at face value" (p.231) . There is also a range of listener factors, such as familiarity (**3.3.2**) and age, which affect whether accents are considered to be easily comprehensible (Munro, Derwing, & Holtby, 2012).

Attitudes to Comprehensibility

When NS employees were asked who they would prefer to talk with, Derwing and Munro (2009a) found that Comprehensibility played a greater role than Accentedness, and the experiment by Lev-Ari and Keysar (2010) described above in **3.2.3** also demonstrated that whether or not someone can be believed is affected by the Comprehensibility of their speech: "statements seemed less credible the harder it was to understand the speaker" (p.1095).

It appears that people infer personal qualities from Comprehensibility. In a study by McBride (2015) it was found that learners of Spanish were judged stereotypically by native speakers: "Speakers judged to be highly comprehensible were imagined to be attractive, nice people, whereas low Comprehensibility scores were associated with negative characteristics" (p.25). Fayer and Krasinski (1995) measured how hesitation phenomena related to judgements of "reported Intelligibility", finding strong correlations with the length of pauses rather than their frequency, but this was a measure of Comprehensibility.

It is necessary to ask whether good Comprehensibility is an adequate goal of the pronunciation component of a language programme. As Comprehensibility is "perceived" there are clearly cases where listeners choose not to perceive it. A speaker whose accent is distracting or irritating may be highly comprehensible but not highly acceptable. This question is addressed in the following section.

3.3   ACCEPTABILITY

Acceptability has had a range of meanings. For Chomsky and Halle (1968, pp. 416–8), acceptability in phonology referred to what was phonologically possible for a language, while for Quirk and Svartvik (1966) it was a way to estimate NS reaction to "linguistic deviance" in grammar (p.12). The term has also been used to mean the extent to which L2 deviates from language norms (Ludwig, 1982, p. 277) or the extent to which those deviations affect Intelligibility (Anderson-Hsieh, Johnson, & Koehler, 1992, p. 550 Notes). These definitions relate to correctness, which in pronunciation means degree of deviance from the norm, now defined as Accentedness. Acceptability has also appeared under the guises of irritation, annoyance and distraction, which are discussed in the following sections.

As a non-technical term, acceptability seems obvious. In each society we may judge certain examples of clothing, language, food and various forms of behaviour in terms of acceptability, which may have a range of grades between what is fully acceptable or completely unacceptable. However, the relevance of acceptability depends on how the response is manifested, which can range from a shrug of the shoulders to beheading. If we can show that the acceptability of forms of language is gradable, at which point are there consequences? Those who act as gatekeepers – such as employers – may translate judgements of acceptability into a job offer or a refusal. Furthermore, they may be acting upon the assumed acceptability of others – for example, would this accent turn my customers away?

If acceptability has consequences for the language learner – particularly in terms of whether they attain their ultimate goals – it is relevant to ask whether acceptability can be measured, and whether it can be distinguished from the other Global Measures.

3.3.1   Irritation and distraction

Matsuda, writing on discrimination against certain accents, commented that actual competence could be overshadowed by factors of attitude and irritation (1990, p. 1381). A number of other researchers have identified features of speech or NS attitudes which involve irritation or distraction.

McBride (2015, p. 25), for example, reported that even highly comprehensible speech amongst learners of Spanish could cause problems: "Some pronunciation inaccuracies, even

when not disrupting comprehensibility, were described as irritating …". Fayer and Krasinki (1987) claimed that irritation could be analysed into distraction and annoyance, but they do not show what value this distinction could have for teaching or learning pronunciation.

Albrechtsen, Henriksen and Faerch (1980), investigating NS reaction to NNSs, concluded that irritation "is directly predictable from the number of errors which an IL text contains, regardless of error type or other linguistic aspects of the text" (p.394); but they were uncertain as to whether this affected Acceptability, and much of the research cited for my thesis demonstrates that some errors, or the frequency of some errors, can be more noticeable than others. Anderson-Hsieh, Johnson and Koehler (1992) considered NS judgements to be based on "overall intelligibility, the irritability of the accent, or its acceptability" (p.530).

Hinofotis and Bailey (1981) found that undergraduates thought speech which was halting (i.e. with excess Pausing) and monotonous (lacking Variety) were the most problematic features of the speech of international training assistants (ITAs), and Fayer and Krasinski (1987) found hesitation phenomena both increased irritation and lowered Intelligibility.

Responses of irritation may also be the result of listener attitudes. Llurda (2000) found that a predictor of irritation was the listeners' belief about the L1 of the speaker, not the speaker's actual L1. Van den Doel evaluated NS reactions to Dutch accents, and concluded that Intelligibility alone cannot explain NS judgements of NNS errors. "Respondents' emotive reactions to certain stigmatised realisations indicate that factors such as irritation or amusement also play a part in prioritising certain errors over others" (2006, p. 287).

Overall, irritation and distraction are judgements which reduce the Acceptability of speech. No separate question was asked about irritation, so judgements in the Employer Survey would be subsumed under Acceptability ratings.

3.3.2   Familiarity

The numerous articles describing discrimination against speakers with accents may give the impression that there is nothing but bad news for NNS migrants. However, it is generally believed that familiarity with a foreign accent will improve its Intelligibility. This familiarity effect, which can make ESOL teachers poor judges of their students' Intelligibility, is well known (Thompson, 1991); and a number of researchers have commented on the extent to

which this may compromise raters' impartiality (Carey, Mannell, & Dunn, 2011; De la Zerda & Hopper, 1979).

The most frequently cited study of familiarity is that by Gass and Varonis (1984), in which the speech of two speakers was played in various forms to 142 NSs in the USA for them to transcribe. The researchers investigated various forms of familiarity, but it is important to note that "being previously familiar with the accent" was not one of them. They found that listeners who had just heard the accent as part of the experiment, and ESL teachers accustomed to foreign accents in general, had fewer errors. Gass and Varonis were therefore reporting on accommodation and professional expertise. This accommodation process, in which listeners can improve their processing of unfamiliar speech over the short term, was also observed by Clarke and Garrett (2004) and Bradlow and Bent (2008). This is different, however, from familiarity over a period of time with a specific non-native accent within the community. If familiarity and accommodation were the same thing, many comprehension problems would be solved after a short conversation. The study by Gass and Varonis was followed up by Kennedy and Trofimovich (2008) who did in fact find that that experienced listeners understood more of L2 speech which they were accustomed to hearing. The listeners' experience improved ratings of Intelligibility, but it is noteworthy that neither Comprehensibility nor effects of semantic context were affected, demonstrating a specific effect of phonological familiarity.

However, there are very few studies which ask whether people (other than professionals in migrant or teaching-related fields) have a better understanding of familiar non-native accents, possibly because this is taken for granted. To demonstrate a familiarity effect for foreign accent, the researcher would have to compare the Intelligibility of a foreign accent (preferably with another as a control) at two different points in time. This has hardly ever been done: as Valles (2015) pointed out, most studies have relied "primarily on the subjective judgments of listeners" (p.57), while many others simply show that there are Comprehensibility problems with foreign speech in general, which is another question.

Studies have shown a correlation between familiarity and Comprehensibility (both self-reported). Ballard (2013) asked university students to rate speakers for their suitability to be teachers, based on their accents, and found that familiarity with the accent correlated positively with both Comprehensibility and Acceptability for US, British and Chinese accented

speakers. Kraut and Wulff (2013, p. 259) found that raters who reported a lower level of familiarity with foreign-accented speech in general found the speakers to be significantly less comprehensible, more highly accented and having less communicative ability.

Some studies have used an objective rating to show that familiarity with non-native speech improves its Intelligibility to native speakers. In a study by Derwing and Munro (1997), native English listeners rated speech by ELLs from four languages in an Intelligibility task, and a correlation was found between familiarity and Intelligibility scores. Familiarity may also increase the listener's willingness to pay attention or their acceptance of what the other is saying, even to the extent of trying to comprehend a message which is not intelligible, as found by Matsuura, Chiba and Fujieda (1999). They reported that pre-existing familiarity with a variety of English predicted higher reports of Comprehensibility by listeners, even though Intelligibility (measured by dictation) did not improve.

It would be useful to have more specific evidence of emergent familiarity with a specific foreign accent. This would involve long-term research in which local NSs rate the Intelligibility of foreign accents heard in the community, and then having the procedure repeated after any of those communities have increased their numbers to the point where the accent was more commonly heard. It does appear that people have the ability to accommodate rapidly to a new accent regardless of prior familiarity, and there is clear evidence that Intelligibility is improved by longer-term familiarity with an accent.

Familiarity was not tested specifically in the Employer Survey, but some likely effects were found. When Employers were asked to guess the origin of the Migrants, they were correct in only 22.6% of responses. However, the two highest correct results were for Chinese (37.6%) and South Asian (mostly Indian) (32.7%). These two ethnic groups have been in New Zealand for longer than most others.

3.3.3 The Global Measure of Acceptability

Researchers have frequently commented on various types of Acceptability. While the Global Measures of Accentedness, Comprehensibility and Intelligibility can be assessed without context (as they are in formal assessments of language proficiency or research into these

measures) this cannot be done for Acceptability. It is therefore always a contextual measure: the term must refer to Acceptability for something which is external to the language itself, such as Acceptability for a manual job, to work in a call centre, to be a doctor, or to undertake postgraduate study.

As used here, Acceptability is defined as **a subjective Global Measure that pronunciation is good enough for a purpose** (from simple communication to offering employment). Bézier and Van Overbeke (1968) described this as "parler correctement, c'est parler à la satisfaction de son entourage immédiat"[1] (quoted in Baetens Beardsmore 1979, p. 127). Lakoff (1977), commenting on the Acceptability judgements of grammaticality found in Chomskyan grammar, wrote "psychological Acceptability outranks purely linguistic Acceptability" (p.76) – from which one could draw a parallel, in terms of pronunciation, that Acceptability outranks Intelligibility.

The idea of being "good enough for a purpose" is found in Van Els and de Bot (1987, p.147): "The degree to which a foreign accent is permissible is determined to a great extent by the aims that one has in mind learning and/or using the foreign language." As explained in **3.2.2**, Intelligibility has become the goal of teaching; but even Levis (2005) admitted that completely intelligible pronunciation may be evaluated negatively (p.376). For example, while Ballard (2013) found all of Comprehensibility, Intelligibility, and Acceptability to be highly correlated (see **3.3.2**), Isaacs (2008) found that teaching assistants rated with high Intelligibility were not necessarily rated as suitable for their jobs, concluding that "there is evidence that Intelligibility is a necessary but not a sufficient condition to be a TA in an undergraduate course" (p.571). Intelligibility is also a minimum target with little room for error. L2 speech is often harder to follow when there is background noise (Munro, 1998) so pronunciation needs to be developed to a point where it is comprehensible even in "less than ideal" conditions (Dauer & Browne, 1992).

As a Global Measure based on proficiency in pronunciation, Acceptability takes Intelligibility a step further because it is contextualised. It is possible that some intelligible and comprehensible speakers will be unacceptable in certain contexts, or that speakers who are less intelligible and comprehensible may be acceptable in different contexts. Therefore, these

---

[1] To speak correctly is to speak to the satisfaction of those around us (my translation).

three Global Measures will probably be correlated to some extent but distinct from one another.

The concept of Acceptability examined here is narrower than simple stereotyping or discrimination. If employers have stereotypical attitudes towards certain groups of NNS, and do not want to employ them, the problem is outside the scope of this study. But it is possible that certain speech traits historically associated with some ethnic groups, or with less competent speakers, may trigger a response regardless of the speaker's actual ethnicity or even perceived ethnicity. Thus the evidence shows that hesitation phenomena affect perception of overall competence (**5.4.4.2**), and that East Asian speakers who have "choppy" rather than smooth speech (**5.4.4.3**) may be stigmatised regardless of the speaker or of the employer's attitudes to other ethnic groups. This then becomes a pronunciation learning issue because teachers and learners may be able to address the problem.

Social Acceptability

Bansal (1990), writing on the pronunciation of English in India, wrote: "Even if a speaker is perfectly intelligible, it is possible that his speech is not acceptable to the listeners because it has marked regional or idiosyncratic features which sound unusual to them" (p.230). Anderson-Hsieh, Johnson and Koehler (1992) suggested that when native speakers evaluate non-native pronunciation, it is usually in terms of the speaker's overall Intelligibility, the irritability of the accent, or its Acceptability. Thus while it may be possible to obtain consistent scores on Intelligibility and other global measures, it is also necessary to ask whether the rater would give the speaker a job.

Acceptability of Pronunciation for Employment

An early investigation by Hopper and Williams (1973) had similar goals and methodology to this thesis: whether an interviewee's features of speech would affect an employer's attitudes and employment decisions. In the final phase, they used typical employment questions on adult males speaking standard US, Black, and Southern US accented English, and the responses were played to 40 employment interviewers. These interviewers responded to semantic differential questions, plus 5-point scales on the likelihood of hiring the speakers for each of seven job categories from executive to manual. The semantic scale reduced to three factors: competence, agreeableness, and self-assurance. They found that employer ratings of speech characteristics, and of competence, predicted only higher level positions.

They surmised (p.301) that speaking ability was seen as unimportant for manual positions. This may no longer be the case, as purely manual work has become less common, while the stricter legislation and rise in safety concerns mean that employers have a greater concern today about their employees' ability to communicate.

Similar research by Hopper (1977) found that judgements of competence and likeability were significant predictors of the hiring decision. Hopper's samples were from trained Black and White bi-dialectal informants, and raters were aware of the race of the speaker. Hopper found that the main effects were due to race and standardness of dialect, but they also interacted so that Black speakers of standard dialect were rated as even more acceptable than White speakers (p.347, the "accommodation hypothesis").

De la Zerda and Hopper (1979) studied Texan employers' responses to recorded speech samples to judge their attitudes towards applicants with Mexican American accents. They found that the more strongly accented speakers were less likely to be hired. For lower level positions, accented speakers were considered more suitable. They also asked questions about employers' attitudes to language use and accent, and found that two responses were significant predictors of employability for semi-skilled (and not higher level) workers: "I would not hire someone with an accent for a position of responsibility" and "speaking with an accent should not matter as long as the speaker can make himself understood", presumably reflecting the diversity of opinion amongst respondents. Given that the accent and speakers belonged to a specific ethnic group, such research provides us with evidence for stereotyping, but not whether particular pronunciation features could trigger these responses.

A study in Texas (Carlson & McHenry, 2006) found that stronger accents correlated with lower employability, and that Spanish speakers were rated as more employable than African American or Asian speakers. As with many other studies, since the raters were able to identify the accents it is impossible to separate reaction to speech from reaction to ethnicity.

A Canadian study by Kalin and Rayko (1978) showed that foreign-accented speakers were discriminated against in favour of English-Canadians. They asked ten male postgraduate students (5 NS, the others with various L1s) to record a 30 second identically scripted speech about current concerns. Student evaluators rated their speech on 9-point Likert scales for efficiency, reliability and other features. Foreign-accented speakers were rated lower for all

job categories, but higher for the lowest status. This supported the Hopper and Williams (1973) report, noting that "the importance of speech cue decreased with the status of the job in question" (p.1208).

Eisenstein, in a review of the literature on native reactions to non-native speech, concluded that employer judgements of job applicants was a crucial area. "In general, research to date shows that non-native speech causes problems for individuals seeking employment due to natives' negative views of the speaker's ability to perform a job" (1983, p. 174). Singer (1988) and Singer and Eder (1989) reported on ethnicity bias in job selection in New Zealand (for a detailed report, see **2.3.3**).

Hyman (2001) used a form of triangulation between migrant speakers, experienced raters and employers. The students were taking accent reduction courses and so were rated on accent, and the employers were asked to judge for occupational fit at the start and end of their accent reduction courses – that is, a judgement of Acceptability. The speakers were rated for their suitability for the roles at three levels. Although experienced raters agreed that the students' accents had generally improved, Hyman found no correlated change in employer judgements of Acceptability. Thomson (2014) would doubtless agree that the problem lay in the teaching of accent reduction as opposed to pronunciation instruction, since it is unlikely that Acceptability is most strongly influenced by Accentedness.

While there is no doubt that employers make judgements based on accent and ethnicity, this does not mean that these judgements are always negative. Cargile (2000), for example, used a matched-guise technique in mock job interviews and asked non-Asian undergraduates to assess whether the Standard American or Mandarin-accented speakers were suitable for different levels of work. Unlike other studies, no preference for speakers of one or the other accent was found in most cases. Cargile suggested that this is due to American attitudes towards Asians in the US as successful, and also that the responses were clearly to the accent and not the ethnicity, since the treatment of speakers was the same whether they were given a Chinese or Anglo-American name.

The central question of this thesis is to discover the features of migrant pronunciation which are acceptable to New Zealand employers, so all the Employers who took the survey were asked to give the migrant speakers an Acceptability rating of their pronunciation for the pur-

pose of employment. If Intelligibility was a sufficient goal, then speakers with a given Intelligibility would be equally acceptable. Often this is the case. As is shown in Table **7.6-2** Employer Acceptability ratings are partially correlated with reported Intelligibility (0.7), Comprehensibility (0.75) and Accentedness (0.5). This partial correlation shows that the Employers were willing to accept some candidates who they did not rate highly for Intelligibility; but they also rate some intelligible candidates as lower for Acceptability, suggesting that there are factors other than Intelligibility which learners need to aim for.

3.4 DISCUSSION

Accent is one of the distinguishing features of a NNS and sometimes the first to be noticed; it can be stigmatised or liked, but of all the global measures it is the least important for language development. We can accommodate rapidly to an unfamiliar accent, and there is evidence that long-term exposure can increase our ability to understand it. Adult migrants are less likely to attain a native-speaker accent than children, but they have different goals and may indeed not wish to. Furthermore, the idea of a standard accent which migrants must aspire to cannot be justified on educational, social, ethical or evidential grounds.

It is useful to measure migrant speech in terms of how different it is from native speech, how much of it can be understood, and how difficult it is to understand. There is considerable evidential support that these measures are different but partially correlated; but there is also evidence that native speakers have a range of attitudes towards non-native speakers which may affect the validity of their responses to these evaluations.

Teachers ESOL want to know how these global measures can assist them in setting long-term goals for pronunciation development, a field of study whose targets are learner and community dependent rather than objective, relative rather than stable. For some time now the standard goal for teaching pronunciation has been the Intelligibility Principle, but there is sufficient evidence now to show that this is neither standard nor necessarily sufficient. A student can develop a level of language which will pass exams (or an IELTS test) but not win a job: pronunciation for a given context must therefore be not merely intelligible but also acceptable. For this reason, Acceptability is needed as an independent Global Measure.

It is also essential for teachers and learners to be working towards this goal from the start rather than assuming it is a product of the higher levels of learning. Many of the migrants interviewed for the survey reported in this thesis were attending work programmes with the intention of finding a job long before they would have had the opportunity to reach a high proficiency in English. The effects of accent carry high stakes, making it essential for teachers and migrants to know what features of pronunciation are crucial for the goal Acceptability. The following chapters describe how different features of pronunciation would be likely to predict Acceptability for employment, the methodology used and the conclusions reached from the Employer Survey.

4   PREDICTING DIFFICULTIES FOR LEARNERS OF ENGLISH

4.1   OVERVIEW

Teaching pronunciation is based on the assumption that the learner's own phonological system is different from that of the target language system which needs to be established. In many cases the goal of Nativeness would demand the development of virtually every feature of pronunciation; in an ESL context, the goal of Intelligibility requires only that features be sufficiently developed to enable native speakers to understand what the learner is saying. Speakers of a given L1 do not all develop English speech in the same way, but teachers need to be prepared for the likely difficulties their learners may face. This has particular relevance in New Zealand, where English qualifications are currently benchmarked against the Common European Framework of Reference (see, for example, documents on NZQA, 2015) even though the CEFR was not designed for the non-European languages spoken by migrants in this country, or for the lack of formal education of some former refugees, or indeed for people lacking a shared European culture.

The aim of this chapter is to identify which features of migrant pronunciation are likely to cause difficulty for learners of English, particularly the migrants recorded for the Employer Survey[1], and consequently affect the Acceptability of migrant pronunciation to New Zealand employers.

In addition, it provides a resource for teachers in the form of the Notes on Recorded Languages (**4.3.2**) which summarise what is known about the phonological structure (sometimes controversial) of the languages of migrants recorded for the Employer Survey and how it could affect their learning of English Pronunciation. A number of the sources used are not widely known or easy to find, and in the cases of Amharic, Arakanese, Assyrian, Burmese, Chin, Karen and Tetum there is hardly anything published. This section therefore offers some information and references which have not previously been readily available.

As is shown in Chapter 7, Analysis, there are limitations on how far the Employers' responses can be predicted by specific pronunciation features. The results of analysis show

---

[1] See **Chapter 1 Introduction** for a summary and **Chapter 6 Methodology** for a detailed explanation of the Employer Survey.

the relative importance of segmental, prosodic and fluency features, but in practice do not permit further resolution into, for example, the relative importance of consonants or vowels. A detailed description of English phonological structure is therefore not presented here, but some comments are made on marked features of English which cause problems for a wide range of learners. The phonology of New Zealand English is likewise not described in any depth, but is referred to when some aspect might be relevant to judgements of Global Measures by native speakers.

4.2 Uncommon features of English pronunciation

Some features of English pronunciation make it an outlier amongst world languages, causing difficulties for many learners. One explanation for this is the universal principles of marked-ness[1], under which language learners have a preference for unmarked features (Eckman, 2008; Rice, 2007). English pronunciation appears to have an unusually high number of marked features.

Segmentals

It is expected that adults learners of any language will struggle to produce some of its pho-nemes well, but the many vowels in English create an unusually large number of distinctions to learn. English also has a word-final obstruent voicing contrast, which is the most marked position for this contrast (Eckman, 1987, pp. 61–2): speakers of many languages struggle with final voiced obstruents for various reasons. English phonotactics is also problematic: it is widely reported that learners of English from a range of other languages are likely to add, delete or reorder segments (Anderson-Hsieh et al., 1992, p. 534). There is a tendency on the part of many learners to reduce more complex syllables to CV [one consonant followed by one vowel] or struggle with consonant clusters, because the CV structure is unmarked: "All descriptive and theoretical studies of the syllable recognize that the CV syllable is an abso-lute universal in the languages of the world" (Carlisle, 2001, p. 2).

Segmental qualities may be influenced by other features. Preceding vowel duration, for ex-ample, influences English speakers' perception of the voicing feature in word-final stops and fricatives. Flege (1993, p. 1589) regarded this as an unusual feature: "The great interest in vowel duration arises from its apparently unique status in English. In prepausal position, the effect of stop voicing on preceding vowel duration is larger in English than in other lan-guages with final stops."

Prosody

"English is indeed almost in a class by itself as regards prosodic complexity, being one of the most complex languages on earth in this respect" (Whorf, 1964, p. 138). English word stress is both crucial for Intelligibility and more complex than in many other languages (Checklin, 2012). As the rules for assigning word stress are highly complex, learners must learn word

---

[1] Markedness is seen by some linguists in terms of frequency, normality or difficulty (Haspelmath, 2006).

stress as part of the lexical information associated with each word, making the task far more complex and time-consuming than for learning a language with a regular stress pattern. However, McCrockin (2012), while acknowledging this complexity, provides evidence that there are learnable predictive rules.

The English realisation of stress, which results in a tendency for vowels to be higher in pitch (particularly in phrasally prominent words), longer (duration) and louder (energy) than vowels in unstressed syllables (Checklin, 2012, p. 2; Dalton & Seidlhofer, 1994, p. 33) differs from that in many other languages. Kang's (2010b) study on the relative salience of supra-segmental features found that native speakers' judgements of Accentedness were most strongly affected by a reduced pitch range; and, according to Dalton and Seidlhofer, "the duration of syllables seems to be a far more important cue than in other languages" (1994, p. 19). Native speakers of English lengthen stressed syllables and shorten weak syllables, contributing to "unstress", which causes difficulties for many learners.

Mennen (2007) lists intonational errors which occur across a wide range of learners:

- pitch range is too narrow
- replacement of rises with falls and vice versa
- incorrect pitch on unstressed syllables
- difference in final pitch rise

Many learners and users of English as a foreign or additional language speak languages which are often argued to be syllable-timed (**5.3.2**), and they find the stress-based rhythm of English difficult to master. Stanfield (1986b, p. 39) suggests that most foreign learners have difficulty with this "essential feature of English".

In addition, the majority of the world's languages are tonal (Dalton & Seidlhofer, 1994, p. 19), and so speakers of these languages may perceive tone as carrying lexical rather than prosodic information. Four tonal languages were spoken by migrants recorded for this study. Juffs (1990, p. 104) suggested that the function of tones is lexical in Chinese but intonational or syntactic in English, and learners may have difficulty perceiving stress. This confusion about the use of stress and intonation, together with the L1 pattern of assigning a tone to every syllable, causes the staccato effect (choppiness) which can be heard in some Chinese-accented speech.

The above evidence suggests that English intonation structure is uncommon. English depends more than many other languages on prosody for emphasis, changing the pitch of the primary stressed syllable of the focus word, and lengthening the vowel. Languages such as German, Japanese and Spanish have a greater tendency to use lexical items or word order for this purpose (Gilbert, 2014, p. 120).

Fluency

The results of the Employer Survey showed that Fluency is a significant predictor of acceptability, and therefore that developing Fluency features is crucial when learning English pronunciation. English makes use of word linking to create a fairly "smooth" effect and speakers use – and prefer to hear – varied intonation. While these features are not easy for some learners to adopt, there is no evidence that they are marked in their difference from those of other languages.

4.3   LANGUAGES OF SPEAKERS RECORDED FOR EMPLOYER SURVEY

The migrants recorded in the Employer Survey were typical of those in New Zealand language schools at the time of writing, and spoke a total of 19 languages (Table **4.3-1**), assigned in this study to geographic and linguistic groups as described in Table **6.3-1** and **Appendix B5.**

4.3.1   English pronunciation compared to recorded languages

The unusual features of English pronunciation noted in **4.2** are likely to cause difficulties for speakers of many of these languages. The Contrastive Analysis Hypothesis (CAH) (Lado, 1957) offered a strong version of this assumption (depending on the L1), but the Structural Conformity Hypothesis (Eckman, 1981, 1991) showed that learners also use IL strategies and follow universal generalisations (markedness) which cannot be predicted from L1 phonology. The revised CAH therefore incorporates markedness in its predictions.

Munro and Derwing (2008, p. 492) suggested that vowel inventory size could be an important factor in vowel acquisition. Table **4.3-1** shows that while English dialects have around 20 vowel phonemes (fewer for Standard American English), the languages which were spoken by the Migrants recorded for the Employer Survey and for which data can be obtained have an average of 11 vowels, around half the English total. This can be problematic because it increases the likelihood that English will have two vowels in the space where the L1 has only one, making it difficult to perceive and produce the new distinction (Flege, 2003; Koffi, 2012). Even the Vietnamese speakers, whose language has far more vowels, face difficulties because their vowels occupy different spaces. There are no noteworthy global problems with English consonant rarity (the [Ө] phone, for example, is rare but has low functional load and is frequently substituted by native speakers) but English has more consonant clusters, and more complex ones, than many of these languages.

Table **4.3-1** compares specific information about languages compared to English, taken (throughout the table and the rest of this section) from Katzner (2002), from Bauer (2010) on language features and from Swan and Smith (2001) (or 1987 for Vietnamese) plus other references as noted. A Linguistic Distance scale from Chiswick and Miller (2005) with a scale of 3 (closest) to 1 (most distant) is added where available. This covers distance in all features, and was designed for English learners of other languages, so may be different when

reversed. The number of vowels given is total vowel phonemes including diphthongs; conso-
nant clusters are marked with * if the type or lower number of consonant clusters in L1 is
likely to create a number of segmental problems in learning English.

Table 4.3-1 Data on Languages of Migrants Recorded

Language	Vowels	Consonant Clusters	Word Stress	Rhythm	Linguistic Distance
English	20		Complex [1]	Stress	-
Amharic	7		Weak, position variable[2]	Stress[3]	2.00
Arabic	8	*	Quantity	Stress	1.50
Arakanese	?	?	?	?	
Assyrian[4]	6	*	Usually penultimate	Stress	
Burmese	11[5]	*	Major syllables stressed	?	1.75
Chin[6]	12				
Chinese	9	*	N/A	?	1.50
German	18		Word-initial	Stress	2.25
Italian	7	*	Variable – final 3 syllables	?	2.50
Japanese	5	*	Pitch only[7]. Variable.	Mora	1.00
Karen	13	*	Major syllables stressed	?	
Sinhalese	14	*	?	?	1.75
Somali[8]	15		?	Syllable	
Spanish	10	*	Penultimate	Syllable	2.25
Tamil	12	*	Weight; most initial	Syllable	
Telugu	12	*	Quantity	?	
Tetum	5		Penultimate	?	
Turkish	8	*	Usually word-final	None[9]	2.00
Vietnamese	31	*	N/A	?	1.50

---

[1] (Checklin, 2012)
[2] Hayward and Hayward (1992)
[3] Haile (1987)
[4] Odisho (1988)
[5] Or, according to Watkins (2001), 2 sets of 9 vowels in largely complementary distribution.
[6] There is very little information on Chin. This is from *Chin Language Pronunciation* (Lind, N.D.)
[7] Ohata (2004, p. 10)
[8] Kahin (2011). In addition to the 10 vowels noted in other sources, Kahin adds 5 diphthongs.
[9] Lees et al. (2013)

Information is scarce for some of the languages in Table **4.3-1**, particularly the Myanmarese languages, and for features such as rhythm, which is a notable problem for a large number of these speakers because of its use of unstress. Many of the features given in the table may be subject to interpretation, and L1 transfer may be subject to other influences.

4.3.2   Notes on Recorded Languages

Below are notes on specific difficulties which could be predicted from speakers of the recorded languages when learning to pronounce English accurately or intelligibly. These are intended very broadly: comments on the phonology of languages may be valid only for canonical or well-studied dialects. In addition, predictions based on contrastive analysis may well fail because of universal effects such as the unmarked universal stop system (Win, 1998, p. 203).

Amharic

One female and two male Amharic speakers were recorded.

Amharic is syllable-timed and has consonant clusters but not word-initially. According to some researchers, stress is uniform but the last syllable of a word tends not to be stressed (UCLA Language Materials Project, 2014). Sentence accent predominates over word stress, which is much weaker than in English (Haile, 1987). From this it is likely that Amharic speakers may conflate some English vowels, have difficulty with final consonant clusters, and use atypical word stress. According to Hayward and Hayward (1992) the stop [p] exists only in foreign borrowings.

Arabic

One female and two male Arabic (Iraqi) speakers were recorded.

English has many more vowels than Arabic, so Arabic speakers may have transfer problems (Munro, 1993); and there are problems producing schwa [in unstress] instead of full vowels (El-Hassan, 1994). Arabic speakers also have difficulty with consonant clusters (Jayaraman, 2010) and the [p] – [b] contrast. Arabic and English have different syllable structure and linkage rules, which can lead to word juncture misinterpretations. Binghadeer (2008) found that

Arabic speakers' pitch range was narrower than that of speakers of British English, which could interfere with intonation.

Assyrian

One male Assyrian speaker was recorded.

Little information is available on problems for Assyrian speakers, but from my experience their speech is similar to speakers of related languages (Arabic, Amharic) who have some noticeable variations but high Intelligibility. The small number of vowels may make some English vowel distinctions difficult.

Chinese

Two female Chinese (Mandarin) speakers were recorded.

Chinese have problems with a number of English phonemes, especially vowels (Liang, 2014, p. 65). I have found that /ɛ/ is routinely replaced by [eɪ], so that "letter" sounds like "later" and "negative" can be misinterpreted as "native". Hansen (2001) found that Chinese speakers often insert a vowel after final consonants, particularly /t/ or /d/, sometimes to avoid appearing "lazy" by dropping consonants (Deterding, 2010). There are few final consonants, and depending on region of origin there can be confusion between [n] and [l] or [ɹ] and [ʒ]. Stops are distinguished by aspiration rather than voicing (Stanfield, 1986a, p. 5) and this may also affect affricates. Chinese /r/ is very similar to the English sound; however, it can be more retroflex and the category in Chinese is shared with [ʒ] so it is common to hear "urally" for "usually".

Chinese is a tonal language. Speakers have a narrower pitch range and do not use English intonation effectively (Pickering, 2004). Syllable stress exists in Chinese but does not have the same importance as in English; often, too many words are stressed when speaking English, creating a "jerky effect" (Stanfield, 1986a), measured in this thesis under Smoothness. Liu (2007) found that Chinese learners need specific assistance to cope with English word stress.

German

One male German speaker was recorded.

The English and German phonological systems are broadly similar (Swan, 2001). It is only in intonation that significant differences appear (Grover, Jamieson, & Dobrovolsky, 1987, p. 279).

Italian

One female and one male Italian speakers were recorded.

The main difficulties for Italian speakers are rhythm and the irregular word stress of English. Italian speakers are also well known for epenthetic vowels after final consonants. Like Spanish and many other languages, Italian tends to use word order for emphasis so speakers may not use English intonation patterns correctly (Duguid, 2001).

Japanese

One female and two male Japanese speakers were recorded.

In Japanese spoken English, words which contain errors of consonant deletion are hardest for native speakers to understand; vowel addition, rather than causing a problem, actually helps Intelligibility (Suenobu et al., 1992). The Japanese /r/ - /l/ problem is well known. [f], [v], [Ɵ] and [đ] do not occur in Japanese and may be replaced by [h], [b], [s] and [z] respectively, and Japanese allophonic rules may affect pronunciation of categories which the languages do have in common. Initial and final consonant clusters cause problems as they are absent in Japanese (Stanfield, 1986b). Syllable duration can affect NS evaluation of Japanese speech, and failure to shorten vowels in unstressed syllables, and lengthen stressed vowels, causes problems (Nakamura, 2010, p. 13).

Word stress in Japanese is, like English, lexically contrastive (Beckman & Pierrehumbert, 1986, p. 2) although Japanese phonology allows utterances which have no stressed syllables. Ohata (2004) reported that Japanese stress realisation is dependent only on pitch, and there is less overall pitch variation, which can sound monotonous to English listeners. Ohata also notes that Japanese does not make the lax/tense distinction found in English, and has only open syllables.

According to Stanfield (1986b, p. 35) suprasegmental features are "considerably less signifi-cant overall in Japanese and contribute to a lack of intelligibility." Stanfield also noted that Japanese words are "cut off separately", creating a staccato effect (lack of Smoothness) which affects the listener's impression of fluency.

Myanmarese languages

Recordings were made of one male Arakanese (or Rakhine), one male and one female Bur-mese, one female Chin, and one male Karen speaker (subgroup unknown). The latter two are names for a large number of different dialects and languages, and they may not be similar in phonological structure. Stanfield (1986a), however, wrote that "all these languages have many features in common, both in phonology and grammatical structure."

Geba Karen is tonal, has 34 consonant phonemes of which 4 glides can appear in initial or medial clusters, and there are 9 vowels (Shee, 2008). Sgaw Karen is a major dialect and lin-gua franca amongst Karens. It is primarily a monosyllabic language and is tonal. It lacks the phones [f], [g], [v], [ð], and [ʒ], has few final consonants, and the voicing contrast of stops in English is problematic. Its initial vowels are always preceded by a glottal stop (Zetterholm, 2014) which if transferred would make speech less smooth. Final stops and nasals are not pronounced and epenthetic vowels are often inserted into English consonant clusters. Zet-terholm describes a Swedish contrast between aspirated unvoiced stops and unaspirated voiced stops in stressed syllables similar to that in English, and notes that this is difficult for Karen speakers.

Burmese is tonal and generally has monosyllabic words. It has few final consonants and no three-segment consonant clusters (A.-C. Cunningham, 2013; Stanfield, 1986a). Only initial C or C+ glide is permitted (Green, 2005, p. 3) which means that Burmese speakers may add [ʔ] to English initial vowels, giving a choppy effect, and they tend to leave large parts of English words unpronounced. Stanfield (op. cit. p.2) reported that Burmese has no central vowels and both schwa and vowel reduction would be a problem, as Burmese is syllable timed, and there would be difficulty with a number of consonants.

The absence of most consonants in syllable codas (other than [ʔ] according to Watkins (2001), both [ʔ] and [n] according to Stanfield) explains the difficulties, often attested by

teachers, which Burmese speakers have with English final consonants. Stanfield also reported that Burmese borrowings from English reduce final nasals to [n] and all stops and fricatives to [ʔ], and (as for Sgaw Karen pronunciation of English) consonant clusters are often altered by assimilation or epenthetic vowels. Green (2005) reported that final obstruents in borrowed words all become [ʔ] in Burmese.

Win (1998, p. 202) found that not all of the predictions in Stanfield (1986a) were correct, as the speakers had little difficulty with the consonants. Also, despite the presence of voiceless aspirated stops in both languages, Burmese speakers of English followed the universal unmarked system and used unaspirated stops, which decreased Intelligibility. Win also found that the speakers gave stressed and unstressed syllables equal length (p.81), and the syllable-timing created a "machine-gun rhythm" (lack of Smoothness). Most speakers also used falling intonation for a range of sentence types including those questions where native speakers would expect rising intonation. She also found (p78) that word stress could be realised in a variety of ways, which could be a significant cause of communication problems.

Somali

Two male Somali speakers were recorded.

Tone in Somali can be used to mark gender, case, and number, but sentence level intonation may be very different from English. Somali lacks [p], [tʃ], [v], [z] and [ʒ] (A.-C. Cunningham, 2013) which may cause difficulty in differentiating these sounds. Somali has a simple CV(C) syllable structure with no consonant clusters, so speakers often insert vowels, particularly in final clusters (Kahin, 2011). Epenthetic formations such as [sɪkɪs] for "six" are common.

Somali pronunciation of /p/ usually sounds like /b/ to English speakers, particularly at the beginning of words. Speakers also confuse /f/ - /v/ and /s/ - /z/, and the vowel in the English -ed suffix tends to be pronounced in all circumstances. This is not seen as spelling pronunciation as it also occurs amongst illiterate Somali speakers of English (Koffi, 2010). Koffi (2012) also measured Somali speech acoustically to show that speakers produce the English lax vowel contrasts /ɪ/ - /ɛ/ and /ɔ/ - /ʌ/ too closely to be distinguished.

Most Somali syllables have the same length and the same stress, and there is no grouping into rhythms units as in English. Weak forms are a problem (Kahin, 2011).

South Asian

Recordings were made of one female speaker of Sinhalese, one female and one male Tamil speaker, and one female and one male Telugu speakers.

This region includes Dravidian (e.g. Tamil and Telugu) and Indo-European (e.g. Sinhalese) languages. As Sri Lanka and South India are part of the Outer Circle of English speakers (Murphy, 2014a) it is likely that migrants from those countries will have learned English as part of an established regional variety, and their interlanguage may be based on either or both of a South Asian L1 and South Asian English. This can create an additional problem, since they may be confident and fluent but not highly comprehensible in New Zealand: Basnayaka (1999) reported that Sri Lankans were often turned down for interviews because of their accent, and Pio (2005) said the same for Indians.

Dravidian languages do not have aspirated consonants, and voicing of consonants is not phonemic – both of which are likely to reduce Intelligibility. A number of the consonants also cause other difficulties. The word stress pattern is very different.

Speakers of South Asian Indo-European languages may have heavily aspirated consonants but tend to use non-aspirated or even retroflex consonants when speaking English. Benrabah (1997) demonstrated the seriousness of the word stress problem in Indian speech (L1 not stated) for comprehension by British native speakers, although his data involved single sentences cut from spontaneous speech, thus removing context which is also used to interpret meaning. Indian English tends to use full vowels and regularly uses a different word stress from other dialects, such as [vɛˈhaɪkɑl] compared to NZ [ˈviːjəko] for "vehicle". South Asian intonation is very different from that of English (Fuchs, 2015).

Spanish (Latin American)

One female and four male Spanish speakers from Colombia and Ecuador were recorded.

Spanish speakers may have problems with final consonant clusters prohibited in Spanish and many vowels, diphthongs and consonants can be conflated. Flege and Eefting (1987) showed that Spanish speakers had shorter voice onset time (VOT) values for unvoiced stops, which may make them harder to distinguish from the voiced equivalents. In Spanish the latter are realised allophonically as fricatives finally and intervocalically, which means that Spanish speakers are likely to perceive and produce /d/ and /ð/ in the same category.

English speakers may use high pitch at phrase boundaries but Spanish speakers do not, and Spanish speakers of English produced less reduction of pitch on redundant information (Wennerstrom, 1994). This could affect Comprehensibility.

Spanish is syllable timed so speakers also have some difficulties with English stress and rhythm (Yavas, 2011, p. 189) and with the changing placement of stress in pairs like *apply / application* (Flege & Bohn, 1989). I have found that Colombian Spanish is spoken rapidly, and this is transferred into English, reducing Comprehensibility.

Tetum

One female and two male Tetum speakers were recorded.

There is little data about Tetum. The Australian Commonwealth Schools Commission's Asian Language Notes 2 (Stanfield, 1986b) admitted that "the data available on Tetum are scanty" and the tentative description suggests that its predictions were drawn from the language itself rather than observation of actual learners. Little has been produced since then. Stanfield noted that [p], [g] and a number of English fricatives and affricates are absent, which is likely to cause problems. However, Williams-van Klinken (2003) pointed out that some of these (e.g. [p], [ʃ] and [dʒ]) exist in Portuguese loan-words and are used by educated speakers while others may replace them with indigenous Tetum alternatives.

Williams-van Klinken (2003) reports that Tetum stress is usually penultimate. Tetum has a largely CV syllable structure but has additional consonant clusters from its large Portuguese vocabulary. Like Maori, Tetum has only five pure vowels, all of which can be long or short and are never reduced, but Portuguese borrowings have diphthongs. Cunningham (2013) reports that there is no consonant aspiration. The influence of loan words and different types of speaker make prediction difficult.

Turkish

One male Turkish speaker was recorded.

The main problems faced by speakers of Turkish are final voicing contrasts, initial consonant clusters, and the limited number of vowels (Yavas, 2011, p. 192) so [ɪ] – [iː], [æ] – [ɛ] and [ɒ]–[ɔː] contrasts do not exist. Arslan and Field (1997) found that Turkish speakers of English showed much less variation in their intonation contours when compared to other accents. According to Altmann (2006), Turkish speakers have difficulty with English stress timing and

unstress. Word stress is usually on the final syllable but may be moved to penultimate for morphological reasons (Zimmer & Orgun, 1992) so irregular English stress may be a problem. Demirezen (2016) also found that coarticulation is a serious problem for Turkish learners.

Vietnamese

Three female and one male Vietnamese speakers were recorded.

Teachers of English find that Vietnamese learners have greater problems mastering English phonology than most others. Cunningham (2009) commented that even Vietnamese listeners have difficulty picking up which English word another Vietnamese speaker has said, partly because consonant clusters are generally simplified to the point of unintelligibility, particularly those in word-final position (N. Nguyen, 2008). Only voiceless stops and nasals occur finally (Stanfield, 1986a, p. 18) and the stops may be unreleased; and /g/ and /dʒ/ in any position are particularly difficult.

Even though Vietnamese has a highly complex vowel system, Cunningham also noted that many English contrasts, including /ɪ/ – /iː/, /æ/ – /ɛ/ and /ɜː/ – /ɔː/, are problematic. However, Macneil (1987, p. 63) claimed that "the differences between the English and Vietnamese sound systems are no greater than for most languages" and the difficulties lie elsewhere: that Vietnamese is monosyllabic, tonal, syllable-timed, pronounced in staccato manner, and has difference phonotactic rules.

Vietnamese also struggle with English pitch prominence and coarticulation. Vietnamese is "unable to express stress by tone, as in English" (Honey, 1987) and speakers tend to stress all syllables equally, producing a "staccato" quality (lack of Smoothness). Contrastive stress is problematic.

Although Vietnamese is a tonal language, its speakers can learn English stress and prosodic features with specific teaching but this is difficult for many (T. Nguyen & Ingram, 2005). This combination of problems often makes Vietnamese pronunciation of English particularly unintelligible.

4.4 PREDICTING ERRORS

Planning a pronunciation syllabus is far more difficult than preparing one for grammar. Firstly, pronunciation is learner-dependent: whatever effect universals may have, the effects of L1 are always noticeable and may require different types and levels of intervention from the teacher. Secondly, as is explained in the conclusions of this thesis, pronunciation is not a sequential process as, for example, the teaching of English tenses is: segmental accuracy, word stress, intonation and coarticulation can and should all be taught from the earliest stages.

Any prior knowledge about what kind of help learners are likely to need would therefore be useful to the teacher. Knowledge of specific languages is useful, but a number of researchers have also sought evidence for a hierarchy of errors to identify which learner difficulties have stronger negative effects on listeners than others, and hence which features teachers should spend most time on.

4.4.1 Error Hierarchy

Error gravity is based on the assumption that "some linguistic errors are more serious than others in terms of disrupting a NS's comprehension of a NNS's message and that these error types can be identified" (Rifkin & Roberts, 1995, p. 512) – that is, that an error hierarchy exists. The concept, if valid, would be useful in determining which features should be emphasised in teaching. To identify possible universal evidence, it is worth asking whether there is any natural hierarchy which would lead us to expect some errors to be more salient.

Each level in Figure **4.4-1** below sets the boundary for the one above. Thus, according to Pennington (1996, p. 157), vowels "spread across stretches of speech" so are at a broader level than consonants. Errors at lower levels will have more effect than those higher up. Fluency has been added to this chart; in the current thesis, fluency is the broadest of three levels of pronunciation features (the others being segmental and prosodic) and includes features such as speed and hesitancy, so it is therefore the lowest level on the chart below. Speakers of different L1s may make different errors, but a hierarchy such as this is based on priorities for English, and so would be relevant regardless of L1.

Figure 4.4-1 Hierarchical relationship of phonological parameters
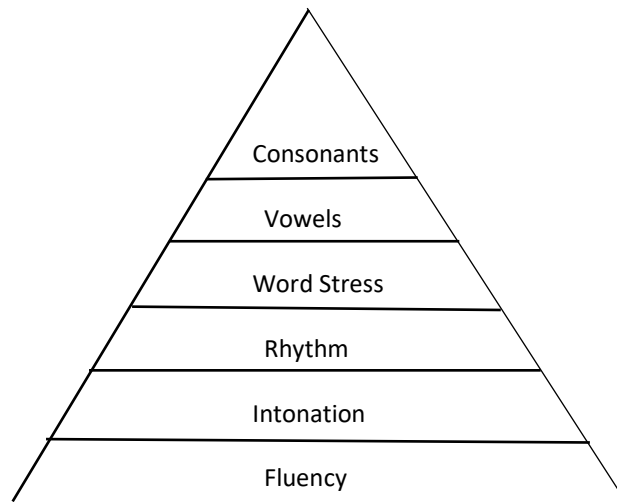
Based on Pennington (1996, p. 157)

This chart does not necessarily predict either Intelligibility or Acceptability, as it is possible that errors in features high on the chart such as frequent missing consonants may be so serious as to outrank the lower levels (Pennington, 1996, p. 253). However, it is useful as a conceptual framework, because it is worth investigating whether improvements in broader levels such as fluency or intonation do in fact have a greater impact on the listener than segmentals or word stress.

4.4.2   Functional load

Functional load (FL) is a limited case of error hierarchy, and the one which has achieved greatest agreement, based on the idea that some segmental distinctions carry more weight than others because of the number of minimal pairs distinguished by the contrast between the two phonemes, and the cumulative frequency of the contrast (Van den Doel, 2006, p. 18). It is also described as a theory under which segmental contrasts can be ranked according to their "performance in English pronunciation" – how likely they are to affect listeners' comprehension (Kang & Moran, 2014, p. 177) – to create a hierarchy in Comprehensibility.

According to Munro and Derwing's Research Timeline (2011, p. 323), Catford (1987) may have been the first to show how this concept could be used to set priorities in pronunciation teaching. Catford provided a ranked list of phoneme contrasts and suggested (p.92) that

pronunciation of consonants ranked higher than that of vowels. Adam Brown (1988) also developed a list of contrasts and recommendations for its use. The theory of functional load was given some confirmation by an analysis by Munro and Derwing (2006) of judgements of Cantonese speakers of English which found that high functional load errors had a greater impact on listeners' perceptions of the Accentedness and Comprehensibility of their speech than did low functional load errors (p.529), although Van den Doel (2006) found that some low functional load errors could also be significant (see **4.4.3**).

Koffi (2013) showed that functional load and perceptual closeness of vowels could be used to assist in explaining the relative difficulty in the Intelligibility of certain vowels in General American English, and the consequent difficulties English learners had when trying to pronounce the sounds distinctly. The theory was further supported by Kang and Moran (2014) who aimed to identify the salient features which distinguish Cambridge ESOL General English Examinations levels. They found that there was a significant drop in high functional load errors between low and high proficiency speakers. High level speakers still made errors but mostly in low functional load segments.

Overall, then, both theoretical descriptions and research have supported the concept and instructional value of a rank ordering of phoneme contrasts in terms of their significance for Comprehensibility.

4.4.3   General Research on error hierarchy

Johansson (1978b) noted the lack of information about which errors were more serious and called for research to "establish the communicative effect of different types of errors". He showed in an error analysis chart (1978a, p. 42) that error gravity (how serious the error is to a native speaker) and frequency (how often the error occurs) could lead to prioritisation for language courses.

Albrechtsen, Henriksen and Faerch, on the other hand, concluded that no hierarchy of errors in relation to irritation (**3.3.1**) could be predicted, because any violation of the formal code could be an irritant (1980, p. 395). Anderson-Hsieh, Johnson and Koehler (1992) considered NS judgements to be based on overall impressions rather than any accumulation of errors.

Other researchers have investigated learner errors and reached different conclusions. Bond and Small (1983), for example, investigated the "salience" (i.e. error gravity) of incorrect voicing of obstruents, the front-back dimension of stressed vowels, and word stress, concluding that errors in the voicing of obstruents were less disruptive than vowel mispronunciations. They found that wrong word stress produced a high level of disruption, but as word stress inevitably affects English vowel quality (Acton & Finney, 1978) two features are involved. Flege and Bohn (1989) studied stress placement of individual words by Spanish-speaking learners of English, and concluded that there could be a "hierarchy of difficulty and order of acquisition for L2 learners" in learning unstress and vowel reduction.

Hinofotis and Bailey (1981) asked a question similar to the basis for this thesis: "What areas of non-native speaker communication do the [US] undergraduates perceive as problematic in prospective TA's?" (p.120). Hinofotis and Bailey tested a wide range of communication features including vocabulary, grammar, pronunciation, speech flow, confidence and content. Their raters agreed that pronunciation was the single most important feature, and that flow of speech (Fluency) was important, suggesting – as did Fayer and Krasinski (1987) – that halting speech (lack of Fluidity) causes irritation. Speech flow is a fluency factor which would appear near the bottom (i.e. broadest and most influential area) of Figure **4.4 1**.

Ludwig (1982) reviewed twelve papers on native speaker reactions to NNS speech, commenting that most of these studies suffered from lack of authenticity (p.276). Schairer (1992) pointed out that the results of papers to date had been diverse, possibly because of the range of cultures and languages, but concluded that there was indeed a hierarchy of errors (p.318). Rifkin and Roberts (1995) examined 28 investigations into error gravity from 1977 to 1995 and critiqued their methodology, using criteria such as authenticity of speech samples, whether they were placed in a wider communicative context, the use of objective versus subjective tests, and control of order effects. Of the five papers which dealt with spoken English (including Albrechtsen et al., 1980; Fayer & Krasinski, 1987; Johansson, 1978b; Varonis & Gass, 1982) none met all of the criteria, and the variety of methodologies made comparison difficult. Rifkin and Roberts also asked whether researchers could be sure that any error was lexical or syntactic without using objective tests.

Munro and Derwing (1995a) were also unconvinced that any hierarchy had been demonstrated, claiming that there was "little empirical evidence regarding the role of pronunciation in determining intelligibility" nor of which features were most crucial (p.288). However, Pennington (1996) includes a "Hierarchical analysis of student pronunciation" (p.253) as a guide to teachers of areas to work on in "descending order of generality and importance". Her level of greatest importance was the fluency feature of Variety, starting with intonation which is monotonic or unusually high or low at times and monotonic stress; then pervasive segmental errors. The second level included the prosodic feature of Word Stress, then segmental substitutions.

A more recent comprehensive study which tested error hierarchy is the doctoral thesis *How Friendly are the Natives* (Van den Doel, 2006), which engaged 577 native speakers from throughout the English-speaking world including New Zealand and used current statistical procedures to examine native speaker reaction to typical pronunciation errors of Dutch learners of English. Its results challenge conclusions from a number of previous studies.

Contrary to Albrechtsen et al. (1980), Van den Doel suggested that there is at least a hierarchy of irritation, noting "respondents' comments indicate that, across different groups of native speakers, some errors are clearly and consistently more irritating than others" (p.287) and concluding that errors could be more or less salient depending on word position (p.288). He found a hierarchy for suprasegmental features under which word stress was most important, failure to weaken unstressed forms less so, and intonation errors least important (p.289).

These results suggest that native speakers are using criteria that go beyond Intelligibility and Comprehensibility and include the Global Measure which Van den Doel describes as irritation. As mentioned in **4.4.2** above, functional load theory was also challenged because some low frequency contrasts are significant. In contrast to Munro and Derwing (1995a), Van den Doel found vowel and consonant errors to be equally significant.

Johansson (1978b) found that phonemic errors were usually more severe than subphonemic ones, but Van den Doel found that this was not always so, recalling Standwell's (1991) argument that allophones are often so distinctive that phonemes are a "red herring".

Van den Doel's research has limitations: the raters were found through open web access, the survey studied responses to Dutch accents only, and it used actors who could reproduce single errors rather than spontaneous speech. It nevertheless shows that error hierarchies can exist.

The availability of technology which can record speech with higher fidelity and enable vast processing power for acoustic analysis has encouraged researchers to use increasingly fine observations to determine which features of pronunciation are rated mostly highly. Ginther, Dimova and Yang (2010) examined temporal features and noted that higher scores were given to those who could speak faster, say more and pause less (p.388), suggesting that fluency features were important.

The most exhaustive work in this area has been done by Okim Kang, who analysed the speech and scores of 120 Cambridge ESOL General English Examination candidates (Kang, 2012). Her analysis included a wide range of features categorised under segmental, fluency, stress and pitch, and tone choices, with detailed measurements of speed and pause length. Kang's analysis suggested a hierarchical priority with stress and pitch the most salient, then fluency, then segmental errors, with tone choice considerably lower (p.13). It should be noted that Kang was measuring features against examination outcomes rather than native speaker judgements in more real-life settings, thus not meeting the authenticity criterion (Rifkin & Roberts, 1995). Also, the predictions are measured against Nativeness, and thus might not be the same as predictions against other global measures which native speakers might use under other circumstances (see comments on the contrast between judgements of Nativeness versus Intelligibility at the end of **3.2.1**).

It is also worth investigating the possibility that phonological universals can at least help to predict likely learner errors. Eckman (1991) claimed that IL consonant clusters may follow L1 or L2 rules, but do not violate universal generalisations; which is supported by research from Koffi (2012) who showed that the common errors in vowels [ɪ, ɛ, ɔ, ʌ] cannot be explained by contrastive analysis. Thus a universal feature hierarchy such as that in Ladefoged (2007, p. 162) may help predict segmental learning difficulty in any language.

Munro's (2011) criteria for a hierarchy are based on:

- Identifying phonetic properties of L2 speech that reduce Intelligibility and Comprehensibility;

- Identifying improvements that genuinely make a difference for communication, and

- Setting teaching priorities, given the limited time available.

His resultant priorities for teaching are (without any order):

- Primary (nuclear) stress

- Rhythmic patterns

- Global prosodic instruction

- Segmental teaching based on functional load

However, a table in Derwing and Munro (2015, p. 73) demonstrated the contradictory conclusions reached by different papers, probably because of the wide range of variables involved, including methodology, tasks, analyses and speakers' L1s. They concluded that "individual differences and variable error gravity" must be taken into account for an evidence-based approach (op. cit. p.76).

From the wide variety of conclusions, both positive and negative, it would appear that different features will rank more highly under different circumstances, and the only consistently identified hierarchy of errors is segmental functional load.

4.5 DISCUSSION

This chapter has shown that English has many segmental and prosodic features which are different from those found in many L1s and are known to cause difficulty to speakers of a wide range of other languages. If these features affect the Comprehensibility and Intelligibility of speakers then they may affect their Acceptability as well. This does not provide evidence for an error hierarchy, but does suggest that pronunciation should be emphasised in any teaching programme since there is a good chance that some of the students will be having difficulties.

Knowing what difficulties speakers of a given L1 might have is useful for teachers, as they will be better prepared to listen for these problems and know how to teach them. However, since errors and their magnitude differ from one learner to another, it is difficult to establish a curriculum which allocates time for teaching every identified feature. The programme must be flexible.

It is important that learners' time should be used effectively, and functional load provides sound evidence that some segmental distinctions are worth spending more time on than others. However, clear evidence of a hierarchy of importance for other features has not been found, and Derwing and Munro (2015, p. 76) found that error salience was not necessarily critical to Comprehensibility or Intelligibility.

The lack of any clear conclusions in terms of what should be taught suggest that perhaps the wrong question is being asked, and instead of considering inputs based on difficulty we should be searching for outcomes based on need.

The following chapter examines the features of pronunciation (plus other factors which might have some influence) which appeared in the Employer Survey, and discusses how those features may affect ratings of the Global Measures.

5   PREDICTING ACCEPTABILITY

5.1   OVERVIEW

This chapter is a review of evidence for those factors which are known or are likely to predict the Global Measures, particularly Intelligibility and Comprehensibility, and therefore which are likely to predict Acceptability. It explains what the feature is and the importance of the feature for learners of English. As this chapter explains, some of the terms for pronunciation features have been used in many different ways, and I have selected definitions and groupings which appear to represent the evidence most clearly and usefully.

There are also other factors which have been shown to predict Comprehensibility, some of which are evident in an audio recording: sex of the speaker, L1 (speaker's actual first language and listener's assumption about it), grammar, lexis, voice quality and noise. These are discussed in section **5.5**.

Pronunciation is often described in terms of two levels: Segmentals and Suprasegmentals. The outcome of this investigation led to a further distinction, in that Suprasegmentals are divided into Prosody and Fluency levels. This is explained in **7.2.3**.

In this chapter, features of pronunciation such as Smoothness are capitalised if they are also variables in the Employer Survey used for this thesis.

5.2   SEGMENTALS

Segments have been defined in various ways. For O'Connor (1973, p. 67ff) segments are phones as opposed to phonemes; for Brown (2014, p. 300) segmental means "to do with vowels and consonants". For the current study, no analysis below the phonemic level is made, and the terms segment or segmental follow Cauldwell (2013, p. 329) to mean **vowel and consonant phonemes**.

There has been some debate about the relative importance of consonants and vowels to Intelligibility. Most writers consider consonants to have greater effect than vowels; Catford (1987, p. 92) suggests that "a reasonably correct pronunciation of consonants is probably more important for intelligible and acceptable English than a correct pronunciation of vowels" and this is supported by Munro and Derwing (1995, p. 76). Suenobu, Kanzaki and Yamane (1992) concluded that consonant deletion was the most important cause of unintelligibility for Japanese speakers of English. Only Van den Doel's thesis (2006) disputes the primacy of consonant errors.

Confusion between two segments can cause problems for listeners – even between native speakers. Koffi (2013, p. 60) reported that a delegate talking about "[ʃɛl] clauses" in a contract was misunderstood until she spelt the word as "shall". I heard a speaker at a meeting ask "What is the colour of [feə]?" Listeners with a variety of L1's were trying to work out the colour of "fair" until the speaker clarified that the word was "fear", as the phonemes /eə/ and /ɪə/ are conflated by many speakers of New Zealand English and either form (usually [ɪə]) may be used (Hay et al., 2008).

Phonotactics, or the rules which permit placement and sequencing of segments within a language, are also important. Bent, Bradlow and Smith (2007) found that only segmental errors in word-initial position correlated strongly with Intelligibility.

Derwing and Munro (1997) used highly intelligible speakers from four languages to investigate which features most strongly predict three Global Measures, and found that generally intelligible speakers could be rated as highly accented and some as hard to understand. The writers noted that segmental errors are "perceptually salient" (p.13) and contribute strongly to Accentedness but not to Comprehensibility. Hayes-Harb (2014) also found a "significant relationship between segmental accuracy and Accentedness".

Correlations have been found between Segmentals and other Global Measures: Saito, Trofimovich and Isaacs (2015) found that Segmentals were more highly correlated with Comprehensibility than prosodic or fluency features. On the other hand, analysing the results of four Cambridge ESOL proficiency level tests, Kang (2012) found that segmental errors were responsible for only 8% of the variance, far less than prosodic or fluency features. This may have been because Saito et al. were using Comprehensibility as a measure while Kang was using Nativeness. Pennington's hierarchy of student pronunciation errors (1996, p. 253) placed "pervasive" segmental errors second only to monotonous Intonation and stress.

It is likely that Segmentals are the easiest feature for untrained listeners to grasp, but not necessarily the most important for predicting the Intelligibility of speakers who are not omitting or mispronouncing too many segments. Trofimovich and Isaacs (2012) suggested that teachers were likely to notice Segmentals when judging Accentedness of French speakers because of the emphasis on teaching Segmentals, and because prosodic errors are difficult to describe.

Segmentals, then, may be among the easier feature for teachers, learners and listeners to observe, the ones which untrained raters feel most able to comment on, and the most commonly taught. They affect Accentedness and Comprehensibility, but (beyond a basic threshold) have less effect on Intelligibility. Unfortunately, much ESL pronunciation teaching offers little beyond the Segmental level. The following chapters show the importance of Prosody and Fluency.

5.3 PROSODY

Crystal (1999, p. 275) defines Prosody as "Variation in pitch, loudness, tempo and rhythm, as encountered in any use of spoken language."

In Pronunciation Revisited (1986), Pennington and Richards called for a review of the way pronunciation was being taught in TESL/TEFL methodology, emphasising the inclusion of prosodic features and voice-setting in addition to the focus on Segmentals. The importance of prosodic features was demonstrated by further research (Derwing et al., 1998). According to Kjellin (1999, p. 387) "...it is the present author's experience with every learner he has taught so far that many remaining segmental mispronunciations will even go completely undetected by the native listener, if the student pronounces the prosody correctly."

Hayes-Harb (2014) also found that overall Prosody correlated with nativelikeness. There is now general agreement amongst researchers on the need to teach suprasegmentals to establish Intelligibility. However, an examination of 12 ESL general-skills textbook series (Derwing, Diepenbroek, & Foote, 2012) found wide disparities in the presentation and emphasis on prosodic features and often little guidance for the teacher.

5.3.1 Word Stress

Checklin (2012) describes Word Stress as "an imperative part of English language learning" for both comprehension and Intelligibility. Speakers of every language must have strategies to distinguish separate words in the flow of speech. In English, the main strategy is to assume that strong syllables are highly likely to be the initial syllables of lexical words (Cutler & McQueen, 1995) which means that variant Word Stress will reduce Intelligibility. This is supported by Sicola (2014, p. 210) who argues that a word-stress problem "directly influences the quality (e.g. clarity, duration, and aspiration) of the vowels and consonants themselves" and so is typically more important to Intelligibility than phonemic contrasts.

Speakers also need repair strategies (Hahn & Watt, 2011), which are used when speech is unclear, and many researchers consider Word Stress to be essential for repair. According to Dalton and Seidlhofer, English speakers use Word Stress to identify which words a speaker is

saying (1994, p. 39). Repair, of course, becomes more important when there are other errors or unfamiliar pronunciations.

Benrabah (1997) showed that Word Stress errors in sentences from speakers of three different languages caused serious misinterpretations by British native speakers because incorrect stress placement can override segmental information. If a word or phrase is unclear, English native speakers mentally replace it with another which has the same Word Stress pattern, on the assumption that this is least likely to have been misspoken – even if the segments are not the same (J. Field, 2005). Zielinsky's studies involving a Vietnamese speaker (2006) and later Chinese, Vietnamese and Korean speakers (2008) also found that NSs rely heavily on syllable stress patterns to identify words. However, both Benrabah and Zielinski used brief utterances for the transcription test, which would isolate the Word Stress problem but divorce the words in question from other contextual clues and thus exaggerate the impact of Word Stress. It would be useful to have evidence from extended speech in context.

Correlations have been found between Word Stress and Global Measures. Witton-Davies (2011, p. 191) found that native speakers "tend to stress fewer words than learners of English". Saito, Trofimovich and Isaacs (2015) found that Word Stress was more highly correlated with Comprehensibility than Intonation or speech rate (Speed), but less so than Segmentals. Kang (2010b) found that low-proficiency speakers, particularly Chinese and Japanese, tended to put incorrect stress on function words and pronouns, and she produced a hierarchical priority list: stress and pitch played the most important role, followed by fluency, segmental errors, and Intonation tone choices (2012, p. 12).

Giles (2011) showed that the prosodic feature *space* (the proportion of unstressed words in a sample of speech) can influence Fluency: he found that speech with fewer stressed words overall correlates with greater Speed and nativelikeness. Thus, according to Giles, "A higher proportion of stressed words correlates with lower speech rates and more clause-internal pausing." Kang (2010b, p. 310) also found that "the more stressed syllables in words ITAs produced, the more accented listeners found the ITAs' lectures." On the other hand, Kormos and Dénes (2009) used the temporal variable *pace* (the number per minute of stressed words in a sample of speech) and found it gave "striking" results when correlated with judgements of Fluency. These writers conclude "if a speaker utters a lot of unstressed words

at a high speed, he or she is not necessarily perceived to be very fluent" (p.158). In this case, the temporal feature of frequency of stress is strongly affected by the prosodic feature of getting Word Stress right in the first place, and this will also affect the Fluency feature Variety.

Given the range of evidence provided, it is likely that misplaced Word Stress will affect ratings of Fluency as well as the Intelligibility and Comprehensibility of speech and predict lower Acceptability.

5.3.2   Rhythm

It is traditionally claimed that English and some other languages are stress timed, and languages such as Spanish are syllable timed (Dalton & Seidlhofer, 1994). This has been strongly disputed (Low, 2008) and alternative theories have been proposed (Nakamura, 2010) but there is nevertheless a strong perception that there is a reality in stress timing, and consequently a challenge for some learners (for a summary of the arguments, see Jenkins, 2000, p. 43 & 149ff). Writers including Deterding (2010) and Trofimovich and Baker (2006) accept the reality of stress versus syllable timing, even if they qualify it as a continuum or a complex phenomenon. More recently, Dickerson (2015) has proposed the "two-peak profile" based on tone units as an alternative to the stress-timing theory.

Regardless of the regularity of the English stress pattern, there is no question that English utterances feature strongly stressed and strongly weakened syllables, which have a strong impact on rhythm. Writers including Nakamura (2011) have described this phenomenon and the difficulty it creates for a wide range of learners, particularly in the failure to reduce unstressed syllables. Chan and Li (2000, p. 82) report that "Cantonese speakers have problems in learning English rhythm" because in Cantonese all syllables are given similar stress and length. Dickerson (2010, p. 12) also showed that sentence rhythm can take priority over other rules such as vowel lengthening before voiced stops.

However, if English stress timing is a problem, a mitigating factor for migrants is that New Zealand English is considered to be less stress-timed than other varieties and has a tendency

to use full vowels more often in unstressed syllables (Hay et al., 2008, p. 30; Warren & Britain, 2000), which suggests that New Zealand employers might evaluate syllable-timed speech less harshly.

5.3.3   Intonation

Researchers are in agreement about what Intonation is. It has been described as "the function of pitch at sentence level" (Lehiste, 1970, p. 54) and "change of pitch over time" (Van Els & De Bot, 1987) but it must also be related to meaning, or "the linguistically meaningful use of vocal pitch level and pitch movement in phrases" (Kang, Rubin, & Pickering, 2010, p. 556). Its importance was summarised by Eskanazi (1999, p. 64): "Intonation is the glue that holds a message together. It indicates which words are important, disambiguates parts of sentences, and enhances the meaning with style and emotion." Gilbert (1987, p. 33) also described Intonation as crucial for communication: "The most powerful signals in English are expressed by intonational devices: pitch patterns and timing" and a number of writers have argued for the importance of Intonation in teaching (Chun, 1988). This underlines both the importance of teaching Intonation and the likelihood that poor Intonation could predict lower Comprehensibility.

According to Grover, Jamieson and Dobrovolski (1987) it can be difficult for native speakers to identify a stretch of speech as unequivocally native or non-native on the basis of Intonation alone without also specifying the speaker's intentions or emotional state. However, Munro (1995) used a low-pass filter which left only prosodic features of the recordings, and native speakers were still able to identify which were non-native. Munro concluded that "there are nonsegmental properties of accented English speech that can cause it to sound inherently non-native regardless of content" (p.31). Warren, Elgort and Crabbe (2009) used the same technique to show that prosodic information is important for Comprehensibility.

Despite the general agreement on what Intonation is and its importance, there is less agreement on what constitutes correct Intonation. Unlike syntax, in which the number of possible correct forms for any string is very limited, an utterance can have a range of possible Intonation patterns depending on the intent overlaid on the message. Thus the sentence "He took the money" could, depending on the emphasis and purpose, have altered pitch on any word

and rising or falling Intonation, so it is hard to judge Intonational accuracy devoid of context (Pirt, 1990, p. 147).

It is not necessary here to describe English Intonation – only to describe those aspects with which learners may have problems. For example, as noted in **5.4.4.4**, it is common for speakers of other languages to use less pitch movement than native speakers of English.

What is clear is that speakers from a range of other languages use intonational patterns which may reduce their Intelligibility in English, make their speech harder to understand and sound more accented. Intonation is therefore a likely predictor of ratings in the Global Measures.

5.4   FLUENCY

Fluency was described by Derwing and Munro (2014b, p. 41) as a "perceptual dimension" and is one of the three broad levels into which pronunciation can be analysed: Segmental, Prosodic, and Fluency. The definition of Fluency in this thesis is "**a group of high-level pronunciation features comprising Speed, Smoothness, Fluidity, Variety, Coarticulation, and Pausing".** It emerges from the literature and the Factor Analysis of the Base Ratings (see **7.2.3**), combining Segalowitz's (2010) *utterance fluency* and *perceived fluency* (see below, **5.4.1**), which Préfontaine (2010) recommended should be studied together.

5.4.1   Definitions of Fluency

As Wood (2001) pointed out, "Fluency is a term used frequently to describe language performance, yet it is often defined vaguely and used as a substitute for a group of aspects of proficiency in general" (p.574). Both Wood (2001) and Romova, Smith and Neville-Barton (2008, p. 13) reviewed numerous definitions from the literature which include one or more of global proficiency, automaticity, use of chunking or "prefabricated language units", lack of Hesitation, use of temporal variables, and strategic competence. Wood (2001) found temporal variables an essential feature of Fluency: "Defining fluency as distinct from other aspects of oral proficiency leads us repeatedly back to temporal variables in speech, such as speed, pauses, hesitations, fillers, and so on" (p.574). Riggenbach (1991, p. 424) commented that the assumption in a number of studies was that Fluency equated to Nativeness, so it is easy to see why describing a speaker of an L2 as "fluent" often means "highly competent." There is no doubt that, as suggested by Pennington's chart (1996, p. 157) any of the broader features, such as Intonation, can contribute to overall speaking competence. However, for purposes of a practical analysis, the term "Fluency" needs to be defined more narrowly.

Segalowitz (2010, p. 65ff) defined three types of Fluency: cognitive, utterance and perceived. *Cognitive fluency* is a measure of the effectiveness and efficiency with which a speaker mobilizes their cognitive system for speaking; *utterance fluency* measures the factors (Speed, Pauses, Hesitations and repair features) which contribute to fluid speech; and *perceived fluency* is a measure of the listener's judgement of the speaker's cognitive fluency. However, he acknowledged that "Normally, all three meanings will more or less coincide

and be mutually supportive" (p.66). For example, Witton-Davies (2011, p. 184) described Fluency as "a perception of the ease and efficiency of speech by the listener," concluding that the development of automatic knowledge (a fundamental attribute of Fluency in any skill) enables the L2 speaker to conceptualise their speech plans, while those who are less fluent often have to hesitate while doing so. This definition incorporates cognitive and perceived Fluency.

5.4.2   Relevance of Fluency

Attempts to define the important areas of pronunciation from the 1990s onward tended to focus on the distinction between segmentals and suprasegmentals. Anderson-Hsieh, Johnson and Koehler (1992), for example, listed four main areas: Segmentals, Prosody, syllable structure and voice quality. The call to teach features above the Segmental level – suprasegmentals, or "global" accent – was at first largely limited to prosodic features such as Stress, rhythm and Intonation (Derwing & Rossiter, 2003). However, the aspects of speech to be studied or taught are now more often described in terms of three levels, in which the Suprasegmental level is further divided into Prosodic and Fluency (Isaacs & Trofimovich, 2012; Kang, 2012). This distinction is developed further in section **7.2.3**.

Self-reporting of fluency has been challenged by arguments that it should also be considered as a performance measure perceived by the listener. Thus, according to Lennon (1990, p. 391), Fluency is "an impression on the listener's part that the psycholinguistic processes of speech planning and speech production are functioning easily and efficiently" and Derwing, Munro, and Thomson (2008, p. 360) saw it as "an automatic procedural skill on the part of the speaker ... and a perceptual phenomenon in the listener" (Segalowitz's *perceived fluency*).

As Levis (2008) argued, taking listener perception into account also raises questions about the validity of purely statistical or acoustic assessments of Fluency (*utterance fluency*), since Fluency is judged differently in an interaction, where disfluent elements can be ignored if the interaction is "appropriate and rewarding"(p.264). Hoekje and Linnell (1994) also argued that language tests should have authenticity as well as statistical validity in order to "engage the speaker's language competence in ways similar to the tasks of the target-use context"

(p.122). This reinforces other calls for pronunciation to be taught and tested in context (Pickering, 2001, p. 249; Pirt, 1990, p. 147).

All of this also justifies the goal of Acceptability, which requires that pronunciation teaching should be aimed at real-life purposes in context (such as suitability for academic study, finding a job or social communication) rather than a decontextualised goal such as Intelligibility.

5.4.3   Measuring Fluency

Fluency is a different measure from Accentedness (Pinget, Bosker, Quené, & de Jong, 2014), as is also shown by Derwing, Rossiter, Munro and Thomson (2004), who found that Fluency is judged by listeners as significantly more closely correlated with Comprehensibility than with Accentedness. Derwing et al. (2004, p. 664) referred to Fluency in terms of "the flow [Fluidity] and Smoothness of speech." Tavakoli and Skehan (2005) included Speed, Fluidity and Pausing in their division of Fluency into three components: Speed; breakdown, or the number and length of pauses; and repair, or the number of false starts, corrections and repetitions. The IELTS examination treats Fluency as a separate measure from pronunciation and includes temporal variables (Hesitation – rated under Fluidity – and length of turn) as well as discourse variables (topic development, use of discourse markers); but examiners struggle to decide whether disfluencies are caused by language or content (Annie Brown, 2006, p. 8).

Kormos and Dénes (2004, p. 146) claimed that a number of studies, including those by Lennon (1990) and Riggenbach (1991), used unreliable methods of analysing pause length and used too few participants. To provide new data they recorded 16 Hungarian learners and correlated a digital measurement of temporal variables with ratings by experienced teachers to establish which variables (accuracy, lexical richness and productivity) were more significant in perceptions of Fluency. They concluded (p.161) that Fluency "is best conceived of as fast, smooth and [grammatically] accurate performance" (Speed and Smoothness). Levis (2008, pp. 263–4) included Smoothness , Speed and Pausing in two definitions: firstly as "smoothness, clarity, rapidity of expression, with little hesitation"; and secondly as "overall spoken proficiency, involving automaticity and forward planning of speech". He added that perception of speech rate is also important. In their recent work, Derwing and Munro (2015,

p. 4) emphasise the importance of Hesitation phenomena to Fluency by defining it as "the rate and the degree of Fluidity of speech, as signalled by the presence or absence of hesitation markers, self-repetitions, and filled and unfilled pauses" (Speed, Fluidity, Pausing).

In a study of L2 speakers of Dutch (Bosker, Pinget, Quené, Sanders, & De Jong, 2012), it was concluded that *perceived fluency* is best predicted (both acoustically and from listener rating) by breakdown (the number of silent pauses), then Speed, and lastly repair.

An additional feature of Fluency is "chunking" – the ability to incorporate formulaic sequences into speech, which, contrary to the Chomskyan model of infinite creativity, increases Speed because each sequence operates as a single lexical item and thus reduces processing time. According to Wray and Perkins (2000, p. 2) as much as 70% of native speech may be formulaic. Use of formulaic language would also demonstrate better Smoothness and Coarticulation since the "chunks" are spoken rapidly and without pauses, but this was not specifically queried in the Employer Survey or Base rating.

The following sections comment on the specific features of Fluency included in the Employer Survey and the likelihood that they are predictors of other Global Measures.

5.4.4   Features of Fluency

The following sections describe the features of Fluency which were included in the Employer Survey or the Base rating. Researchers have written about other features which may be combinations of those described here. For example, automaticity (Gatbonton & Segalowitz, 1988) is the ability to produce utterances rapidly [Speed] and smoothly [Pausing], while thought groups or paratones (Cauldwell, 2013; Murphy, 2013) are the flexible and meaningful use of brief pauses [Pausing] between groups of words.

5.4.4.1   Speed

Speed is frequently described in terms of a number of temporal phenomena, and definitions can be found in various sources (Derwing et al., 2004; Grosjean, 1980; Kormos & Dénes, 2004; Riggenbach, 1991). It is measured in different ways, including articulation rate or the

mean number of syllables per second excluding pauses (Kang et al., 2010, p. 302) or speaking rate, defined as the average number of syllables per minute produced within each utterance (Grosjean, 1980, p. 40).

The major difference between researchers is that the minimum length of pauses ranges from 80 ms to 1.5 seconds (Kenny, 1996, p. 38) – although more common levels are around 200 ms and the unit measured may be syllables or phonemes.

While slow speech is normally associated with lower language competence, Anderson-Hsieh and Koehler (1988) reported that NNS became less comprehensible when they increased their speaking rate. According to Hayes-Harb (2014), speech rate is important: speech which is too fast or too slow is judged as more accented. Munro and Derwing (2001) concluded that Comprehensibility and Accentedness could be correlated with speaking rate, and also that the optimal speaking rate for a NNS was somewhat faster than that of the speakers they rated but slower than that of a native speaker. Derwing, Rossiter, Munro and Thomson (2004) asked untrained listeners to rate the Fluency of low-proficiency L2 speakers, and found that, out of five temporal measures, pruned syllables per second[1] accounted for 65% of the variance.

Cucchiarini, Strik and Boves (2000) also found that articulation rate was one of two major predictors of Fluency (frequency of pauses was the other), and Moyer (2014) claimed that speech rate was the most important criterion for Intelligibility. While untrained listeners can judge speech rate, researchers have been using technology which can measure a wide range of temporal features. Kang (2010b) measured syllables per second, articulation rate, mean length of run, and phonation-time ratio (p.305) to show that these features correlated with judgements of the Comprehensibility of ITAs.

Ginther et al. (2010) found that increased Speed was a predictor of increased Oral English Proficiency Test (OEPT) scores. Saito, Trofimovich and Isaacs (2015) found that optimality of speech rate had a higher correlation with Comprehensibility than Intonation, but lower than that of Word Stress or Segmental errors.

---

[1] Total number of syllables minus self-corrections, self-repetitions, false starts, nonlexical filled pauses, and asides, divided by the total number of seconds. (Derwing, Rossiter, Munro, & Thomson, 2004, p. 665)

A wide range of research strongly correlates speech rate with various measures of spoken competence, although not always in a linear manner.

5.4.4.2   Fluidity

Measures of Hesitation phenomena are defined in the same sources listed under Speed (**5.4.4.1**).

I have reserved the term *Fluency* for the third level of pronunciation above segmental and prosodic features, which leaves a lack of any generally used term to cover "absence of dysfluencies such as filled and unfilled pauses, false starts, self-repetitions" which is Munro and Derwing's definition of Fluency (2015, p. 14). However, they helpfully equate this with "fluidity of speech", which gives *fluid* as an antonym for *hesitant*.

Fluidity is defined in this thesis as **a measure of speech flow based on appropriate use of pausing and lack of disfluencies.**

Fluidity is likely to be highly correlated with other features. Ginther et al. (2010) used recordings of speakers taking the Oral English Proficiency Test (OEPT) from a number of different L1s to analyse the relationship between 15 temporal speech variables and OEPT scores. Broadly speaking, they found that "as examinees speak faster, say more, and pause less, their scores increase."

According to M. Watanabe and Rose (2012, p. 481), "the current consensus on pauses and hesitations in first language production is that speakers are making processing decisions (brought upon by high cognitive load or by error, for example) leading to a delay" and as there is a greater cognitive load for second language speakers there is a greater likelihood of the use of Hesitation phenomena.

A similar measure to Fluidity is *utterance fluency* (Segalowitz, 2010) which refers to the actual properties of speech which "render the speech quite fluid" (p.66). It was described by Rose (2015) as "smoothness of articulation", but it is evident that this definition is not in opposition to "choppiness" (as in **5.4.4.3**) but refers to the presence or absence of temporal features such as filled pauses, repetitions, corrections and prolongations (or drawls) which both Rose (op. cit.) and Wiese (1984) include under Hesitation phenomena. Native speech

may also include any of these, but at a much lower rate (Wiese p.19) and also includes Pausing, which is usually excluded from analysis of L2 Hesitation by setting a minimum duration (Rose, 1998, p. 7), although the duration and placement of pauses in non-native speech can contribute to judgements of Accentedness and Comprehensibility. Inappropriate Pausing, whether through frequency, length, or position of pauses, is heard as more accented (Hayes-Harb, 2014), although an analysis by Trofimovich and Baker (2006) found that, once cross-correlations were removed, only duration of pauses correlated significantly with Accentedness. Kang (2010b) used a range of pause measures to determine that long pauses, and those in inappropriate places, predicted judgements of Accentedness and to a small extent Comprehensibility (p.309).

There is a general consensus that the mean length of silent pauses, but not filled pauses, shows a positive association with ratings of Accentedness (M. Watanabe & Rose, 2012, p. 3). Cenoz (2000) found that low proficiency speakers had a higher percentage of silent pauses to filled (69% to 31%) while high proficiency speakers used both more evenly (46% to 53%); but the total number of pauses does not predict lower proficiency. Kormos and Dénes (2004, p. 161) also reported that "the number of filled and unfilled pauses and other disfluency phenomena were not found to influence perceptions of fluency."

Grosjean (1980, p. 42) considered attempts to distinguish between "fluent" and "non-fluent" silent pauses had been unsuccessful, while Kenny (1996, p. 36ff) evaluated silent pauses aurally rather than by a measuring device because he believed that only if the speaker's intent was known would it be possible to distinguish between hesitations, pauses which indicate planning (such as emphasizing or inviting a response), and grammatical pauses, which mark the boundaries between syntactic units. The latter are normal features of native speech and not disfluencies.

Cucchiarini et al. (2000) found that Pausing (along with articulation rate) was one of the two major predictors of Fluency, and Rossiter (2009) reported that "unfilled or non-lexical filled pauses, self-repetitions, and speech rate accounted for over three-quarters of the negative temporal Fluency impressions recorded by the listeners" (p.407).

Pausing has been reported to affect proficiency scores. Kitagawa, A. Watanabe and Kumaki (2011), using a range of East Asian speakers, studied the effect of three Fluency features (speaking rate, fillers and silent pauses) and found that only the frequency of silent pauses

was a critical predictor of global oral Fluency. As noted above under Speed (**5.4.4.1**), Ginther et al (2010) found that less Pausing was correlated with higher score on the OEPT.

Non-native filled pauses can also be an Accentedness-related feature, as not everyone says *er* or *um.* As well as actual words, each language has an autonomous filler, a hesitation noise equivalent to English *er* or *um* such as the Chinese *mm,* Korean *aw*, and French *euh* (see Vasilescu, Candea, & Adda-Decker, 2005, for a study of language discrimination using autonomous fillers).

Pausing in native speech

It is important to note that Pausing is a normal part of native speech. Silent pauses are found for various purposes, such as before infrequent words (Warren, 2012) and around thought groups. Filled pauses are also common – so much so that Toastmasters meetings appoint someone to count "um's and ah's" in order to dissuade speakers from using them. However, they are not necessarily evidence of disfluency. Rose (2008, p. 45) gave the following example:

> A: *Would you like to see a movie?*
> B: *No thank you.*
> B': *Uh, no thank you.*

The filled pause here is socially necessary to express a considered response rather than an abrupt refusal. A silent pause at that point would be discourteous. Browning (1982) found that use of "uh" was more prevalent in competent L2 speakers, and Clark and Fox Tree (2002, p. 73) demonstrated that "uh and um are conventional English words, and speakers plan for, formulate, and produce them just as they would any word."

Rose (1998, p. 2) also pointed out that filled pauses are one of the most frequent features of everyday English as listed in the Collins Cobuild Spoken Language Corpus, and so should not necessarily be interpreted as evidence of poor second language competence.

Given the effect of pauses and phrasing on proficiency scores, it is likely that length and placement of pauses, especially silent pauses, would affect judgements of Comprehensibility and Acceptability.

5.4.4.3   Smoothness and Coarticulation

Like most Fluency features, Coarticulation and Smoothness are not often taught in ESOL classes, or at least not to lower level students.

Coarticulation refers to the features of connected speech, which are listed by Akram and Qureshi (2014) as weak/strong forms, assimilation, elision, liaison, linking and intrusive /r/, /w/ and /j/, and juncture. Learners unaware of Coarticulation or unwilling to make use of it are likely to utter words in their citation forms rather than in the reduced manner expected in everyday speech.

Coarticulation involves effects on segmentals such as neutralising vowels and dropping of both vowels and consonants in conjunction with the sentence stress pattern (Pennington, 1996, pp. 142–3). It is likely that failure to reduce vowels would show up in Accentedness but not Comprehensibility, but the overall effect on Smoothness and linkages would have most impact on the Fluency ratings.

Lack of Smoothness is evident in a number of accents, particularly East Asian. McIntyre (2014, p. 306) described how her "staccato-like, choppy" Malaysian Chinese accent was perceived as abrupt. Chan and Li (2000, p. 82) argued that, since reduced forms do not exist in Cantonese, speakers tend to separate all words and fail to make linkages, so the lack of Coarticulation creates a staccato effect. While working on patterns of voice quality, groups of NS students thought that this effect could be interpreted as anger or irritation in English (Jones & Evans, 1995, p. 249). According to Ohata (2004, p. 15) Japanese also lacks reduction and speakers tend to stress syllables evenly with the same effect, which "can adversely affect the comprehensibility of their English to the native speakers." German speakers (and, to a lesser extent, Dutch) often place a glottal stop before initial vowels rather than linking the words (Koster & Koet, 1993, p. 78).

Macneil (1987, pp. 64–5) suggested that the staccato speech pattern of Vietnamese speakers could be related to the short pulses of air pressure used to produce Vietnamese words as opposed to the English requirement to maintain even air pressure over long stretches of speech, and that this staccato delivery, combined with other features of Vietnamese English, was responsible for a great deal of information loss. She found (in a small sample) that

only the most clear-speaking Vietnamese students identified Coarticulation ("linking") as important.

This is not invariably a serious problem, but if a speaker articulates each word separately it can have a staccato effect which can be irritating or hard to understand, and in some case interpreted as abrupt or aggressive. It may well make a speaker sound less acceptable for a job.

5.4.4.4   Variety

Variety is related both to Intonation and the speaker's global competence. It is a judgement of a speaker's ability to use different Intonation patterns to reflect the content or emotional impact of what they are saying. There is no evidence that being monotonous makes a speaker unintelligible, but this feature is likely to affect ratings of irritation, Comprehensibility and Acceptability.

Researching the dissatisfaction with ITAs, Hinofotis and Bailey (1981) asked undergraduate students to rate videotapes of ITAs. They found that the most common problem was a boring presentation style, which suggests that Variety should be important in assessing speakers' Acceptability. Pennington (1996, p. 253) ranked monotonous Intonation or stress as the most general and important level for teachers' attention.

Munro (2011, p. 10) reported that an utterance of a Cantonese speaker was unintelligible because of pronunciation faults "including a high pitched monotone intonation and staccato rhythm" and Kang (2010b) noted that "Typically, NNS speech sound is considered as somewhat monotonous," the reason being a restricted pitch range: Kang (p.310) cites research findings concerning ELLs with different L1s in support of this claim.

As noted under the prosodic feature of Rhythm, Cantonese speakers tend to give each syllable the same stress and length (Chan & Li, 2000, p. 82), leading to their comment "No wonder native speakers of English, and even Cantonese speakers of English like the present authors, would sometimes perceive such an English pronunciation pattern as 'flat and boring'." A. Cunningham's thesis (2013, p. 17) made a similar point about Korean speakers, and Pickering (2001, p. 249) reported that NNS "employed long sequences of falling and level

tones regardless of the informational or social value of the matter contained within the tone units" which tended to create a "flat, monotonic pitch structure unfamiliar to NS hearers."

The reference to content is important as it demonstrates the contextual nature of Fluency features. While statements can generally be judged grammatical or ungrammatical in isolation, the tone of a statement out of context cannot be judged as wrong, and may be perfectly intelligible, but used inappropriately may reduce the listener's desire to pay attention and affect their judgement of the speaker's overall competence.

5.5   Other factors

Section **3.2** showed that the Global Measures, particularly Comprehensibility, can be predicted by many different factors. Kraut and Wulff's research (2013) suggested that it may be difficult to isolate these factors specifically when predicting global measures of pronunciation. They found that a range of variables significantly correlated with NS rater judgements of NNS Accentedness, Comprehensibility and communicative ability. The speaker's sex, language (treated as Asian, Hispanic or Middle Eastern), proficiency (low, intermediate, high), and listeners' self-reported familiarity with the languages spoken, not only correlated with judgements of the three Global Measures but also interacted with one another.

Shih (2002, p. 102) claimed that employers who offer low-skill jobs "*are more likely to desire workers who are manageable, obedient and pliable; i.e. those who are least likely to contest their direct authority*" (original emphasis) and so the racial or gender attitude on the employers' part may be driven by beliefs about how manageable different groups are. However, this was a US-based study, and the employment culture may be different in New Zealand.

Sex of the speaker

Research has not produced any consistent effects from sex of the speaker. Piske, MacKay and Flege's review of the literature (2001) concluded that "Most studies have not identified gender as a significant predictor of degree of L2 foreign accent" (p.200) and their own study had the same result. On the other hand, Kraut and Wulff (op. cit.) found that male speakers consistently received more favourable ratings than female speakers at low and intermediate proficiency levels regardless of the language family of their L1 (p.259), but not at high proficiency levels. In their study all the speakers read the same passage, which was displayed to the raters, so it is not possible to tell whether there were further effects on all speakers from the prosody of reading and expectations brought about by the visible text. There is also the risk that speakers would be using language inappropriate for their gender. A review of customer satisfaction ratings by Hekman et al. (2010, p. 259) also found evidence which they describe as "fairly robust findings across jobs, contexts, raters, and ratees" that "customer ratings are biased against women and racial minorities."

In an Australian study by Gallois and Callan (1981), NS students listened to recordings of written material read aloud by one male and one female speaker from each of four L1s plus British and Australian NS. They found significant effects for sex of the speaker: males were perceived less favourably on a range of semantic differentials than females, while females were seen as less dynamic. Timming (2016) found that female foreign job applicants were rated far lower than males.

There is therefore a possibility that males and females will be rated differently by employers.

Actual L1 / Assumed L1

Two possibilities arise when considering the effect of L1: the speaker's actual L1, and what the listener assumes it to be. The "assumed L1" is used here very loosely, as listeners may be able to identify a specific L1 (Vietnamese), a region (South East Asian) or only a catch-all category (Asian). Effects can be due to attitudes towards specific groups (listeners have no problem with the accent but ratings are affected by beliefs about the speakers) or actual responses to pronunciation (there is no problem with any ethnic group but ratings are affected by irritating pronunciation features).

In Kraut and Wulff's (2013) research cited above, low-proficiency Asian speakers (meaning those from China, South Korea, Japan and Taiwan) received the lowest ratings at all levels. Middle Eastern speakers were given the highest general rankings, followed by Hispanic, with speakers of Asian L1s in general rated lowest throughout.

Raters in the study by Gallois and Callan (1981) described above were poor at identifying Greek and Vietnamese accents. This may be a reflection of the familiarity of accents to Australians.

In New Zealand, Watts (1981) reported highly negative attitudes to foreign accents, but more recently Y. Watanabe (2008) found that New Zealand listeners were positive towards nonnative accents, although the assessors were students who had contacts with NNS and therefore may be been more favourably inclined towards them than the general population.

It is possible that some effects will be found for actual groups, or for speakers believed to be in a particular group.

Grammar and Lexis

As has been described in **3.2.5**, Grammar, and to a lesser extent Lexis, have been identified as predictors of Comprehensibility; it makes sense that faulty words and jumbled grammar will make speech harder to follow. As the focus of this thesis is on pronunciation there is no description of these features, but they were included in the Base Rating in order to test them against other features.

Voice Quality and Intrusive Noise

These features were included in the survey to ensure that results were not distorted by them. Voice Quality was not used in the sense of Honikman's (1964) articulatory settings, but identified speakers whose voice might be shrill, monotonous etc. and therefore irritating. Intrusive Noise referred to any additional noises which affected the recording.

5.6 DISCUSSION

This chapter has asked what predictions can be made about the Employers' likely Acceptability ratings of the features of migrant pronunciation. Most of the Migrants recorded for the Employer Survey speak languages which are phonologically very different from English, and (as shown in Chapter 4) this creates identifiable pronunciation difficulties, particularly for those speaking South Asian, Vietnamese and Myanmarese languages.

Researchers have shown that many different features of pronunciation can predict judgements of Global Measures, and some features such as Word Stress, nuclear stress and weak forms have been identified as highly salient. However, it does not appear that any level (Segmental, Prosodic or Fluency) can be ignored. Given that Prosody and Fluency are, in fact, often poorly taught, it is more likely that speakers will be deficient in these areas.

Learners' Segmental problems will depend largely on their L1, IL and markedness effects. These are certain to affect ratings of Accentedness and are likely to be noticed. Where segmental problems are extreme, such as frequent omission of final consonants, they will affect Comprehensibility, but otherwise they are likely to have less importance than other levels of pronunciation.

Prosodic features, particularly Word Stress, are often associated with Comprehensibility and other Global Measures. Speed and Fluidity often affect judgements of overall competence, which may be reflected in Acceptability ratings.

Non-phonological features such as Lexis and Grammar may also affect Comprehensibility ratings, and attitudes to migrants or ways of speaking are influential. A wide range of phonological features have been implicated in predictions of all the Global Measures.

The actual rating of Acceptability will also depend on its correlation with the other Global Measures, and this has not been measured before.

6 METHODOLOGY

6.1 OVERVIEW

My primary goal was to answer the research questions:

1. Does pronunciation affect employers' assessments of the Acceptability of migrants for employment?

2. If so, which pronunciation features are the strongest predictors of the assessment?

3. Once pronunciation features are taken into account, are there other employer or speaker factors which affect employer ratings of Acceptability?

In addition, the employers' responses would provide evidence for the value of including Acceptability as a Global Measure of speech.

The migrants I was concerned with had an elementary to low intermediate level of English, were often former refugees, and (at the time) were generally looking for low or unskilled work. As described in sections **2.2** and **2.3** there is substantial research on employment challenges for skilled migrants to New Zealand but considerably less information about the even greater difficulties facing unskilled migrants, many of whom are former refugees with additional challenges such as poverty, trauma, low education, lack of English skills, and health or family problems.

Unskilled migrants may still have a low level of English when they need to look for work, so many study English to attain the minimum level needed. Given the frequent references by employers to migrants' lack of English (see Chapter 2) it would be useful for ESOL teachers to have better information about which pronunciation features contribute most to employers' decisions, and therefore what the teachers should concentrate on in their classes.

To obtain this information I decided to make recordings of 40 migrants (the **Migrants**, sometimes the **speakers**) to be used in an online survey (the **Employer Survey**) so that their speech could be rated by employers, or people involved in hiring staff, in industries where unskilled migrants would typically apply for work. In addition, professionals with experience in teaching migrants and assessing their speech (the **Experienced Raters**) would be asked to

rate the recordings in order to provide a baseline or model against which the employers'
ratings could be compared. Inter-rater reliability was found for the Global Measures, but not
for the specific pronunciation features, so to provide a baseline for these I prepared the final
Base Rating myself.

Questions from the full surveys can be found in **Appendix B**.

The data from the Base Rating and Employer survey was used to answer all three Research
Questions. The phonological variables in the surveys were used for RQ 1 and 2, while the
data about the Employers and Migrants, and additional Rater questions, were used to an-
swer RQ3. The statistical approach was the same throughout, with the exception of some
qualitative responses reported in Section **7.4.5**.

6.1.1   Comparison with previous research methodology

Few studies have focused on employer Acceptability of migrant pronunciation, all of them in
the US apart from that of Singer and Eder (1989) in New Zealand. These studies (detailed
in **2.3.3** and **3.3.3**) are listed and their methodologies summarised in Table **6.1-1**. This thesis
used the most effective methodologies from these studies and incorporated features from
other research. It also aimed to meet authenticity criteria (Rifkin & Roberts, 1995).

Respondents

Seven of the ten employer studies used actual employers, or people involved in hiring. Em-
ployers may rate some speakers more harshly than non-employers because they believe
they cannot be employed to deal with customers, or they may rate speakers more gener-
ously because they cannot afford to be too picky. Using respondents other than employers
may therefore give unrealistic Acceptability ratings. Some of the researchers took the re-
cordings to the employers at their places of work, which is the best real-life option but
rarely practicable. Many had a mixture of ethnic groups and both male and female speakers.
One (Timming, 2016) used an online crowdsourcing platform which achieved a high number
of respondents but provided less control over who they were.

For the present thesis I contacted employers personally to gain their agreement to take
part, which meant I knew that they were employers, what their workplace was, and

whether they were male or female. Their industries employed low-skilled workers who were highly likely to include migrants, and most of the employers said they had in fact done so, which largely eliminated effects from employers who might have a bias against migrants or foreign accents (or at least had learnt to put it aside from necessity). Since the employers were located in main centres throughout New Zealand, an online survey platform was the only practicable means of surveying them.

Speakers

Table 6.1-1 Methodology of Previous Employment Acceptability Research

Authors	Respondents	Speakers / Sex	Data Type	Criteria	Context
Hopper & Williams (1973)	40 E (phase II)	3 NS: 2 x W US accents + AA. 1 NNS: H. All M	90 second free responses (Job questions)	Likelihood of hiring 15 SDQs.	5 levels
Hopper (1977)	105 E I in situ	All NS: AA, W Standard/Non Std. All M	Same data as previous. Matched-guise 80 sec, read by bidialectals	Likelihood of hiring 15 SDQs.	3 levels
Rey (1977)	E, 20 W, 11 AA,12 Cuban	8 NS W, AA, F&M; 11 NNS Cuban F&M	20-40 second Free response (describe pet)	Likelihood of hiring 4 SDQs + guesses on personal info.	7 levels
Kalin & Rayko (1978)	203 M/F students	5 US NS, 5 different NNS All M	30s scripted speech on "current concerns"	Suitability 8 SDQs.	4 levels
De la Zerda & Hopper (1979)	67 (24F, 43M) ELLs at their workplace	8 Mexican M (NS?)	Same data as #1, 60-90s, but read as scripts	Likelihood of hiring 7 SDQs + language attitudes	3 levels
Singer & Eder (1989)	210 Pakeha students	NS: Pakeha Maori (NS?) NNS: Chinese, Dutch All M	Videotaped full job interview of 3-4m with memorised script	Suitability 7 social status characteristics	High, low status
Cargile (2000)	192 students 105 W, 57 H, 30 AA.	M & F Chinese NNS F & M	Matched-guise US and Chinese-accented 52s. read by bidialectals	Suitability 5 SDQs	High, low levels
Hyman (2001)	51 business managers / owners	6 F NNS 3 Chinese 3 Russian	Phone. Prep. 20s message. Pre and post accent correction	3 SDQs	3 Levels
Carlson & McHenry (2006)	60 HR specialists 12 AA, 3 H, 45 W, 14M 46F.	3 F bi-dialectal actors: H, Asian, AA	2 versions (max & min accented) of prepared scripts 35-40s	Suitability Comprehensibility Accentedness	Entry level job
Timming (2016)	Online (open) managers. 108M, 115F Born in USA	NS US, British 2F, 2M; NNS Chinese, Indian Mexican, 2F 2M	Phone. Prepared 7s self-intro	Likelihood of hiring. 7 SDQs.	Customer, Non-Customer

***Key: W**hite, **A**frican **A**merican, **H**ispanic, **M**ale, **F**emale, **E**mployers, **I**nterviewers.*
SDQ=Semantic differential question.

Three of the employer studies in Table **6.1-1** above used actors or bidialectal speakers, while others had people using their natural accents. One had a "matched guise" using the same speaker pre and post accent modification. Some studies included native speakers, including some with various accents. Only three of the studies used both male and female speakers.

I wanted to focus on migrant speech only and I used natural accents to ensure authenticity. My aim was to have equal numbers so that effects of sex of the speaker could be tested, but in the end I found 24 male and 16 female volunteers. This was more representative than selecting only males or females.

Speech content and elicitation

Eight of the ten employer studies in Table **6.1-1** used memorised or read out scripts. Various methods of eliciting samples of nonnative speech for analysis have been employed by other researchers, such as:

- Reading aloud, including delayed repetition (Juffs, 1990; Piske et al., 2001; Southwood & Flege, 1999; Trofimovich & Baker, 2006; Warren et al., 2009)

- Unrehearsed natural speech (Kang, 2008a)

- Combination of free and read speech (Levis, 2011)

- Constrained elicitation and role-play (Gynan, 1985; Hinofotis & Bailey, 1981).

Speer, Warren & Schafer (2011, p. 38) give a number of reasons why read-out speech, even when delayed, is likely to differ from spontaneous speech and therefore lacks authenticity for a study of pronunciation features. As my purpose was to rate Acceptability of migrants as potential employees, a read-out sample would pose two problems. Firstly, the samples – although more uniform – would not reflect the authentic communication task of an applicant speaking to an employer. Secondly, the prosody of reading aloud is not the same as that of natural speech. Munro and Derwing (1994) found no difference between read out and extemporaneous speech in evaluations of Accentedness, but the English proficiency of these speakers was very high (p.258). Some speakers who are comfortable with the written word may be advantaged by their fluency, while others may lack reading ability (a likely possibility for a migrant group with a low level of English) or be disadvantaged by a monotonous reading style.

More importantly, a study of a young Vietnamese speaker (Sato, 1985) showed that inter-language production can be affected by task type (such as free conversation or reading aloud) so any task which differed greatly from the real-life interaction between an employer and a potential employee would run the risk of collecting questionable data.

Content

The content of the audio material was important, since language has context, purpose and meaning. Most of the content in the ten employer studies was job-related, but one had speakers talking about current events and another about pets. These topics created a mismatch with the communication purpose and could have affected the semantic differential questions as a speaker's personality will be revealed both by how they speak and the content of their speech. Any task evaluating Acceptability must be relevant to the context.

My aim was to present employers with samples which

- were realistic in terms of content and context for migrants applying for work

- could be rated for pronunciation features, and

- minimised distracters from the task of rating pronunciation.

Since an authentic job interview, or any approach for work, is largely impromptu, I considered it essential that that this should apply to the audio samples presented to the respondents.

In five of the employer studies, recordings were used which were fairly short (7 to 40 seconds) which is long enough to judge Accentedness but would not allow for a range of features to be used or observed. One used a lengthy video, which would have been impractical. I used prior testing to determine the length of recording needed to be sufficient for rating: short enough for survey use, but with sufficient time to offer a variety of features for analysis. I settled on audio recordings of two minutes (where the Migrant had spoken long enough) and requested respondents to listen to at least one minute. I used free responses to job interview questions because they were relevant to the context.

As Tarone (1979) pointed out, all sampling procedures are likely to cause speakers to monitor their speech so the result may not be fully natural, but in this case the speakers would

be in a similar situation to a real job interview. The option of a second interview (as described in **6.3.6**) also meant that the speakers would be more confident and relaxed in the second recording which was usually the one chosen for the Employer Survey. To make the recorded situation as realistic as possible, I wrote an interview question script (**Appendix B1**) in the interview to provide relevant subject matter and allow free responses. Although the "candidate" may have had a specific job in mind, the questions all had to be general in order to be the same for all speakers. They were selected from a range of online interview questions as being both typical and likely to produce responses from the speakers.

Awareness of Ethnicity

In all cases except Hyman (2001) the respondents for the previous employer studies were aware of the ethnicity of the speakers – in some cases they were specifically told, and in others (such as those with speakers described as African American or Hispanic) the accents were prominent and known by the community. This means that pronunciation features could not be separated from attitudes to ethnicity in the Acceptability judgements.

The survey reported in this thesis had speakers of 19 different languages which I categorised into 9 language groups (**6.3.2**), with no information provided to employers about the speakers' origins. Employers were not good at guessing these (**7.4.7**) and their guesses were not found to be a significant predictor of Acceptability. We can therefore be confident that the results of this survey are not affected by attitudes towards specific ethnic groups.

Criteria / Context

All of the previous employer studies had Acceptability as their DV, expressed as "likelihood of hiring" or "suitability" and most had semantic differential questions. From these questions they were able to find that employers ranked different ethnic groups as more or less acceptable for employment, and also found effects from the context of different levels of work (recorded in the final column of Table **6.1-1**).

The focus of the current thesis was Acceptability on the basis of pronunciation for potential applicants at one (low-skilled) level of work, which is probably why no effect from type of work offered was found. However, context was included in the Employer Survey in terms of communications with Supervisors, Co-workers and Customers, similar to the Customer/Non-Customer variables described by Timming (2016) as a "previously unexplored question"

(p.2), and significant differences were found in the Acceptability ratings for these contexts. This thesis therefore supports previous research which demonstrates that Acceptability is contextual.

Experienced Raters

Experienced Raters are often used to evaluate learner pronunciation (Munro, Derwing, & Thomson, 2015). Two of the previous employer studies used experienced Raters: Hyman (2001), who used three experienced ESL instructors to rate speakers for proficiency and Accentedness; and Carlson and McHenry (2006), who had experts transcribe and add coding to note the number of dialectal markers so they could identify speech as minimally or maximally accented.

The use of Experienced Raters, experienced university teachers, was central to the analysis of the Employer Survey in the current thesis, because they provided an external standpoint from which to discover predictions of Employer ratings of Acceptability.

6.1.2   Survey Process

Employer Survey Process

1. I recruited Migrants, mostly from language schools, or by word of mouth. Many were former refugees but this was not recorded.

2. I recorded speech samples by interviewing the Migrants, using free responses constrained by function (job interview) and content (specific questions posed). Excerpts were taken of sections with the speakers' most clear and confident speech.

3. I designed the Employer Survey in two parts and loaded them into Qualtrics (2015, accessed 2013), including the speech samples, to enable employers to rate the samples online.

4. I ran a trial of the Employer Survey with a group of volunteer employers and improvements were made.

5. I contacted Employers and invited them to participate in order to rate the Migrants' Acceptability for employment on the basis of their pronunciation, and to provide specific pronunciation and attitudinal ratings.

6. Employers assessed (ideally) five samples each, from either Part A or Part B of the Survey, by rating each sample on 1-9 point ordinal scales for Acceptability, pronunciation features and other factors (for specific results per employer see **6.4**).

Base Rating Process

1. I designed the Base Rating process to assess the same samples using a set of pronunciation and other features similar to those in the Employer Survey, but with more specific responses to pronunciation features.

2. I recruited five academic professionals (Experienced Raters) to rate the Migrants' speech and create a baseline against which the Employers' ratings could be predicted.

3. I met with the Experienced Raters to agree on guidelines for rating.

4. I trialled the Base Rating and then uploaded it as a separate survey in Qualtrics.

Statistical Analysis

I carried out statistical analyses of the results using the R platform (R Core Team, 2015b) to determine the answers to the thesis questions, and reach conclusions to a range of questions about the data.

6.2   SELECTION OF PARTICIPANTS

Surveys for various purposes related to pronunciation have used varying numbers of non-native speakers and raters in different ways, as is shown by these examples:

- Hopper and Williams (1973): 4 speakers + 40 employers

- Singer and Eder (1989): 4 speakers + 210 undergraduates

- Munro and Derwing (1995a): 10 speakers + 18 raters

- Cargile (2000): 1 M and 1 F speakers, each with 3 personas, + 192 undergraduates

- Kang (2008a) 11 speakers + 70 then 63 raters

- Jun and Li (2010): 7 speakers + 7 raters

- Levis (2011): 2 speakers + 9 experienced Raters

The participants in the Employer Survey are described below.

6.2.1   Migrant speakers

Recruitment

My aim was to find 40 male and female migrants studying English at elementary to low intermediate level (around IELTS 4), so there would be few highly proficient speakers in the group. I also aimed to find speakers representative of migrant groups currently likely to be applying for low-skilled or unskilled work. I therefore recruited students from local language schools. There was a limited pool available and willing, so there were some gaps (e.g. no Koreans or speakers of Hindi, although there are many amongst recent migrants) and more male than female migrants (m=24, f=16). It is possible that, for family and cultural reasons, there were more males studying English on "get to work" programmes. Although there are few Italian and German migrants in language schools, some were included to obtain employer responses to European languages for comparison. I did not conduct any test of English level, so some of the speakers had a higher overall competency, which would have been reflected in the ratings. Low proficiency speakers might be more affected by problems with lexis and grammar, but this too would show up in the ratings, so neither case would affect the integrity or usefulness of the data.

I knew each Migrant's native language, country, address and sex from their acceptance forms. I did not ask about their education or age. It is possible that these factors could have

had an effect on acceptability, but they were outside my aim of discovering the pronunciation features which were acceptable (or not) to employers and they formed no part of the survey.

The distribution is shown in Table **6.3-1**.

6.2.2 Employers

Actual employers (the term covers human resources managers or anyone responsible for hiring employees) were chosen as the ideal raters of the Acceptability of migrant pronunciation for employment.

The various studies cited in the previous section often used undergraduates to rate non-native speech for a range of purposes. Of particular interest is the use made by Hinofotis and Bailey (1981) of native-speaker first-year students to rate foreign training assistants (TAs) because they were the "ultimate judges" of non-native speaking TAs (p.123). This makes use of the Global Measure of Acceptability, as the TAs were being assessed not only for their Intelligibility but for their Acceptability for the purpose of teaching (the most common problem being a boring style of speaking – the fluency feature Variety). Hsieh (2011, p. 66) suggested that undergraduates might be "self-servingly harsh" in their ratings because of their fear of being taught by an unintelligible TA, but that is after all the contextual nature of Acceptability. Employers can be as self-servingly harsh as they wish, as they serve the interests of their businesses, and as they have a gatekeeping role towards migrant access to work (Carlson & McHenry, 2006, p. 79) they were the appropriate raters of migrant speech for the Employer Survey.

I needed a minimum of five ratings per speaker over the two parts of the survey to obtain 400 observations, which required at least 80 employers. I recruited employers by emailing and then following up or cold-calling a total of 495 businesses in the main centres – Auckland, Hamilton, Palmerston North, Wellington, Dunedin and Christchurch. Most employers responded favourably to the idea of the survey even if they were unwilling to take it themselves. 302 respondents offered to take the survey, of whom ultimately 95 successfully completed at least one full response (that is, answered all the questions about one of the speakers). The Employers were selected from 13 different industry types likely to be hiring

migrants for low-skilled or unskilled work, such as supermarkets, hotels, rest homes for the elderly, security, removals and cleaning companies.

Table **6.2-1** shows the types of work offered by Employers who took the survey (the number of Employers was 95; many offered more than one type of work, or the job included more than one type of activity).

Table 6.2-1 Employment Types

Employment type	Number of Employers
Caregiving	10
Cleaning	24
Driving	14
Hairdressing	2
Handling goods	18
Hospitality & customer	15
Kitchen	2
Labouring	14
Nursing	3
Office	4
Retail	37
Security	7
Other	5
TOTAL: 13	**TOTAL: 155**

6.2.3   Experienced Raters

Research Question 2 asked "Which pronunciation features are the strongest predictors of the assessment?" The first stage was to obtain ratings on factors such as global measures such as Comprehensibility or Acceptability from the Employers, which could then be predicted by a measure of the performance of the speakers. It was therefore essential to perform a rating of the speakers' pronunciation features to create the independent (predictor) variables.

Rating can be done through instrumental analysis, and this has been used by some of the studies cited in this thesis (Kang, 2010b; Koffi, 2012; Trofimovich & Isaacs, 2012). This can capture temporal features far more accurately but is not able to report on (for example) appropriate use of contrastive stress. As shown by the examples at the start of section **6.2**,

studies have often involved raters to assess the actual performance of speakers. The rating for this thesis would best be done by experienced raters (Isaacs & Thomson, 2013, p. 156), so I asked five TESOL professionals with expertise in judging speaking performance (the **Experienced Raters**) to rate the speech samples in order to create a rating which would serve as a baseline to discover the actual predictors of the Global Measures. This was named the **Base Rating** as a neutral term to avoid implications that it could be objective or that the raters were phonetic experts as opposed to experienced professionals. All of the Experienced Raters had extensive backgrounds in teaching and assessing learners of English. They therefore all had some familiarity with the accents concerned and were accustomed to making allowances for that familiarity.

The process for developing the Base Rating is described in section **6.5.**

6.3   EMPLOYER SURVEY DESIGN

The Employer Survey was designed in order to find answers to the research questions. Four steps were taken to make it more likely that the data gained by this survey would be trustworthy. Firstly, as seen in Chapters 3-5, the survey was based on literature from the 1970s to the present: it continues the concerns of an established body of research into predictors of global measures of speech (Munro & Derwing, 2011). Secondly, sampling methods (Dörnyei, 2007, p. 95ff) were used to gather information from the widest group of migrant speakers and employers which it was practicable to call upon in the time available. Thirdly, a trial run of the survey was made to ensure that it worked and to obtain feedback from the testers; and finally the data from the actual final survey was analysed and interpreted using established statistical techniques.

In any survey, the questions asked determine what data will be found. The phrasing of questions, their inclusion or omission, will alter participants' responses. In addition, there are concerns on the use of an online survey because of the lack of control over the conditions under which the speech samples are rated (Munro & Derwing, 2015, p. 25). However, Gosling, Vazire, Srivastava and John (2004) reported that a range of studies demonstrate that internet responses are "typically consistent with the effects from studies using traditional methods" (p.101). Some of the statistical data obtained from the Employer Survey are remarkably similar to those obtained from face-to-face methods in previous research (see **7.6.2**) which provides verification that the respondents were acting in a similar manner and that the data has not been devalued by random input.

6.3.1   Question Design

Much of the key research on Global Measures referred to in this thesis was produced by Munro and Derwing, who used nine-point scales on the grounds that "it was better to overestimate the listeners' ability to resolve Accentedness than to underestimate it" (1994, p. 259). Southwood and Flege (1999) concluded that a scale of at least 9 was needed to avoid ceiling effects from shorter scales, and also that Accentedness was a "metathetic continuum" (p.337) which justified interval scaling.

Isaacs and Thomson (2013) noted that 1-9 numerical rating scales are commonly used, and they are normally labelled only at the endpoints. They are useful because they are practical, highly versatile and need little explanation (p.136). However, the writers raise some caveats. For example, a given score may have different meanings for different raters, and there may be too many distinctions. When they tested a 9-point against a 5-point scale, the writers found them equally reliable but the 9-point judgements were more consistent.

Each step of the scales could be labelled to clarify what they mean, but this does not necessarily improve rating performance. For example, the scales in Morley's (1991) Speech Intelligibility Index include rubrics such as "largely unintelligible" and "reasonably intelligible", but Isaacs (2014, p. 10) commented that "these semantic differences do little to guide raters on how the qualities manifested in test-takers' performance samples align with the scale levels."

I used the 9-point scale for comparability with previous research, the only difference being that I used the low end for low competence (e.g. low Intelligibility) and 9 for highest competence, as this seemed more intuitive for respondents. This was used for all questions dealing with Global Measures, Acceptability, personal qualities or pronunciation features. Dörnyei (2007, p. 106) recommended varying the poles to avoid "superficial responding", but I decided this would be more likely to lead to wrong answers by confusing the respondents.

Some of the questions provided lists or yes/no choices where the possibilities were limited (length of time in NZ, employed NNS, had problems with them) and the question "Which of the following best describes your view about migrant employees using their own language?" had only six possible answers so that they could be included in the statistical analysis. Others had a list with a final box for "other" (kind of work, native language).

"Which country do you think the migrant comes from?" was open-ended as any list would have narrowed the possibilities. If the respondent replied "yes" to having pronunciation problems with NNS employees, they were asked to describe the problem. These were the only qualitative responses.

6.3.2   Questions

While the goal was to discover how the Employers' judgements of Acceptability could be predicted by the Base rating, it was necessary to consider what other factors might influence that rating. Kang (2008a) found that the raters' background characteristics (contact and attitude towards the NNS groups) accounted for 20-30% of the variance in their ratings of NNS proficiency, and Hsieh (2011) also found effects from rater background. To take account of this, information about employer attitudes towards migrants and their speech were included. As the Employers were asked to focus solely on pronunciation when rating Acceptability, it was also possible that asking overt questions about their attitudes might enable respondents to "compartmentalise" their answers and focus more closely on pronunciation. A question on overall language ability could have been asked, but might have distracted the respondents from focusing on pronunciation. The exact wording of questions from the full surveys can be found in **Appendices B2-4**.

Employer data

Sex of employer was known as I had spoken to them all personally. Questions on the following topics were included:

1. L1 and origin of employer
2. Type of work offered
3. Attitudes to when English ought to be used
4. Problems experienced with migrant employees

Employer beliefs about Speakers

The Employers would have no other information about the speakers than their voices and the content of their responses (since these were unscripted), but the accents might influence attitudes towards certain speakers:

1. Specific ethnicities
2. Specific accents
3. Males or females
4. Attitudinal judgements, such as whether the speaker sounded polite.

Employers were not told about the speakers' origins, so a question was asked about the Employers' beliefs about their origins from which L1 could be deduced. All of this information

could be further correlated with the objective data: the speaker's actual L1 and sex. All Employers were asked:

Which country do you think the migrant comes from? [Free response]

How acceptable is the migrant's pronunciation for a job with your company? [1-9 scale]

Language

Language appeared in three variables. It was a fixed variable as the L1 of each speaker, as stated by the Migrants themselves on their information forms (referred to as *Actual Language* below); it appeared in response to the Employer Information question "What is your native language?" (Response: English / Other – please state); and it was interpreted in response to the Employer Survey question "Which country do you think the migrant comes from?" asked in both surveys. There was no drop-down box so Employers were free to state any country. The question stated "country" because asking for the speaker's language would have produced many more non-responses (e.g. unwillingness or inability to be specific about which South East Asian or Indian/Sri Lankan language) while the country responses could be more readily categorised.

After the Employer Survey returns had been examined, "which country" responses were classified into groups for consistency and to reduce the number of variables (Table **6.3-1**). The groupings were based on:

1. Specific languages which New Zealanders would usually be aware of: Chinese, Japanese.

2. Well-known geographic/linguistic groupings: Western European, Pacific, South American

3. Geographic groupings of related and unrelated languages with many cultural and phonological features in common: South East Asian, South Asian

4. Convenient geographic groupings of languages which New Zealanders would generally be unable to distinguish: African, Middle Eastern.

The categories used for Employer language guess are given in Table **6.3-2**.

Table 6.3-1 Language Category, Actual language, Sex and Number of Speakers

Category	Language	Male	Female	Total	Total per Category
African	Amharic	2	1	3	5
	Somali	2	0	2	
Chinese	Chinese	0	2	2	2
Japanese	Japanese	2	1	3	3
South East Asian	Arakanese	1	0	1	9
	Burmese	1	1	2	
	Chin	0	1	1	
	Karen	1	0	1	
	Vietnamese	1	3	4	
Western European	German	1	0	1	3
	Italian	1	1	2	
Middle Eastern	Arabic	2	1	3	5
	Assyrian	1	0	1	
	Turkish	1	0	1	
Pacific	Tetum	2	1	3	3
South American	Spanish	4	1	5	5
South Asian	Sinhala	0	1	1	5
	Tamil	1	1	2	
	Telugu	1	1	2	
Total Speakers		**24**	**16**	**40**	**40**

The placing of Indo-European and Dravidian languages in South Asia into one group is justified by Anderson-Hsieh, Johnson and Koehler (1992, p. 550) on the grounds of "similar nonnative patterns of rhythm and intonation and errors in aspiration". Similar reasons apply to the South East Asian grouping. Spanish was grouped under South American because all the Spanish speaking speakers, and the major group which employers are likely to encounter, were from Colombia or Ecuador. The dialect is similar but the people are not "Western Europeans". Further rationale for these groupings can be found in **Appendix B5**.

Table 6.3-2 Categories for Employer Language Guess

Code	Language Group
1	Pacific
2	South East Asian
3	Chinese
4	Japanese
5	Korean**
6	South Asian
7	African
8	Middle Eastern
9	Eastern European**
10	Western European
11	South American
12	Could not tell*
13	Other*
14	Asian*

*Not used in Actual Language. **No speakers of this language or group

Attitudinal Responses

As described in **3.2.3**, accents are often stigmatised, and are often viewed across the dimensions of competence such as intelligence and power or attractiveness, which can include kindness and solidarity (Hosoda et al., 2012). As this survey was aimed at employers I chose judgements which they might take into account when selecting an employee for low status work. Semantic differential questions (Gardner & MacIntyre, 1993) were therefore included on how friendly, polite and hard-working respondents thought the speakers were. Friendliness was chosen as an "attractiveness" quality, Politeness as both "attractiveness" and suitability for work, and Hardworking as a criterion which employers would find important. Employers would be able to make these judgements on both the Migrants' Accentedness and the content of their responses, so it would not be possible to disentangle these.

Employers in Part A were asked:

From the way this person speaks, how would you place them on the following scales?

- *Work   [1= Lazy =Hard worker]*

- *Friendliness   [1= Unfriendly 9=Very Friendly]*

- *Politeness   [1= Impolite 9=Very Polite]*

Type of work and communication

A number of studies of Acceptability for employment referred to in this thesis asked how the speakers would be rated for jobs of varying status to correlate groups or accents with high or low status jobs (see Table **6.1-1**). As the focus of this thesis was low-status work this distinction was not relevant. However, employers might make different decisions depending on who the employee would have to communicate with. The Acceptability of any job applicant may be different if they have to communicate with colleagues or customers as part of their work. Separate contextual questions were therefore added on Acceptability for communication with colleagues, supervisors and customers. Employers in Part A were asked:

How acceptable is the migrant's pronunciation for communication with the following people?

- Their supervisor?

- Other employees?

- Customers?

Pronunciation Features

Although the Base Rating would be use to predict the Employers' responses, I wanted to gain some understanding of how Employer judgements of specific features such as Intonation would predict their ratings of Acceptability. I therefore included a number of rating scales on the Migrants' pronunciation features in the Employer Survey. Employers were asked about

- *"Sounds" (Vowels, consonants, and added or missing)*

- *Stress on words*

- *Intonation*

- *"General impression" (Speed, Variety, Smoothness, Pausing and Hesitation).*

Question Groups

The information I wanted to collect from Employers came under the following groups:

- Employer: personal details, attitudes to language use, experience with migrant employees.
- Response to speech sample: guess of language; Acceptability for a job.
- Acceptability of speaker for communication with supervisors, other employees and customers.
- Attitudinal responses to speaker (e.g. hard worker)
- Rating of speaker for Intelligibility[1], Comprehensibility and Accentedness.
- Rating of speakers for specific pronunciation features: segmental, prosodic, fluency.

However, two problems arose. Firstly, the sheer amount of data sought meant that each survey would take a long time and probably create resistance to continuing with further samples. I needed each respondent to assess five samples. Secondly, the results of one type of question might have an effect on the responses to other questions. The survey had to focus on Employers' Acceptability responses on pronunciation alone and not have these confounded with other attitudinal or evaluative factors.

My solution was to ask some of the questions in each of two separate surveys to which the respondents were randomly assigned. After entering the "Employer data" such as type of work offered, all Employers would then be randomly allocated by the Qualtrics survey platform to Survey A or B. They would all hear speech samples from the same set, and they would all be asked to guess the speaker's language and rate their overall Acceptability for a job. The following questions would then be either on pronunciation features or on global measures, speaker qualities, and Acceptability for different types of communication.

---

[1] As noted in **7.6.2**, Intelligibility was measured by impression and not with a transcription task.

6.3.3   Structure

The structure of the Employer Survey, in the order encountered by respondents, was as follows:

Figure 6.3-1 Structure of the Employer Survey

```
          ┌─────────────────────────────────────┐
          │    Introduction and information      │
          └─────────────────────────────────────┘
                            │
                            ▼
          ┌─────────────────────────────────────┐
          │            Consent form              │
          └─────────────────────────────────────┘
                            │
                            ▼
          ┌─────────────────────────────────────┐
          │        Employer information          │
          │   Questions about work offered etc.  │
          └─────────────────────────────────────┘
                            │
                            ▼
          ┌─────────────────────────────────────┐
          │            Instructions              │
          └─────────────────────────────────────┘
                            │
                            ▼
          ┌─────────────────────────────────────┐
          │     Randomly assigned to Part A or B │
          └─────────────────────────────────────┘
                   ╱                    ╲
                  ▼                      ▼
```

Part A	**Part B**
Click player to hear randomly assigned sample.	Click player to hear randomly assigned sample.
Common Questions:	**Common Questions:**
• Guess language	• Guess language
• Acceptability (overall)	• Acceptability (overall)
Part A Questions:	**Part B Questions:**
• Acceptability (contexts)	• Pronunciation features
• Speaker characteristics	Repeat to total 5 samples.
• Global Measures	
Repeat to total 5 samples.	

All Employers responded to up to 5 samples from either Part A or Part B. Most completed the set 5 samples (from the same set of 40), some stopped before completing (and were included in the final analyses), but others were prevented from taking the survey by problems with the survey platform or slow response speeds.

6.3.4   Rater Effects

The structure of the survey also needed to take into account possible effects on respondents, as factors such as exposure and fatigue can affect rater reliability.

Exposure: A rater's judgement can change during a rating exercise as the first ratings can be re-evaluated in the light of later samples. My intention was to make the process as realistic as possible. In real life any given interviewee might be first or last, and any effects would be randomly distributed amongst the 40 speakers, so I did not consider this would affect the outcomes.

Fatigue: Reliability can be reduced if the session is too long: there is a balance between gathering a lot of information and having the quality diminish over time. Firstly I limited the survey length to five observations, and secondly the Employers were given the option of stopping part way through and restarting, provided they used the same computer (for technical reasons to do with Qualtrics).

Identity: As the survey was online I had little control over the identity of the person taking the survey. I contacted every employer personally by phone and explained the survey and the reasons for it using a standard script. They were not offered payment and volunteered to take the survey. I sent a link to the email address they supplied, and this provided their identity once they completed it. Theoretically they could then have asked someone else to take the survey, so I could only trust that this did not happen. The people concerned were employed in positions of trust and held responsibility within their organisations, and they would have had nothing to gain from any deception.

6.3.5   Employer Survey Trial

After testing the survey myself I sent a version of it to some employers, colleagues and my supervisors and asked for their responses about any problem encountered or queries about the questions themselves. The following are some unedited comments made by those trialling the survey.

- *It would have been helpful to know what criteria we were going to be assessing the speakers on beforehand.*

- *Could have a stab, but it would have felt fairer on the speakers if we could listen more than once. But that is balanced with the time taken to do the survey ie you don't want it to take too long.*

- *It is very difficult to judge a person's character based solely on an audio snippet, what I was recalling here was tone of voice.*

- *Knowledge and understanding of the English language was an obvious issue so I rated these people as difficult to follow even though their accents were not too bad. I can see that if they learnt more English it would be fine.*

- *Some sliders default in odd places, most default at 0 but two did not. I did leave some at 0 and then was confused as to why the questions would not submit, perhaps the error message could be clearer.*

I considered each of these responses. Some of them were due to technical problems which I could repair, while others called for reconsideration of the questions or the way they were expressed. Clearly not every idea was practical or was accepted. Specifically, I wanted all employers to listen to each sample once only as otherwise the resultant data would contain some responses from one listening and others from multiple listenings, which could influence the outcomes in different ways.

Once I had made all the changes which I thought would improve the survey I opened it to employers who had expressed an interest. The whole response process took about 6 months.

6.3.6   Recording and management of speech samples

The recordings were all made in an acoustically treated recording room at Victoria University of Wellington. The speakers wore a head-mounted microphone which recorded directly to computer.

Speakers were told the full purpose and procedure for the recordings (see Permission form, **Appendix A1**) and were recorded once they had signed the agreement. They knew that the

recording would be a mock job interview which I would use to find out what features of migrant speech employers were happy or not happy with, and they knew that it would be made available online to selected employers. I first asked speakers to identify themselves so that I could keep track of the recordings, but (as explained to the speakers) this and any other identifying information was always deleted from the version used.

Some speakers, particularly those with more limited English, were very unprepared for a job interview and I was concerned that they often had little to say, which restricted the phonological content. For example, if Word Stress was a problem, this would not show up with a speaker using few multisyllabic words. However, as a quid pro quo for the speakers' assistance, I commented on their job interview technique after the first recording and offered them the opportunity of making a second recording. I offered suggestions such as "Don't just give a quick answer. Talk about some examples." The result was that I had two interviews from most participants, from which I could choose the more copious sample. This process had no bearing on the validity of the evidence as speakers were given advice on how to improve the content of their responses but not their pronunciation.

Once I had the recordings I used Audacity[1] to perform the following adjustments:

- Export as MP3 file.

- Remove of all my speech and long gaps between questions.

- Remove any references to the speaker's identity, language or country of origin.

- Reduce recording to a maximum of 120 seconds (usually from the start of the interview, or by cutting out repetition).

- Reduce background noise, and remove clicking or tapping, by using the Compressor or Noise Reduction functions in Audacity.

- Increase volume, where needed, to ensure audibility.

One problem was that on rare occasions there were unfilled pauses so long that listeners might conclude the interview had finished, or simply give up listening, so these were cut

---

[1] Audacity(R) software is copyright (c) 1999-2012 Audacity Team. The name Audacity(R) is a registered trademark of Dominic Mazzoni.

down to the length of long unfilled pauses. These pauses were usually evidence of the un-preparedness of some of the speakers for a job interview. It is true that this would affect the result, since a speaker with extremely long pauses is highly likely to be judged as having low fluency. However, I decided that reducing the pauses was a necessary trade-off to ensure that the Employers continued listening. The remaining long pauses preserved hesitation phenomena, and the speakers could then be judged for their ability in other features as well.

Also on rare occasions more fluent speakers anticipated later questions so their answers were in a different order from other speakers. I used clear breaks in their responses for cutting and pasting to re-order the answers for consistency, since the questions themselves were always erased. Apart from the changes described above the speech itself was not adjusted in any way.

All of the sound files (including second recordings as described above) were numbered for identification. The 40 files selected were uploaded to the survey platform, Qualtrics, so they could be accessed for rating as part of the survey.

6.4   EMPLOYER SURVEY COMPLETION

Technical Problems

Once the survey was designed I loaded it into Qualtrics, the survey platform made available by Victoria University of Wellington. It eventually became evident – and was later acknowledged by Qualtrics technicians (personal email 8/11/2013) – that the platform was not capable of handling a survey of this size and complexity, and this led to a range of problems while setting it up and then using it with the respondents.

The extremely slow response rate of Qualtrics meant that some survey takers gave up, thinking there was no connection. Some respondents reported they were unable to complete the survey for technical reasons.

Data obtained

Observations per Employer

The numbers in Table **6.4-1** show that 38 Employers completed the full 5 ratings for Survey A and another 38 did so for Survey B. For Survey A, one Employer completed only 1 sample, while 6 completed 2-4 samples. In Survey B, 8 completed 1-2 samples and 4 completed 3-4. This produced 209 observations for Survey A and 215 for Survey B, totalling 424 observations.

Table 6.4-1 Number of Observations Obtained

Survey A			Survey B	
No. Samples	*No. Employers*		*No. Samples*	*No. Employers*
1	1		1	4
2	2		2	4
3	2		3	3
4	2		4	1
5	38		5	38
Total	**209**		**Total**	**215**

Observations per Speaker by Survey

My aim was to have at least five observations per speaker, with not too many above that number. However, as the Qualtrics function to distribute responses evenly failed, the two parts of the survey had unequal numbers by the time all likely respondents had finished, and responses per speaker ranged from 4 to 8. The distribution is shown in Figure **6.4-1** below.

Figure 6.4-1 Responses per speaker
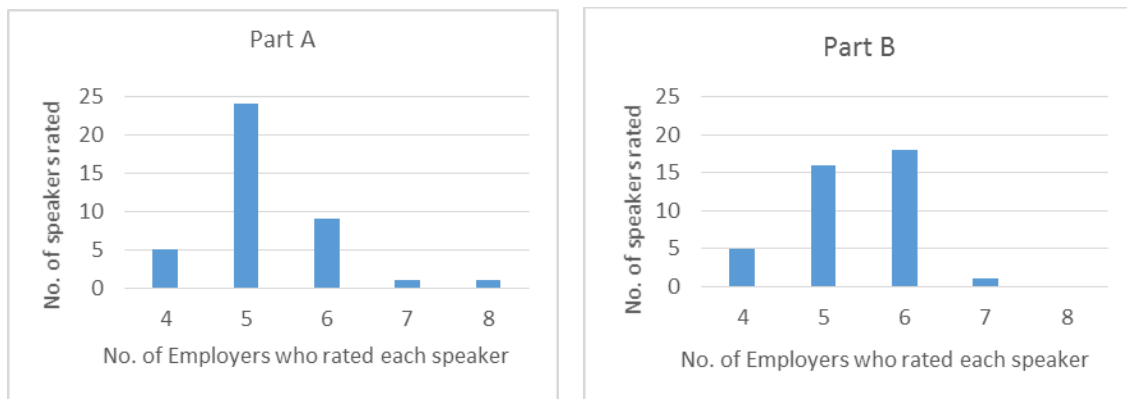
In Figure **6.4-1** each column represents the number of speakers who were rated the number of times (4-8) given on the x-axis. Thus, five speakers in both parts A and B were rated by only 4 respondents; the majority of speakers in both parts were rated by 5 or 6 respondents; one speaker in each was rated 7 times and one in A had 8 ratings.

No action was taken on the number of observations per speaker, but Employers who had completed fewer than 3 samples were removed from some analyses to find out whether there was any effect on the results. No effect was found.

6.5   BASE RATING

6.5.1   Design and Trial

As explained in section **6.2.3**, five Experienced Raters were asked to assess specific features of the Migrants' speech to create the Base rating. The process was as follows:

1. I designed a draft rating table and guidelines
2. I drafted the purpose and procedure for discussion
3. I recruited the Experienced Raters by word of mouth from the university
4. We met to discuss procedures and guidelines
5. We trialled the rating table with 3 samples
6. We discussed variations in our ratings and modifications to the questions
7. I produced rater guidelines (**Appendix C1**) and the final versions of the rating table (**Appendix C2**).
8. I transcribed the samples and gave the transcriptions to the Experienced Raters
9. The 5 Experienced Raters rated the 40 samples.

Questions - Summary

The questions in detail are in **Appendix C2**. The scales were from 1 (feature is omitted or used wrongly) to 9 (feature is used correctly). The questions were in the following groups:

Global Measures: Intelligibility, Comprehensibility, Accentedness

Segmentals: Consonants, vowels, added or missing finals, added or missing others.

Suprasegmentals: Word stress used, word stress placement, intonation, contrastive intonation, coarticulation, pauses and phrasing.

General impression: Slow/fast, monotonous/interesting, choppy/smooth, hesitant/fluent.

Non-phonological: Lexis, grammar, voice quality, intrusive noise.

General L2 proficiency rating could have been included in order to test whether Employers were in fact focussing on pronunciation features, but such a rating would either have been highly impressionistic or would have required a detailed test. If necessary, an overall pronunciation rating could have been created by averaging the Base Rating Results. The degree

of correlation among many variables shows that few variables were assessed in isolation. However, as noted at the end of section **8.2**, *The lack of effects from most other features suggests strongly that the Employers did what they were asked to do: rate the Migrants for Acceptability based on their pronunciation.*

Base Rating Trial and Results

Initially I met with four of the Experienced Raters who were available to discuss the rating procedure. We agreed that we were looking for phonological deviance which would be noticeable in a normal New Zealand employment situation. Features which were not normal amongst New Zealanders but acceptable from overseas dialects, such as a flapped or rolled [r], would be ignored in Comprehensibility ratings unless causing some interference. We examined each of the pronunciation features from the Employer Survey which we thought might influence Accentedness and Comprehensibility (e.g. Intonation) and decided whether more specific ratings should be used where they were likely to be used in teaching (e.g. Contrastive Intonation). We decided that the Experienced Raters' scales would therefore be more specific than the Employers' but use the same 1-9 ordinal scales. I then ensured that the fifth rater understood these decisions before running a trial.

From the trial on three samples there was broad agreement on the Global Measures, Segmentals, Fluidity, Speed, Voice Quality and Noise; less agreement on Variety and Smoothness; and low agreement on Stress, Prosody, Lexis and Grammar. The raters were given the results to see how they compared with one another and modify their ratings.

We then agreed on a number of overall procedures on pronunciation features:

1. Native speakers would have used the feature, but speaker didn't: treat as wrong use.

If the speaker uses little Coarticulation, then their speech can probably be described as "choppy" and they get low score for that feature.

2. Rarely used or needed, but correct when it appears: treat as correct use.

Even if rarely used, treat as correct if not wrongly omitted elsewhere.

3. No opportunities to use the feature: treat as mid-point (i.e. 5 on scale)

For example, if the content did not require Contrastive Stress, a 5 on the 9-point scale would be used, as it would not be recorded as positive or negative when correlating with Employer Acceptability. On the other hand, Word Stress must be used unless the speaker uses mono-syllables throughout, so I added a "Word Stress Not Evident" scale.

Base Rating Process

In order to reduce the Experienced Raters' workload, each of them rated 16 of the 40 speech samples which were distributed by sex of speaker and language, so all speakers were assessed by two raters and each rater overlapped 8 samples from each of two others. Table **6.5-1** below shows how the speakers were distributed.

Table 6.5-1 Distribution of Speaker Origin and Sex by Rater

RATER	R1	R1	R3	R4	R5
Males	9	11	9	9	10
Females	7	5	7	7	6
Middle Eastern	2	2	2	2	2
East Asian	6	5	6	6	5
African	2	2	2	2	2
South American	2	2	2	2	2
Western European	1	2	1	1	1
South Asian	2	2	2	2	2
Pacific	1	1	1	1	2
	Samples allocated				

While the Employers were to be asked for their opinions, the Experienced Raters were attempting – as far as possible – to provide an objective observation. They would therefore be able to listen to each sample as often as they wished.

6.5.2   Transcription of speech samples

The Experienced Raters agreed that a transcription of each sample would assist them with the rating. There is a trade-off in using a transcription: it can be directive in terms of pointing the rating in a certain way, but also encourages consistency, so that the same principles of interpretation were used throughout, and raters would not have to spend time deciphering what was said.

I transcribed all of the speech samples myself, having previous experience of speech transcription. The recordings were mostly quite clear and did not present many transcription difficulties, although there were some words or phrases which I was unable to decipher, usually because of the speaker's unintelligibility. The transcription was intended to make an orthographic record of the words which I believed speakers intended to say, so words were all transcribed in normal spelling and no attempt was made to represent variant pronunciation, since that is what the Raters were attending to.

Words

When the speaker chose an incorrect word (e.g. "I employment myself") I transcribed it as "employment" as this is a lexical alternative, albeit incorrect. However, when I was satisfied that attempts at words and fillers were not meaningful words, they were transcribed as heard, to save the raters from repeating my efforts to decode sounds. For example:

> *Ah, I think mm … they think me, they think, um, I'm a good person*

> *Also my **customer** sometimes **am** ch- some old people*

In this case I did not transcribe "customer" with a final 's' as it was not pronounced; I am assuming "am" is the intended word even though it is grammatically incorrect (this can be verified by other examples from the speaker); "ch-" is an attempt to start a word which is then altered.

Where speakers had errors which were clearly lexical I made a note to that effect. For example, one speaker always said "customent" instead of "customer" (perhaps analysing the word as "custom + man" or "custom + ment") so I made a note *"Speaker uses "customen" or "customent" throughout for "customer".* This is then a lexical, not a pronunciation error. My interpretations could of course have been wrong in some cases and raters were instructed to use their own judgement.

Punctuation

I used minimal sentence punctuation (capitals, full stops, commas, occasional semicolons, and paragraphs) for ease of reading and to identify noticeable tone boundaries and pauses. I used a comma in an abnormal position to indicate a brief pause, three dots to indicate a long pause and two sets of three dots for a very long pause.

Interpretation

Transcription inevitably creates questions of interpretation and decisions about grammatical versus pronunciation errors. For example:

> **I am very <u>organise</u>:** Wrong tense or omitted /d/? (Speaker uses final -ed for past so can say it but may not know adjectival use). Decision: I wrote what I heard.

> **I <u>meet</u> a lot of people** (referring to past time)**:** Wrong tense, wrong pronunciation of "met", or correct NZ pronunciation of "meet"? This has to be judged against speaker's pronunciation of other words with this sound. Decision: I wrote "meet".

Some decisions were more obvious, e.g.

> **I have clean in hospital** this speaker does not use the past tense. Decision: I wrote what I heard.

> **I can cleans, er clean kitchenhand:** Speaker self-corrects so "cleans" can be taken as a mistake not an error. "Kitchenhand" is strange – surely he means "kitchen" – but uses the same word when describing his job, so this is a lexical error. Decision: I wrote what I heard.

There was a challenge with more unintelligible speakers:

> **I foll-, mm falis, falit, ah I [polite]:** Speaker actually said [fəˈlais]. Given the context ("What makes you a good worker?) and accent (Vietnamese) I believe it is reasonable to assume the speaker is trying to say "polite". Decision: I wrote [polite] to indicate my opinion and lack of certainty.

> **[aim ʧeɪn ɗə aː pˀɪlo kˀes, mæʧɹəʃif]** which in context (describing work in hotel) and language (Arakanese) can be reasonably interpreted as "I'm change the, ah, pillow case, mattress sheet". "I'm" is grammatically wrong; "[ʧeɪn]" must be "change"; "[pˀɪlo kˀes]" is quite intelligible; but the final word has to be interpreted, I assume it is "mattress sheet(s)". Decision: I wrote "I'm change the, ah, pillow case, mattress [sheet]" indicating my confidence in all but the last word.

Figure **6.5-1** is an excerpt from one of the transcriptions to show how one of the less intelligible samples was transcribed.

Figure 6.5-1 Example of transcription: Recording 022

NOTE: much of this speech is problematic.

Ah yup. I was manager in coffee shop, and I [work] [kitchenette]. Ah mm I [gone] manager and [kitchenette] ah and, when I were manager I can … fix some problem [with] ah computer, ah, ah, um I, I know I made some food, for people, [?] when it's people hungry like him, and I can cleans, er clean [kitchenette]and keep ah [kitchenette]ah clear because I likes ah work to work with ah another people because when I work with another people I can ask more information more advice, and when I have some problem the people can help me fix it.

I got up early, I ah clever, I foll-, mm falis, falit, ah I [polite] and I have I has lot of idea ah for ah my s- [?] come to shop in [rush hour] mm my shop ah my boss ah very happy ah and some people like ah, some people are, are, I must, ah some people are like me in shop.

6.5.3   Base Rating completion and output

The Experienced Raters' rating forms were accessed through a separate Qualtrics survey (**6.5.1**). This survey had various restrictions removed: there was nothing to prevent the Raters listening to the samples as many times as they wished, and the question validations were removed so that, unlike the Employers, they could go back and forth through the questions. One problem resulting from this was that raters omitted a small number of answers, so imputed answers were inserted as many analyses are unable to handle rows with missing responses (see **7.2.1** for details).

Unfortunately, inter-rater reliability using ICC (intraclass correlation coefficient) was too low between some of the raters. Since each rater covered only 16 samples it was not possible to eliminate any one of them without rendering all the ratings unusable. In the end I rated all the samples myself, which although lacking the benefit of multiple ratings was at least consistent. An ICC of my rating plus any of the pairs generally produced higher reliability, and

when rating only the three Global Measures the ICC was 0.750 to 0.937, indicating that we had strong agreement on the overall ability of each speaker even if we could not agree on which features of pronunciation were the cause. The analyses and their development appear in section **7.3**.

7   RESULTS AND ANALYSIS

7.1   OVERVIEW

This chapter describes how a Base Rating of the Migrant speech samples was developed, and how a 3-Factor Model was established to analyse the data obtained from the Employer Survey and to test the thesis research questions. Further hypotheses were tested, arising from questions asked over the whole Survey, or in each of the two parts.

Chapter Organisation

```
┌─────────────────────────────────────────┐
│  7.1    Overview                        │
│                                         │
│  7.2    Building the Base rating        │
│                                         │
│  7.3    Building the 3-Factor Model     │
└─────────────────────────────────────────┘
                    │
                    ▼
┌─────────────────────────────────────────┐
│  7.4    Employer and Migrant Data       │
│                                         │
│  7.4.1-7.4.5      Employer Data         │
│                                         │
│  7.4.6  Sex of Speaker and Employer     │
│                                         │
│  7.4.7  Language – Actual and Guessed   │
└─────────────────────────────────────────┘
                    │
                    ▼
        ┌─────────────────────────────┐
        │  7.5    Employer Surveys:    │
        │         Parts A and B        │
        └─────────────────────────────┘
           ↙                      ↘
┌────────────────────────────┐   ┌──────────────────────────┐
│ 7.6   Employer Survey       │   │ 7.7   Employer Survey    │
│       Part A                │   │       Part B             │
│                            │   │                          │
│ 7.6.1 Supervisors,         │   │ 7.7.1  Employer          │
│       Co-workers and        │   │        Pronunciation     │
│       Customers             │   │        Responses         │
│                            │   │                          │
│ 7.6.2 Global Measures       │   │                          │
│                            │   │                          │
│ 7.6.3 Personal Qualities    │   │                          │
└────────────────────────────┘   └──────────────────────────┘
           ↘                      ↙
        ┌─────────────────────────────┐
        │  7.8    Summary of Results   │
        └─────────────────────────────┘
```

7.1.1   Explanations of Statistical Tools

The data from the survey, and reliability of raters, were interpreted and analysed by a number of statistical tools run through the *R* platform (R Core Team, 2015a). Some terms are explained briefly below.

Cumulative Link Mixed Model

A statistical model was needed to show how Acceptability and other variables were predicted by responses in the Employer Survey and by the Base rating. The Cumulative Link Mixed Model (clmm) is a regression model which was used as the data was *ordinal* – that is, responses to the 1-9 scales were ordered (6 means a higher response than 3) but not numeric (we cannot be sure that 6 means a response twice as high as 3). "Ordered categorical data, or simply *ordinal* data, are commonplace in scientific disciplines where humans are used as measurement instruments. Cumulative link models … are a powerful model class for such data since observations are treated rightfully as categorical, the ordered nature is exploited and the flexible regression framework allows in-depth analyses" (Christensen, 2011). Clmm can be found in the R package *ordinal* (Christensen, 2015) and its input is the same as that for the more commonly used lmer.

The output of a clmm analysis includes an estimate, the standard error of that estimate, a z-value and p-value for each predictor. An example output is given in Model **7.1-1** in which Acceptability is predicted by the factors Prosody, Fluency and Segmental (these factors are explained in **7.2.2**).

Model 7.1-1 Example of a clm model output predicting Acceptability

```
           Estimate Std. Error z value Pr(>|z|) Odds
Prosody    0.663      0.197     3.361   p<0.001  1.94
Fluency    0.865      0.208     4.162   p<0.001  2.38
Segmental  0.683      0.216     3.161   p<0.01   1.98
```

The *estimate* can be used to calculate the log-odds (added to each model as *Odds*) of a unit change in the predictor resulting in a shift from one response category to a higher response category. The odds of moving from one response category to the next can be obtained using the exponential of the estimate. Thus for the example in Model **7.1-1** the odds of a unit change in the Prosody factor resulting in shift of Acceptability from one response category to a higher one are 1.94:1 (=exp(0.663)), while the odds of a change in Fluency having this result are 2.38:1 and those for Segmental are 1.98:1. Therefore, it is more likely that a unit

increase in the Fluency factor will result in a shift in Acceptability than it is for a unit increase in Segmental or Prosody. The statistical strength of the estimate depends also on its variance. Accordingly, the estimate is divided by the standard error of the estimate to derive a z-value, for which a probability (p-value) can then be obtained. In Model **7.1-1**, each of the factors is a significant predictor of Acceptability (i.e. the p-value is below the conventional 0.05 level).

ANOVA

The statistical significance of including predictors in clm models was tested by comparison of models with and without one predictor (IV). The comparison returns a likelihood ratio statistic with a chi-square distribution. This is reported in the form (ANOVA, χ^2 (*n1*)=*n2*, p<*n3*), where n1 = degrees of freedom, n2 = likelihood ratio statistic, and n3 relates to the p-value, which needs to be less than 0.05 for the result to be significant, indicating that the models are different and the model with the higher log likelihood is to be preferred as a better fit (if they are the same, the simpler model is preferred). The value is given as p<0.05 (or lower) if significant, or the actual p-value (e.g. p=0.74) is reported where the result is not significant.

Factor Analysis

Factor analysis is "useful for exploring and verifying patterns" by exploring the underlying variance structure in a set of correlation coefficients – that is, the responses to each question (J. Brown, 2001, p. 15); or, according to Rummel (2002) "a means by which the regularity and order in phenomena can be discerned." The Employer and Expert surveys asked for ratings of pronunciation features such as Word Stress, Intonation and Variety, but it is likely that responses to these variables will be correlated with one another. The extent to which a variable is independent of other variables is reflected in its uniqueness score, which can range from 0.0 (can be fully predicted by other variables) to 1.0 (no variation in common, therefore unique). A factor analysis shows which groups of variables are most strongly correlated with one another. Each variable is weighted proportionally to its involvement in the pattern and the resulting factors consist of highly interrelated variables. The factors are also ranked in order of the proportion of the data they account for – the *relative proportional variance* (Rummel, op. cit.). This process may appear to reduce the ability to pinpoint which

variables are significant, but the correlations makes them impossible to separate, and according to Dörnyei (2007, p. 253) "it reduces the number of variables submitted to the analysis to a few values that will still contain most of the information contained in the original variables." Factor analysis was used on the Experienced Rater responses.

Intraclass Correlation Coefficient (ICC)

ICC was used to test inter-rater reliability – the degree of consensus amongst raters. Shrout and Fleiss (1979, p. 427) commented that "it is important to assess the reliability of judgments made by observers in order to know the extent that measurements are measuring anything". ICC is a method of checking the consistency of two or more ratings (op. cit. p. 420).

Prediction

Much of the following discussion is in the form "X predicts Y", e.g. "Migrant's native language predicts Acceptability", proposing that Employers would give different Acceptability ratings to Migrants with different languages. The predicting variable (X, here language) is the Independent Variable or IV, and the variable to be predicted (Y, here Acceptability) is the Dependent Variable (DV).

Note

In this chapter, to distinguish survey variables from references to pronunciation and other features in general, variables are always capitalised.

7.2 BUILDING THE BASE RATING

The purpose of the Base Rating is explained in section **6.2.3** and its design is described in section **6.5.1**. This section describes the statistical analysis of the data obtained and how the final Base Rating was produced.

7.2.1 Imputation and Inter-Rater Reliability

The five Experienced Raters assessed all the samples through a Qualtrics survey (**Appendix C2**). Of a possible 1,680 responses, 12 (0.07%) were (by oversight) missing. These gaps could have had a serious effect on some analyses which discard an entire row of data if gaps are encountered, so an imputation code was used (Gelman & Hill, 2007 Ch.25) to estimate the missing data, and these estimates remained in all further models using the 5 Experienced Raters.

To combine a number of ratings into a single set there needs to be a high level of inter-rater reliability (IRR). Previous researchers have found satisfactory levels of reliability amongst un-trained raters for global measures of pronunciation (Pinget et al., 2014; Saito et al., 2015) and amongst experienced raters on a range of pronunciation features (Munro et al., 2015; Saito et al., 2015). However, this is not always the case: Mullen (1980) found considerable disagreement between judges, and Levis (2011) reported that even experienced raters could not agree on what features were relevant to Intelligibility.

Establishing IRR in this case was more difficult because each rater dealt with only 16 samples, overlapping two others by 8 samples (as described in **6.5.3**). This meant that all IRR tests had to be done in pairs. A series of Intraclass Correlation Coefficient (ICC) tests (**7.1.1**) was performed on the 13 variables (**7.3.2**) used in the general analyses, producing the following ICC ratings for each of the five rater pairs (Rater 1 to Rater 5):

Table 7.2-1 ICC Ratings between 5 Experienced Rater pairs

Raters included in 2-way comparison	ICC (consistency score for 2 raters)	F-value (df:167,167 in all cases)	p-value
R1/R4	0.719	3.55	p<0.0001
R1/R5	0.216	1.28	p<0.06
R2/R3	0.520	2.08	p<0.0001
R2/R4	0.497	1.99	p<0.0001
R3/R5	0.696	3.29	p<0.0001

Although most of the p-values give the scores as significant, Hallgren (2012, p. 6) noted that a conservative assessment will discount ICC values less than 0.67 although 0.4-0.6 has been seen as "moderate agreement". In this case, if any one pair failed then the ring was broken – none of the ratings could be used. The lower ratings, and particularly one of 0.216, indicated a low IRR. This rating also has marginal significance ($p < 0.06$). This result confirms that even experts may not agree on how speakers should be rated, which is deserving of further study.

As inter-rater reliability could not be fully established, I rated all the samples myself (without reference to the other ratings) to create the Base rating. As confirmation of my own consistency with the other ratings I then ran a further 3-way ICC with each pair of raters plus myself on the 13 pronunciation variables, which produced correlations of between 0.397 and 0.790 with those of the Experienced Raters (Table **7.2-2**), somewhat higher than the original correlations. All of the results were statistically significant. This confirmed that my ratings were consistent with the other 5 Experienced Raters.

Table 7.2-2 Three-way ICC with 6 Experienced Raters

3-way comparison including raters	ICC (consistency score for 3 raters)	F-value (df:103, 206 in all cases)	p-value
R1/R4	0.790	4.76	p<0.0001
R1/R5	0.397	1.66	p<0.01
R2/R3	0.672	3.04	p<0.0001
R2/R4	0.647	2.83	p<0.0001
R3/R5	0.659	2.93	p<0.0001

In addition, an ICC using all 6 Experienced Raters' responses to the three Global Measures (Intelligibility, Comprehensibility and Accentedness) gave correlations which were all higher than those drawn from the 13 pronunciation variables, ranging from ICC 0.493 to 0.800 (Table **7.2-3**). All results were statistically highly significant ($p < 0001$). This indicated that the 6 Experienced Raters might not agree on the specific factors but did broadly agree on the Global Measures.

Table 7.2-3 Three-way ICC with 6 Experienced Raters: Global Measures

3-way comparison in-cluding raters	ICC (consistency score for 3 raters)	F-value (df: 127, 254 in all cases)	p-value
R1/R4	0.800	1.97	p<0.0001
R1/R5	0.493	1.97	p<0.0001
R2/R3	0.705	3.34	p<0.0001
R2/R4	0.663	2.97	p<0.0001
R3/R5	0.733	3.75	p<0.0001

My rating of all Migrant pronunciation features then became the Base Rating used throughout this thesis to predict Employer responses. However, since the ICC 3-way comparisons of the 6 Experienced Raters on the Global Measures reached the "moderate" level described in Hallgren (2012, p. 6), a model was also run using the mean of all 6 Experienced Raters to compare Experienced Rater and Employer judgements of the Global Measures (**7.6.2**).

7.2.2 Variables and Factor Analysis

List of Variables

Rater Survey (Appendix C2): Intelligibility, Comprehensibility, Accentedness, Consonant, Vowels, Added or Missing Finals, Other Added or Missing Sounds, Intonation, Word Stress Used, Word Stress placement, Intonation, Contrastive Intonation, Coarticulation, Pauses and Phrasing, Speed, Variety, Smoothness, Fluidity, Lexis, Grammar, Voice Quality, Intrusive Noise.

Employer Survey – All Employers Information about Employer Appendix B2: Job, Time in NZ, Language, Attitude to employees speaking English, Problems with NNS employees,

Employer Survey on Migrant Speech – All Employers (Appendix B3): Country (Language Guess); Acceptability

Employer Survey on Migrant Speech – Survey A (Appendix B3): Acceptability for communication with Supervisor / Co-workers / Customers; Hard worker; Friendliness; Politeness.

Employer Survey on Migrant Speech – Survey B (Appendix B4): Vowels, consonants, Word stress, Intonation, Added/Missing sounds, Speed, Variety, Smoothness, Fluidity.

Answering Research Questions 1 & 2

Firstly, in order to answer the first two research questions, the four non-phonological features (Grammar, Lexis, Voice Quality and Noise) were set aside. These variables could be added to the model later in order to answer RQ3.

Factor Analysis

Before the Base Ratings could be used to predict Employer responses, it was necessary to use a Factor Analysis to reduce the number of variables, as the number (14) was too high for only 40 samples (A. Field, Miles, & Field, 2012, p. 273). There were also significant correlations (Pearson's r) between the variables. For example, Intonation was correlated at r>0.5 with six 6 other variables, three with an r-value of 0.7 or more. Uniqueness ranged from 0 to 0.57 with mean of 0.33. The number of variables was therefore reduced by running a factor analysis (Dörnyei, 2007, p. 233) on the 14 pronunciation variables. This produced three factors, but I discovered a problem with one of the pronunciation variables, Speed.

Speed

The variable Speed was found to be problematic because its scale did not conform to those of the other variables. While all the latter could be interpreted as 1= "uses the feature very poorly" to 9= "uses the feature very well", Speed had 1= "too slow" and 9= "too fast" leaving 5 as the implied "uses the feature well". Analyses assuming that this variable operated in the same way as the others would therefore produce invalid results for Speed. To accommodate Speed as a variable, I recalculated its responses so both 1 and 9 became 1, both 2 and 8 became 2 and so on, giving a poor-to-good scale similar to the others in a new "Speed Deviance" variable *SpeedDev* with a range of 1-5 (see Figures **7.2-1** and **7.2-2**). This removed the fast/slow contrast and left a scale running from 1 (most deviant) to 5 (least deviant). However, even with the modified variable, responses to Speed did not behave like others. While other variable responses dropped off at the highest levels, most of the SpeedDev ratings (both Employer and Experienced Rater) were at the higher levels (see Figure **7.2-2**). Warren, Elgort and Crabbe (2009) found (contrary to many other studies) that speech rate showed "no significant association with the comprehensibility rating" (p.95) and surmised that this was due to the "generally slow rate" of the utterances. This may also have been true of the Migrants in the Employer Survey as their proficiency was not high.

Figure 7.2-1 Original Employer Speed Ratings
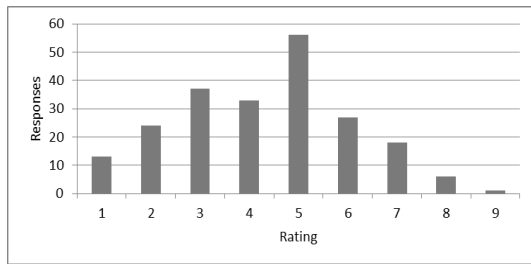
1 = Too slow   9 = Too fast

Figure 7.2-2 Speed ratings converted to SpeedDev
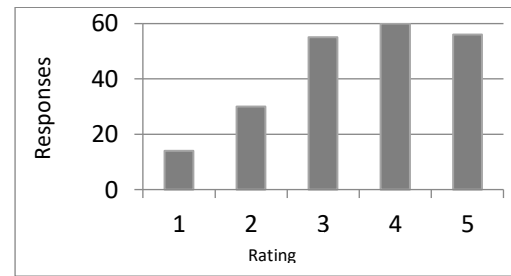
1 = Most Deviant   5 = Least Deviant

As a result, when a Factor Analysis was produced using *SpeedDev*, even this version was found to be unusable because the variable now sat outside of these factors with a uniqueness of 0.87. As the variable in the form Speed or SpeedDev did not contribute to the Base Rating model it had to be left out altogether. A similar decision was made by Rey (1977, p. 9) because the values of three variables "could not be considered as true polarities of positive-negative reactions". This left 13 variables for which I ran a further factor analysis on my ratings alone (**Appendix D9**) which produced 3 factors from the final 13 variables as shown in Table **7.2-4**.

Table 7.2-4 Factor Analysis: 1 rater & 13 variables

Factor	PV	Description	Survey variables included[1]
1	0.25	Prosody	**Main weightings:** Intonation (0.98), Contrastive Stress (0.76), Stress placement (0.67), Word stress (0.53) **Shared weightings:** Variety (0.61)*, Vowels (0.44), Pausing (0.42)*
2	0.24	Fluency	**Main weightings:** Pausing (0.82), Fluidity (0.78), Coarticulation (0.71), Smoothness (0.70) **Shared weighting:** Variety (0.62)
3	0.19	Segmental	**Main weightings:** Added or Missing Other Segmentals (0.80), Added or Missing Finals (0.77), Consonants (0.66), **Shared weighting:** Vowels (0.61)

Weightings in italics were treated as belonging to a different factor.

The ranking of the three factors reflects the relative proportional variance (PV) of each (see Factor Analysis in **7.1.1**). The number after each variable is its weighting within its factor.

The resulting three factors fell out from the statistical model, but were also compatible with phonological descriptions of English as shown in Chapters 3-5, fitting neatly into a descrip-

---

[1] Some of the terms used in the original survey so as to be understood by respondents have been changed in this thesis to positive descriptions of the features (Smoothness for Choppiness, Variety for Monotony and Fluidity for Hesitation).

tion of three levels of phonology: Fluency, Prosody and Segmental. These levels, and the researcher's involvement in their construction, are discussed in section **7.2.3**. When the three factors are used as predictors in the statistical models reported below, they are represented by the relevant factor scores for each speaker.

These three factors formed the *Base Rating* which was used throughout the ensuing analyses, in which other variables in the Experienced Rater and Employer Surveys were added to the model to find out whether they could help to answer RQ3. This initial model, in which Acceptability is predicted by the 3 factors of the Base rating, is referred to throughout as the **3-Factor Model**. Factors 1, 2 and 3 are henceforth referred to as **Prosody**, **Fluency** and **Segmental** respectively and as the **Base Rating** collectively.

7.2.3 Segmental, Prosody, Fluency

The result of the Factor Analysis may raise the question of circularity: Did the researcher create data which would then, inevitably, appeared in the statistics? In the sense that this was an analysis of the researcher's own ratings, yes; but in the sense that these were outcomes which were expected or hoped for, no.

Firstly, a factor analysis of the original 5 raters put the variables into very similar groupings as mine (**Appendix D9**), as follows (note that the order of factors 1 and 2 is different, suggesting a different weighting for these two):

```
                  Factor1 Factor2 Factor3
Smoothness        0.82
Fluidity          0.80
Variety           0.75
Pausing           0.60    0.55
Coarticulation    0.43    0.58
Intonation        0.42    0.84
StressMarked              0.56    0.51
StressPlace               0.68    0.39
AddMissFinal                      0.53
AddMissOther                      0.59
Consonants                        0.72
Vowels                            0.72
Contrastive Stress                0.40
```

Note that Speed had to be removed from the Factor Analysis for reasons given in **7.2.2.**

Secondly, the Factor which I subsequently named "Fluency" mostly contained variables described as "general impression" in the surveys, as the construct of Fluency had not been discussed. The result of the Factor Analysis was thus the effect of statistics making sense of the way six experienced raters interpreted what they heard. However, I acknowledge that independent ratings would have been preferable.

How, then are the three factors above to be interpreted? Phonological features have long been described in two levels: segmental (consonants, vowels) and suprasegmental (everything else). If a factor analysis of our ratings finds that suprasegmentals can be divided into two groups, is this just a statistical construct or a phonological reality?

In the factor analysis of my ratings (**Appendix D9**), **Factor 3**, consisting of the variables **Added or Missing Other Segmentals, Added or Missing Finals, Consonants and Vowels**, is indisputably the Segmental level.

Factor 1 includes **Intonation, Contrastive Stress, Stress placement** and **Word stress**, all of which are unequivocally prosodic features. **Vowels** also appears here, but with a lower weighting than it has in Segmental, which makes sense: it is understandable that accuracy of vowel use would have some effect on prosody, such as lack of unstress resulting in schwa pronounced as a full vowel. **Factor 1** also shared **Variety** equally with the remaining factor and **Pausing** with a lower weighting.

There now needed to be an explanation of what **Factor 2** was. It includes **Pausing** (with greater weighting than in **Factor 1**), **Fluidity**, **Coarticulation** and **Smoothness**, and shares **Variety** equally with **Factor 2**. Suprasegmentals have been divided into two levels (Isaacs & Trofimovich, 2012; Kang, 2012) and while Fluency usually involves automaticity (which could not be tested here) researchers tend to include Speed, Pausing, Hesitation, use of fillers, and Fluidity, while some add Smoothness (Kormos & Dénes, 2004, p. 146; Levis, 2008, pp. 263–4) which is affected by good use of Coarticulation. The remaining feature was Variety, the only variable with an almost equal weighting in two Factors: **Factor 1** (0.61) and **Factor 2** (0.62). Variety can be understood as a prosodic feature based on Intonation whose effective use gives the impression of greater Fluency.

Since a model based on these 3 factors would form the basis of all subsequent analyses (as detailed in **7.3** below), and since the features fitted into previously described constructs, it made sense to divide Suprasegmentals into two levels: Prosody and Fluency.

Dividing Suprasegmentals in this way does not imply that the component levels have nothing to do with one another. Many features interact with one another, as is shown by the three variables which appear in more than one factor. But there is a phonological reality, supported by the literature and the statistics, in putting features centred around the use of pitch into one level and those centred on temporal measures and ease of speech into another. In terms of this research, the separation serves to highlight those features which are commonly taught overtly and those which are not.

The inclusion of features such as Smoothness in the Fluency level requires further investigation. However, for consistency, the terms Segmentals, Prosody and Fluency are used throughout this thesis to mean these three pronunciation levels which include the features as grouped by the Factor Analysis.

7.3 BUILDING THE 3-FACTOR MODEL

7.3.1 Test for Normality

Once the survey was closed off I converted all the responses into a form which could be used with the R platform (R Core Team, 2015a) to analyse the data.

Examining the data obtained from the Employer Survey showed that Acceptability ratings were not normally distributed. Figure **7.3-1** gives the number of Employers (n=95, n.observations=424) who used each of the 1-9 rating scales in response to the question *"How acceptable is the migrant's pronunciation for a job with your company?"* (Mean = 5.40, SD = 2.23). The histogram shows that, although Employers used the full range of the rating scale, the survey had two types of skewness (both a tendency towards a greater number of high ratings and an unusual rise at the lowest level) and high kurtosis (a high peak with rapid falloff, in this case to the right).

Figure 7.3-1 Distribution of Responses to Acceptability – whole survey
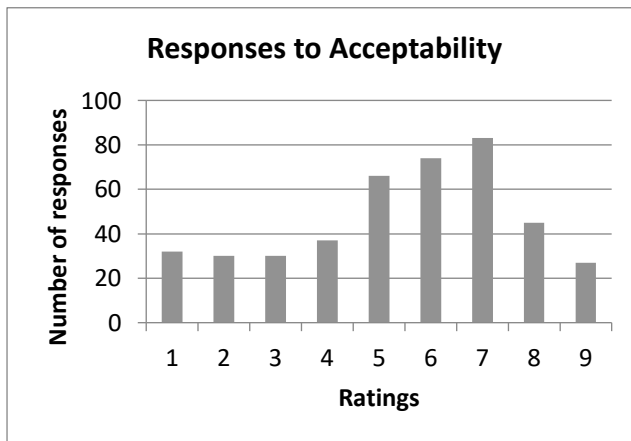
Figure 7.3-2 Distribution of Responses to Acceptability in Parts A and B
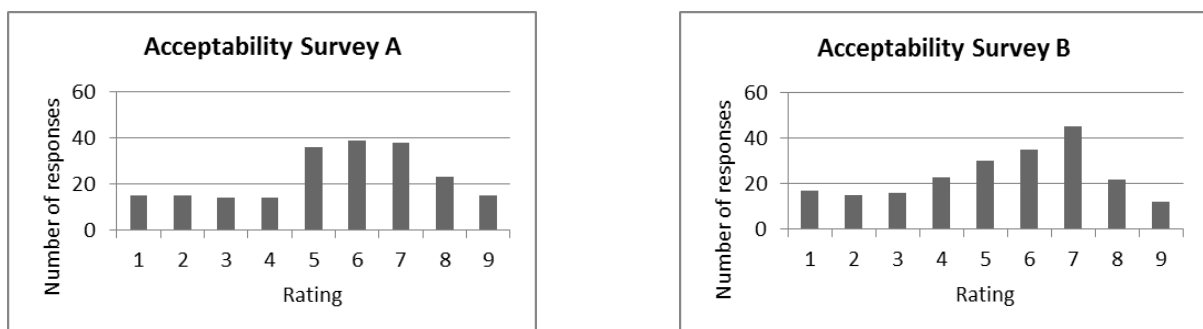
As the Employer Survey had two separate parts, with a different group of employers in each (Part A n=45; Part B n=50), the tests were run for the survey as a whole and then for each of the parts. Figure **7.3-2** (above) shows the raw numbers of responses for the two parts of the survey.

Kurtosis was tested using the Anscombe-Glynn kurtosis test (Komsta, N.D.a) which found that kurtosis was not equal to 3 and therefore it exists in the data. Skewness was tested using the D'Agostino skewness test (Komsta, N.D.b) and deviates from 0 by more than 1, so the data is highly skewed (Table **7.3-1**). Two Employers had a tendency to give many low scores, so a model was run without these to find whether these deviations were caused by outliers, but kurtosis and skewness remained.

Table 7.3-1 Tests for normality

Survey	Kurtosis	z-value	p-value	Skewness	z-value	p-value
Full	2.298	-4.523	<0.0001	-0.422	-3.472	p<0.001
Part A	2.369	-2.550	<0.05	-0.447	-2.614	p<0.01
Part B	2.234	-3.588	<0.001	-0.399	-2.386	p<0.05

A standard general linear model (GLM) would not work because GLMs have an assumption of normality (A. Field et al., 2012, p. 169). This provided a further reason for the use of the cumulative link mixed model (clmm) to analyse the rating scales used in the survey (see **7.1.1**).

It should be noted that non-normality means that different approaches to analysis are required, not that there is some fault or even strangeness about the data. Micceri (1989) examined 440 real-life distributions and found that "even amongst the bounded measures of psychometry and achievement, extremes of asymmetry and lumpiness are more the rule than the exception" (p.161). Non-normality in the data from the Employer Survey may arise from the speakers themselves as they were selected from those attending English language school classes at elementary to low intermediate, so there was not a full range of speaking abilities. It may also come from the employers and their attitudes: they were selected from those offering low or unskilled work, so they may regard a wide range of candidates as acceptable but have a cutoff point when intelligibility is too low or irritation too high. The questions themselves may also have made some answers more likely than others.

7.3.2   The 3-Factor Model

The research questions for this thesis were:

1. Does pronunciation affect employers' assessments of the Acceptability of migrants for employment?

2. If so, which pronunciation features are the strongest predictors of the assessment?

3. Once pronunciation features are taken into account, are there other employer or speaker factors which affect employer ratings of acceptability?

To answer these questions, I asked employers to rate the pronunciation of migrants for acceptability for employment, created a Base Rating for comparison, and then used a series of regression models to find out to what extent the judgements in the Base Rating were predictors of Acceptability.

This chapter will show that answers to the research questions are:

1 Yes. The Base Rating of pronunciation features is found to be a strong predictor of Employers' Acceptability ratings.

2 All of the three Factors of the Base Rating are significant predictors, but Fluency features are the strongest.

3 Once added to the 3-Factor Model, most other Employer and Migrant-related variables do not contribute to judgements of acceptability, with the exception of some Employer judgements of speaker qualities and variables which are discussed further in **7.6.3** and **7.7.1**.

These results have implications for teaching pronunciation, which are discussed in Chapter 8.

As shown in **7.2.2**, it was not possible to use all variables from the Experienced Rater Survey as distinct predictors, so the final 13 were grouped into three factors. I used a clm model (**7.1.1**) with these three factors as the independent variables (IVs) and Acceptability as the dependent variable (DV). The model includes two random intercepts: Employers and Speakers, namely the Employers (n=95) who took the surveys and the Migrants (n=40) who were recorded. Including these as random intercepts removed the random variation of responses

amongst members of each of these groups so that predictions of Acceptability by the three factors (or other variables) could be seen more clearly. This, as shown in Model **7.3-1** below, produced the following results over the whole survey (this is the same model as Model **7.1-1** where the meaning of the numbers is explained; see **Appendix D3** for full model output).

Model 7.3-1 The 3-Factor Model

```
          Estimate Std. Error z value Pr(>|z|) Odds
Prosody    0.663      0.197    3.361  p<0.001  1.94
Fluency    0.865      0.208    4.162  p<0.001  2.38
Segmental  0.683      0.216    3.161  p<0.01   1.98
```

Based on the odds in the model, all three factors are significant predictors of Employer ratings of Acceptability ($p < 0.01$), with Fluency highest. This was the model against which other models were compared by introducing other variables gleaned from the survey. It has three versions: the full version, used for whole Employer Survey questions; the Part A version, using only the subset of data from the 45 Employers who responded to it; and the Part B version with the subset of data from the 50 Employers who responded to that part.

There are different approaches to developing a final descriptive model of data. Because of the structure of the Survey, consisting of three sections (questions asked of all Employers, and Parts A and B with questions for separate subsets of Employers), the variables from all-Employer questions were tested one by one against the 3-Factor Model to create a final model which could then be used as the basis to examine the responses to Parts A and B separately. While the samples of Migrant speech were the same throughout, the two parts had different groups of Employers, so a single model containing all variables could not be created.

7.3.3   Voice Quality and Noise

Voice Quality and Noise were included in the Rater Survey to test whether potentially distracting factors of the Migrants' speech or recordings influenced the responses. These non-phonological variables were left out of the factor analysis in order to answer RQs 1 and 2. Once the Base Rating and 3-Factor Model were established, these variables could be added as IVs to answer RQ3.

Voice Quality

Hypothesis:

> *Predictions of Acceptability will be affected by Voice Quality*

There is a possibility that the individual quality of some speakers' voices might cause them to be given low ratings of Acceptability, even though their proficiency is adequate.

The term *Voice Quality* as used here does not refer to Honikman's (1964) articulatory settings which form the basis of pronunciation for a language, so any effects from such settings would appear under Accentedness and pronunciation features. Experienced Raters were more concerned with problems of individual voice production (shrill, husky, boring, etc.) A model was run adding Voice Quality to the 3-Factor Model, but it was not a better fit than that model (ANOVA, $p=0.681$). The hypothesis was therefore not supported.

Noise

Hypothesis:

> *Predictions of Acceptability will be affected by Noise*

Scott (1999) found that background noise had a detrimental effect on the intelligibility of accented speakers, although Munro (1998) concluded that it was unsafe to generalise as individuals are affected differently. This feature was added in order to detect problems with the survey itself. If the recordings included tapping noises, feedback or other distractors, they might influence skew results.

A model was run adding Noise to the 3-Factor Model, but it was not a better fit (ANOVA, $p=0.280$). The hypothesis was not supported.

7.3.4   Grammar and Lexis

Hypothesis:

> *Predictions of Acceptability will be affected by Grammar and Lexis.*

Data: **Appendix D5**

Grammar and Lexis were included in the Rater Survey because (as shown by Section **3.2.5**) they are found amongst the predictors of Comprehensibility, which is likely to be correlated to some extent with Acceptability. These non-phonological variables were left out of the factor analysis in order to answer RQs 1 and 2. Once the Base Rating and 3-Factor Model were established, these variables could be added as IVs to answer RQ3.

The Rater Guidelines (**Appendix C1**) instructed Experienced Raters to treat these features as context dependent (i.e. taking into account their effect on comprehensibility) rather than independently, which means that it was difficult to distinguish unique effects of Grammar and Lexis from those which are related to pronunciation features or to comprehensibility. The first problem, shown in Table **7.3-2**, was that Grammar and Lexis are strongly correlated (shown by *) with each other and with Fluency.

Table 7.3-2 Grammar and Lexis: correlations (Method=Spearman)

	Lexis	Grammar	Prosody	Fluency
Grammar	0.667*			
Prosody	0.240	0.213		
Fluency	0.405 *	0.456 *	0.006	
Segmental	0.293	0.412 *	-0.097	0.040

Grammar and Lexis were added to the model separately and then together to test the effect of these correlations.

Grammar

When Grammar alone is added to the 3-Factor Model (Model **7.3-2**), the ranking of the Factors changes and Segmental is no longer a significant predictor (p=0.149). As it is unlikely that that Segmentals do not predict Acceptability (or any other Global Measure) this is probably a product of the model, in which the effects of the other variables are too strong for the Segmental effect to be measurable.

The ranking of estimates is now Fluency – Prosody – Grammar although the z-value for Prosody is highest. Grammar shows collinearity with all three factors in this model (Prosody -0.287; Fluency -0.557; Segmental -0.575).

The model including Grammar was a better fit than the 3-Factor Model (ANOVA, χ^2 (1)=4.55, p<0.05).

Model 7.3-2 The 3-Factor Model + Grammar

```
          Estimate Std. Error z value Pr(>|z|)
    Grammar   0.333      0.153    2.183  p<0.05
    Prosody   0.538      0.194    2.745  p<0.01
    Fluency   0.566      0.236    2.393  p<0.05
    Segmental 0.360      0.250    1.443  p=0.149
```

Lexis

The results for Lexis were similar, with Segmental no longer a significant predictor (p=0.148) (see Model **7.3-3**) and collinearity was similar to Grammar (Prosody -0.348; Fluency -0.475; Segmental -0.509). The estimates rank the IVs as Fluency – Prosody – Lexis but the z-value of Lexis is the highest. The model including Lexis was a better fit than the 3-Factor Model (ANOVA, χ^2 (1)=7.67, p<0.01).

Model 7.3-3 The 3-Factor Model + Lexis

```
          Estimate Std. Error z value Pr(>|z|)
    Lexis     0.407      0.142    2.879  p<0.01
    Prosody   0.459      0.191    2.401  p<0.05
    Fluency   0.551      0.214    2.570  p<0.05
    Segmental 0.331      0.229    1.448  p=0.148
```

Grammar + Lexis

Each of the models with one of these variables was a better fit than the 3-Factor Model, in both cases displacing Segmental as a predictor. However, with both added (Model **7.3-4**), both Grammar and Segmental cease to be significant predictors; the ranking of estimates remains at Fluency – Prosody – Lexis but the z-value of Prosody is highest.

Model 7.3-4 The 3-Factor Model with Lexis + Grammar

```
          Estimate Std. Error z value Pr(>|z|)
    Grammar   0.161      0.166    0.970  p=0.333
    Lexis     0.330      0.161    2.053  p<0.05
    Prosody   0.435      0.190    2.289  p<0.05
    Fluency   0.466      0.229    2.039  p<0.05
    Segmental 0.242      0.243    0.995  p=0.320
```

The model including the 3 Factors, Grammar and Lexis was compared with a model with Grammar alone, and found to be a better fit (ANOVA, χ^2 (1)=4.042, p<0.05). The same model compared with Lexis alone was not found to be a better fit (p=0.335). However, there is collinearity between Lexis and Grammar (-0.492), between Grammar and each of Prosody and Fluency (approx. -0.375), and between Lexis and both Prosody (-0.24) and Segmental

(-0.227). In addition, the three Factors, which normally show virtually no collinearity in the 3-Factor Model, now have scores of between 0.24 and 0.33. Outside of the model, Grammar correlated most strongly with Fluency.

A further model tested the interaction between Grammar and Lexis. Comparison of the 3-Factor Model with Grammar and Lexis as simple effects with a model that also included the interaction of Grammar and Lexis showed that the latter was not a better fit (p=0.844).

From this evidence it is likely that both the Grammar and Lexis variables have some predictive effect. However, there is so much collinearity within the model amongst the three factors and two variables that it is difficult to distinguish where the effects come from. For example, missing final *-s* or *-ed* could be categorised as "missing segments" or "grammatical errors". It should also be noted that the Employers were asked to focus on pronunciation in their ratings.

Models were run using centering, which is claimed to reduce collinearity (Robinson & Schumacker, 2009). Centering all three Factors, Grammar and Lexis produced a model in which Lexis was the only predictive variable, but variables were still correlated between 0.21 and 0.55. Lexis plus the 3 Factors produced Model **7.3-5** in which Prosody, Fluency and Lexis were significant but the p-value was too low in Segmentals. However, correlation between the 3 Factors and Lexis remained at between 0.35 and 0.51. Wurm and Fisicaro (2014, p. 46) state that centering does not reduce "essential collinearity" and it appears that this method did not succeed here.

Model 7.3-5 The 3 Factors + Lexis, Centered

```
            Estimate Std. Error  z value  Pr(>|z|)   log
 cProsody    0.4585     0.1909     2.401   p<0.05    1.58
 cFluency    0.5507     0.2143     2.570   p<0.05    1.73
 cSegmental  0.3309     0.2285     1.448   p=0.148   1.39
 cLexisR     0.4073     0.1415     2.879   p<0.01    1.50
```

From the above results, noting the Raters' instructions and the collinearity, it is difficult to make any certain statements about the roles of Grammar and Lexis. However, if these objections can be put aside, the ANOVA results prefer models with Lexis and Grammar than without either variable, and once both are included they indicate that only Lexis is a significant predictor of Acceptability.

7.4 Employer and Migrant Data

In both parts of the survey, Employers were asked the same six questions about themselves and the work they offered (see **Appendix B2**). Responses to each of these questions were added to the 3-Factor Model to see if they contributed to the prediction of acceptability, and the results are discussed in sections **7.4.1** to **7.4.4**.

There was also a qualitative question to give Employers the opportunity to comment on pronunciation problems they may have had with employees. The responses to these questions are discussed in section **7.4.5**

Possible effects of the sex of the Migrant or Employer are discussed in section **7.4.6**.

After listening to the sample, all Employers were asked to make a guess at the Migrants' country of origin, and these responses are discussed and compared with the Migrants' actual languages in section **7.4.7**.

7.4.1   Work offered

Hypothesis:

Predictions of Acceptability will differ depending on the type of work being offered.

Acceptability is a context-based Global Measure. This is demonstrated in section **7.6.1** where the ratings vary depending on who the Migrants are expected to communicate with. It is therefore possible that the context of *kind of work* ("Job") offered will also influence Acceptability. Employers were asked what kind of work they offered, and this is listed in Table **7.4-1** (below). The "Other" heading consisted of a small number of jobs which did not fit into any category.

Employers could provide multiple responses to "What work do you offer?" which resulted in 59 employers listing only one kind of work and the remaining 36 listing between 2 and 5 kinds, which was difficult to analyse. It was therefore decided to use a Principal Components

Analysis (PCA) to statistically group the variable Job, and hence reduce the number of employers who fitted into more than one level of this variable. [1] The conclusion from the PCA below was that only a weak predictive effect of Acceptability could be found from type of work offered.

Table 7.4-1 Work types used in analysis

Type of work	No. Employers offering
Caregiving	10
Cleaning	24
Driving	14
Hairdressing	2
Handling goods	18
Hospitality, customer	15
Kitchen	2
Labouring, removals	14
Nursing[2]	3
Office	4
Retail	37
Security	7
Other	5
TOTAL	**155**

Principal Components Analysis

A Principal Components Analysis (**Appendix D4**) was used to obtain statistical groupings from the variable Job. This resulted in four factors as shown in Table **7.4-2**.

Table 7.4-2 Grouping by Principal Components Analysis

Factor	**Components**
C1	Caregiving, Cleaning, Nursing, Kitchen
C2	Driving, Handling goods, Labouring, Removals, Office
C3	Hairdressing, Hospitality, Customer, Security, Other
C4	Retail

However, some Employers still had responses in more than one category, which might have biased the model towards those responses. A series of four models was therefore used as follows:

1. Only 1 response allowed per Employer per Factor. Thus, if an employer offered both caregiving and cleaning, this counted as 1 under C1.

---

[1] Other methods of analysing Job were attempted but were not successful.
[2] "Nursing" refers to unskilled care of the elderly or similar work.

2. In each of the 4 models in turn, allowing only one response per Employer. Thus in Model 1, which ordered the responses C1-2-3-4, all responses from C1 were kept, but Employers with responses in more than one factor would keep the first response only.

3. The other models were run with each Factor favoured in turn.

Factor C3 was the reference level as it had the highest number of data items (224). Summaries of these models are listed below. Of these models, all p-values for Job variables C1, C2 and C4 in models 2 and 3 were greater than 0.05, and the result of all the models run was that little evidence could be found that Job was a predictor of acceptability.

Model 1 (Factor order C1-2-3-4): only JobC3 had $p<0.05$, but comparison of the 3-Factor Model with Model 1 showed no statistical difference and the 3-Factor Model was preferred ($p=0.063$).

Model 2 (Factor order C2-3-4-1): Comparison of the 3-Factor Model with Model 2 found no statistical difference so the 3-Factor Model was preferred (ANOVA $p=0.197$).

Model 3 (Factor order C3-4-1-2): Comparison of the 3-Factor Model with Model 3 found no statistical difference so the 3-Factor Model was preferred (ANOVA $p=0.250$).

Model 4 (Factor order C4-1-2-3): No Job variable had a p-value less than 0.05. Comparison of the 3-Factor Model with Model 4 showed no statistical difference ($p=0.056$).

None of the models rejected the null hypothesis so the hypothesis was rejected.

The Employers who took the Employer Survey were from industries likely to employ migrants with a low level of English, and the work was low skilled or unskilled. It is therefore probable that there was insufficient difference between the levels of English required for these kinds of work to have a significant effect.

7.4.2   Non-native Speaking Employers

Hypothesis:

Employers who are non-native speakers of English will give lower Acceptability ratings than those who are native speakers.

Various reports have suggested that NNS rate the speech of other NNS more harshly than NS (Fayer & Krasinski, 1987; Rossiter, 2009) although Lima (2012) found little difference. Munro, Derwing and Morton (2006) found that listeners' responses to NNS pronunciation were not heavily influenced by their L1.

Employers were asked

> *What is your native language?*
>
> 1. English
> 2. Other _____

Out of 95 Employers, 10 (10.5%) were NNS, as shown in Table **7.4-3**.

Table 7.4-3 Non-native speaking Employers

Origin	No.
Chinese	1
Eastern European	1
Middle Eastern	2
Pacific	1
South Asian	1
South East Asian	3
Western European	1
TOTAL	**10**

With such small numbers, and diverse backgrounds, it was highly unlikely that any effects from specific languages or regions would be noticeable. The respondents were therefore assigned to two levels (English/Other). Comparison of Employer Language using this new variable with the 3-Factor Model found no statistical difference so the 3-Factor Model was preferred (ANOVA, p=0.184).

7.4.3   Length of Residence

Hypothesis:

> *Employers' length of residence in New Zealand will affect predictions of acceptability.*

Data: **Appendix D7**

This question was a further check on the question reported in **7.4.2** *Non-native Speaking Employers*. Since accent itself is subject to attitudes and experience, it is possible that length

of residence may affect attitudes to other NNS. If either hypothesis were supported, interactions between these responses could provide further data.

Employers were asked

How long have you been in New Zealand?

> *(1) I was born here*
>
> *(2) Less than 2 years*
>
> *(3) 2 years or more*

Only one employer responded with (2), so responses were recategorised to two levels: NZ (born in NZ) and Overseas (not born in NZ). 28 Employers (29.4%) were born outside of New Zealand.

ANOVA found no statistical difference between the models, so the 3-Factor Model was preferred to one including Residence as a factor (p=0.340), and the hypothesis was not supported: Length of Residence does not significantly affect Acceptability.

7.4.4   Employee Use of English

Hypothesis:

> *Employers' attitudes towards when and how often employees ought to use English will affect their ratings of acceptability.*

Data: **Appendix D8**

Section **2.3** discussed the problems faced by migrants looking for work because of employers' attitudes towards foreign accents. If employers are antagonistic towards people who continue to use their own language, this may be reflected in their rating of Acceptability.

Employers were asked

> *Which of the following best describes your view about migrant employees using their own language?*
> 1. *They can use their own language whenever they like.*
> 2. *They must use English with co-workers and customers.*
> 3. *They must use English at work when it is appropriate*
> 4. *They must use English while actually on the job.*

5. *They must use English at all times during work and breaks.*
6. *They should learn English and use it all the time.*

I have often encountered anecdotal statements that migrants ought to speak English at home and give up their own language, and Portes (2002, p. 10) reports that the powerful English-only movement in the USA seeks "linguistic assimilation of foreigners". While there is no strong movement of this type in New Zealand, it is possible that individual attitudes to migrants' use of other languages may make some employers more judgemental about those with low English proficiency. The question presented six choices and respondents could se-lect only one in order to force the responses into analysable categories.

Responses to the questions are in Table **7.4-4,** and there is also a comparison of NS and NNS Employers in Figure **7.4-1** showing that there is little distinction between the two groups.

Table 7.4-4 Responses: Use of own language and English

Question	Number	%
Use own language whenever they like	3	3.2%
Use English with co-workers and customers	49	51.6%
Use English when appropriate	13	13.7%
Always use English when working	4	4.2%
Always use English at work and in breaks	6	6.3%
Use English all the time	20	21.0%
TOTAL	**95**	100%

Figure 7.4-1 Percentage comparison of NS & NNS response to English use
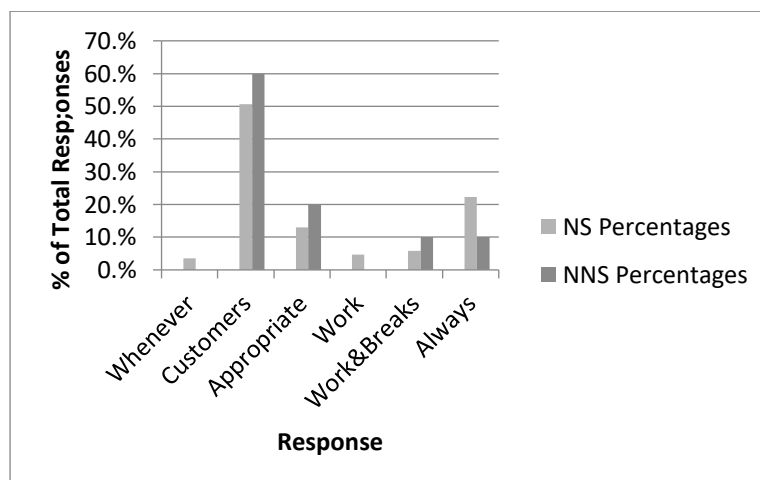
(NS=85; NNS=10). It is worth noting that a fifth of the Employers believed that migrants should always use only English, which reinforces some of the concerns noted above.

Model **7.4-1** (below) tested the variable "Employee Use of English" against the 3-Factor Model.

Model 7.4-1 Employee Use of English

```
                        Estimate Std. Error z value Pr(>|z|)
UseEnglishAlways         -0.17152    0.48412  -0.354 0.723114
UseEnglishAppropriate    -0.09119    0.56820  -0.160 0.872497
UseEnglishWhenever       -0.69341    1.09155  -0.635 0.525263
UseEnglishWork            0.06107    0.95987   0.064 0.949267
UseEnglishWork&Breaks    -1.71282    0.82675  -2.072 0.038289 *
Prosody                   0.66507    0.19564   3.399 0.000675 ***
Fluency                   0.85372    0.20633   4.138 3.51e-05 ***
Segmental                 0.67410    0.21458   3.141 0.001681 **
```
Reference level: Customers (highest response)

Apart from the 3 Factors, one variable, "use English in Work and Breaks", was significant (p<0.05) and lower than the reference level. However, comparison of the 3-Factor Model including Employee Use of English with the same Model without this variable found no statistical difference so the 3-Factor Model was preferred (ANOVA, p=0.470). The hypothesis was therefore not supported.

7.4.5   Employee Pronunciation Problems

Hypothesis:

> *Having had an employee whose pronunciation caused problems will negatively affect predictions of Acceptability.*

Section **2.3.3** described the resistance to hiring migrants because of their accents. It is likely that problems with migrant employees would reinforce any stereotypical attitudes.

Employers were asked

> *Have you ever employed people who are not native speakers of English?*
>
> > *Yes (1) No (2)*
>
> *[If Yes] Have any of your non-native speakers had pronunciation problems which affected their work? Yes (1) No (2) (Describe problem)*

Only 4 employers had never employed NNSs and 28 reported no pronunciation problems with the migrants they had employed, so I recategorised all of these as "no" since the hypothesis referred to having an employee with pronunciation problems. The question was therefore recategorised in 2 levels, with 63 and 32 employers respectively:

      1=Yes (stated that they had had problems)

      2=No (stated that they had not had problems)

The variable "Pronunciation Problems" was then tested against the 3-Factor Model. Comparison of the 3-Factor Model including Pronunciation Problems with the 3-Factor Model without this variable found no statistical difference so the latter was preferred (ANOVA, p=0.188). The hypothesis was therefore not supported.

Qualitative Data

Employers were given the opportunity to record any problems they had observed, and these are listed in **Appendix E1.** There was a range of comments, and I have put some words in bold to show common topics.

Many specified the difficulty with customers:

> *Some of my **clients** are not always supportive of staff who have some difficulty with English.*

> *Elderly deaf and dementia **residents** can have difficulty understanding the accents.*

Only a few specified the pronunciation problem:

> *Difficulty in understanding what they are trying to pronounce, ie the sound they make is not what I expect the word they are trying to say to sound like especially with v's and w's*

Which can of course be amusing:

> *Some accents are too strong - funny case - ice cream in South African accent can sound like arse cream!*

Some described effects on colleagues:

> *One worker gets ignored by **other workers** fairly often because his accent is so bad we can't understand what he is saying. We smile and nod.*

> *The level of English spoken by this person caused problems in communicating tasks and standards required **in their department**. This led to performance issues.*

Others mentioned a solution they were using:

> *Problems when ordering supplies for jobs. I now tell them to text me.*

As there were only 64 varied and brief responses it was not possible to extract any detailed themes, but I found four main employment-related topics amongst their concerns:

- ***Customers*** *at checkout not understanding response from staff member.*

- ***Co-workers*** *not understanding verbal response.*

- ***Telephone conversations*** *where person ringing in cannot fully understand the reply.*

- *Inability to understand* ***instructions***.

Table 7.4-5 Pronunciation Problems: Common Responses

Topics	Customers	Co-workers	Telephone	Instructions
Number of Employers commenting	28	13	15	10
% of Total	44%	20%	23%	16%

As the Employers were not prompted as to the type of problem, we can say that everyday workplace communication and speaking with customers were uppermost in the minds of these Employers. It is not surprising that using the telephone would be a source of difficulty as there is less context and no body language to support the message.

These responses open the possibility of gathering further data to clarify the needs of employers. Section **7.6.1** shows that Acceptability is contextualised to different types of work, so it would also be useful for teachers to understand employers' actual concerns about employee pronunciation and the situations in which they arise.

7.4.6 Sex of Speaker and Employer

Hypotheses:

1. *Employers will rate male and female speakers differently.*

2. *Male and female Employers will rate speakers differently.*

3. *Effects will be found from the interaction of male/female Employers and Speakers.*

Data: **Appendix D2**

Section **5.5** showed that there is evidence that Accentedness ratings can be affected by sex of the speaker. Timming (2016) found "severe discrimination" against foreign accented female job applicants but no effect from the sex of the hiring manager (p.15). All the models below use the full survey, so number of Employers = 95, number of Migrants = 40, and total observations = 424.

Sex of Migrants

Migrants were asked to record Male or Female on their consent forms (**Appendix A1**). The mean raw Acceptability rating by all employers on the 1-9 scale was 5.77 (SD=2.11) for females and 5.15 (SD=2.29) for males, suggesting at least that there was no overall bias against the female Migrants. A model was run adding sex of the migrant as a predictor of Acceptability, but no statistical difference was found between this and the 3-Factor Model so the latter was preferred (ANOVA, p=0.296). The hypothesis that the sex of the Migrant would predict Acceptability was therefore not supported.

Sex of Employers

Sex of Employers was not one of the questions asked, but since I had contacted all employers personally I was able to note this information. A model was run adding Sex of Employer as a variable, but no statistical difference was found between this and the 3-Factor Model so the latter was preferred (ANOVA, p=0.708).

Further models tested whether both variables (sex of Employers and sex of Migrants) could be included, and whether there was any interaction between them. No significant effects were found and no significant difference was found between the models with and without the interaction (ANOVA, p=0.293). The hypotheses that the sex of the Employers or that an interaction between sex of Employers and Migrants would predict Acceptability were therefore not supported.

While some researchers referred to in **5.5** found gender bias under certain conditions (such as customers rating sales assistants) the criteria for hiring low-skilled employees may well be different. These employers may find that there is no advantage in favouring one type of candidate over another, provided they can do the job.

7.4.7   Language – Actual and Guessed

Hypotheses:

1. *The group of L1s classified under* Language Actual: South East Asian *will have lower Acceptability than others.*
2. *Europeans and Middle Eastern speakers will have higher Acceptability than all others.*

Research has shown that East Asian speakers can be treated more harshly than others (**5.5**), while speakers of European and Middle Eastern languages tend to be more intelligible (**4.3.2**).

Data: **Appendix D1**

Language as a predictor of acceptability

The Migrants were asked to note their native languages on their consent forms (**Appendix A1**) so these could be tested as predictors of acceptability. Employers were not given any information about the ethnicity or language of the speakers, but after hearing the speech sample they were asked *"Which country do you think the migrant comes* from?" These answers were recoded as languages, and I fitted both these and the Migrants' actual languages into groups (as explained in **6.3.2**). Ballard (2013) found that students had a poor ability to identify accents, and with the exception of accents likely to be familiar to New Zealanders (e.g. South Asian), Employers in the present study were not expected to make many correct identifications of the speakers' origins. Table **7.4-6** shows how many actual speakers of each language there were, and how many times each language / origin was guessed.

"Language Guess" used more groups than "Language Actual" as the Employers could guess whatever they liked, so for example Language Guess "Korean" and "Eastern European" appeared although no speakers matched these descriptions, and "Asian" was included as a category often used by Employers but was too wide to be matched with any actual grouping when judging accuracy of the guesses. Minor groups and "don't know" with no attempt at a guess were recategorised as "Other".

Table 7.4-6 Groupings of Actual Migrant Language and Guessed by Employers

Region	Short name	Actual and Guessed Origins included	Actual	Guess
Pacific	Pacific	Polynesian, Tetum, Fijian	3	13
South East Asian	SEAsian	Vietnamese, Cambodian, Thai, Fili-pino, Myanmarese	9	67
Chinese	Chinese	Chinese	2	41
Japanese	Japanese	Japanese	3	10
Korean	Korean	Korean	0	8
South Asian	SAsian	Indian, Sri Lankan, Bengali	5	41
African	African	Non-Arabic: Somali, Amharic	5	69
Middle Eastern	Mid-Eastern	Arabic, Assyrian, Farsi, Turkish, Israeli	5	35
Eastern European	EEuropean	Russian	0	41
Western Euro-pean	WEuro-pean	European, Greek, Italian; Spain, Portu-guese	3	32
South American	SAmerican	Spanish and Portuguese (Latin Ameri-can)	5	18
Other	Other	South African, unknown	0	37
Asian	Asian	Only if "Asian" is the only response	0	11
		TOTAL	40	424

A model was run adding Language Actual to the 3-Factor Model, but no statistical difference was found so the latter was preferred (ANOVA, p=0.723). A further model adding Language Guess to the 3-Factor Model also found no statistical difference (ANOVA, p=0.210). Hypotheses 1 and 2 required evidence that the speakers' languages would predict Acceptability to some extent. The two hypotheses are therefore not supported, and there is no evidence that the language of the speakers (actual or guessed) had any influence on Employer ratings of Acceptability.

Accuracy of Employer Guesses

Accuracy of Language Guess was tested through a manual observation of the spreadsheet, comparing raw ratings of Language Guess with Language Actual. Employers guessed the correct region or origin for 22.6% of 424 observations. The percentages of correct guesses per language group are shown in Table **7.4-7**. South Asian (32.7%) and Chinese (37.6%) were most commonly recognised, which supports the assumption made at the start of this section as people from China and India are amongst the most common NNS migrants to New Zealand. Few recognised Japanese, which was interpreted with a range of different guesses.

Table 7.4-7 Percentage of each language group guessed correctly

Language Group	% Correct Guesses
African	22.6%
Chinese	37.6%
Japanese	6.3%
Middle Eastern	18.4%
Pacific	13.5%
South American	13.6%
South Asian	32.7%
South East Asian	27.2%
Western European	27.2%

An interesting effect was found in the Language Guess model. *South East Asian*, which had a high number of actual speakers (9) and guesses (67) was the reference level, and apart from the 3 Factors there was only one significant effect: the guess *Korean* (7 speakers, one of them on 3 occasions) had by far the highest estimate (p<0.05) although (as noted above) there were no actual speakers. To investigate this suggestion further, the speakers guessed to be Korean were checked for proficiency using the raw figures for global measures as assessed in the Base rating. Most of the "Koreans" actually had higher than average ratings (105%-109% of average Global Measure ratings), which suggests that that the Employers had a positive view of Korean speakers and therefore assumed that good speakers were more likely to be Korean.

7.5   EMPLOYER SURVEYS: PARTS A AND B

Once Employers had entered the Employer Data analysed in **7.4** they were randomly assigned to Part A or Part B of the Survey. They would click to hear a migrant's responses in a mock job interview which they were allowed to hear once. For each of 5 random samples, all Employers were then asked about the Migrant's acceptability for a job (based on pronunciation alone) and guess where they came from, before being asked specific questions about each of the samples.

The speakers (N=40) were the same for both parts, but the Employers were different. 45 Employers answered Part A, with 209 observations, and 50 Employers answered Part B, giving 215 observations. At 5 observations per Employer these should have been 225 and 250 respectively, but some Employers failed to complete the whole set or were prevented by technical problems.

Section **6.4** gives the completion rates and Figure **7.3-2** (in section **7.3.1**) shows histograms comparing the two parts. The 3-Factor Models below (**7.5-2** and **7.5-3**) which test variables from Parts A and B are therefore subsets of the full model.

Model 7.5-1 Prediction of Acceptability by the Base Rating (whole survey)

```
          Estimate Std. Error z value Pr(>|z|)  Odds
Prosody   0.663    0.197      3.361   p<0.001   1.94
Fluency   0.865    0.208      4.162   p<0.0001  2.38
Segmental 0.683    0.216      3.161   p<0.01    1.98
```

Model 7.5-2 Acceptability as predicted by the Base Rating (survey Part A)

```
          Estimate Std. Error z value Pr(>|z|)  Odds
Prosody   0.597    0.212      2.826   p<0.01    1.82
Fluency   0.807    0.222      3.636   p<0.001   2.24
Segmental 0.787    0.237      3.316   p<0.001   2.17
```

Model 7.5-3 Acceptability as predicted by the Base Rating (survey Part B)

```
          Estimate Std. Error z value  Pr(>|z|)  Odds
Prosody   0.765    0.250      3.057    p<0.01    2.15
Fluency   0.965    0.267      3.614    p<0.001   2.62
Segmental 0.599    0.271      2.209    p<0.05    1.82
```

The relative importance of the 3 Factors varies in these three models. Noting from the p-values that all results are significant, we can take from the Odds (the likelihood that a higher Base Rating for the Factor will increase Acceptability) that the order of effects for the whole survey is as follows (the underlining indicates that all the values of the two Factors are clearly closer to each other than to the third):

Whole survey:	Fluency	<u>Segmental</u>	<u>Prosody</u>
Part A:	<u>Fluency</u>	<u>Segmental</u>	Prosody
Part B:	Fluency	Prosody	Segmental

This can be seen in Figures **7.5-1** and **7.5-2** where the bars for Prosody and Segmental are reversed in Parts A and B.

Figure 7.5-1 Acceptability: Whole Survey
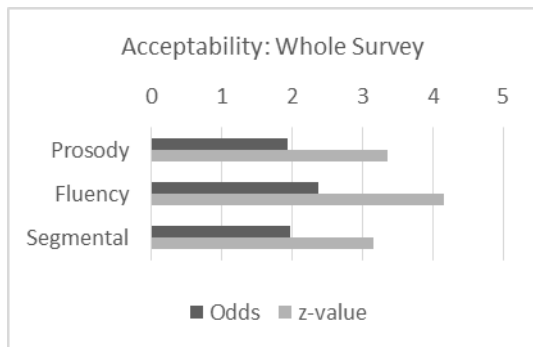
Figure 7.5-2 Acceptability: Parts A and B
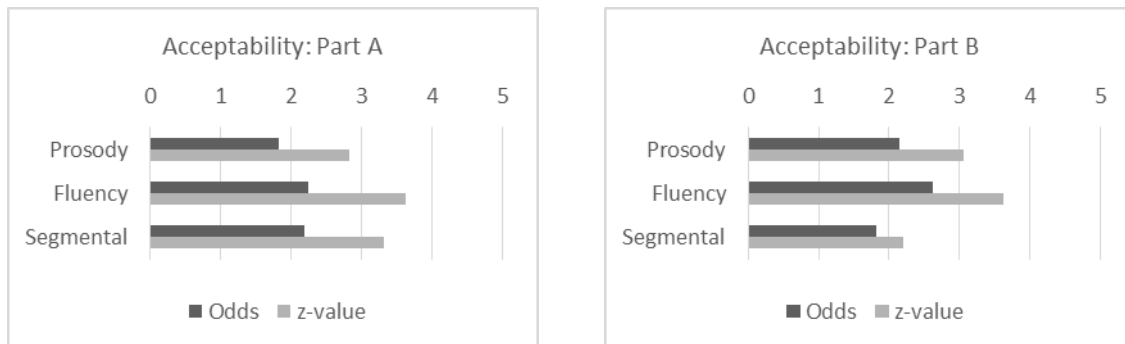
Fluency is the strongest predictor and had the highest z-value throughout the survey, but the relative strength of Prosody and Segmental differs considerably in the two Parts. The Employers were randomly directed to one part or another, so it is unlikely that their composition was markedly different: Employer details for parts A and B are given in Table **7.5-1**. More Employers in Part A were born overseas, but the same number were NNS, and none of these factors predicted Acceptability in the survey as a whole.

In order to test the effect of taking either Part A or Part B, a model was run adding Survey as an IV. This was compared with the 3-Factor Model but no statistical difference was found so the model without Survey was preferred (ANOVA, p=0.760). An interaction between Survey

and the 3 Factors was also tested, and none of the interactions were significant. We can therefore be confident that the allocation to Survey A or B did not in itself affect predictions of Acceptability.

Table 7.5-1 Composition of Employers in Parts A and B

Employer Data	Part A	Part B
No. Employers	**45**	**50**
Born outside NZ	18 (40%)	11 (22%)
Origin of NNS	1 x Chinese, Pacific, 2 x Mid Eastern; SE Asian	1 x E European, S Asian, W European; 2 x SE Asian
Sex	F 42% M 58%	F 46% M 54%
Work offered:		
Caregiving	11%	10%
Cleaning	27%	26%
Driving	11%	18%
Hairdressing	4%	0%
Handling goods	22%	20%
Labouring, removals.	11%	18%
Retail	44%	36%
Hospitality/customer	22%	20%
Security	7%	8%
Nursing	2%	4%
Office	4%	6%
Kitchen	4%	0%
Other	7%	4%

7.6   EMPLOYER SURVEY PART A

Respondents to Part A were asked three groups of questions.

The first group asked the Employers to place Acceptability in three communication contexts in order to test whether their responses would differ (**7.6.1**).

Secondly, Employers were asked to make global assessments of the Migrants' speech (**7.6.2**).

In the third group, semantic differential questions were used to test whether the Employers had formed any judgements about the personal qualities of the Migrants, and whether this had any effect on Acceptability ratings (**7.6.3**).

Conclusions from Part A contribute to our understanding of Employer responses but cannot be incorporated into the full 3-Factor Model since the Employers in Parts A and B were separate groups.

The 3-Factor Model for the full Employer Survey, and for each of the two parts, is given in **7.5**.

7.6.1   Supervisors, Co-workers and Customers

Hypotheses:

1. *As Acceptability is context-based, Employers will rate the Migrants differently when asked to consider different communication situations.*

2. *Employers will demand higher standards of pronunciation for Acceptability when their potential employees must communicate with customers.*

Employers responding to Part A (n=45) were asked the following questions:

> *How acceptable is the migrant's pronunciation for communication with the following people?*
>
> - *Their supervisor?*
> - *Other employees?*
> - *Customers?*

These terms were not defined further, but I assumed that anyone in a human resources or employer role in these industries would interpret "their supervisor" to mean the immediate supervisor of the employee, "other employees" to mean co-workers on a similar level, and

"customers" as any people making use of the services of the business. Contextual effects have been found in other studies. De la Zerda and Hopper (1979) and other studies found that accent is a less important predictor of Acceptability for low-status positions than for high-status ones. Timming (2016) used a binary context distinction, presence or absence of customers, and found that foreign accents were scored significantly lower than those of white British or Americans for customer-facing jobs. Timming concluded that managers will hire on the basis of what they think customers will think. This is supported in this thesis by the results to the above questions.

The raw responses to the three questions appear in Figure **7.6-1**, showing a similarity between Supervisors and Co-workers in the number of responses for each of the scales, and a different distribution for Customers with many more responses at the lower end. These charts can be compared with the distribution of Employer responses in Part A of the Employer Survey in Figure **7.3-2**.

Figure 7.6-1 Distribution of Employer responses to Supervisors / Employees / Customers
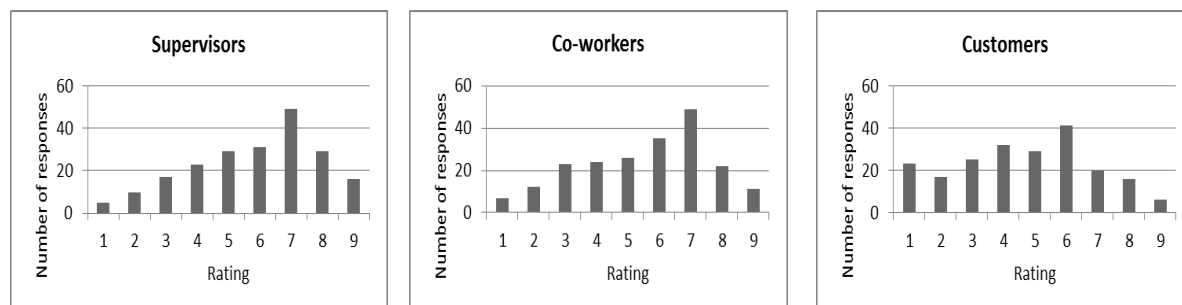
Models **7.6-1** to **7.6-3** below test the 3 Factors as predictors of judgements of acceptability for communication with Supervisors, Co-workers and Customers. These are, in effect, more specific contexts of Acceptability. All 3 Factors remain significant in these three contexts (p<0.01), with the z-value of Fluency always the highest. Fluency has a higher Estimate than the other Factors in Supervisors and is 0.907 higher than Prosody, but all 3 Factors are only 0.159 apart in their estimates for Co-workers, and even closer (0.89 apart) for Customers. However, the ranking of Estimates for Customers is different from the other two. Supervisors and Co-workers have the same order of factors (Fluency – Segmental – Prosody) as the 3-Factor Model and Comprehensibility (Model **7.6-6**), while the order for Customers is Segmental – Fluency – Prosody, the same order as for Intelligibility (Model **7.6-5**).

Model 7.6-1 Supervisors

```
           Estimate Std. Error z value Pr(>|z|) Odds
Prosody    0.685      0.233    2.936  p<0.01   1.98
Fluency    0.972      0.251    3.879  p<0.001  2.64
Segmental  0.740      0.256    2.895  p<0.01   2.10
```

Model 7.6-2 Co-workers

```
           Estimate Std. Error z value Pr(>|z|) Odds
Prosody    0.695      0.231    3.014  p<0.01   2.00
Fluency    0.854      0.244    3.502  p<0.001  2.35
Segmental  0.808      0.254    3.181  p<0.01   2.24
```

Model 7.6-3 Customers

```
           Estimate Std. Error z value Pr(>|z|) Odds
Prosody    0.761      0.219    3.482  p<0.001  2.14
Fluency    0.843      0.229    3.676  p<0.001  2.32
Segmental  0.850      0.240    3.535  p<0.001  2.34
```

Figure 7.6-2 Z-values and Odds for Acceptability and Subsets
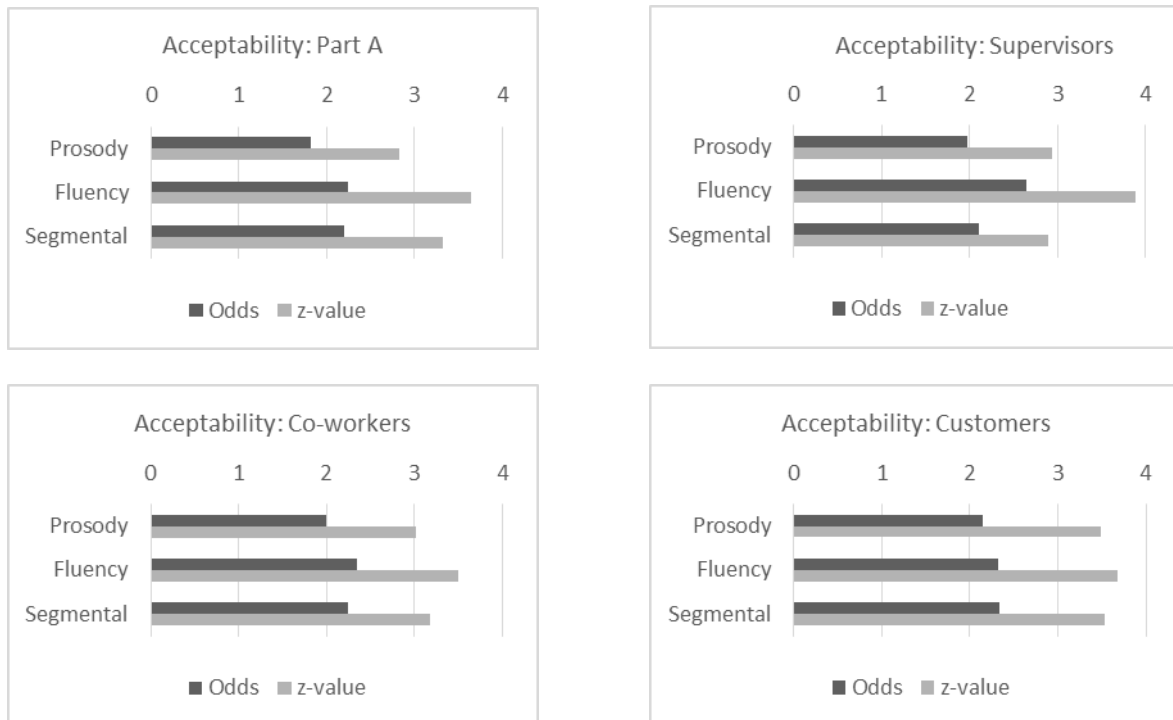
Suitability

To compare the contexts with one another, a model was created using the 3 Factors from the Base rating, a new DV *Suitability* formed by combining acceptability scores from the three contexts, plus another new IV *Context* with three levels: Supervisors, Co-workers and Customers. Comparison of models with and without Context showed that the model including Context was a significantly better fit (ANOVA, χ^2 (2)=98.043, p<0.0001). Further models

testing the interaction between Context and each of the three Factors in turn were not significantly different from the non-interactive model (ANOVA, p=0.398, 0.884 and 0.156 respectively) so the latter was preferred. The final non-interactive model appears as Model **7.6-4** below. This model shows that all variables are strong predictors of Suitability. With Customers as the reference level (lowest values), the estimate is, as expected, higher for Co-workers and higher still for Supervisors, and the z-values follow the same pattern. In this model, the Context variables are stronger predictors of Suitability than the 3 Factors in both z-values and Odds.

Model 7.6-4 Context + Base Rating predicting Suitability

```
Coefficients:
                   Estimate Std. Error z value Pr(>|z|)  Odds
ContextCo-Workers    1.336     0.186    7.185   p<0.0001  3.81
ContextSupervisors   1.790     0.192    9.326   p<0.0001  5.99
Prosody              0.820     0.248    3.300   p<0.001   2.27
Fluency              1.013     0.261    3.878   p<0.0001  2.75
Segmental            0.873     0.271    3.220   p<0.001   2.39
(Reference level: Customers)
```

The overall mean raw Acceptability rating by Employers who completed Part A of the survey was 5.46 (SD 2.23). However, in the three contexts discussed here, the mean score for Supervisors was 5.83 (SD 2.036); for Co-workers it was 5.52 (SD 2.062); and for Customers it was 4.66 (SD 2.181). This suggests that when asked about communication with supervisors and co-workers, the employer is thinking of a supportive context – one in which allowances can be made, further explanation can be offered, and poorer communication skills can be tolerated. However, when Employers are asked to consider the context of dealing with customers, the average Acceptability ratings fall. There is more risk in employing someone with poor pronunciation in this context.

The results show that Employers require different standards of pronunciation when asked to rate acceptability for specific communication contexts, so both hypotheses are supported. This also provides further evidence for Acceptability as a Global Measure which needs to be considered in pronunciation teaching: speech cannot be effectively developed in isolation because its intended context must be taken into account.

7.6.2 Global Measures

Hypothesis:

> *Responses to Acceptability will have a partial positive correlation with Intelligibility, Comprehensibility and Accentedness.*

Derwing and Munro (2015, p. 5) describe the "partial independence" of Accentedness, Intelligibility and Comprehensibility. If Acceptability is a Global Measure, then it too should correlate with these three measures. Of course, a very close correlation with any of them would mean that Acceptability is not an independent measure.

Both Employers (Survey Part A, n=45) and Experienced Raters (n=6) were asked to rate the Migrants (n=40) on the three Global Measures, using 1-9 ordinal scales for the questions:

What is this migrant's accent like?

> *How much of the sample did you understand? (1=nothing, 9=all)*
> *How easy was it to follow what they were saying? (1=very difficult, 9=very easy)*
> *How strong was their accent? [1= very strongly accented, 9=very little accent]*

While the three questions ask for ratings of Intelligibility, Comprehensibility and Accentedness respectively, it was impracticable to measure Intelligibility using transcription or comprehension tasks as described by Derwing and Munro (2014a) and the responses are therefore impressionistic rather than objective. Nevertheless, the Employers' responses to Intelligibility (Figure **7.6-4**) are very different from those to Comprehensibility (Figure **7.6-7**) or Accentedness (Figure **7.6-9**). This is also shown in Table **7.6-1**.

This section begins with predictions of Employer ratings of the three Global Measures by the 3-Factor Model, and then compares Employer ratings of the three Global Measures with the average of all 6 Experienced Raters (3 ratings per sample). As reported at the end of **7.2.1**, adequate Experienced Rater ICC scores were obtained for the three Global Measures alone: with all 6 raters the range was 0.493 to 0.793, with $p<0.0001$ for all scores, so averages taken from these ratings were usable for the comparison of Employer ratings of Global Measures with those of all 6 Experts. To avoid confusion, ratings using all the Experienced Raters are described below as the "Full Base rating".

Predictors of Employer Ratings of the Global Measures

The models below test to what extent the 3 Factors predict Employer ratings of each of the three Global Measures. They demonstrate that these Global Measures are strongly predicted by the 3 Factors and all responses are significant.

Intelligibility

Model 7.6-5 Employer Intelligibility ratings predicted by the 3 Factors

```
           Estimate Std. Error z value Pr(>|z|) Odds
   Prosody   0.512      0.225   2.280   p<0.05   1.67
   Fluency   0.614      0.232   2.642   p<0.01   1.85
   Segmental 0.739      0.249   2.969   p<0.01   2.09
```

(N=45, observations=209. Raw mean=6.58, SD=2.15.) The order of odds and z-values here is Segmental-Fluency-Prosody. Research on predictors of intelligibility have identified the importance of segmentals: see **3.2.4**. Even with the shift towards teaching Prosody, Segmentals clearly remain an important aspect of intelligible speech.

Comprehensibility

Model 7.6-6 Employer Comprehensibility ratings predicted by the 3 Factors

```
           Estimate Std. Error z value  Pr(>|z|) Odds
   Prosody    0.582      0.247   2.357  p<0.05   1.79
   Fluency    0.917      0.263   3.479  p<0.001  2.50
   Segmental  0.645      0.273   2.362  p<0.05   1.90
```

(N=45, observations=209. Raw mean=5.51, SD=2.26.) The order of odds and z-values here is Fluency-Segmental-Prosody – although the last two are close. Fluency is clearly the strongest predictor of Employer ratings of Comprehensibility, as for Acceptability, but it is interesting that this should contrast so strongly with the ratings of Intelligibility. The Employers were able to make a consistent distinction between understanding and the difficulty of understanding, and the Fluency Factor was the most important influence.

Comprehensibility is also the Global Measure most strongly correlated with Acceptability (see Table **7.6-2**). Researchers have identified a wide range of features which predict comprehensibility (**3.2.5**), so rather than asking which features predict comprehensibility it may be more useful to ask whether there are features of the research itself, such as the context of communication, which may emphasise some features over others.

Accentedness

Model 7.6-7 Employer Accentedness ratings predicted by the 3 Factors

```
          Estimate Std. Error z value Pr(>|z|) Odds
Prosody     0.608     0.235    2.585   p<0.01   1.84
Fluency     0.580     0.246    2.358   p<0.05   1.79
Segmental   0.750     0.264    2.840   p<0.01   2.12
```

(N=45, observations=209. Raw mean=3.64, SD=1.88.) The order of odds and z-values here is Segmental-Prosody-Fluency. As was noted in **3.2.3**, ratings of Accentedness are predicted by the speaker's own accent, although pitch range and word stress are important, and the overall proficiency of speakers affects the rating.

Figure **7.6-3** illustrates the differing degrees of prediction for each of the four Global Measures by the 3 Factors. Acceptability is most strongly predicted, and most resembles Comprehensibility in the high significance of Fluency, which is least important for Accentedness.

Figure 7.6-3 Four Global Measures as predicted by the Base rating
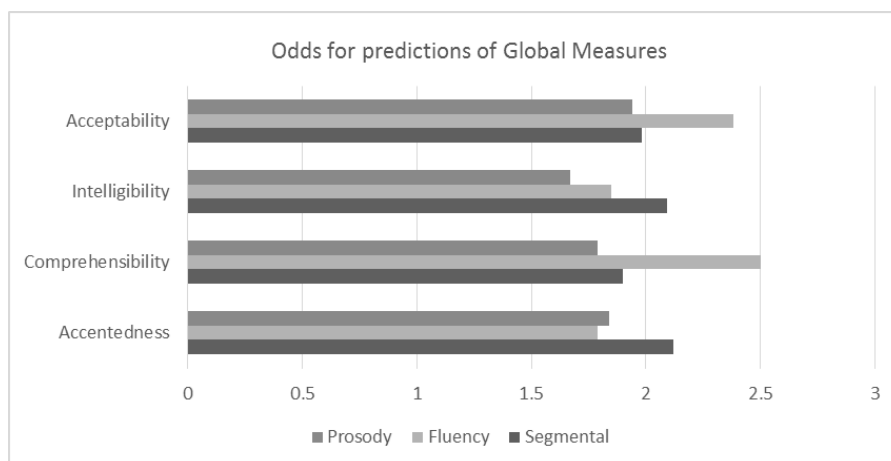
All of the 3 Factors are significant and strong predictors of these four Global Measures. That is, the greater the Migrants' proficiency at any level of pronunciation, the more likely it is that Employers will give a higher rating for the Global Measures. Fluency has the greatest impact on ratings of both Comprehensibility and Acceptability. It is the strongest predictor in many of the other models described in this chapter, and suggests that it is worthwhile for learners of English to spend time developing the features usually associated with Fluency. The Segmental Factor has the strongest effect on Intelligibility (as identified above) and Accentedness.

While there are differences, none of the 3 Factors can be discounted as unimportant. It appears that all features of pronunciation are implicated in judgements which native speakers make of non-native speakers.

Comparison of responses from Experienced Raters and Employers

Munro (2008, p. 200) commented that "phonetically trained evaluators do not necessarily respond to L2 speech in the same way as unsophisticated listeners", and an analysis by Trofimovich and Isaacs (2012) showed that novice and experienced raters focus on different aspects of L2 phonology when rating speakers. In order to determine the difference between the Experienced Rater and Employer responses to the Global Measures, a comparison was made between the Employer ratings (n=45, observations=209) and an average based on all 6 Experienced Raters (observations=120) to create the Full Base Rating as described above. Each average consists of my rating plus two others; since I rated every Migrant while the other Experienced Raters scored 16 each, the overall result is more affected by my ratings. The Experienced Raters' averages were rounded to integers to make it possible to create the histograms, which show the average number of responses to each of the 1-9 scores. The Employer histograms are based on all observations. It should be remembered that the Employer ratings were based on a single listening, while Experienced Raters could listen as many times as they wished.

Intelligibility

Figure 7.6-1 Distributions of responses to Intelligibility: 6 Experienced Raters compared to Employers
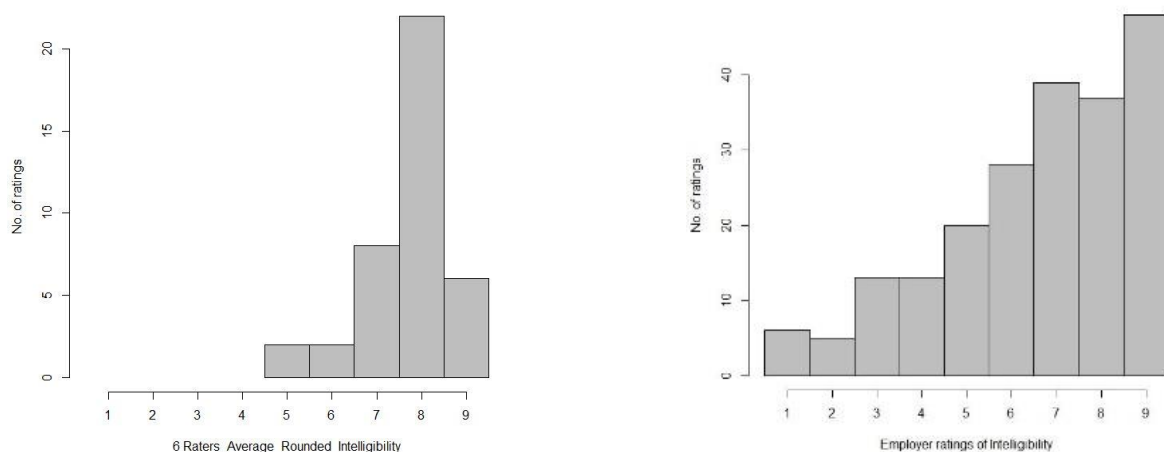
(1=unintelligible, 9=fully intelligible)

The raw mean for the original 5 raters was 8.11, SD 1.17. For 6 raters the mean was 7.74, SD 1.35. Figure **7.6-4** shows that the Full Base Rating creates a limited range of responses with no average under 5 and most around 8, while the Employers used the full scale. This may be the result of the role of familiarity when professionals rate NNS, but it is interesting that Employers put many more speakers at the high end. The Pearson correlation between the Full Expert and Employer ratings is 0.445 (p<0.0001), showing a fair degree of correlation.

Given that the current goal for pronunciation teaching is Intelligibility, it is worth asking how well this outcome also correlates with Acceptability for employment. Figure **7.6-5** below shows the correlation between Employers' ratings of Acceptability with their own ratings of Intelligibility. The jitter plot shows that Acceptability ratings can in some cases be higher than Intelligibility (thus, applicants with an Intelligibility rating of only 4 can get an Acceptability rating of 6) but a given Intelligibility score also correlates with Acceptability scores ranging down to 1. Thus, an Intelligibility score of 8 is found with similar Acceptability scores, but there are also many who are far less acceptable. Briefly, considering only pronunciation, Intelligibility doesn't necessarily get you a job.

Figure 7.6-5 Employer ratings of Acceptability / Intelligibility
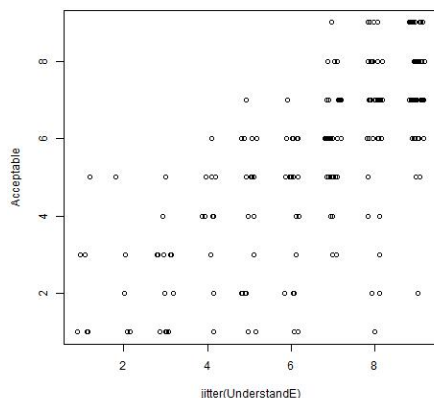
Comprehensibility

The raw mean for the original 5 raters was 7.24, SD 1.54. For 6 raters the mean was 6.88, SD 1.67. Figure **7.6-6** shows distribution of accent ratings of 10 proficient speakers of Chinese by 18 listeners from Munro and Derwing (1995a, p. 82). The histogram has been flipped horizontally so the trend can be compared with the result from Employer Survey Part A in the

second chart of Figure **7.6-7** (although Derwing and Munro were reporting on frequency, not count of responses).

The rating frequency increases steadily towards the highest end. The numerical result from Employer Survey A in Figure **7.6-7** (1=very hard to understand, 9=very easy to understand) is less smooth and drops at the highest level, but follows the same slope.

Figure 7.6-6 Distribution of listener comprehensibility ratings (Derwing and Munro 1995a)
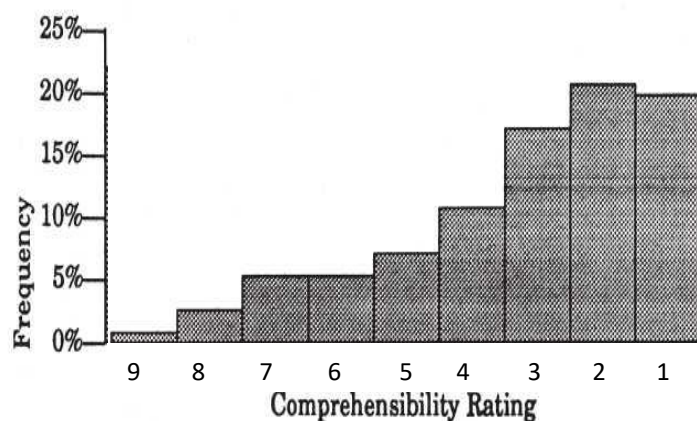
Printed with permission of the publishers

Figure 7.6-7 Distributions of Reponses to Comprehensibility: 6 Experienced Raters compared to Employers
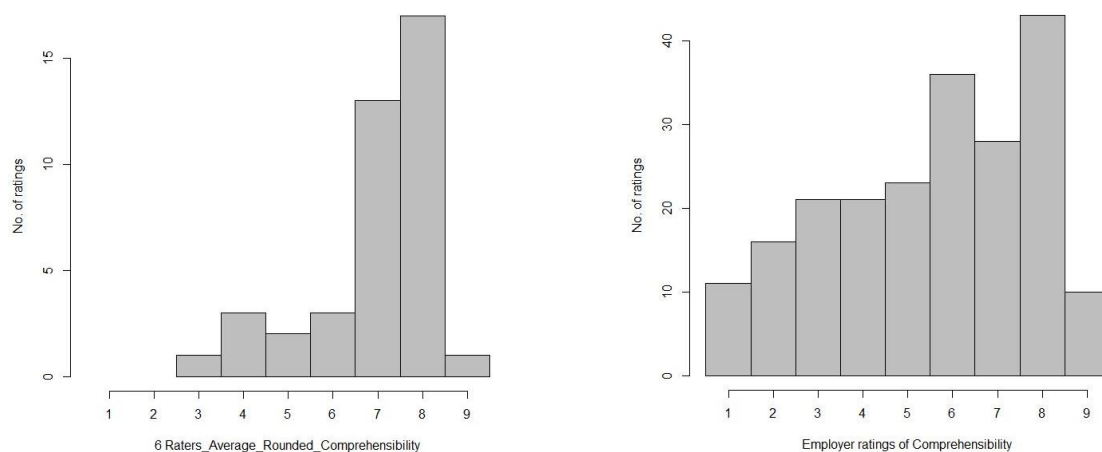
The Full Base Rating results show some spread, with a number of speakers in the hard-to-follow range, none under 3, and most at 7-8. This is less smooth than other charts of Comprehensibility (Munro & Derwing, 1995a, p. 87). The Employer chart is similar, but shows more steady increase towards the "high comprehensibility" end. The Pearson correlation between the Full Expert and Employer ratings is 0.504 (p<0.0001).

Accentedness

The raw mean for the original 5 raters was 5.49, SD 1.40. For 6 raters the mean was 5.25, SD 1.51**.** Figure **7.6-8** shows the distribution of accent ratings of 48 speakers by 26 listeners in Derwing and Munro (1997, p. 7). The histogram has been flipped horizontally so it can be compared with the almost identical results from Employer Survey Part A in Figure **7.6-9** (although Derwing and Munro were reporting on frequency, not count of responses).

Figure 7.6-8 Distributions of Reponses to Accentedness (Derwing and Munro 1997)
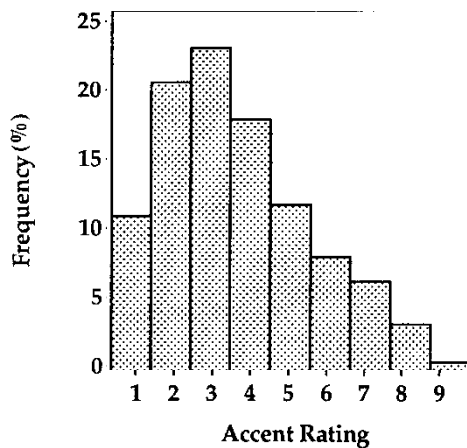
Printed with permission of the publishers

Figure 7.6-9 Distributions of Reponses to Accentedness: 6 Experienced Raters compared to Employers (Survey Part A)
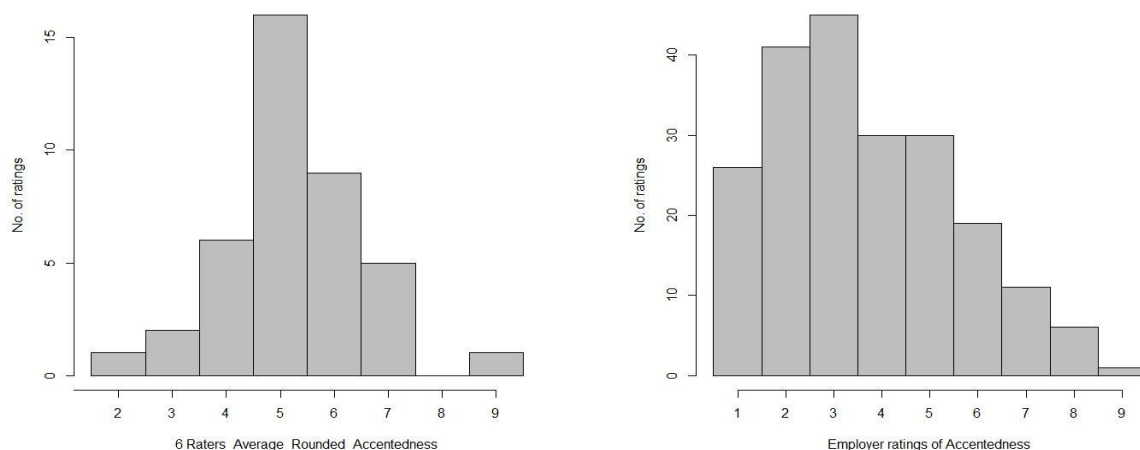
(1=very highly accented, 9=very little accent.) The Full Base Rating is closer to a bell curve, with no average responses for 1 or 8, most responses at 4-6 and few at the extremes. Employers, on the other hand, found most of the speakers highly accented. The Pearson correlation between the Full Expert and Employer ratings is 0.418 (p<0.0001).

These results show that the full group of Experienced Raters and the Employers responded to the Migrants' speech in somewhat different ways, but there is nevertheless some similarity and correlation between the ratings.

Accentedness and Comprehensibility ratings: Comparison

Munro and Derwing (1995a, p. 92) reported that "a strong foreign accent does not necessarily cause L2 speech to be low in comprehensibility or intelligibility." Figure **7.6-10** below, from Derwing and Munro (2009b, p. 480), shows percentages from raw Accentedness and Comprehensibility ratings assigned by a group of listeners to a set of utterances.

Figure 7.6-2 Raw percentages for Accentedness & Comprehensibility (from Derwing and Munro, 2009b)
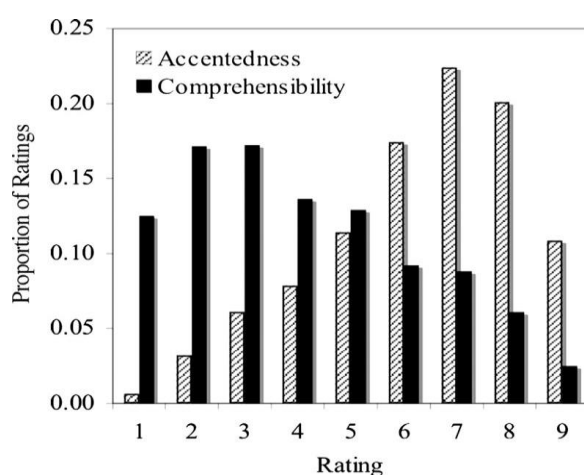
Printed with permission of the publishers

Figure **7.6-11** is from Part A of the Employer Survey, and shows very similar opposing slopes.

Figure 7.6-11 Raw percentages for Employer ratings of Accentedness & Comprehensibility (reversed)
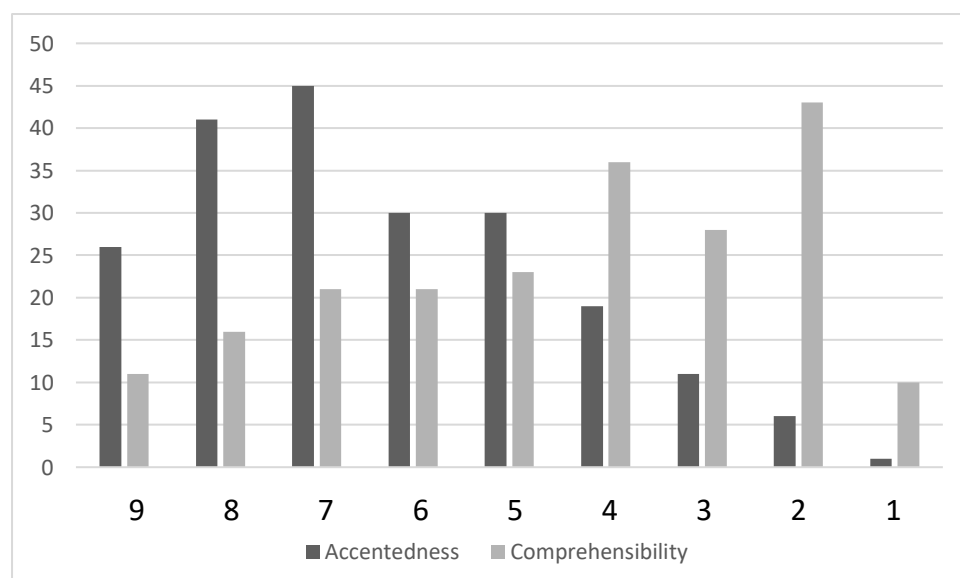
In Figure **7.6-11** the x-axis is flipped left to right to match Figure **7.6-10** as Derwing and Munro used 9 = most accented / least comprehensible whereas my scales ran in the opposite direction. The difference in slope can be accounted for by the level of the speakers: many of the Migrants had lower proficiency than Munro's intermediate speakers.

Warren, Elgort and Crabbe (2009, p. 95) commented that "perceived nativeness [i.e. Accentedness] provides a baseline on top of which comprehensibility appears to be built." Figure **7.6-12** shows that the average ratings of Accentedness and Comprehensibility in Warren et al. and the Employer Survey have similar correlation patterns: Comprehensibility ratings can generally be equal to or higher than Accentedness but not lower.

Figure 7.6-3 Accentedness / Comprehensibility: Warren et al (2009) (left) and Employer Survey (right)
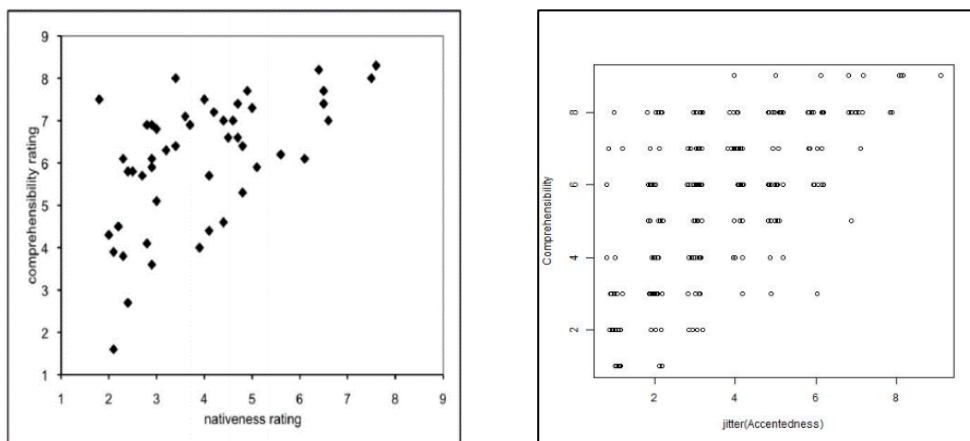
Printed with permission of the publishers

The similarity between the Employer Survey raw results for Accentedness in comparison to Comprehensibility and those reported in previous research demonstrates the likelihood that data from the Employer Survey are valid.

Correlations between three Global Measures

Table **7.6-1** shows the correlations between Employer ratings of three Global measures from the Employer Survey (method: Pearson, p<0.0001 in all cases).

Table 7.6-1 Correlations between Employer ratings of 3 Global Measures

Measures Correlated	Correlation
Intelligibility + Comprehensibility	0.838
Intelligibility + Accentedness	0.544
Comprehensibility + Accentedness	0.704

There is clearly a high correlation between Intelligibility and Comprehensibility. Accentedness is correlated far less strongly with Intelligibility than with Comprehensibility – which is to be expected, as accent has an effect on the difficulty of understanding but less effect on whether the listener can understand at all.

Acceptability places these three Global Measures into a specific context: given the proficiency of pronunciation, how is the speaker rated as suitable for a job? Some correlation with Acceptability is therefore to be expected. Table **7.6-2** below shows the correlations between Employer ratings of Acceptability and each of the other three Global Measures. It shows that these correlations are very much in the same range as the correlations found in Table **7.6-1**.

Table 7.6-2 Correlations between Employer ratings of Acceptability with 3 Global Measures

Measures Correlated	Correlation
Acceptability + Intelligibility	0.704
Acceptability + Comprehensibility	0.751
Acceptability + Accentedness	0.524

Figure **7.6-3** above shows some similarity between Acceptability and Comprehensibility in terms of the Factors which predict each of them, particularly in the strength of Fluency, and Table **7.6-2** shows that Comprehensibility is the most closely correlated with Acceptability, although not far ahead of Intelligibility (method: Pearson, $p < 0.0001$ in all cases). However, Acceptability is similar to Intelligibility in having a weaker correlation with Accentedness. These four Global Measures are related in how they are predicted and correlated, but each remains distinct.

The evidence from Employer Survey Part A therefore supports the hypothesis that responses to Acceptability will have a partial positive correlation with Intelligibility, Comprehensibility and Accentedness. They also confirm research conclusions (**3.2.3**) that Accentedness is the least important of the Global Measures, and demonstrate that Comprehensibility is a key predictor of Acceptability.

7.6.3   Personal Qualities

Hypothesis:

> *Employers' impressions of the personal qualities of speakers will influence their rat-*
> *ings of acceptability.*

Data: **Appendix D6**

As explained in **3.2.3**, listeners often make judgements about the personality of a speaker on the basis of their accent, and such reactions may affect judgements of acceptability for employment, so the Employers' ratings of Acceptability could be affected by these impressions. Employers responding to Part A (n=45) were asked the following questions:

Impressions
From the way this person speaks, how would you place them on the following scales?
       Work   *[1= Lazy → 9=Hard worker]*
       Friendliness   *[1= Unfriendly → 9=Very Friendly]*
       Politeness   *[1= Impolite → 9=Very Polite]*

The raw responses to personal qualities show that the Employers awarded a large number of high scores for all three personal qualities. Their judgements would have been based on the way the Migrants spoke and of course the content of the mock job interviews, using their human resources skills to pick those who seemed keen to work.

Figure 7.6-13 Distribution of Employer responses to Personal Qualities
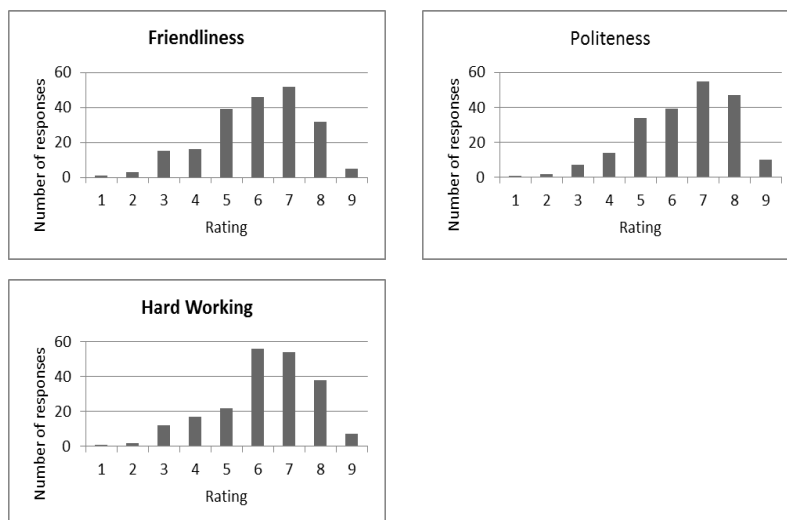
However, the similarity between these histograms masks the differences in the extent to which they predict Acceptability for employment. A model was run with the 3 Factors (Part

A only) and the three Personal Quality variables. There was collinearity in the model between the three Personal Qualities, particularly Politeness and Friendliness (0.567), so a centered version of the model was run, but this made no difference. The variable "Hardworking" was the only significant Personal Quality and it had higher log odds (2.09) and z-value (4.101, p<0.0001) than the 3 Factors (Model **7.6-8**).

Model 7.6-8 Acceptability predicted by Personal Qualities

```
                Estimate  Std.Error  z-value  Pr(>|z|)   Odds
  Friendly        0.211     0.182     1.157   p=0.247    1.28
  Hardworking     0.738     0.180     4.101   p<0.0001   2.09
  Polite          0.228     0.184     1.240   p=0.215    1.26
  Prosody         0.486     0.197     2.466   p<0.05     1.63
  Fluency         0.646     0.202     3.202   p<0.01     1.91
  Segmental       0.699     0.219     3.189   p<0.01     2.01
```

As Friendliness and Politeness were not significant predictors, Model **7.6-9** was run with Hardworking alone plus the 3 Factors, and this model was found to be a better fit than the 3-Factor Model (ANOVA, χ^2 (1)=67.559, p<0.0001).

Model 7.6-9 The 3-Factor Model plus Workhard

```
                Estimate  Std. Error  z value  Pr(>|z|)   Odds
  Hardworking    1.041      0.142      7.349   p<0.0001   2.83
  Prosody        0.460      0.197      2.335   p<0.05     1.59
  Fluency        0.664      0.204      3.263   p<0.01     1.94
  Segmental      0.739      0.221      3.349   p<0.001    2.09
```

For Part A, then, Hardworking is the strongest predictor of Acceptability, outranking the 3 Factors. This may be associated with the level of employment. For professional employees there may be a stronger need to fit into a team or work well with customers, but for work such as packing shelves in a supermarket the employer may be mostly concerned with whether the employee will turn up regularly and do the job.

The 3 Factors all remain significant predictors, with Segmental the strongest followed by Fluency and then Prosody. It is understandable that addition of this variable to the 3-Factor Model (Part A) has reduced the estimate weighting of the strongest predictor (Fluency) by 0.143 and Prosody by 0.137 but has reduced Segmental by only 0.048. A judgement that someone is hardworking can be gained from *what* they say, but if pronunciation is involved then *how* they say it is likely to be based on Prosody and Fluency features, which give more evidence of intention and personality. These two Factors, and not Segmental, would therefore be expected to share some of the weighting of this judgement.

This analysis assumes that Employer judgements of personal qualities are predictors of Acceptability along with the three rater factors. However, there is no simple statistical standpoint to determine the direction of causality. That is, should we assume that the Employers listen to the speakers, make a judgement from what they say and how they say it that they are hard-working, and therefore rate them as more acceptable (Qualities as IVs) or do the pronunciation features of the speakers also predict whether Employers consider them to be friendly, polite or hard-working? To test this I ran further models with each of the three qualities in turn as DVs and the 3 Experienced Rater factors as predictors.

Model 7.6-10 Friendliness predicted by Rater Factors

```
          Estimate Std. Error z value Pr(>|z|)
Prosody   0.269     0.210     1.281    p=0.200
Fluency   0.420     0.219     1.920    p<0.06
Segmental 0.421     0.235     1.789    p=0.737
```

Model 7.6-11 Politeness predicted by Rater Factors

```
          Estimate Std. Error z value Pr(>|z|)
Prosody   0.199     0.218     0.912    p=0.362
Fluency   0.438     0.225     1.942    p<0.06
Segmental 0.498     0.244     2.046    p<0.05
```

Model 7.6-12 Hardworking predicted by Rater Factors

```
          Estimate Std. Error z value Pr(>|z|)
Prosody   0.463     0.226     2.046    p<0.05
Fluency   0.448     0.235     1.910    p<0.06
Segmental 0.339     0.246     1.378    p=0.168
```

Evidence that Fluency predicts each of the three personal qualities is marginal (p<0.06); Segmental predicts Politeness (p<0.05) and Prosody predicts Hardworking (p<0.05). No general pattern emerges from this. There is insufficient evidence from this data to be sure about how these variables influence one another: the 3 Factors influence both Acceptability and – to a more limited extent – the Qualities, but the Qualities themselves also influence Acceptability.

The hypothesis that Employers' impressions of the personal qualities of speakers will influence their ratings of acceptability is therefore supported. The prediction of Acceptability by "Hardworking" arises from Employer Survey Part A only, so the result cannot be incorporated into the full 3-Factor Model. That is, a single model cannot include personal qualities and all the Employers or the Employer Pronunciation variables dealt with below in section **7.7.1**, since the respondents were two different Employer groups.

7.7   EMPLOYER SURVEY PART B

Respondents to Part B were asked about specific pronunciation features (Segmental, Prosodic) and for their general impression (Fluency) of the speech of 5 Migrants.

The 3-Factor Model which tests variables from Part B is a subset of the full model, using only data from the 50 Employers who responded to Part B and from the Base Rating for the same questions. Conclusions from Part B contribute to our understanding of Employer responses but cannot be incorporated into the full 3-Factor Model since the respondents in Parts A and B were different. The 3-Factor Model for the full Employer Survey, and for each of the two parts, appears in **7.5**.

7.7.1   Employer Pronunciation Responses

Hypothesis

Employers' perceptions of Migrant pronunciation (as opposed to their actual pronunciation) will affect predictions of Acceptability.

As explained in **3.2**, the Global Measures are all subjective judgements on the part of the listener. Section **3.2.4** reports on research which shows that what is heard is, in part, constructed by the listener, and must therefore affect ratings of Acceptability.

Employers who took Employer Survey Part B (n=50) were asked the following questions:

Specific Features: *How would you rate these specific features of their pronunciation?*
1. Individual sounds
     Vowels – e.g. first *pronounced correctly, or as* farst *or* feerst
     Consonants – e.g. job *pronounced correctly or as* chob *or* jop
     [1= Many sounds wrong →9= Generally correct]
   Stress on words
      For example, eleMENtary (correct stress) not eLEMentary (incorrect stress)
     [1= Many wrong stresses →9= Generally correct]
    Intonation
     How the voice rises and falls in a sentence, e.g. My ↑last job was quite
     ↑inte↓resting, or ↑Where are you ↑go↓ing?
     [1= Intonation often wrong →9= Generally correct]
   Added or missing sounds
     For example, saying ba' instead of back (missing), or sikis instead of six (added).
     [1= Many added/missing sounds →9= Generally correct]

General Impression: *Finally, how would you describe the migrant's speech?*
[Labels following each question are the variable names used in this thesis]

1. *1= Slow → 9= Fast*   [Speed]
2. *1= Monotonous → 9= Interesting* [Variety]
3. *1= Choppy → 9= Smooth*   [Smoothness]
4. *1= Hesitant → 9= Fluent*   [Fluidity / Pausing]

The distributions of raw Employer ratings of all these features are shown in Figure **7.7-1**.

There is a tendency for these histograms to show a curve which falls away at each end, but many of the variables have high numbers of low ratings. As previously explained, *Speed* was removed from the model, leaving seven pronunciation variables.

Figure 7.7-1 Distribution of Employer pronunciation ratings
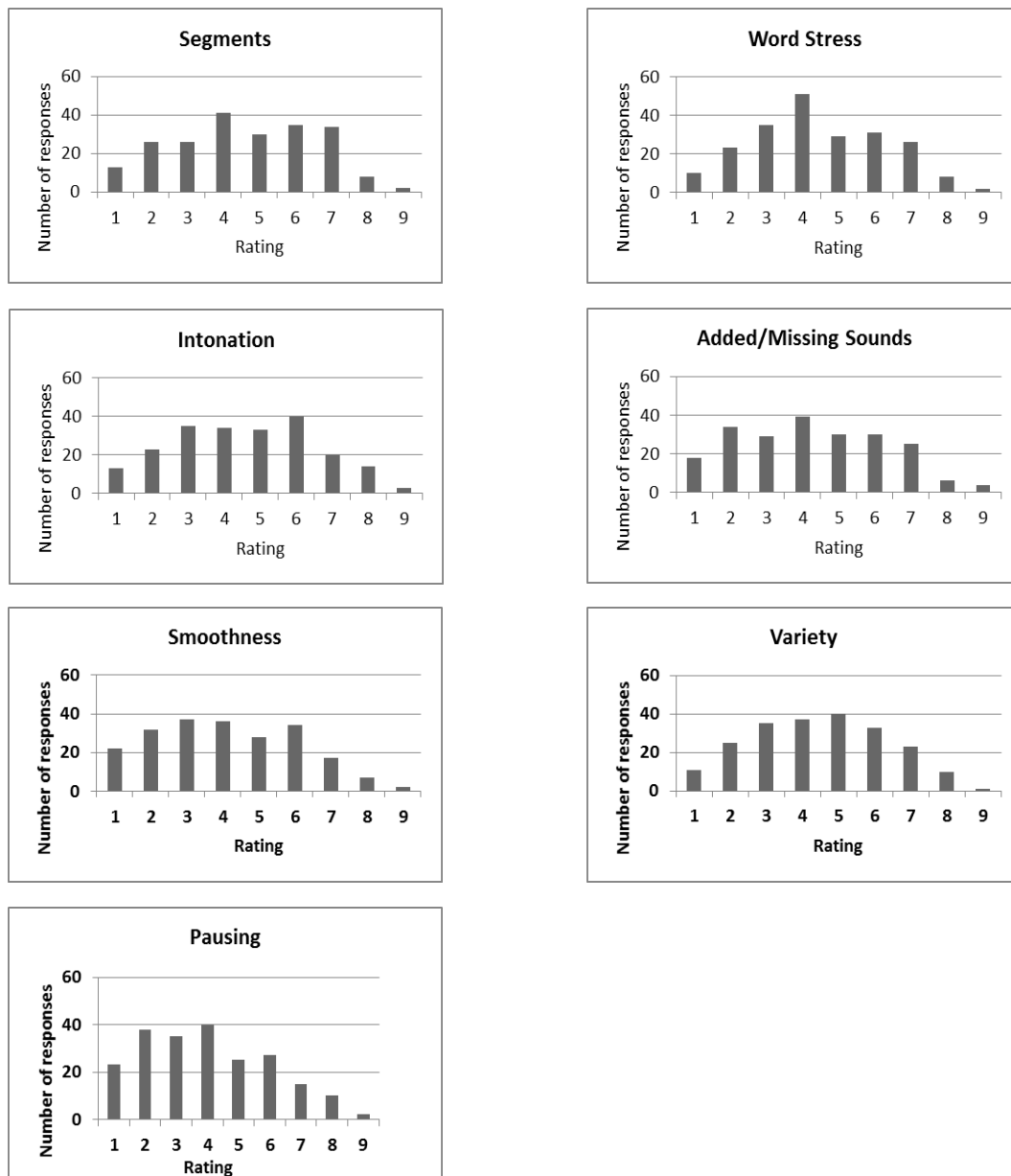
Comparison of Models

In order to compare models with and without each variable I made use of *addterm*, part of the Ordinal package in R which runs with clm2, an earlier version of clmm. It enables a series of variables to be added to the model without having to run a separate model for each or for each combination. This process was used to test all predictions of Acceptability by the Employer pronunciation variables (e.g. Fluidity). Each iteration had one variable removed, and addterm compared that with a model including the variable. A χ^2 test was used to show whether the model including the variable was a better fit. The model also examined the relative ability of each variable to predict Acceptability. The results are shown in Table **7.7-1**.

Table 7.7-1 Test of Employer Pronunciation Variables: Predictions of Acceptability

Tested→ Removed ↓	Fluidity	Smooth- ness	Variety	Added/ Missing Seg- ments	Intona- tion	Stress	Seg- ments
Fluidity	<0.05	<0.01	X	X	<0.01	X	<0.001
Smoothness	<0.001	<0.05	X	X	<0.01	X	<0.001
Variety	<0.05	X	X	X	<0.05	X	<0.001
Added/Miss-ing Seg-ments	<0.05	X	X	X	<0.05	X	<0.001
Intonation	<0.05	<0.05	X	X	<0.01	<0.05	<0.001
Stress	<0.05	<0.05	X	X	<0.01	X	<0.001
Segments	*0.05	<0.05	X	X	<0.001	<0.001	<0.001

Key
Figures are all p-values   * = marginal   X = p>0.05
Columns = p-values of the variable in the header when the variables in the left column are removed.
Rows = p-values of other variables when the named variable (left column) is removed.
Variable with itself = χ^2 likelihood that a model with this variable is different from one without it.

Table **7.7-1** shows that removing other variables does not generally make Variety, Added / Missing Segments or Stress better predictors, and the χ^2 tests report that models with these three variables are no different. On the other hand, other variables remain significant when Variety, Added or Missing Segments or Stress are removed. Segments and Intonation are highly significant whichever variables are removed, and Fluidity and Smoothness are usually significant. This suggests that these four variables are significant predictors of Employer ratings of Acceptability, but Variety, Added or Missing Segments or Stress are not.

A clm model was run to test the prediction of Acceptability by all the Employer pronunciation variables (except Speed) and the 3 Factor Model. Segments, Intonation and Smoothness and Fluidity were the only significant variables (p<0.05).

Model **7.7-1** retains only the significant Employer pronunciation variables. In this model these are all significant predictors (p<0.05) but none of the 3 Experienced Rater Factors are. There is collinearity between some of the variables: Intonation + Segments (0.284), Intonation + Smoothness (0.206), and Smoothness + Fluidity (0.380).

Model 7.7-1 Prediction of Acceptability by Employer Pronunciation Variables

```
            Estimate Std. Error z value Pr(>|z|)  Odds
Segments    0.906    0.136      6.671   p<0.0001  2.47
Intonation  0.426    0.136      3.186   p<0.01    1.53
Smoothness  0.411    0.139      2.966   p<0.01    1.51
Fluidity    0.259    0.121      2.147   p<0.05    1.30
Fluency     0.225    0.157      1.436   p=0.151   1.25
Prosody     0.053    0.169      0.314   p=0.754   1.05
Segmental   0.165    0.160      1.025   p=0.305   1.18
```

From these models it would appear that by far the strongest predictor of Acceptability in Survey Part B is the Employer rating of Segments (z-value = 6.671), followed by Intonation, Smoothness and then Fluidity (variables which, in the Base rating, were factored into Segmental, Prosody, Fluency/Prosody, and Fluency respectively). This presents a similar result to that in **7.6.3** *Personal Qualities* where the 3 Factors were weaker predictors than an Employer rating – although, in this case, the 3 Factors are completely eclipsed.

The evidence here is that the Employers are most likely to find any of the speakers more acceptable for employment if they believe that the speakers pronounce segmentals well, even though the Base Rating showed that Fluency is in fact the strongest predictor. It is therefore worth asking whether there is external support for the Employers' ratings by looking at prediction in the opposition direction: how well does the Base Rating predict Employer ratings of pronunciation variables? A series of models was run, each testing one of the Employer pronunciation variable ratings by the Base rating.

Table **7.7-2** shows the results, showing that all the variables are strongly predicted by Prosody and Fluency and less so by Segmental, but not necessarily by the expected Factor. Segmental has the most consistent results: the only variable ratings significantly predicted by Segmental are Segments, Added or Missing Segments and Stress, as would be expected. Flu-

idity is most strongly predicted by Fluency, which is again consistent. Intonation, Smoothness and Variety are all significantly predicted by both Prosody and Fluency. However, Base ratings of both Fluency and Prosody significantly predict *all* Employer ratings including segmental variables.

Table 7.7-2 Employer Pronunciation Variables predicted by the Base rating

3 Factors	Prosody		Fluency		Segmental	
Emp. Ratings	**Odds**	**p-value**	**Odds**	**p-value**	**Odds**	**p-value**
Add/Miss	1.46	p<0.05	1.75	p<0.01	1.54	p<0.05
Fluidity	1.83	p<0.001	3.24	p<0.0001	1.23	p=0.246
Intonation	1.89	p<0.01	2.43	p<0.001	1.51	p=0.087
Segments	1.67	p<0.05	1.90	p<0.01	1.65	p<0.05
Smoothness	1.95	p<0.01	2.08	p<0.01	1.35	p=0.235
Stress	1.75	p<0.05	2.17	p<0.01	1.88	p<0.05
Variety	1.81	p<0.01	1.66	p<0.01	1.21	p=0.220

From this we can conclude that Employers paid closest attention to Segmental proficiency and produced accurate ratings, but for all remaining pronunciation features they rated more proficient speakers highly without being able to discriminate the feature concerned. The exception may be Fluidity, which is predicted by Fluency with higher odds and p-value than any other variable receives for any other prediction.

While the 3-Factor Model shows that all of the 3 Factors are significant predictors of Acceptability, generally with Segmental less so, the results in Model **7.7-1** make Employer ratings of Segments the most significant predictor. The difference probably lies in the extent to which the Employer ratings are *not* supported by the Base rating: that Employers place far more attention on non-native segmental pronunciation, while trained raters are able to detect more nuance in the Fluency and Prosody areas.

As a result, the hypothesis is supported to the extent that Employers' perceptions of Migrant pronunciation do affect predictions of acceptability. The assumption that this would be distinctly different from the Migrants' actual pronunciation was also true for suprasegmentals – the Prosody and Fluency Factors. Only the Employer judgements of Segmental proficiency are predicted by the Base rating.

7.8 SUMMARY OF RESULTS

7.8.1 Findings

Eighteen hypotheses arising from the Employer Survey were tested, and five of them were supported. No hypothesis predicting that Employer judgements would be influenced by their own characteristics, experience, sex or language were supported. Hypotheses concerning the effects of context (Acceptability) were supported, as were those predicting effects from some judgements of the speakers' pronunciation and (for one variable) personal qualities.

Table 7.8-1 Hypotheses and Findings

Acceptability will be affected by the following factors:	Finding
Voice Quality	Not supported
Noise	Not supported
Grammar and Lexis	Not supported
Type of work being offered	Not supported
Employers who are non-native v native speakers of English	Not supported
Employers' length of residence in New Zealand	Not supported
Employers' attitudes towards when and how often employees ought to use English	Not supported
Having had an employee whose pronunciation caused problems	Not supported
Sex of speaker	Not supported
Sex of Employers	Not supported
Interaction of male/female Employers and Speakers.	Not supported
Actual Language of speakers	Not supported
Different communication situations (Supervisor, Co-worker, Customer)	Supported
Communication with customers	Supported
Personal qualities of speakers	Supported for 1 variable (Workhard)
Employers' perceptions of Migrant pronunciation (as opposed to their actual pronunciation)	Supported for 4 variables (Segments, Intonation, Smoothness and Fluidity)
Responses to Acceptability will have a partial positive correlation with Intelligibility, Comprehensibility and Accentedness.	Supported

7.8.2 Summary of Process and Results

The 40 samples, mock job interviews with Migrants, were rated in two ways. Firstly, Experienced Raters judged the speakers' proficiency so that predictions could be made from a standpoint which would be as objective as could reasonably be attained. Their ratings of 13 pronunciation variables were reduced to three factors: Fluency, Prosody and Segmental.

Secondly, Employers were asked to rate the same Migrant interviews for a variety of features, including Acceptability for employment in both parts of the Survey, so the prediction of Employer Acceptability by the Base Rating became the 3-Factor Model which was used as the basis against which all other variables could be tested.

Other Experienced Rater variables were then introduced. Voice Quality and Noise were found to have no significance. Grammar and Lexis were correlated both with one another and with the Base rating, and the Rater Guidelines may have compromised these as independent variables. They could therefore not be included in the model.

Data about Employers and Migrants was then tested against the 3-Factor Model. Employer (and some Migrant) data was:

- Work offered by Employer
- Employer's Length of Residence
- Employer's Native language
- Employer's opinion on Employee use of English
- Employer's report on employee pronunciation problems
- Sex of Migrant and Employer
- Migrant origin guessed by Employer
- Actual language of Migrant

None of the Employer or Migrant data variables were found to have any significant effect on predictions of Acceptability. In some cases (such as Employer's native language) the population size may have been too small for any effects to show. However, the overall lack of any effect from the Employer or Migrant variables suggests that the Employers were responding, as they were asked, by basing their ratings on the pronunciation they heard rather than stereotypes or personal background. This supports the conclusion in Cargile (2000, p. 173) that "listeners respond to accented speech and not simply to the identification of a speaker's ethnic group membership."

Beyond these questions, Employers took either Survey Part A or Part B, so separate models were run on each of these.

In Survey Part A, Employers were asked to rate the acceptability of the Migrants for work in the context of communication with supervisors, co-workers and customers. The raw rating means for supervisors and co-workers were higher than the Part A Acceptability mean, but lower for customers, which suggested that the Employers gave harsher Acceptability ratings when considering the Migrants for positions where communication had higher stakes.

In the three clm regression models for these contexts all 3 Factors remained significant, but the distinction between them lessened. A model combining these contexts into a single Suitability variable produced similar results. This evidence suggests that Acceptability is a Global Measure which varies according to context.

Examination of the relationship between the 3 Factors, Acceptability and the standard Global Measures showed that the 3 Factors were strong predictors of Employer ratings of Intelligibility, Comprehensibility and Accentedness as well as of Acceptability. All of the Global Measures were correlated to some extent with one another, and similar results were found to raw scores from previous research.

Parts A and B also found that specific non-pronunciation variables were strong predictors of Acceptability.

Part A tested three Personal Qualities: Friendliness, Politeness, and Hardworking. Of these, only Hardworking was found to be significant, predicting Acceptability more strongly than any of the 3 Factors. Employers could have formed their conclusions about Personal Qualities from both the content of the mock interview and the Migrant's way of speaking. The weighting of Hardworking reduced those of Fluency and Prosody but not Segmental, suggesting that the suprasegmental features are important for conveying personality.

Part B comprised Employer ratings of pronunciation features, of which four were found to be significant predictors of Acceptability: Segments, Intonation, Smoothness and Fluidity, with Segments a far stronger predictor than any other and the 3 Factors no longer significant at all. The Employer ratings of pronunciation features were then tested for prediction by the Base ratings, and consistent ratings were found in only two areas: the Segmental Fac-

tor derived from the Base Rating predicted the variables associated with segments (Segments, Added or Missing Segments, Word Stress, all p<0.05) and Fluency was a strong predictor of Fluidity (p<0.0001). However, all Employer variables were predicted indiscriminately by Prosody and Fluency, suggesting that Employers gave high ratings in all pronunciation variables to all speakers with good proficiency, which then fed into a high rating of Acceptability. This also sheds light on the extent to which untrained raters can identify pronunciation features.

7.8.3   Answers to Research Questions

As stated at the outset of this chapter, the research questions were:

1. Does pronunciation affect employers' assessments of the Acceptability of migrants for employment?
2. If so, which pronunciation features are the strongest predictors of the assessment?
3. Once pronunciation features are taken into account, are there other employer or speaker factors which affect employer ratings of acceptability?

The answer to **RQ1** is an unqualified Yes: Both Experienced Rater Factors and Employer ratings of pronunciation variables were significant predictors of Acceptability.

Because of the correlations it is not practicable to identify which specific pronunciation features (**RQ2**) are the strongest predictors. However, it was possible to test features in combinations which were both statistically and phonologically motivated: Fluency, Prosody and Segmental. In all cases mentioned below, the results are statistically significant (p<0.05 in all cases) and the strength of prediction is shown by the relative z-values and odds.

 Overall, the 3-Factor Model (Model **7.3-1**) showed that Fluency was the strongest predictor of Acceptability, and this remained true for Part A (Model **7.5-2**) and Part B (Model **7.5-3**), for the Supervisors and Co-workers (Model **7.6-1** and Model **7.6-2**), and for Employer Comprehensibility (Model **7.6-6**).

While Fluency was the strongest predictor, the positions of Prosody and Segmental often changed. In the 3-Factor Model (Model **7.3-1**) they were not far apart and well behind Fluency; for Part A (Model **7.5-2**) Segmental was closer to Fluency, and for Part B (Model **7.5-2**) the order switched and Prosody came second.

As Fluency is the strongest predictor overall it is worth looking at those models where it was surpassed by one of the other Factors. While the order for Supervisors and Co-workers was Fluency – Segmental – Prosody, for Customers (Model **7.6-3**) the order was (marginally) Segmental – Fluency – Prosody. This is evidence that Employers think of this situation differently, and focus on different areas of speech when considering an employee who needs to communicate in this more critical context. The difference in odds between the three Factors is minimal so it would be more appropriate to say that *all* areas of pronunciation have become important.

Where the 3 Factors were eclipsed by an Employer rating (Personal Quality and pronunciation features) the order was also different. In Model **7.6-9** the variable Hardworking from Part A is easily the strongest predictor of Acceptability, and the order of the 3 Factors is Segmental-Fluency-Prosody. In this case the first two are close but Prosody is lower. As suggested earlier, this may be because the judgement of Hardworking is partially drawn from suprasegmental rather than segmental aspects of speech. In Part B, four pronunciation variables were found to be significant predictors, but the Factor order is not relevant as none were significant.

Predictions, by the 3 Factors, of Employer ratings of the three standard Global Measures were also variable. While Fluency remained the strongest predictor for Comprehensibility, the order was Segmental-Fluency-Prosody for Intelligibility Model **7.6-5**) and Segmental-Prosody-Fluency for Accentedness (Model **7.6-7**). The order is less important than the differences: Factors estimates are separated by of 0.125 and 0.102 for the former and 0.142 and 0.28 for the latter, suggesting that Segmentals are clearly the most significant and there is not much to distinguish Fluency and Prosody in the Accentedness model.

Overall, then, a wide range of models found that Fluency was the strongest predictor of Acceptability. The positions of Prosody and Segmental varied so often that a clear order cannot be given for these two Factors, and on occasion either of them could appear as the

strongest predictor. This suggests that the Employers found different features more significant in different contexts, and that all pronunciation features should be included in a language programme. It is not possible to say with certainty whether some variables contributed more to these predictors than others, but the weightings of variables for the Factor Analysis are given in Table **7.2-4**, and all were at least 0.5.

The answer to **RQ3** (other factors) is, firstly, that no Migrant or Employer-related factors were reliably found to predict Acceptability, but it is likely that Grammar and Lexis are significant. However, two types of variable were found to be significant. In Part A the Employers' attitudes towards the Migrants from their speech (or from what they said) meant that "sounding like a hard worker" predicted Acceptability. Further research is needed to answer the question "What makes a speaker sound like a hard worker?" It is likely that Migrants who were able to describe their previous jobs effectively and confidently would rate higher. Employer ratings of pronunciation in Part B were also very significant: Model **7.7-1** shows that Segments, Intonation and Smoothness and Fluidity were all very strong predictors, with Segments by far the strongest.

These results, especially for **RQ3**, raised interesting questions about the relationship between the beliefs of linguistically naïve raters and how they make decisions. The Employers' own ratings of Segmentals most strongly predicted their decisions about Acceptability. However, as measured by the Experienced Raters, Fluency is the factor which most strongly affected the Employers' actual judgements of Acceptability. Employers were therefore making "employability" decisions on high-level speech ability, not low-level errors. Employers themselves judged Smoothness and Fluidity to be important features for Acceptability

The statistical results also supported the validity of the data itself. If the process of data collection (the migrant interviews, selection of respondents, responses directly from Employers *in situ* rather than in laboratory conditions) was flawed, then a high degree of randomness could be expected from the responses. This was not the case. Figures **7.6-10** and **7.6-11** show how the raw Accentedness and Comprehensibility ratings from Employers mirror those in Munro (2008, p. 204) and, as shown in Figure **7.6-12**, have a similar pattern to those in Warren et al. (2009, p. 96).

Two important conclusions arise from the analysis of the Employer survey. The first is the grouping of pronunciation features. Much comment has been made about the need to

teach suprasegmentals, and there is no question that pronunciation teaching has been limited by an excessive focus on segmentals. However, teachers also need to be aware of a level beyond word stress and intonation. The Employers responding to this survey tended to respond most favourably to those speakers with the highest levels of Fluency, which means that these features must also be taught.

Finally, Acceptability emerges from this study as a Global Measure and a more appropriate goal than Intelligibility or Comprehensibility. Migrants learn English to attain other goals, whether study or work, and in terms of pronunciation is it not sufficient that they become intelligible. Their speech also needs to be acceptable for the context and the speech community.

8 CONCLUSIONS

8.1 OVERVIEW

My reasons for undertaking this research were my own experiences as a migrant and later as an ESOL teacher and manager of language schools, together with my lifelong interest in language and especially pronunciation (**1.3**). We are identified by our accents, but we can also be judged as untrustworthy, incompetent or unintelligent if our proficiency in pronunciation is low. I believed – and have found that many other professionals also believe – that knowing which features of pronunciation are more or less acceptable to employers would benefit both teachers and learners of English.

Chapter 2 detailed the barriers migrants face to gaining employment in New Zealand. Effects of ethnicity and language are hard to separate in many studies, but there can be no question that there is discrimination against non-ESANA migrants (**2.3.2**) or that lack of adequate English is one of the reasons often given for the difficulty migrants have in finding employment. There is not a great deal of research into the importance of pronunciation to finding employment because it tends to be lumped in with "poor English", but there is evidence of accent discrimination against specific migrant groups (**2.3, 2.4**).

While there cannot be a quick fix to negative attitudes, there is also evidence of positive attitudes to migrants, many of whom have been successful in finding employment (**2.3.4**). Clearly the ESOL industry has a responsibility to ensure that the methodology and content of English programmes will enable migrants to reach the level of Acceptability required by employers. I therefore wanted evidence on what pronunciation features it would be most useful to teach.

Evidence used in Applied Linguistics research is often gathered under controlled conditions in order to have confidence in the data and how it has been collected. For example, groups of students are taught using alternative methods and their proficiency is compared at the start and finish of the observation. The outcome is therefore tied to the Global Measure being tested, such as Intelligibility; but this is not necessarily the desired outcome for the student, who is not learning English to become intelligible but to be able to gain further qualifications or employment – real-world outcomes. To investigate whether pronunciation

teaching has been successful for the student, it is necessary to go outside the classroom and find out from the supervisors or gatekeepers relevant to the learners' goals whether their speech is acceptable.

For this thesis, the obvious solution was to ask employers themselves what they thought of migrant accents: how acceptable they would be for employment. A range of migrants, mostly attending language schools, were asked to take part by providing speech samples; and to ensure the samples were appropriate for the context, I made use of spontaneous speech – mock job interviews.

My literature review covered the need for ESOL in the New Zealand context (Chapter 2), the establishment and use of Global Measures as yardsticks for pronunciation teaching (Chapter 3), the specific difficulties which could be predicted for learners of English and their impact on teaching (Chapter 4), and a study of the features of pronunciation which were most likely to predict Acceptability (Chapter 5).

Arising from this research and the subsequent survey was the Global Measure of Acceptability itself. All Global Measures are to some extent contextual: they refer to a specific speech community, because they may change if transposed from, say, New Zealand to Texas. Acceptability, however, adds the context of purpose and situation. The evidence from this thesis is that a speaker with a given level of Comprehensibility may have varying levels of Acceptability for a job depending on its communication requirements and probably other criteria in the mind of the employer. Since the full range of pronunciation features were found to be significant predictors of Acceptability, the implication for ESOL teachers is that the full range should be taught from the earliest stages.

8.2 ANSWERS TO THE RESEARCH QUESTIONS

My aims were summarised in the three research questions:

1. Does pronunciation affect employers' assessments of the Acceptability of migrants for employment?

2. If so, which pronunciation features are the strongest predictors of the assessment?

3. Once pronunciation features are taken into account, are there other employer or speaker factors which affect employer ratings of Acceptability?

Research Question 1

The outcome of the analysis of the Employer Survey was that, when Employers were asked to think in terms of pronunciation and rate Migrants for Acceptability, the 3 Factors of Fluency, Prosody and Segmentals were strong and significant predictors of Acceptability (**7.3.2**). This is supported by the answers to questions 2 and 3.

The answer to Question 1 is therefore, unequivocally, that pronunciation affects employers' assessments of the Acceptability of migrants for employment.

Research Question 2

When I designed the Employer Survey I expected that results would be found for the features described in Chapter 5: that Word Stress, for example, would be highly important, and Smoothness perhaps less so. This proved impracticable. Firstly, even though I found 95 Employers who were prepared to rate around 5 samples each, the total number of observations did not allow, statistically, for some 20 separate features to be rated. Secondly, the features themselves do not act in isolation. Any rater will hear patterns which will cause responses to (say) Fluidity and Pausing to be correlated, but not Fluidity and Consonants. A factor analysis (**7.2.2**) was therefore used, producing three statistically-determined factors which broadly agreed with three levels of pronunciation: Segmental, Prosody and Fluency. No effect was subsequently found from Voice Quality or Noise. When the 3 Factor Model was used to predict Employer Acceptability of Migrants for Employment in various contexts and in each part of the Employer Survey, Fluency was the strongest predictor of Acceptability, although all three factors were significant (**7.8.3**).

The answer to Question 2 is therefore that Fluency features appear to be the strongest predictors of Employer ratings of Acceptability. The pronunciation features included in this factor (in order of weighting – see Table **7.2-4**) were Pausing, Fluidity, Coarticulation, Smoothness, and Variety (which was shared equally with Prosody). If Speed could have been included it would probably have had a high weighting. However, the Prosody and Segmental factors were also significant, meaning that all three levels are important components of migrant speech.

In Survey B, the Employers' ratings of pronunciation features, even when they did not agree with those of the Experienced Raters, also strongly predicted Acceptability (**7.7.1**). From the Employers' perspective, someone who appears to pronounce segmentals well is highly acceptable in terms of pronunciation; and Smoothness and Fluidity (Fluency features) and Intonation (a Prosodic feature) were also significant. Thus we can conclude that Acceptability is predicted not only by how well the Migrants *actually* used most pronunciation features but also how Employers *believed* they used the above four features. As Table **7.7-2** shows, while Employer Segmental ratings were predicted by Experienced Rater judgements of Segmentals, both Fluency and Prosodic factors from the Base Rating predicted all of the Employer ratings of Smoothness, Fluidity and Intonation. It is likely that Employers were (as expected) using those three features with far less discrimination than Experienced Raters. This therefore sheds light on the extent to which untrained raters can identify and comment on pronunciation features: Segmentals are easy to notice and comment on while Prosody and Fluency features are less consciously distinguishable.

Research Question 3

Clearly the decision to hire someone is based on many different factors, which may include their background, appearance, experience, qualifications, personality and so on. The survey was designed to eliminate most of these factors so as to focus on pronunciation, but some could not be eliminated. Data about the type of work, background of the employer or the speaker's sex are evident, so these questions were added to test whether they played a role in predicting Acceptability.

All such variables from the Employer Survey were added in turn to the 3 Factor Model and found not to be significant. Models included the following Employer information: work of-

fered, length of residence, attitudes to Migrant use of English, experience of employee pronunciation problems. In addition, the sex and language of both Employers and Migrants (**7.4**) were tested. In every one of these, pronunciation was the only significant predictor of Employers' assessments of the Acceptability of Migrants for employment.

The remaining additional variables rated by the Experienced Raters were Voice Quality, Noise, Grammar and Lexis. Voice Quality and distracting noise were not significant (**7.3.3**). When Grammar and Lexis were added, Lexis appeared to be significant and the Segmentals Factor ceased to be. As explained in **7.3.4** the guidelines for Experienced Raters described these features in terms of comprehensibility which meant that they were not being rated as independent variables. Possibly as a result of this the two features were highly correlated with the 3 Factors. When each of these variables was tested alone against the 3-Factor Model it was found to be a significant predictor of Acceptability. It is very likely that both of these features are involved in judgements of acceptability for employment, but further evidence is needed to be certain.

One of the Personal Qualities from Survey A was significant when added to the 3 Factor Model. There was no consistent prediction of Acceptability of Migrants who were thought of as polite or friendly, but unsurprisingly those who Employers thought were hard workers were highly acceptable (**7.6.3**). This may have been an effect of the methodology: spontaneous speech in a mock job interview gave the Migrants the opportunity to put some of their personality and experience into their responses in addition to what was shown by their pronunciation skills.

The lack of effects from most other features suggests strongly that the Employers did what they were asked to do: rate the Migrants for Acceptability based on their pronunciation. The answer to Question 3 is, therefore, that the Employers were not influenced significantly by their own background or attitudes to the Migrants, only by their perception that certain pronunciation features were used well and the impression that the speaker would work hard – reasonable criteria for a NNS employee. The implications of these outcomes is discussed further in **8.6.**

8.3 Contribution

This thesis makes three main contributions to the study of pronunciation teaching and learning: a methodology which has enabled new questions to be asked and new information to be discovered, evidence of Acceptability as a Global Measure, and evidence for the importance of Fluency to learners of English and therefore to ESOL programmes.

8.3.1 Methodology

As explained in **6.1.1**, very few studies, most of them in the US, have focused specifically on employer acceptability of migrant pronunciation, and no preceding study has been able to reduce the impact of ethnic stereotyping to the point where no predictive effects could be found. This thesis is the first piece of research to go beyond the question of which groups of migrants are considered more acceptable for different types of employment, and instead to identify which aspects of migrant pronunciation predict Acceptability for employment, and to consider whether the prediction varies in different contexts. The contexts here can be seen in terms of degrees of importance for communication (with co-workers, supervisors or customers).

Underlying the methodology was the criterion of authenticity (Rifkin & Roberts, 1995) as discussed in **5.4.2**. While classroom-based studies are useful for comparing techniques and outcomes in language development, Acceptability requires real-world evidence. Validity of the results was therefore strengthened by using the most authentic aspects of previous research of this type:

- Respondents who were actually employers or hirers of staff
- Migrants using their own way of speaking
- Participants individually recruited
- Unscripted speech samples
- Speech length adequate to form judgements about pronunciation features
- Speech relevant to the context of work
- Both male and female respondents
- Removal of clues (other than pronunciation) about language, ethnicity and origin
- Contextualisation of Acceptability (Supervisors etc.)
- Use of semantic differential questions.

Techniques not used by previous research into acceptability for employment include:

- Wide range of migrant groups and languages currently found in New Zealand

- The opportunity for "Second draft" speech samples after interview advice

- Focus on low or unskilled migrants and the types of employment they seek

- Focus on pronunciation features

- Employer rating of specific pronunciation features

- Creation of a Base Rating as a baseline for comparison

- Random assignment to a two-part survey to test pronunciation ratings separately from other variables.

Using the above methodology has enabled me to come far closer than any previous study to answering the question: "Which pronunciation features are the strongest predictors of employers' assessments of the Acceptability of migrants for employment?"

8.3.2   Acceptability

Chapter 3 described the development over the last half century of beliefs about learning pronunciation, in which the need for the overt teaching of L2 pronunciation was gradually accepted and a deeper understanding of learner goals and attainment was gained. The review showed how evidence had been produced for separate Global Measures such as Comprehensibility in place of the catch-all concept of "accent", and how the goal of Intelligibility had been advocated to replace that of Nativeness. Finally, Chapter 3 described research involving Acceptability, and explained its relevance as a further Global Measure. Chapters 6 and 7 of this thesis described the research and statistical evidence for the role of Acceptability, its effect on migrants seeking employment, and the need for teachers and course planners to take it into account when designing the pronunciation component of ESOL programmes.

When considering the goals of learners of English, it is essential to realise that they rarely include learning English for its own sake, any more than taking a flight is the goal for reaching a destination. English, for most migrants, is a means to an end: integration into an English-speaking society, work in an English-speaking environment, or further study in an

English-language institution. Thus the classroom goals of Nativeness or Intelligibility have never been more than learning goals *with further goals in mind*. The learning goal is negotiated – overtly or not – between each teacher and learner. The teacher determines what content to teach, which techniques to use, what to emphasise, and how to motivate the learner; the learner decides what they will try to learn, what learning methods to use, how much effort to put in, and at what point they will stop learning (**3.1.2**). Any of these may be a default position rather than a conscious decision, but the outcome is the same. Above all, the teacher has a crucial role in clarifying to the learner what level of each skill they will need to have in order to achieve their further goal, and how they can achieve that level. As a very broad generalisation, this clarification is made brutally clear in the marking of every piece of grammar or writing, but is largely absent from feedback on pronunciation. This is one of many reasons why the former is well developed in learners while the latter is not.

The pronunciation goal for English teaching which has been advocated for the past quarter-century has been the Intelligibility principle. Teachers therefore ought to be enabling learners to understand whether people in the community outside the classroom will be able to understand what they are saying, and without too much difficulty. For the adult learner to have any motivation to make progress, the teacher must ignore their own familiarity with non-native speech and feign incomprehension, or arrange conversations with native speakers. I have seen little evidence of this in classrooms: teachers are rarely using any consistent approach to meet the goal of Intelligibility.

However, this thesis demonstrates that the situation is probably even worse, because Intelligibility is not enough. If the ultimate goal of many learners of English is employment, then the learning goal must be to speak in whatever way employers find acceptable. Intelligibility is a good starting point because according to the Employer Survey it correlates at around 0.7 with Acceptability (Table **7.6-2**). As shown by Figure **7.6-5**, there may be some job applicants who are not highly intelligible but whose fluency makes them acceptable. On the other hand, there will be a substantial number of learners who have achieved a good level of Intelligibility but not of Acceptability, and they will be left with no idea what needs to be done about their pronunciation. Perhaps they will go to accent reduction classes, unaware that prominent researchers regard them as highly dubious (**3.2.1**) and the correlation between Accentedness and Acceptability is only 0.5 (Table **7.6-2**). In other words, it is possible for

"improvements" to be made to a learner's accent without, for example, improving any of the prosodic or fluency features sufficiently to make them acceptable.

Acceptability is subjective, context-based, and varies from one dialect community to another. For it to be a usable pronunciation goal, teachers need a description of what it is and how to achieve it. Fortunately, this thesis provides such a description: it involves a comprehensive development of pronunciation features, Fluency above all, and this is the third contribution made by this thesis to the study of teaching and learning pronunciation. It is discussed in the following section.

8.3.3   Fluency

Chapter 5 described the features of pronunciation under the headings Segmentals, Prosody and Fluency, plus other factors investigated for this thesis. It summarised the evidence for predictions of the Global Measures by these features, and from this extrapolated the likely relationship between these various features and the Employers' ratings of Acceptability.

The comparatively recent shift from Nativeness to Intelligibility in ESOL teaching has been paralleled by an equally important change of emphasis in the focus of pronunciation teaching. Calls to extend pronunciation teaching beyond the segmental level often referred to suprasegmentals, mostly meaning word stress, intonation and rhythm (**5.3**) with the justification of an improvement in Intelligibility.

However, there still seemed to be an assumption that Fluency was solely an emergent phenomenon – "fluent" often being a synonym for "proficient". In typical communicative teaching, language features are taught in the presentation phase and then tasks used to allow Fluency to develop naturally, with the features which constitute fluency remaining untaught.

As noted above as the answer to Question 2, Fluency was found to be the strongest predictor of Acceptability. This suggests that if migrants, even those at a low level of proficiency, develop the Fluency features (**5.4**) which include Pausing, Fluidity, Coarticulation, Smoothness, Variety (and, although not reported on in this thesis, Speed) to a high enough level,

they will be better equipped to achieve their ultimate purpose for studying. There is evidence, examined in section **8.4**, that the features of Fluency can and should be explicitly taught.

This focus on Fluency therefore requires some changes in pronunciation teaching and a shift in the target which accompanies the focus on Acceptability. As noted in **4.4.3**, researchers have found that high ratings are given to speakers for their overall impression and proficiency. The role of Fluency in predicting Accuracy, in addition to Segmentals and Prosody, shows there is considerable justification for these conclusions.

8.4   THE GOALS OF PRONUNCIATION TEACHING

There are many internationally recognised and local tests of English proficiency, but the global measures introduced in Chapter 3 provide criteria to underpin such tests and set learners' goals. Historically these measures were lumped together as *accent*, but research demonstrated that Intelligibility, Comprehensibility and Accentedness are related but partially independent measures (**3.2.3**). The goal of pronunciation teaching also shifted from Nativeness (i.e. elimination of accent) to Intelligibility, and the evidence-based distinctions between the Global Measures provided further support for this approach.

The justification for this shift in emphasis came not only from evidence of the great difficulty which adults have in achieving native speaker-like pronunciation but also from the personal and cultural associations with accent, and the ability which native speakers have to accommodate a range of accents. For most adult learners, a native-like L2 accent is not only unattainable but even undesirable (**3.1.2**) and accents can become more comprehensible through familiarity (**3.3.2**).

However, there are strong reasons why, as a goal for pronunciation teaching, Intelligibility may not be enough. Firstly, Intelligibility is no more fixed as a target than Nativeness because Intelligibility itself is subjective (**3.2.4**). There may be other reasons why a listener chooses not to find a speaker fully intelligible, which include racial stereotyping and unwillingness to compromise with foreign speech, or features of NNS speech which cause irritation or are stigmatised (**3.3.1**).

The question "intelligible to whom?" is also valid, given that there is no such thing as a standard English native speaker. Many ESOL teachers are poor at judging Intelligibility or Comprehensibility because of their professional familiarity with accents and the known context of the ESOL classroom, which makes these measures even harder for teachers to assess. In addition, there has been little specific research on what makes a speaker intelligible, while research on Comprehensibility has produced a list of contributing factors so comprehensive that it is difficult to point at a feature which does not predict it to some extent (**3.2.5**).

Furthermore, Intelligibility is a more limited goal than it might appear, as it is possible for a speaker who is deficient in Fluency or even Prosody features to be highly intelligible. One

who speaks with a lack of Smoothness or has poor coarticulation, monotonously, slowly, and with long pauses, could be well understood, but would not impress the listener as a competent or effective user of English. Listeners can accommodate to other accents by ignoring them and focusing on the message, but features which cause irritation can divert attention from the message and its intent (**3.3.1**).

Of course, a large number of students in classes aiming for Intelligibility will end up as competent speakers, just as many who learned under grammar-translation or audiolingual methods did. Language learners, especially those who make use of the language outside of class, are not bound by the restrictions, accent or goals of the teacher. Nevertheless, if some aspects of competent pronunciation are untaught for whatever reason, there will be learners who fail to pick them up. This is why many learners can spend years in language schools and still have noticeable deficiencies in some areas of pronunciation.

Chapter 4 identified the features which were most likely to cause difficulty to different groups of learners, and examined the hypothesis that there is a hierarchy of errors which can be useful for teachers. It showed that speakers from any given L1 may have specific areas which may need emphasis, and that some typically conflated phoneme contrasts have greater frequency and importance than others.

Any ESOL teacher in New Zealand already knows that learners can be divided into three very broad categories. Those from Europe, Africa and the Middle East can usually develop a reasonably comprehensible pronunciation fairly rapidly. East Asian learners may need more assistance and attention to aspects of English pronunciation because of more substantial differences between the phonological systems of their L1s and English phonology, and associated markedness effects, while those from Outer Circle countries, particularly South Asians, may be using established dialects which cause comprehension problems for New Zealand listeners. The predictions in Chapter 4 show that these categories plus more detailed information on each L1 can provide some understanding of what needs to be emphasised in teaching, but an excessive focus on troublesome phonemes or other features may obscure the need, demonstrated by the evidence in this thesis and elsewhere, for all learners to develop skills across the full range of pronunciation features because they have similar ultimate goals.

8.5   LIMITATIONS

Survey Construction

While the separation of the survey into two parts (**6.3.3**) ensured that respondents were not answering too many questions and could focus on one area or another, the separation created three surveys: the whole survey, Part A and Part B, so results had to be analysed separately. Given the small number of respondents for the number of samples, it would have been preferable to have fewer questions and a single survey. This does not, however, invalidate any of the findings. Problems with the platform used to implement the survey raised the likelihood of errors appearing, and its inability to distribute respondents evenly resulted in a varied number of observations per sample (**6.4**). If these two problems could have been solved, the same number of respondents would have produced a statistically more comfortable ten observations per sample.

Employer Ratings

Use of uncontrolled responses has been criticised because of a lack of control over the conditions (**6.3**). This resulted in differing audio quality for the respondents and technical difficulties when using a variety of equipment, platforms and firewall policies. The effect is to reduce the quality of the data, although I see no practicable alternative to this method for gathering this type of data.

Variables

The question on the variable Speed did not use the 1-9 scale in the same way as the others so this variable was eliminated from the analysis (**7.2.2**). It would need either one question (1 = appropriate speed, 9 = too fast/slow) or two separate questions. The removal of this variable weakened the force of the Fluency factor.

Instructions for the questions on Grammar and Lexis were given in a way which created high collinearity with the pronunciation factors and rendered the outcome less secure (**7.3.4**). From the evidence given in **3.2.5**, it is highly likely that both Grammar and Lexis would have been predictors of Acceptability.

While Intelligibility was given as a variable, it was measured subjectively rather than with a transcription task, which would have been impractical. Any results for Intelligibility are

therefore based on perception rather than measurement. Results for Comprehensibility, on the other hand, were gained in the usual way.

Analytical Models

Although the estimates and odds appeared to put Fluency as a stronger predictor than the other factors in most of the models, it is difficult to make an unequivocal interpretation of a clmm ranking. We can, however, state with confidence that all three factors were significant.

Base Rating

Experienced Raters were allowed access to the survey without question validation (i.e., preventing a survey taker from moving on without answering the question) so they could revisit the speech sample, but this meant that a small number of questions were missed (**7.2.1**). Some means of picking this up should have been used.

Also, the Base rating, while supported by those of five other experts, was mostly the work of one rater – myself. A collective rating would have preferable, but inter-rater reliability (IRR) could be not found amongst the five Experienced Raters (**7.2.1**). Since they had rated only 16 samples each, this meant that the whole Base Rating could not be used (except for models using only the Global Measures). However, the addition of my rating resulted in an improvement in the IRR scores and demonstrated that I was rating along the same general lines. Future surveys of this types should have independent raters for all the data, and more preparation should be used to ensure raters are using the same benchmarks throughout.

8.6 IMPLICATIONS

A lack of guidance in pronunciation teaching in general and work-focused ESOL classes and a focus on segmentals has meant that learners have received very limited support to develop skills beyond some phonemic accuracy. Evidence has been presented previously that supra-segmentals should be included in pronunciation teaching, but the answer to RQ2 suggests a need to include not only segmental, intonation and word stress features but also those associated with Fluency.

There are many textbooks on pronunciation which explain what to teach and how to teach it, but above all the teacher needs guidance on what emphasis to place on the various features of pronunciation and at what stage they should be taught. An examination of 12 ESL general-skills textbook series (Derwing et al., 2012) found that there is now some emphasis on Prosodic features, but little support for the teaching of Fluency features other than linking (included in Coarticulation). The outcome of the Employer Survey was that all features of pronunciation are likely to be predictors of Acceptability. These include Fluency, which is currently given little emphasis in teaching.

The Employers' conclusions about Personal Qualities may also offer insight into the importance of teaching the Prosodic and Fluency levels. Only the quality "Hardworking" was found to be a predictor of Acceptability, based on what the Migrants said and how they said it. When this variable was included in the model (**7.6.3**), the odds for Fluency and Prosody were reduced, suggesting that ratings were shifted from them to "hardworking". Speakers express their personality through higher level phonological features such as Variety and Intonation, which goes beyond mere intelligibility into being able to express enthusiasm and other emotions and intentions through appropriate pronunciation. It is likely that the Employers were responding to speakers who could express their enthusiasm, highlighting the importance of Prosody and Fluency features.

Teachers are accustomed to presenting material according to a hierarchy of difficulty, with the simplest features of grammar and vocabulary presented first and more complex features much later. Many teachers are likely to believe that this hierarchical approach applies to pronunciation as well, so it is crucial that this should be disposed of as another myth.

As Brown (2014, p. 6) pointed out, "the whole concept of pronunciation can be analysed as being composed of many components, all of which are present at the same time in speech." A speaker might use only the present tense and the first three hundred words, but cannot speak without simultaneously using segmentals, intonation, word stress and (by virtue of its presence or absence) *every other feature* of pronunciation. Pausing, for example, may be obligatory or not, appropriate or not, but is nevertheless a feature contributing to accentedness and intelligibility. The evidence of this thesis is that all levels of pronunciation are significant predictors of Acceptability, which suggests that they all need to be included in teaching from the earliest stages. Learners who plan to seek employment as soon as possible cannot afford to wait until they have reached the Advanced level before they are given advice on developing Fluency features, and as pronunciation is partly a matter of mechanical habits it is far better if they start developing them at the earliest stages.

Teaching Fluency – adding a top-down approach

The key questions for teaching higher level features such as Prosody and Fluency are, therefore, whether they can be introduced to low-level learners and, if so, how they can be included in a communicative context. Hieke (1985, p. 135) noted that Fluency was "not generally well learned or effectively taught" and this is still likely to be the case. Researchers have often suggested that pronunciation can be taught at the lowest levels, but there may be little or no consideration given to a formal development of Fluency (e.g. Pennington, 1996; Zielinski & Yates, 2014).

Those who reject the idea of teaching suprasegmentals to beginners are in a minority. G. Brown (1990, p. 158) opposed the idea of teaching Coarticulation, claiming that beginners should work from the bottom up and produce "single words or short sentences carefully and slowly enunciated." Dalton and Seidlhofer (1994, p. 115) dealt only with Coarticulation, warning against the "sudden occurrence" of learned assimilations in contrast to the learner's normal speech. This advice is valid, but does not obviate the overall development of Fluency features. On the other hand, Reed and Michaud (2015) illustrate the negative effect of introducing suprasegmentals too late in a language programme. Students "expressly rejected" the role of intonation (p.461) and refused to use the patterns outside of the classroom because they believed they were successfully communicating without them.

Many other writers have supported teaching or Prosody and Fluency from the start. Eskanazi (1999, p. 64) argues that "intonation is the glue that holds a message together" and so "It follows that prosody should be taught from the beginning." Nation and Newton (2009) explained that Fluency teaching can take place in all skills at any stage and should take up a quarter to a third of the programme. Sicola and Darcy (2015, pp. 483–4) concluded that "it is crucial to address the pronunciation needs of students at an early stage" and advocated "promoting automaticity of phonological processing" – in other words, actively teaching Fluency.

Derwing and Munro (2015, p. 101) reported on the evidence that exposure improves Fluency, but also noted "until L2 speakers are made aware of their difficulties, they do not know where to begin". Saito, Trofimovich and Isaacs (2015) found that ratings of low-level learners' comprehensibility correlated with features including Speed and Variety, suggesting the importance of teaching Fluency features at an early stage. Dauer and Browne (1992, p. 1) recommended that "the model of English we must teach [learners] is the fluent language as it is actually spoken in conversation" or "connected speech", and Demirezen (2016, p. 501) argued that faulty assimilation "spoils pronunciation" but his experimental work showed that assimilation can be learnt.

This thesis has not examined the specific role of thought groups or paratones (the flexible and meaningful use of brief pauses between groups of words). Paratones are a feature of Fluency, and according to Murphy (2013, p. 3), they are the highest priority of pronunciation teaching, "anchoring" most other components of pronunciation, and should be taught "early on".

Attempts have been made to show how Fluency teaching can be taught communicatively. Gatbonton and Segalowitz (1988), for example, describe automaticity, a component of Fluency – the ability to produce utterances rapidly (Speed) and smoothly (Fluidity) which can be learned through training to use formulaic speech or learned chunks. They fit this into a communicative framework by using multiple repetitions (not drills) followed by focused exercises. They also consider that such exercises can be used effectively with beginners. Temperley (1987) provided recommendations for teaching linking (Coarticulation), and Nation (1989) suggested activities which would improve Speed and Fluidity including the well-

known 4/3/2 in which each student gives the same recount to their partner in decreasing periods of time.

It is beyond the scope of this thesis to discuss teaching techniques. However, it does provide evidence that teachers should be aware of what would make their learners' speech acceptable for their communication goals, and suggests that covering all features of pronunciation rather than only a few would make their success more likely.

8.7 FURTHER RESEARCH

This thesis has investigated Acceptability in the context of New Zealand low or unskilled employment, but its results are not dependent on any feature of New Zealand pronunciation. It supports overseas researchers who have asserted that listeners pay attention to overall proficiency rather than individual features, which explains why Fluency appears to be the most significant predictor of Acceptability.

As noted in **6.1.1**, there have been few studies of this kind into employer acceptance of non-native speech. There is therefore scope for this kind of investigation to be repeated elsewhere in order to discover whether there are broad descriptions of Acceptability which can be applied in other countries or contexts. It is highly likely that the results would be different when rating migrants for professional positions (just as they were for dealing with customers): that when the communication stakes are higher, the bar for Acceptability will be raised, and all levels (Segmental, Prosody, Fluency) will become more important, because the level of performance needs to be raised for more complex communication.

However, I would offer the hypothesis that the results of this thesis will be found anywhere in the same employment context. While it would be ideal for all migrants and former refugees to become better qualified and enhance their status, income and contribution to society, the reality is that many cannot afford to wait before earning whatever they can. The ESOL teaching profession owes it to these learners to develop the language skills they urgently need, and my hope is that these findings will encourage teachers to review how they teach pronunciation.

This thesis proposes Fluency as a separate suprasegmental construct from Prosody, and claims that it consists not only of clearly temporal features such as Pausing but also features such as Variety, which clearly relate to Prosody as well. Labels and constructs should have some practical application, and in this thesis Fluency appears as one of the three distinct Factors significantly predicting Acceptability. It also captures an area of pronunciation teaching which is largely neglected. Practical research into the effects of including Fluency features in pronunciation teaching might confirm the proposal that they are necessary to raise learners' acceptability when applying for work. In addition, teacher guides frequently include teaching techniques for segmental and prosodic features, but rarely offer techniques

for developing fluency, perhaps on the assumption Fluency is simply automaticity gained by practice rather than through overt teaching. ESOL teachers would appreciate more support with Fluency techniques.

Researchers may wish to note the limitations in this research to fill in some of the gaps, such as the role of Speed in predicting Acceptability.

As suggested in **7.7.1**, qualitative research into actual pronunciation problems might contribute to the language learning goals of ESOL classes. The choice of accent, as described in **3.1.2** and **3.1.2**, is also worth investigating, since there is little point in making the effort to teach features which students do not want to learn.

9   GLOSSARY

9.1   DEFINITIONS WITH REFERENCES

Term	Definition
Accent	**Definition used in this thesis: *A set of auditory speech features which identify the speaker's geographic, social or language origin or their difference from local speakers.*** *Second language (L2) 'accent' comprises the noticeable differences between L2 users' speech and a local language variety* (Derwing & Munro, 2014a). *For most people, accent is a dustbin category: it includes all the technical meanings, and a more general and subjective one: accent is how the other speaks* (Lippi-Green, 2009, p. 165).
Accentedness	**A Global Measure** **Definition used in this thesis: *A global measure of how a native speaker rates an accent in terms of degree of difference from local accents.*** *Accentedness is a judgment of how much one's speech differs phonologically from the local variety. It is often measured on a Likert-like scale (e.g., 1 = no accent; 9 = extremely heavy accent)* (Derwing, 2010, p. 29)
Acceptability	**A Global Measure** **Definition used in this thesis: *The extent to which a NS considers that a NNS's pronunciation does not hinder participation in communication for a specific purpose.*** *The extent to which language deviates from native speaker norms* (Ludwig, 1982, p. 277). *Acceptability is generally considered to relate to a norm or a mutually agreed standard specified by a particular social group in a social context. Both group and context determine how far from the norm something (e.g. an accent) can deviate in order to still be acceptable. It is essential to define both group and context in the study.* (Beinhoff, 2013, p. 37) *Acceptability emphasises the norms associated with standards forms, and, by extension, the norms associated with social power* (Levis, 2008). *The degree to which a NS speaker judges a language sample as meeting implicit or explicit target norms* (Rifkin & Roberts, 1995, p. 522).
Basis of Articulation	*A language-specific tongue posture and a key factor systematically influencing vowel quality* (Borissoff, 2012, p. 9) *as distinguished from Articulatory setting or Voice Quality which are mutable.*
Category	*A "universal phonetic category" is the same as a feature such as voicing or nasality* (Chomsky & Halle, 1968, p. 14). *A phonetic distinction along a continuum such as VOT* (Flege & Eefting, 1987).
Coarticulation	*Blending and overlapping of sounds in fluent speech; shortenings, weakenings, and assimilations* (Pennington & Richards, 1986, pp. 210, 218).
Communicative ability	*The level of success a non-native speaker would experience in clearly communicating with a native speaker* (Kraut & Wulff, 2013, p. 250).

Term	Definition
Comprehensi-bility	**A Global Measure** **Definition used in this thesis:** *The listener's experience of how difficult the speech is to understand* (Munro, 2011, p. 9). Comprehensibility = (in order of weighting) pronunciation + grammar + familiarity with topic + familiarity with person + familiarity with speaker's native language + fluency + social factors … (Varonis & Gass, 1982, p. 132). *The extent to which a speaker's message is <u>understandable</u>* [from global language features] (J. Field, 2003, p. 35).
Error gravity / Error Hierar-chy	*The ranking of errors by comprehensibility, irritation or other criteria* (Gynan, 1985, p. 160). Error gravity research is based on the assumption that *some linguistic errors are more serious than others in terms of disrupting a NS's comprehension of a nonnative speaker's (NNS's) message and that these error types can be identified* (Rifkin & Roberts, 1995, p. 512).
Fluency	**One of the three levels of pronunciation.** **Definition used in this thesis:** ***A global measure of speech competence comprising speed, smoothness, flow, variety, coarticulation, and pausing.*** *An automatic procedural skill on the part of the speaker … and a perceptual phenomenon in the listener* (Derwing et al., 2008, p. 360).
Fluidity	**Definition used in this thesis:** ***A measure of speech flow based on appropriate use of pausing and lack of disfluencies.*** Munro and Derwing (2015, p. 14) define Fluency as *absence of dysfluencies such as filled and unfilled pauses, false starts, self-repetitions*. The term Fluency is reserved in this thesis for the highest level of pronunciation, but Munro and Derwing (op. cit.) equate fluency with "fluidity of speech" so the term *fluidity* is used in this thesis to avoid ambiguity.
Functional Load	(1) *A measure of the number of minimal pairs which can be found for a given opposition.* (2) *A measure of the work which two phonemes (or a distinctive feature) do in keeping utterances apart* (R. King, 1967, p. 831)
Hesitation phenomena	Silent pauses, filled pauses, repairs, repeats and false starts (Rose, 2013)
Intelligibility	**A Global Measure** **Definition used in this thesis:** *the extent to which a speaker's message is actually understood by a listener* (Derwing & Munro, 2005, p. 385, 2009b, p. 478). *The extent to which the speech signal produced by the speaker can be identified by the listener as the words the speaker intended to produce* (Zielinski, 2006, p. 3).
Intonation	*The linguistically meaningful use of vocal pitch level and pitch movement in phrases* (Kang et al., 2010). *Intonation is the glue that holds a message together. It indicates which words are important, disambiguates parts of sentences, and enhances the meaning with style and emotion* (Eskenazi, 1999, p. 64).

Term	Definition
Markedness	*Any phenomenon A in some language is more marked relative to some other phenomenon B if, cross-linguistically, the presence of A in a language implies the presence of B, but the presence of B does not necessarily imply the presence of A* (Eckman, 1981, p. 19).
Nativeness	**A Global Measure** *Global foreign accent may be defined as overall pronunciation proficiency in a second language, or how native-like the accent is. Foreign accent ratings are necessarily based on native speaker judgments of authenticity (nativeness)* (Major, 1987, p. 157).
Pace	*The number of prominent syllables per run – a measure of acoustic fluency* (Kang, 2008b, p. 108).
Phonation-time ratio	*The percentage of time from the sample producing audible speech, including filled pauses (a measure of speaking rate)* (Kang et al., 2010, p. 302).
Pitch	*... pitch is the property of a sound which allows a listener to place it on a scale from high to low. ... . It rises when the air flows out of the lungs at a greater rate than normal* (Scarcella & Oxford, 1994, p. 222).
Pronunciation	*The practice and meaningful use of TL [target language] phonological features in speaking, supported by practice in interpreting those phonological features in TL discourse that one hears* (Burgess & Spencer, 2000, p. 191).
Prosody	**One of the three levels of pronunciation.** *Variation in pitch, loudness, tempo and rhythm, as encountered in any use of spoken language* (Crystal, 1999, p. 275).
Pruned syllables	*the total number of syllables produced excluding dysfluencies (e.g., filled pauses, repetitions, self-corrections, false starts)* (Isaacs & Trofimovich, 2012, p. 484).
Reverse Linguistic Stereotyping (RLS)	*RLS suggests that listeners ascribe stereotyped characteristics to speech (i.e., listeners hear nonstandard accent where none may be present, for example) based on information supplied about the speaker's social identity even before hearing the speech* (Kang & Rubin, 2014) .
Rhythm	*The measured movement or musical flow of language. English has a rhythm in which stressed syllables normally occur at regular time intervals* (Scarcella & Oxford, 1994, p. 222).
Segmental	**One of the three levels of pronunciation.** *To do with vowels and consonants* (A Brown, 2014, p. 300).
Space	*The proportion of stressed words to the total number of words* (Kormos & Dénes, 2004, p. 152).
Speaking rate / Speech rate	*Average number of syllables per second produced within each utterance* (Baese-Berk & Morrill, 2015, p. 2).

Term	Definition
Stress	*... the use of extra respiratory energy. It applies to both individual sounds as well as to whole syllables, and involves pushing out air from the lungs and sometimes increasing the activity of the laryngeal muscles. There are several different speech signals for stress. These include loudness, pitch, and vowel length.* (Scarcella & Oxford, 1994, p. 221). *A stressed syllable is louder, longer, clearer, and higher pitched than un-stressed syllables. .... It is the vowel reduction that makes the difference between stressed and unstressed syllables very clear in English.* (C. F. Chen, Fan, & Lin, 1996)
Stress-timed rhythm	*… said to occur when stressed syllables occur in the stream of speech isochronously, i.e. at equal intervals of time* (Taylor, 1981, p. 219).
Suprasegmen-tal	*To do with features including voice quality, rhythm, intonation and stress. Suprasegmentals extend over stretches longer than a segment* (A Brown, 2014, p. 300). In this thesis, suprasegmentals are divided into Prosody and Fluency.
Unstress	The opposite of stress (Barrera-Pardo, 2008). *The ability to unstress vowels is a necessary, but not sufficient, condition for vowel reduction* (Flege & Bohn, 1989, p. 35).
Voice quality Articulatory setting (AS)	*Those features of accent that result from the characteristic disposition & use of the articulatory organs by speakers of a particular language* (Thornbury, 1993). However, this is distinguished by Borissoff (2012) as separate from **Basis of articulation** q.v.
Willingness to communicate (WTC)	*… the stable predisposition to talk in various circumstances ... a personal-ity trait* (MacIntyre et al., 1998, p. 546)

9.2   ABBREVIATIONS AND COMMONLY USED TERMS.

Terms specific to this thesis are in bold.

Term	Definition
3-Factor Model	The basic statistical model produced by the data resulting from the Employer Survey. Pronunciation variables were grouped into three factors, which were used to predict Employer Acceptability.
Asian	New Zealanders with only a vague idea of someone's origins tend to group all peoples from Eastern Asia (Chinese, Japanese, South East Asian, even Indonesian or Philippine) together as "Asians". This term does not include South Asians, who are likely to be seen as "Indians".
Base rating	The ratings by the group of Experienced Raters of the speech samples used in this thesis.
EAL	English as an additional language; another term for ESOL
EFL	English as a foreign language: English (courses) learnt in a place where it is a foreign language
ELL	English language learner

Term	Definition
Employer Survey	The survey used in this thesis to obtain ratings of Migrants by Employers.
Employers	Respondents in the survey used for this thesis.
ESANA	Immigrants from Europe (including Russia), South Africa and North America
ESL	English as a second language: English (courses) learnt in an English-speaking environment.
ESOL	English for Speakers of Other Languages: Usually English (courses) learnt in an English-speaking environment. This is the general term used in New Zealand for any English courses for non-native speakers.
Experienced Raters	Experienced professionals used to produce a baseline rating of the Migrants' speech samples used in this thesis.
IL	Interlanguage: *A language system created by someone who is in the process of learning a foreign language* (Crystal, 1999, p. 165).
ITA	(In the United States) International Teaching Assistant
L1	Language one: A person's mother tongue or language spoken at home
L2	Language two: The language being learnt, whether it is a second or addition language.
Maori	The indigenous race of New Zealand, also written as Māori.
Migrants	The participants interviewed in the survey used for this thesis to provide samples of non-native speech.
NNS	Non-native speaker.
NS	Native speaker.
NZ	New Zealand
NZCEL	New Zealand Certificate in English Language, the national ESOL qualification.
Pakeha	Fair-skinned New Zealander of European origin.
Respondent	In a survey, the person answering the survey questions.
RP	Received Pronunciation, a prestige British accent often taught or described as a model.
SAE	Standard American English, often taught or described as a model.
SD	Standard Deviation, a measure of how widely spread the responses are.
TA	(In the United States) Teaching Assistant
TL	Target language: The language being learnt, the same as L2.

10 REFERENCES

Abercrombie, D. (1991). Teaching Pronunciation. In A. Brown (Ed.), *Teaching English pronunciation: a book of readings* (pp. 92–95). London ; New York: Routledge.

Acton, W. (1984). Changing fossilized pronunciation. *TESOL Quarterly*, *18*(1), 71–85. https://doi.org/10.2307/3586336

Acton, W., & Finney, R. (1978). Spelling in TESL: stress cues to vowel quality. *TESOL Quarterly*, *12*(2), 163–175. https://doi.org/10.2307/3585607

Akram, A., & Qureshi, A. (2014). The Role of Features of Connected Speech in Teaching English Pronunciation. *International Journal of English and Education*, *3*(3), 230–240.

Albrechtsen, D., Henriksen, B., & Faerch, C. (1980). Native speaker reactions to learners' spoken interlanguage. *Language Learning*, *30*(2), 365–396. https://doi.org/10.1111/j.1467-1770.1980.tb00324.x

Altmann, H. (2006). *The perception and production of second language stress: A cross-linguistic experimental study* (PhD Thesis). University of Delaware, Newark.

Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of non-native pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning*, *42*, 529–555. https://doi.org/10.1111/j.1467-1770.1992.tb01043.x

Anderson-Hsieh, J., & Koehler, K. (1988). The effect of foreign accent and speaking rate on native speaker comprehension. *Language Learning*, *38*(4), 561–613. https://doi.org/10.1111/j.1467-1770.1988.tb00167.x

Arslan, L., & Field, J. (1997). A study of temporal features and frequency characteristics in American English foreign accent. *Journal of the Acoustical Society of America*, *102*(1), 28–40. https://doi.org/10.1121/1.419608

Aslin, R. (2014). Phonetic category learning and its influence on speech production. *Ecological Psychology*, *26*(1–2), 4–15. https://doi.org/10.1080/10407413.2014.874839

Baese-Berk, M., & Morrill, T. (2015). Speaking rate consistency in native and non-native speakers of English. *The Journal of the Acoustical Society of America*, *138*(3), EL223-EL228. https://doi.org/10.1121/1.4929622

Baetens Beardsmore, H. (1979). The recognition and tolerance level of bilingual speech. *Working Papers on Bilingualism*, (19), 116–128.

Ballard, L. (2013). *Student attitudes toward accentedness of native and non-native speaking English teachers* (M.A.). Michigan State University, Michigan, USA. Retrieved from http://search.proquest.com.helicon.vuw.ac.nz/pqdtglobal/docview/1353660942/abstract/C70900BC6B9045B8PQ/11?

Bansal, R. (1990). The pronunciation of English in India. In S. Ramsaran (Ed.), *Studies in the Pronunciation of English: A Commemorative Volume in Honour of A.C. Gimson* (pp. 219–230). London: Routledge.

Barrera-Pardo, D. (2008). The reality of stress-timing. *ELT Journal*, *62*(1), 11–17. https://doi.org/10.1093/elt/ccm073

Basnayake, A. (1999). *Employment experiences of Sri Lankan immigrants in New Zealand*. Auckland, N.Z.: Equal Employment Opportunities Trust. Retrieved from http://www.eeotrust.org.nz/content/docs/reports/emplmnt_srilankan_migrant.pdf

Bauer, L. (2010). *The linguistics student's handbook*. Edinburgh: Edinburgh Univ. Press.

Bauer, L., & Warren, P. (2008). New Zealand English: phonology. In K. Burridge & B. Kortmann (Eds.), *Varieties of English, Volume 3: The Pacific and Australasia* (1st ed., Vol. 3, pp. 39–63). Berlin ;New York: Mouton de Gruyter.

Bayard, D., Weatherall, A., Gallois, C., & Pittam, J. (2001). Pax Americana? Accent attitudinal evaluations in New Zealand, Australia and America. *Journal of Sociolinguistics*, *5*(1), 22–49. https://doi.org/10.1111/1467-9481.00136

Beckman, M., & Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonology*, *3*(1), 255–309.

Beebe, L. M. (1984). Myths about interlanguage phonology. *Studies in Descriptive Linguistics*, *13*, 51–61.

Beinhoff, B. (2013). *Perceiving identity through accent: attitudes towards non-native speakers and their accents in English*. New York: Peter Lang.

Bell, A., Harlow, R., & Starks, D. (Eds.). (2005). *Languages of New Zealand*. Wellington, N.Z.: Victoria University Press.

Bell, A., & Kuiper, K. (Eds.). (2000). *New Zealand English*. Wellington, N.Z.: Victoria University Press.

Benrabah, M. (1997). Word-stress-a source of unintelligibility in English. *IRAL*, *35*(3), 157–165. https://doi.org/10.1515/iral.1997.35.3.157

Bent, T., Bradlow, A., & Smith, B. (2007). Segmental errors in different words positions and their effects on intelligibility of non-native speech. In O. Bohn & M. Munro (Eds.), *Language experience in second language speech learning : In honor of James Emil Flege* (pp. 331–347). Amsterdam; Philadelphia: John Benjamins.

Best, C., McRoberts, G., & Sithole, N. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(3), 345–360. https://doi.org/10.1037/0096-1523.14.3.345

Binghadeer, N. (2008). An acoustic analysis of pitch range in the production of native and nonnative speakers of English. *Asian EFL Journal*, *10*(4), 96–113.

Bohn, O., & Flege, J. (1997). Perception and production of a new vowel category by second-language learners. In A. James & J. Leather (Eds.), *Second-language speech : structure and process* (pp. 53–74). Berlin; New York: Mouton de Gruyter.

Bond, Z., & Small, L. (1983). Voicing, vowel, and stress mispronunciations in continuous speech. *Attention, Perception, & Psychophysics*, *34*(5), 470–474. https://doi.org/10.3758/BF03203063

Borissoff, C. (2012). Basis of articulation or articulatory setting? *Speak Out!*, (46), 9–12.

Bosker, H. R., Pinget, A.-F., Quené, H., Sanders, T., & De Jong, N. H. (2012). What makes speech sound fluent? The contributions of pauses, speed and repairs. *Language Testing*, 159–175.

Boyer, T. (1996). Problems in paradise: Taiwanese immigrants to Auckland, NZ. *Asia Pacific Viewpoint*, *37*(1), 59–79.

Bradlow, A., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, *106*(2), 707–729. https://doi.org/10.1016/j.cognition.2007.04.005

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

Brown, A. (2006). *An examination of the rating process in the revised IELTS Speaking Test* (Vol. 6). IELTS Australia and British Council. Retrieved from https://band-score.ielts.org/pdf/Volume%206,%20Report%202.pdf

Brown, A. (2014). *Pronunciation and phonetics: a practical guide for English language teachers*. New York ; London: Routledge.

Brown, G. (1990). *Listening to spoken English* (2nd ed.). London: Longman.

Brown, J. (2001). What is an eigenvalue? *Shiken: JALT Testing & Evaluation SIG Newsletter*, *5*(1), 15–19.

Browning, G. (1982). *A Listener-Based Hierarchy of Acceptability for Non-Native Features of Oral English* (Ph.D.). University of California, Los Angeles, California USA.

Burgess, J., & Spencer, S. (2000). Phonology and pronunciation in integrated language teaching and teacher education. *System*, *28*, 191–215. https://doi.org/10.1016/S0346-251X(00)00007-5

Butcher, A., Spoonley, P., & Trlin, A. (2006). *Being accepted: The experience of discrimination and social exclusion by immigrants and refugees in New Zealand* (Research report No. 13) (p. 90). Palmerston North, N.Z.: Massey University. Retrieved from http://newsettlers.massey.ac.nz/publications.php?tab=0&PHPSESSID=4d0d077790b9db56f4e8f85a0072d8fd#

Carey, M., Mannell, R., & Dunn, P. (2011). Does a rater's familiarity with a candidate's pronunciation affect the rating in oral proficiency interviews? *Language Testing*, *28*(2), 201–219. https://doi.org/10.1177/0265532210393704

Cargile, A. C. (2000). Evaluations of employment suitability: Does accent always matter? *Journal of Employment Counseling*, *37*(3), 165–77. https://doi.org/10.1002/j.2161-1920.2000.tb00483.x

Carlisle, R. (2001). Syllable structure universals and second language acquisition. *IJES, International Journal of English Studies*, *1*(1), 1–19.

Carlson, H., & McHenry, M. (2006). Effect of accent and dialect on employability. *Journal of Employment Counseling*, *43*(2), 70–83. https://doi.org/10.1002/j.2161-1920.2006.tb00008.x

Catford, J. (1987). Phonetics and the teaching of pronunciation: a systemic description of English phonology. In J. Morley (Ed.), *Current perspectives on pronunciation: practices anchored in theory.* (pp. 87–100). Washington D.C.: TESOL. Retrieved from DOI:10.1017/s0025100303221514

Cauldwell, R. (2013). *Phonology for listening: teaching the stream of speech*. Birmingham: Speech in Action.

Cenoz, J. (2000). Pauses and Hesitation Phenomena in Second Language Production. *ITL, Review of Applied Linguistics*, *127–128*(Jan), 53–69.

Chan, A., & Li, D. (2000). English and Cantonese phonology in contrast: explaining Cantonese ESL learners' English pronunciation problems. *Language*, *13*(1), 67–85. https://doi.org/10.1080/07908310008666590

Chang, S., Morris, C., & Vokes, R. (2006). *Korean migrant families in Christchurch: Expectations and experiences*. Wellington, N.Z.: Families Commission. Retrieved from http://www.superu.govt.nz/sites/default/files/downloads/BS-korean-migrant-families.pdf

Changemakers Refugee Forum. (2012). *People with Refugee Backgrounds Can Do the Job: Refugee-background experiences of employment in Wellington.* (p. 40). Wellington, N.Z.: Changemakers Refugee Forum. Retrieved from http://crf.org.nz/sites/crf.org.nz/files/staff/ChangeMakers%20Employment%20Research%20Final_0.pdf

Checklin, M. (2012). What in the world do we know about word stress? A review of what it is and how to teach it. *TESOL In Context*, (Special Edition S3). Retrieved from http://www.tesol.org.au/Publications/Special-Editions

Chen, C. F., Fan, C., & Lin, H. (1996). A new perspective on teaching English pronunciation: rhythm. In *The proceedings of the fourth international symposium on English teaching* (pp. 24–41). Retrieved from http://eric.ed.gov/?id=ED397638

Chen, E., & Ward, R. (2013). *Employers' Role and Influence in Migration: A literature review*. Wellington, N.Z.: Ministry of Business, Innovation and Employment. Retrieved from http://www.dol.govt.nz/publications/research/employers-role-influence-migration-lit-review/lit-review.pdf

Chiswick, B., & Miller, P. (2005). Linguistic distance: A quantitative measure of the distance between English and other languages. *Journal of Multilingual and Multicultural Development*, *26*(1), 1–11. https://doi.org/10.1080/14790710508668395

Choi, L. J. (2016). Revisiting the issue of native speakerism: "I don"t want to speak like a native speaker of English'. *Language and Education*, *30*(1), 72–85. https://doi.org/10.1080/09500782.2015.1089887

Chomsky, N., & Halle, M. (1968). *Sound pattern of English*. New York: Harper & Row.

Christensen, R. (2011). Regression Models for Ordinal Data: Introducing R-package ordinal. r-project.org. Retrieved from ftp://ftp.musicbrainz.org/pub/cran/web/packages/ordinal/vignettes/clm_tutorial.pdf

Christiansen, T. (2014). Putting the accent on intelligibility: What constitutes "good" pronunciation in the context of English as a lingua franca? A case study of learners of different L1s. *Textus*, (1/2014). https://doi.org/10.7370/77484

Chun, D. M. (1988). The neglected role of intonation in communicative competence and proficiency. *The Modern Language Journal*, *72*(3), 295–303.

Clark, H., & Fox Tree, J. (2002). Using uh and um in spontaneous speaking. *Cognition*, *84*(1), 73–111. https://doi.org/10.1016/S0010-0277(02)00017-3

Clarke, C., & Garrett, M. (2004). Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America*, *116*(6), 3647. https://doi.org/10.1121/1.1815131

Cotterall, S. (2008). Autonomy and good language learners. In C. Griffiths (Ed.), *Lessons from good language learners* (pp. 110–120). Cambridge University Press.

Couper, G. (2003). The value of an explicit pronunciation syllabus in ESOL teaching. *Prospect*, *18*(3), 53–70.

Couper, G. (2006). The short and long-term effects of pronunciation instruction. *Prospect*, *21*(1), 46–66.

Crowther, D., Trofimovich, P., Saito, K., & Isaacs, T. (2015). Second language comprehensibility revisited: Investigating the effects of learner background. *TESOL Quarterly*, *49*(4), 814–837. https://doi.org/10.1002/tesq.203

Crystal, D. (1999). *The Penguin dictionary of language* (2. ed). London: Penguin Books.

Cucchiarini, C., Strik, H., & Boves, L. (2000). Quantitative assessment of second language learners' fluency by means of automatic speech recognition technology. *Journal of the Acoustical Society of America*, *107*(2), 989–999. https://doi.org/10.1121/1.428279

Cummins, J. (1999). BICS and CALP: Clarifying the Distinction. Retrieved from http://eric.ed.gov/?id=ED438551

Cunningham, A.-C. (2013). *Increasing teachers' knowledge of ELL's linguistic and cultural backgrounds through the use of a flip chart describing the differences between English and ten languages* (Masters). State University of New York, Fredonia, New York USA. Retrieved from https://dspace.sunyconnect.suny.edu/bitstream/handle/1951/63032/Amy_Catherine_Cunningham_Master_Project_May2013.pdf?sequence=1&isAllowed=y

Cunningham, U. (2009). Phonetic correlates of unintelligibility in Vietnamese-accented English. In *Proceedings, FONETIK 2009* (p. 108). Dept. of Linguistics, Stockholm University. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.160.7742&rep=rep1&type=pdf#page=108

Cutler, A., & McQueen, J. (1995). The recognition of lexical units in speech. In B. de Gelder & J. Morais (Eds.), *Speech and reading: a comparative approach* (pp. 33–48). Hove, East Sussex: Erlbaum Taylor and Francis.

Daldy, B., Poot, J., Roskruge, M., & van Dyk, M. (2011). Immigrant integration and workplace discrimination in New Zealand. *Integration of Immigrants Programme Working Paper*. Retrieved from http://econfin.massey.ac.nz/school/seminar%20papers/Albany/2012/roskruge-1.pdf

Dalton, C., & Seidlhofer, B. (1994). *Pronunciation*. Oxford: OUP.

Dauer, R., & Browne, S. (1992). Teaching the pronunciation of connected speech. Presented at the Annual Meeting of TESOL, Vancouver: TESOL International Association. Retrieved from http://eric.ed.gov/?id=ED354777

Davies, A. (2005). *A glossary of applied linguistics*. Edinburgh: Edinburgh University Press.

De Bot, K., & Mailfert, K. (1982). The teaching of intonation: fundamental research and classroom applications. *TESOL Quarterly*, *16*(1), 71–77. https://doi.org/10.2307/3586564

De la Zerda, N., & Hopper, R. (1979). Employment interviewers' reactions to Mexican American speech. *Communications Monographs*, *46*(2), 126–134. https://doi.org/10.1080/03637757909375998

Demirezen, M. (2016). Assimilation as a co-articulation producer in words and pronunciation problems for Turkish English teachers. *Educational Sciences: Theory & Practice*, *16*(2), 477–509. https://doi.org/10.12738/estp.2016.2.0235

Derwing, T. (2010). Utopian goals for pronunciation teaching. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 1st Pronunciation in Second Language Learning and Teaching Conference,* (pp. 24–37). Ames, Iowa: Iowa State University.

Derwing, T., Diepenbroek, L., & Foote, J. (2012). How well do general-skills ESL textbooks address pronunciation? *TESL Canada Journal*, *30*(1), 22–44.

Derwing, T., & Munro, M. (1997). Accent, intelligibility, and comprehensibility: Evidence from four l1s. *Studies in Second Language Acquisition*, *19*(1), 1–16. https://doi.org/10.1017/S0272263197001010

Derwing, T., & Munro, M. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly*, *39*, 379–397. https://doi.org/10.2307/3588486

Derwing, T., & Munro, M. (2009a). Comprehensibility as a factor in listener interaction preferences: implications for the workplace. *The Canadian Modern Language Review*, *66*(2), 181–202. https://doi.org/10.3138%2Fcmlr.66.2.181

Derwing, T., & Munro, M. (2009b). Putting accent in its place: Rethinking obstacles to communication. *Language Teaching*, *42*(4), 476–490. https://doi.org/10.1017/S026144480800551X

Derwing, T., & Munro, M. (2011). Intelligibility, comprehensibility and accent: their relevance to pronunciation teaching. *Speak Out!*, (45), 4–8.

Derwing, T., & Munro, M. (2014a). Accent and intelligibility: cracking the conundrum. *Speak Out!*, (50), 12–16.

Derwing, T., & Munro, M. (2014b). Once you have been speaking a second language for years, it's too late to change your pronunciation. In L. Grant & D. Brinton (Eds.), *Pronunciation myths: applying second language research to classroom teaching* (pp. 34–55). Ann Arbor: University of Michigan Press.

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

Derwing, T., Munro, M., & Thomson, R. (2008). A longitudinal study of ESL learners' fluency and comprehensibility development. *Applied Linguistics*, *29*(3). https://doi.org/10.1093/applin/amm041

Derwing, T., Munro, M., & Wiebe, G. (1997). Pronunciation instruction for fossilized learners: Can it help? *Applied Language Learning*, *8*(2), 217–35.

Derwing, T., Munro, M., & Wiebe, G. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language Learning*, *48*(3), 393–410. https://doi.org/10.1111/0023-8333.00047

Derwing, T., & Rossiter, M. (2002). ESL learners' perceptions of their pronunciation needs and strategies. *System*, *30*(2), 155–166. https://doi.org/10.1016/S0346-251X(02)00012-X

Derwing, T., & Rossiter, M. (2003). The effects of pronunciation instruction on the accuracy, fluency, and complexity of L2 accented speech. *Applied Language Learning*, *13*, 1–17.

Derwing, T., Rossiter, M., Munro, M., & Thomson, R. (2004). Second language fluency: Judgments on different tasks. *Language Learning*, *54*(4), 655–679. https://doi.org/10.1111/j.1467-9922.2004.00282.x/abstract

Deterding, D. (2010). Norms for pronunciation in Southeast Asia. *World Englishes*, *29*(3), 364–377. https://doi.org/10.1111/j.1467-971X.2010.01660.x

Devereux, M. (2004). Just an accent triggers abuse. *New Zealand Herald*. Auckland, N.Z. Retrieved from http://www.nzherald.co.nz/nz/news/article.cfm?c_id=1&objectid=3565160

Deverson, T., & Kennedy, G. (Eds.). (2005). *The New Zealand Oxford dictionary*. Auckland, N.Z..;South Melbourne, Vic.: OUP.

Dickerson, W. (2010). Walking the walk: Integrating the story of English phonology. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 1st Pronunciation in Second Language Learning and Teaching Conference* (pp. 10–23). Ames, Iowa: Iowa State University.

Dickerson, W. (2015). A nail in the coffin of stress-timed rhythm. In J. Levis, R. Mohammed, M. Qian, & Z. Zhou (Eds.), *Proceedings of the 6th annual Pronunciation in Second Language Learning and Teaching Conference* (pp. 186–197). Santa Barbara, CA, USA: Iowa State University. Retrieved from https://www.academia.edu/16793567/Proceedings_of_the_6th_Pronunciation_in_Second_Language_Learning_and_Teaching_Conference?auto=download&campaign=weekly_digest

Dörnyei, Z. (2007). *Research methods in applied linguistics: quantitative, qualitative, and mixed methodologies*. Oxford ; New York, N.Y: OUP.

Duguid, A. (2001). Italian Speakers. In M. Swan & B. Smith (Eds.), *Learner English: a teacher's guide to interference and other problems* (2nd ed, pp. 73–89). Cambridge ; New York: CUP.

Dunstan, S., Shorland, P., & Dibley, R. (2004). *Refugee Voices: A journey towards resettlement*. Wellington, N.Z.: NZ Immigration Service: Department of Labour. Retrieved from http://www.dol.govt.nz/research/migration/refugees/refugeevoices/RefugeevoicesFullDocument.pdf

Earle, D. (2009). *The effect of first language and education on literacy, employment and income: an analysis from the Adult Literacy and Life Skills survey*. Wellington, N.Z.: Tertiary Sector Performance Analysis and Reporting, Strategy and System Performance, Ministry of Education.

Eckman, F. (1981). On predicting phonological difficulty in second language acquisition. *Studies in Second Language Acquisition*, *4*(1), 18–30. https://doi.org/10.1017/S0272263100004253

Eckman, F. (1987). Markedness and the contrastive analysis hypothesis. In G. Ioup & S. Weinberger (Eds.), *Interlanguage phonology: the acquisition of a second language sound system* (pp. 55–69). Cambridge, MA: Newbury House Publishers.

Eckman, F. (1991). The structural conformity hypothesis and the acquisition of consonant clusters in the interlanguage of ESL learners. *Studies in Second Language Acquisition*, *13*(1), 23–41. https://doi.org/10.1017/S0272263100009700

Eckman, F. (2008). Typological markedness and second language phonology. In *Phonology and second language acquisition* (Vol. 36, pp. 95–115).

Eisenstein, M. (1983). Native reactions to non-native speech: A review of empirical research. *Studies in Second Language Acquisition*, *5*(2), 160–176. https://doi.org/10.1017/S0272263100004836

El-Hassan, S. (1994). English accentuation and vowel quality as pronounced by Arabs: a ped-agogic statement. *Papers and Studies in Contrastive Linguistics*, *29*, 205–216.

Eskenazi, M. (1999). Using automatic speech processing for foreign language pronunciation tutoring: some issues and a prototype. *Language Learning and Technology*, *2*(2), 62–76.

Fayer, J., & Krasinski, E. (1987). Native and nonnative judgments of intelligibility and irrita-tion. *Language Learning*, *37*(3), 313–326. https://doi.org/10.1111/j.1467-1770.1987.tb00573.x

Fayer, J., & Krasinski, E. (1995). Perception of hesitation in nonnative speech. *Bilingual Re-view/La Revista Bilingüe*, 114–121.

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R* (Kindle). London ; Thou-sand Oaks, Calif: Sage.

Field, J. (2003). The fuzzy notion of "intelligibility": A headache for pronunciation teachers and oral testers. *IATEFL Special Interest Groups Newsletter*, 34–38.

Field, J. (2005). Intelligibility and the listener: the role of lexical stress. *TESOL Quarterly*, *39*(3), 399–423. https://doi.org/10.2307/3588487

Flege, J. (1980). Phonetic approximation in second language acquisition. *Language Learning*, *30*(1), 117–134. https://doi.org/10.1111/j.1467-1770.1980.tb00154.x

Flege, J. (1989). Chinese subjects' perception of the word-final English/t/-/d/contrast: Per-formance before and after training. *Journal of the Acoustical Society of America*, *86*(5), 1684–1697. https://doi.org/10.1121/1.398599

Flege, J. (1993). Production and perception of a novel, second-language phonetic contrast. *Journal of the Acoustical Society of America*, *93*(3), 1589–1608. https://doi.org/10.1121/1.406818

Flege, J. (1995). Two procedures for training a novel second language phonetic contrast. *Ap-plied Psycholinguistics*, *16*(4), 425–42.

Flege, J. (2003). A method for assessing the perception of vowels in a second language. In *Issues in clinical Linguistics*. Padova, Italy: Unipress. Retrieved from http://www.jim-flege.com/files/Flege_in_Fava_2003.pdf

Flege, J., & Bohn, O. (1989). An instrumental study of vowel reduction and stress placement in Spanish-accented English. *Studies in Second Language Acquisition*, *11*, 35–92. https://doi.org/10.1017/S0272263100007828

Flege, J., & Eefting, W. (1987). Production and perception of English stops by native Spanish speakers. *Journal of Phonetics*.

Fletcher, M. (1999). *Migrant settlement: a review of the literature and its relevance to New Zealand* (Immigration Research programme). Wellington, N.Z.: NZ Immigration Ser-vice. Retrieved from http://www.immigration.govt.nz/NR/rdonlyres/D98843C7-1DEA-4558-AC4B-380616EF7D68/0/migrant_settlement_literature_review_sept_1999.pdf

Forsyte Research. (1998). *Migrants' experiences of the English language bond*. Wellington, N.Z.: NZ Immigration Service. Retrieved from http://victoria.lconz.ac.nz/vwebv/hold-ingsInfo?bibId=686817

Fraser, H. (2000). Coordinating improvements in pronunciation teaching for adult learners of English as a second language. *DETYA (Australia National Training Authority Adult Literacy Project)*, *31*. Retrieved from http://metz.une.edu.au/%7ehfraser/docs/HF%5fANTA%5fREPORT.pdf

Fuchs, R. (2015). *Speech rhythm in varieties of English: evidence from educated Indian English and British English*. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=1072276

Gallego, J. (1990). The intelligibility of three nonnative English-speaking teaching assistants: an analysis of student-reported communication breakdowns. *Issues in Applied Linguistics*, *1*(2), 219–237.

Gallois, C., & Callan, V. (1981). Personality impressions elicited by accented English speech. *Journal of Cross-Cultural Psychology*, *12*(3), 347–359. https://doi.org/10.1177/0022022181123006

Gardner, R., & Maclntyre, P. (1993). On the measurement of affective variables in second language learning. *Language Learning*, *43*(2), 157–794. https://doi.org/10.1111/j.1467-1770.1992.tb00714.x

Gass, S., & Varonis, E. (1984). The effect of familiarity on the comprehensibility of nonnative speech. *Language Learning*, *34*, 65–89. https://doi.org/10.1111/j.1467-1770.1984.tb00996.x

Gatbonton, E., & Segalowitz, N. (1988). Creative automatization: Principles for promoting fluency within a communicative framework. *TESOL Quarterly*, *22*(3), 473–492. https://doi.org/10.2307/3587290

Gatbonton, E., Trofimovich, P., & Magid, M. (2005). Learners' ethnic group affiliation and L2 pronunciation accuracy: a sociolinguistic investigation. *TESOL Quarterly*, *39*(3), 489–511. https://doi.org/10.2307/3588491

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge: CUP.

Gifford, P. (2009). Bickerstaff set rules for sports talk on radio. Retrieved March 20, 2016, from http://www.stuff.co.nz/sunday-star-times/sport/3048700/Bickerstaff-set-rules-for-sports-talk-on-radio

Gilbert, J. (1987). Pronunciation and listening comprehension. In J. Morley (Ed.), *Current perspectives on pronunciation: practices anchored in theory.* (pp. 29–39). Washington D. C.: TESOL.

Gilbert, J. (2014). Intonation is hard to teach. In L. J. Grant & D. Brinton (Eds.), *Pronunciation myths: applying second language research to classroom teaching* (pp. 107–136). Ann Arbor: University of Michigan Press.

Gill, M. (1994). Accent and stereotypes: their effect on perceptions of teachers and lecture comprehension. *Journal of Applied Communication Research*, *22*(4), 348. https://doi.org/10.1080/00909889409365409

Ginther, A., Dimova, S., & Yang, R. (2010). Conceptual and empirical relationships between temporal measures of fluency and oral English proficiency with implications for automated scoring. *Language Testing*, *27*(3), 379–399. https://doi.org/10.1177/0265532210364407

Gluszek, A., & Dovidio, J. (2010). The way they speak: a social psychological perspective on the stigma of nonnative accents in communication. *Personality and Social Psychology Review*, *14*(2), 214–237. https://doi.org/10.1177/1088868309359288

Gluszek, A., & Hansen, K. (2016). What does speaking with a foreign accent mean? Content analysis of newspaper articles. *ResearchGate*. https://doi.org/10.13140/RG.2.1.1912.3281

Gnevsheva, K. (2015). Style-shifting and intra-speaker variation in the vowel production of nonnative speakers of New Zealand English. *Journal of Second Language Pronunciation*, *1*(2), 135–156. https://doi.org/10.1075/jslp.1.2.01gne

Golombek, P., & Jordan, S. (2005). Becoming "black lambs" not "parrots": a poststructuralist orientation to intelligibility and identity. *TESOL Quarterly*, *39*(3), 513–533. https://doi.org/10.2307/3588492

Gosling, S., Vazire, S., Srivastava, S., & John, O. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. *American Psychologist*, *59*(2), 93–104. https://doi.org/10.1037/0003-066X.59.2.93

Green, A. (2005). Word, foot, and syllable structure in Burmese. *Studies in Burmese Linguistics*, *570*, 1–24.

Griffen, T. (1991). A non-segmental approach to the teaching of pronunciation. In A. Brown (Ed.), *Teaching English pronunciation: a book of readings* (pp. 178–193). London ; New York: Routledge.

Griffiths, C. (Ed.). (2008). *Lessons from good language learners*. Cambridge: CUP.

Grosjean, F. (1980). Temporal variables between languages. In H. Dechert & M. Raupach (Eds.), *Towards a Cross-Linguistic Assessment of Speech Productions* (pp. 39–53). Frankfurt: Lang.

Grover, C., Jamieson, D., & Dobrovolsky, M. (1987). Intonation in English, French and German: Perception and production. *Language and Speech*, *30*(3), 277–295. https://doi.org/10.1177/002383098703000307

Guion, S., & Pederson, E. (2007). The role of attention in phonetic learning. In O. Bohn & M. Munro (Eds.), *Language experience in second language speech learning : in honor of James Emil Flege* (pp. 57–77). Amsterdam; Philadelphia: John Benjamins.

Guiora, A., Beit-Hallahmi, B., Brannon, R., Dull, C., & Scovel, T. (1972). The effects of experimentally induced changes in ego states on pronunciation ability in a second language: An exploratory study. *Comprehensive Psychiatry*, *13*(5), 421–428. https://doi.org/10.1016/0010-440X(72)90083-1

Gynan, S. (1985). Comprehension, irritation and error hierarchies. *Hispania*, *68*(1), 160–165. https://doi.org/10.2307/341633

Hahn, L. (2004). Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly*, *38*, 201–233. https://doi.org/10.2307/3588378

Hahn, L., & Watt, P. (2011). (Un)intelligibility tales. In K. LeVelle & J. Levis (Eds.), *Proceedings of the 2nd Annual Pronunciation In Second Language Learning and Teaching Conference* (pp. 17–29). Ames, Iowa: Iowa State University. Retrieved from http://pslltconference.com/index.php?option=com_content&view=article&id=18&Itemid=16

Haile, A. (1987). Lexical Stress in Amharic. *Journal of Ethiopian Studies*, *20*, 19–43.

Hallgren, K. (2012). Computing inter-rater reliability for observational data: an overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, *8*(1), 23–34.

Hansen, J. (2001). Linguistics constraints on the acquisition of English syllable codas by native speakers of Mandarin Chinese. *Applied Linguistics*, *22*(3), 338–365. https://doi.org/10.1093/applin/22.3.338

Harrison, G. (2014). Accent and "othering" in the workplace. In J. Levis & A. Moyer (Eds.), *Social dynamics in second language accent* (pp. 255–272). Boston: De Gruyter Mouton.

Haspelmath, M. (2006). Against markedness (and what to replace it with). *Journal of Linguistics*, *42*(1), 25–70. https://doi.org/http://dx.doi.org/10.1017/S0022226705003683

Hay, J., Maclagan, M., & Gordon, E. (2008). *New Zealand English*. Edinburgh: Edinburgh University Press.

Hay, J., Nolan, A., & Drager, K. (2006). From fush to feesh: Exemplar priming in speech perception. *Linguistic Review*, *23*(3), 351–379. https://doi.org/10.1515/TLR.2006.014

Hayes-Harb, R. (2014). Acoustic-phonetic parameters in the perception of accent. In J. Levis & A. Moyer (Eds.), *Social dynamics in second language accent* (pp. 31–51). Boston: De Gruyter Mouton.

Hayward, K., & Hayward, R. (1992). Amharic. *Journal of the International Phonetic Association*, *22*(1–2), 48–52. https://doi.org/10.1017/S0025100300004606

Hekman, D., Aquino, K., Owens, B., Mitchell, T., Schilpzand, P., & Leavitt, K. (2010). An examination of whether and how racial and gender biases influence customer satisfaction. *Academy of Management Journal*, *53*(2), 238–264. https://doi.org/10.5465/AMJ.2010.49388763

Henderson, A. (2004). *The settlement experiences of immigrants (excluding refugees) in New Zealand:* Auckland, N.Z. Retrieved from http://www.immigration.govt.nz/NR/rdonlyres/D9BA1C72-E6BA-4C6C-A8FD-66373D55C3E2/0/MigrantLiteratureOverview.pdf

Henderson, A., Trlin, A., & Watts, N. (2001). Squandered skills? The employment problems of skilled Chinese immigants in New Zealand. In *Asian nationalism in an age of globalization* (1st ed., pp. 106–123). Richmond UK: Japan Library.

Hieke, A. (1985). A componental approach to oral fluency evaluation. *The Modern Language Journal*, *69*(2), 135–142. https://doi.org/10.2307/326503

Hinofotis, F., & Bailey, K. (1981). American undergraduates' reactions to the communication skills of foreign teaching assistants. In J. Fisher, M. Clarke, & J. Schachter (Eds.), *On TESOL '80: Building Bridges: Research and Practice in Teaching English as a Second Language.* (pp. 120–133). Washington, D.C: TESOL.

Ho, C., & Alcorso, C. (2004). Migrants and employment: challenging the success story. *Journal of Sociology*, *40*(3), 237–259. https://doi.org/10.1177/1440783304045721

Hoekje, B., & Linnell, K. (1994). "Authenticity" in language testing: Evaluating spoken language tests for international teaching assistants. *TESOL Quarterly*, *28*(1), 103–126. https://doi.org/10.2307/3587201

Hogan, D. (2013). *Transitioning difficulties of overseas trained nurses in New Zealand: What difficulties are experienced by overseas trained nurses in their tranisition in the NZ practice context?* Auckland University of Technology, Auckland, N.Z. Retrieved from

http://aut.researchgateway.ac.nz/bitstream/handle/10292/7047/Ho-ganD.pdf%3Fsequence%3D3&sa=X&scisig=AAGBfm28sjaGetVoQ8VpZzrr-wuGRV9bx9w&oi=scholaralrt

Honey, P. (1987). Vietnamese Speakers. In M. Swan & B. Smith (Eds.), *Learner English: a teacher's guide to interference and other problems* (1st ed., pp. 238–251). Cambridge ; New York: Cambridge University Press.

Honikman, B. (1964). Articulatory settings. In D. Abercrombie, D. Fry, P. MacCarthy, N. Scott, & J. Trim (Eds.), *In Honour of Daniel Jones* (pp. 73–84). London: Longman.

Hopper, R. (1977). Language Attitudes in the Employment Interview. *Communication Monographs*, *44*(4), 346. https://doi.org/10.1080/03637757709390145

Hopper, R., & Williams, F. (1973). Speech characteristics and employability. *Communications Monographs*, *40*(4), 296–302. https://doi.org/10.1080/03637757309375807

Hosoda, M., Nguyen, L., & Stone-Romero, E. (2012). The effect of Hispanic accents on employment decisions. *Journal of Managerial Psychology*, *27*(4), 347–364. https://doi.org/http://dx.doi.org.helicon.vuw.ac.nz/10.1108/02683941211220162

Hosoda, M., & Stone-Romero, E. (2010). The effects of foreign accents on employment-related decisions. *Journal of Managerial Psychology*, *25*(2), 113–132. https://doi.org/http://dx.doi.org.helicon.vuw.ac.nz/10.1108/02683941011019339

Hsieh, C. (2011). Rater effects in ITA testing: ESL teachers' versus American undergraduates' judgments of accentedness, comprehensibility, and oral proficiency. *SPAAN Fellow Working Papers in Second or Foreign Language Assessment*, *9*, 47–74.

Huang, M., & Singer, A. (1984). Ethnic group stereotyping by police and university students in New Zealand. *Journal of Psychology*, *116*(1), 89. https://doi.org/10.1080/00223980.1984.9923621

Hunt, G. (1999). Kiwis still favour White migrants over Asians and Pacific Islanders. *National Business Review*, *14*(24 January), 14.

Hunter, J. (2007). *Workplace language and communication needs : employers and immigrant employees*. Auckland, N.Z.: Massey University.

Hyman, H. (2001). *Foreign-accented adult ESL learners: Perceptions of their accent changes and employability qualifications* (Ph.D.). New York University, New York. Retrieved from http://search.proquest.com.helicon.vuw.ac.nz/pqdtglobal/docview/251142562/abstract/65DA844DD7B04D83PQ/36?accountid=14782

Immigration New Zealand. (2015a). English language requirements. Retrieved January 29, 2015, from http://www.dol.govt.nz/immigration/knowledgebase/item/1179

Immigration New Zealand. (2015b). *Refugee and protection status statistics pack*. Wellington, N.Z.: Immigration New Zealand. Retrieved from http://www.immigration.govt.nz/migrant/general/generalinformation/refugee-protection/refugee-statistics.htm

Immigration New Zealand. (2015c). *Refugee Family Support Category Statistics*. Wellington, N.Z.: Immigration New Zealand. Retrieved from http://www.immigration.govt.nz/migrant/general/generalinformation/refugee-protection/refugee-statistics.htm

Immigration New Zealand. (2015d). *Refugee Quota Arrivals Statistics Pack*. Wellington, N.Z.: Immigration New Zealand. Retrieved from http://www.immigration.govt.nz/migrant/general/generalinformation/refugee-protection/refugee-statistics.htm

Isaacs, T. (2008). Towards defining a valid assessment criterion of pronunciation proficiency in non-native English-speaking graduate students. *The Canadian Modern Language Review*, *64*(4), 555–580. https://doi.org/10.3138/cmlr.64.4.555

Isaacs, T. (2014). Assessing pronunciation. In A. Kunnan (Ed.), *The companion to language assessment* (pp. 140–155). Hoboken: John Wiley & Sons. Retrieved from 10.1002/9781118411360.wbcla012

Isaacs, T., & Thomson, R. (2013). Rater experience, rating scale length, and judgments of L2 pronunciation: Revisiting research conventions. *Language Assessment Quarterly*, *10*(2), 135–159. https://doi.org/10.1080/15434303.2013.769545

Isaacs, T., & Trofimovich, P. (2012). Deconstructing comprehensibility: Identifying the linguistic influences on listeners' L2 comprehensibility ratings. *Studies in Second Language Acquisition*, *34*(3), 475–505. https://doi.org/10.1017/S0272263112000150

Islam, A., & Parasnis, J. (2014). Native-immigrant wage gap across occupations: Evidence from Australia. *Discussion Paper*, *14*(14), 1–33.

Jayaraman, S. (2010). Acquisition of pronunciation of consonant clusters by Arabic speakers of English as a second language. *Sino-US English Teaching, 7*(1), 46–56.

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

Johansson, S. (1978a). Problems in studying the communicative effect of learners errors. *Studies in Second Language Acquisition*, *1*(1), 41–52. https://doi.org/10.1017/S027226310000070X

Johansson, S. (1978b). *Studies of error gravity: Native reactions to errors produced by Swedish learners of English*. Goteborg: Acta Universitatis Gothoburgensis.

Johnston, R., Adank, P., Trlin, A., & Spoonley, P. (2010). Immigration, multiculturalism and geography: Inter-group contact and attitudes to immigrants and cultural diversity in New Zealand. *Asian and Pacific Migration Journal*, *19*(3), 343–369. https://doi.org/10.1177/011719681001900303

Jones, R., & Evans, S. (1995). Teaching pronunciation through voice quality. *ELT Journal*, *49*(3), 244–251. https://doi.org/10.1093/elt/49.3.244

Juffs, A. (1990). Tone, syllable structure and interlanguage phonology: Chinese learners' stress errors. *IRAL*, *XXVIII*(2), 99–115. https://doi.org/10.1515/iral.1990.28.2.99

Jun, H. G., & Li, J. (2010). Factors in raters' perceptions of comprehensibility and accentedness. In K. LeVelle & Levis (Eds.), *Proceedings of the 1st Pronunciation in Second Language Learning and Teaching Conference,* (pp. 53–66). Ames, Iowa: Iowa State University.

Kachru, B. (1985). Standards, codification and sociolinguistic realism: the English language in the outer circle. In R. Quirk & H. Widdowson (Eds.), *English in the world : teaching and learning the language and literatures : papers of an international conference entitled "Progress in English Studies" held in London, 17-21 September 1984* (pp. 11–30). Cambridge ;New York: Cambridge University Press for the British Council.

Kahin, M. (2011). Areas of difficulty for the speakers of Somali in learning English as a foreign language. Retrieved from http://www.awdalpress.com/index/archives/6488

Kalin, R., & Rayko, D. (1978). Discrimination in evaluative judgments against foreign-accented job candidates. *Psychological Reports*. https://doi.org/10.2466/pr0.1978.43.3f.1203

Kang, O. (2008a). Ratings of L2 oral performance in English: Relative impact of rater characteristics and acoustic measures of accentedness. *SPAAN Fellow Working Papers in Second or Foreign Language Assessment*, *6*, 181.

Kang, O. (2008b). *Ratings of L2 oral performance in English: relative impact of rater characteristics and acoustic measures of accentedness*. University of Athens, Georgia, Georgia, USA.

Kang, O. (2010a). ESL learners' attitudes toward pronunciation instruction and varieties of English. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 1st Pronunciation in Second Language Learning and Teaching Conference,* (pp. 105–118). Ames, Iowa: Iowa State University.

Kang, O. (2010b). Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System*, *38*(2), 301–315. https://doi.org/10.1016/j.system.2010.01.005

Kang, O. (2012). Relative impact of pronunciation features on ratings of non-native speakers' oral proficiency. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference* (pp. 10–15). Simon Fraser University Vancouver: Ames, Iowa State University. Retrieved from http://jlevis.public.iastate.edu/pslltconference/4th%20Proceedings/Kang%20PSLLT%202012.pdf

Kang, O., & Moran, M. (2014). Functional loads of pronunciation features in nonnative speakers' oral assessment. *TESOL Quarterly*, *48*(1), 176–187. https://doi.org/10.1002/tesq.152

Kang, O., & Pickering, L. (2011). The role of objective measures of suprasegmental features in judgments of comprehensibility and oral proficiency in L2 spoken discourse. *Speak Out!*, (44). Retrieved from http://faculty.tamuc.edu/lpickering/Pdfs/Publish_1.pdf

Kang, O., & Rubin, D. (2014). Listener expectations, reverse linguistic stereotyping, and individual background factors in social judgments and oral performance assessment. In J. Levis & A. Moyer (Eds.), *Social dynamics in second language accent* (pp. 239–253). Boston: De Gruyter Mouton.

Kang, O., Rubin, D., & Pickering, L. (2010). Suprasegmental measures of accentedness and judgments of language learner proficiency in oral English. *The Modern Language Journal*, *94*(4), 554–566. https://doi.org/10.1111/j.1540-4781.2010.01091.x

Katzner, K. (2002). *The languages of the world* (3rd ed). London ; New York: Routledge.

Kennedy, S., & Trofimovich, P. (2008). Intelligibility, comprehensibility, and accentedness of l2 speech: The role of listener experience and semantic context. *Canadian Modern Language Review*, *64*(3), 459–489. https://doi.org/10.3138/cmlr.64.3.459

Kenny, K. (1996). *Language loss and the crisis of cognition between socio- and psycholinguistics*. Berlin ; New York: Mouton de Gruyter. Retrieved from http://site.ebrary.com.helicon.vuw.ac.nz/lib/vuw/detail.action?docID=10588325

King, M. (2003). *The Penguin history of New Zealand*. Auckland, N.Z.: Penguin Books.

King, R. (1967). Functional Load and Sound Change. *Language*, *43*(4), 831–852. https://doi.org/10.2307/411969

Kissling, E. (2014). What predicts the effectiveness of foreign-language pronunciation instruction? Investigating the role of perception and other individual differences. *Canadian Modern Language Review*, *70*(4), 532–558. https://doi.org/10.3138/cmlr.2161

Kitagawa, A., Watanabe, A., & Kumaki, H. (2011). A path to developing oral proficiency: Speaking rate, silent pauses and fillers. In *Proceedings of The 16th Conference of Pan-Pacific Association of Applied Linguistics*. Hong Kong: Pan-Pacific Association of Applied Linguistics.

Kjellin, O. (1999). Accent addition: Prosody and perception facilitates second langusge learning. In O. Fujimura, B. Joseph, & B. Palek (Eds.), *Proceedings of LP 98* (pp. 373–398). Prague: The Karolinum Press.

Klein, W. (1998). The contribution of second language acquisition research. *Language Learning*, *48*(4), 527–550. https://doi.org/10.1111/0023-8333.00057

Koffi, E. (2010). The pronunciation of <-ED> in coda clusters in Somali-accented English. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 1st Pronunciation in Second Language Learning and Teaching Conference,* (pp. 119–134). Ames, Iowa: Iowa State University.

Koffi, E. (2012). Intelligibility assessment and the acoustic vowel space: An instrumental phonetic account of the production of English lax vowels by Somali speakers. In J. Levis & K. LeVelle (Eds.), *Social factors in pronunciation acquisition* (pp. 216–232). Ames, Iowa: Iowa State University. Retrieved from http://jlevis.public.iastate.edu/Proceedingsfrom3rdPSLLT%20updated.pdf#page=16

Koffi, E. (2013). Confusion as a complement to intelligibility research. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference* (pp. 54–65). Simon Fraser University, Vancouver: Ames, Iowa State University. Retrieved from http://jlevis.public.iastate.edu/pslltconference/4th%20Proceedings/Kang%20PSLLT%202012.pdf

Komsta, L. (N.D.a). R: Anscombe-Glynn test of kurtosis. Retrieved March 6, 2016, from http://127.0.0.1:29294/library/moments/html/anscombe.test.html

Komsta, L. (N.D.b). R: D'Agostino test of skewness. Retrieved March 6, 2016, from http://127.0.0.1:29294/library/moments/html/agostino.test.html

Kormos, J., & Dénes. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, *32*(2), 145–164. https://doi.org/10.1016/j.system.2004.01.001

Koster, C., & Koet, T. (1993). The evaluation of accent in the English of Dutchmen. *Language Learning*, *43*(1), 69–92. https://doi.org/10.1111/j.1467-1770.1993.tb00173.x

Krahn, H., Derwing, T., Mulder, M., & Wilkinson, L. (2000). Educated and underemployed: Refugee integration into the Canadian labour market. *Journal of International Migration and Integration*, *1*(1), 59–84. https://doi.org/10.1007/s12134-000-1008-2

Krashen, S. (1979). Adult second language acquisition as post-critical period learning. *Itl: International Journal of Applied Linguistics*, *43*, 39–52.

Kraut, R., & Wulff, S. (2013). Foreign-accented speech perception ratings: a multifactorial case study. *Journal of Multilingual and Multicultural Development*, *34*(3), 249–263. https://doi.org/10.1080/01434632.2013.767340

Kuhl, P., Williams, K., Lacerda, F., Stevens, K., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, *255*(5044), 606–608. https://doi.org/10.1126/science.1736364

Kuiper, K., & Bell, A. (2000). New Zealand and New Zealand English. In A. Bell & Kuiper (Eds.), *New Zealand English* (pp. 11–22). Wellington, N.Z.: Victoria University Press.

Ladefoged, P. (2007). Articulatory features for describing lexical distinctions. *Language*, *83*(1), 161–180. https://doi.org/10.1353/lan.2007.0026

Lado, R. (1957). *Linguistics across cultures; applied linguistics for language teachers*. Ann Arbor: University of Michigan Press.

Lakoff, R. (1977). You say what you are: Acceptability and gender-related language. In S. Greenbaum (Ed.), *Acceptability in language* (pp. 73–86). The Hague: Mouton.

Leather, J. (1999). Second-language speech research: An introduction. *Language Learning*, *49*(supplement 1), 1–56. https://doi.org/DOI: 10.1111/0023-8333.49.s1.1

Lees, Sezer, Kornfitt, & Underhill. (2013). Language Turkish. In M. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Retrieved from http://wals.info/languoid/lect/wals_code_tur

Lefkowitz, N., & Hedgcock, J. (2002). Sound barriers: Influences of social prestige, peer pressure and teacher (dis)approval on FL oral performance. *Language Teaching Research*, *6*(3), 223–244. https://doi.org/10.1191/1362168802lr107oa

Lehiste, I. (1970). *Suprasegmentals*. Cambridge MA USA: The M.I.T Press.

Lenneberg, E. (1967). *Biological foundations of language*. New York; London: Wiley.

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, *40*(3), 387–417. https://doi.org/10.1111/j.1467-1770.1990.tb00669.x

Lev-Ari, S., & Keysar, B. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology*, *46*(6), 1093–1096. https://doi.org/10.1016/j.jesp.2010.05.025

Levis, J. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly*, *39*(3), 369–378. https://doi.org/10.2307/3588485

Levis, J. (2008). Pronunciation and the assessment of spoken language. In R. Hughes (Ed.), *Spoken English, TESOL, and applied linguistics: challenges for theory and practice* (pp. 245–270). New York: Palgrave Macmillan.

Levis, J. (2011). Assessing speech intelligibility: Experts listen to two students. In J. Levis & K. LeVelle (Eds.), *Proceedings Of The 2nd Annual Pronunciation In Second Language Learning and Teaching Conference* (pp. 56–59). Ames, Iowa: Iowa State University. Retrieved from http://pslltconference.com/index.php?option=com_content&view=article&id=18&Itemid=16

Lewin, J., Meares, C., Cain, T., Spoonley, P., Peace, R., & Ho, E. (2011). *Namasté New Zealand: Indian employers and employees in Auckland* (Research report No. 5). North Shore City: Massey University/University Of Waikato. Retrieved from http://newsettlers.massey.ac.nz/publications_pdfs/Namaste%20New%20Zealand.pdf?PHPSESSID=19539e246dab126eeb9c6ee05cfe1a56

Lewis, M., & Gravatt, B. (1999). The importance of talk to learning: ESOL students' viewpoints. *New Zealand Journal of Adult Learning*, *27*(1), 32–40.

Li, M., & Campbell, J. (2009). Accessing employment: Challenges faced by non-native English-speaking professional migrants. *Asian and Pacific Migration Journal*, *18*(3), 371. https://doi.org/10.1177/011719680901800303

Liang, E. (2014). Pronunciation of English consonants, vowels and diphthongs of Mandarin Chinese speakers. *CSCanada*, *8*(1), 62–65. https://doi.org/10.3968/j.sll.1923156320140801.4012

Lima, E. (2011). Language and nonlanguage factors affecting nonnative undergraduate students' reactions to ITAs. In J. Levis & K. LeVelle (Eds.), *Proceedings Of The 2nd Annual Pronunciation In Second Language Learning And Teaching Conference* (pp. 43–55). Ames, Iowa: Iowa State University. Retrieved from http://pslltconference.com/index.php?option=com_content&view=article&id=18&Itemid=16

Lima, E. (2012). A comparative study of the perception of ITAs by native and nonnative undergraduate students. In J. Levis & K. LeVelle (Eds.), *Social Factors In Pronunciation Acquisition* (pp. 54–64). Ames, Iowa: Iowa State University. Retrieved from http://jlevis.public.iastate.edu/Proceedingsfrom3rdPSLLT%20updated.pdf

Lind, J. (N.D.). Chin Language: Pronunciation. Retrieved July 2, 2015, from http://hobugt.dk/ordbog/pronunciation.htm

Lindemann, S. (2002). Listening with an attitude: A model of native-speaker comprehension of non-native speakers in the United States. *Language in Society*, *31*(3), 419–441. https://doi.org/10.1017/S0047404502020286

Lindemann, S. (2010). Who's "unintelligible"? The perceiver's role. *Issues in Applied Linguistics*, *18*(2). Retrieved from http://escholarship.org/uc/item/89f0w1ch

Lippi-Green, R. (2009). Accent, standard language ideology, and discriminatory pretext in the courts. *Language in Society*, *23*(2), 163–198. https://doi.org/10.1017/S0047404500017826

Lippi-Green, R. (2012). *English with an accent : Language, ideology and discrimination in the United States.* (2nd ed.). London; New York: Routledge.

Liu, L. W. (2007). The English deaf: An orthoepical study of the effectiveness of stress placement intervention for Chinese English speakers. *Dissertation Abstracts International: Section A, The Humanities and Social Sciences*, *68(01),0114*. Retrieved from http://search.proquest.com.ezproxy'library.ubc.ca/docview/304907050

Llurda, E. (2000). Effects of intelligibility and speaking rate on judgements of non-native speakers' personalities. *IRAL*, *38*(3–4), 289–299. https://doi.org/10.1515/iral.2000.38.3-4.289

Long, M. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition*, *12*(3), 251–285. https://doi.org/10.1017/S0272263100009165

Low, E. L. (2008). A review of recent research on speech rhythm: Some insights for language acquisition, language disorders and language teaching. In R. Hughes (Ed.), *Spoken English, TESOL, and applied linguistics: Challenges for theory and practice* (pp. 99–125). New York: Palgrave Macmillan.

Lucas, J. (1984). Communication apprehension in the ESL classroom: Getting our students to talk. *Foreign Language Annals*, *17*(6), 593–598. https://doi.org/10.1111/j.1944-9720.1984.tb01748.x

Ludwig, J. (1982). Native-speaker judgments of second-language learners' efforts at communication: a review. *The Modern Language Journal*, *66*(3), 274–283. https://doi.org/10.2307/326629

Macalister, J. (2010). Emerging voices or linguistic silence?: Examining a New Zealand linguistic landscape. *Multilingua*, *29*(1), 55–75. https://doi.org/10.1515/mult.2010.003

MacIntyre, P., Dörnyei, Z., Clément, R., & Noels, K. (1998). Conceptualizing willingness to communicate in a L2: a situational model of L2 confidence and affiliation. *The Modern Language Journal*, *82*(4), 545–562. https://doi.org/10.1111/j.1540-4781.1998.tb05543.x

Macneil, D. (1987). Some characteristic aspects of Vietnamese English pronunciation. *Prospect*, *3*(1), 61–72.

Madill, M. (2016). Speaking anxiety in the South Korean classroom. *Voices*, (248), 4–5.

Major, R. (1987). Measuring pronunciation accuracy using computerized techniques. *Language Testing*, *4*(2), 155–169. https://doi.org/10.1177/026553228700400203

Major, R. (1997). L2 acquisition, L1 loss, and the critical period hypothesis. In A. James & J. Leather (Eds.), *Second-language speech: Structure and process* (pp. 147–159). Berlin ;New York: Mouton de Gruyter.

Major, R. (2008). Transfer in second language phonology. In J. Hansen Edwards & M. Zampini (Eds.), *Phonology and second language acquisition* (pp. 63–94). Amsterdam / Philadelphia: John Benjamins.

Marlowe, J., Bartley, A., & Hibtit, A. (2014). The New Zealand Refugee Resettlement Strategy: Implications for identity, acculturation and civic participation. *Kōtuitui: New Zealand Journal of Social Sciences Online*, *9*(2), 60–69. https://doi.org/10.1080/1177083X.2014.934847

Matsuda, M. (1990). Voices of America: Accent, antidiscrimination law, and a jurisprudence for the last reconstruction. *Yale Law Journal*, *100*, 1329–1407. https://doi.org/10.2307/796694

Matsuura, H., Chiba, R., & Fujieda, M. (1999). Intelligibility and comprehensibility of American and Irish Englishes in Japan. *World Englishes*, *18*(1), 49–62. https://doi.org/10.1111/1467-971X.00121

Matthews, P. (1997). *The Concise Oxford Dictionary of Linguistics*. Oxford : New York: Oxford University Press.

McAvinue, S. (2012). Migrant labourers face tough hurdles. *Stuff.co.nz*. Retrieved from cassandra.pokoney@stl.co.nz

McBride, K. (2015). Which features of Spanish learners' pronunciation most impact listener evaluations? *Hispania*, *98*(1), 14–30. https://doi.org/10.1353/hpn.2015.0001

McBrien, J. (2014). *Competing priorities in the New Zealand Refugee Resettlement Strategy*. Wellington, N.Z.: Fulbright New Zealand. Retrieved from http://www.fulbright.org.nz/wp-content/uploads/2014/08/axford2014_mcbrien.pdf

McCrocklin, S. (2012). The role of word stress in English as a lingua franca. In J. Levis & K. LeVelle (Eds.), *Social Factors In Pronunciation Acquisition* (pp. 249–256). Ames, Iowa:

Iowa State University. Retrieved from http://jlevis.public.iastate.edu/Proceedings-from3rdPSLLT%20updated.pdf#page=16

McIntyre, N. (2014). *Acculturation experiences and workplace cultural diversity dynamics: A comparative study of Chinese, Indian and Eastern European migrants in New Zealand.* (PhD Thesis). Auckland University of Technology, Auckland, N.Z. Retrieved from http://aut.researchgateway.ac.nz/bitstream/handle/10292/7926/McIntyreNCF.pdf?sequence=3&isAllowed=y

Meares, C., Spoonley, P., Peace, R., & Ho, E. (2010a). *Bamboo networks: Chinese employers and employees in Auckland* (Research report No. 1). North Shore City: Massey University/University Of Waikato. Retrieved from http://newsettlers.massey.ac.nz/publications_pdfs/Namaste%20New%20Zealand.pdf?PHPSESSID=19539e246dab126eeb9c6ee05cfe1a56

Meares, C., Spoonley, P., Peace, R., & Ho, E. (2010b). *Kimchi networks: Korean employers and employees in Auckland* (Research report No. 2). North Shore City: Massey University/University Of Waikato. Retrieved from http://newsettlers.massey.ac.nz/publications_pdfs/Namaste%20New%20Zealand.pdf?PHPSESSID=19539e246dab126eeb9c6ee05cfe1a56

Micceri, T. (1989). The Unicorn, The Normal Curve, and Other Improbable Creatures. *Psychological Bulletin*, *105*(1), 156–166. https://doi.org/10.1037/0033-2909.105.1.156

Migone, P. (2012). Migrants like NZ, but don't make friends easily. *Stuff.co.nz*. Retrieved from http://www.stuff.co.nz/national/7454024/Migrants-like-NZ-but-don-t-make-friends-easily

Miller, J. (2003). *Audible difference: ESL and social identity in schools*. Clevedon, England ; Buffalo, N.Y: Multilingual Matters.

Milroy, J., & Milroy, L. (1999). *Authority in language: Investigating standard English* (3rd ed). London [England] ; New York: Routledge.

Ministry of Business, Innovation and Employment. (2015). *Migration trends and outlook 2014 to 2015* (p. 1). Retrieved from http://www.mbie.govt.nz/info-services/immigration/migration-trends/trends-and-outlook/2014-15

Ministry of Social Development. (2014). Lifetime limit - StudyLink. Retrieved March 12, 2015, from http://www.studylink.govt.nz/about-studylink/glossary/lifetime-limit.html

Morley, J. (1991). The pronunciation component of teaching English to speakers of other languages. *TESOL Quarterly*, *25*(3), 481–520. https://doi.org/10.2307/3586981

Morley, J. (1996). Second language speech/pronunciation: Acquisition, instruction, standards, variation, and accent. In J. Alatis (Ed.), *Georgetown University Round Table on Languages and Linguistics (GURT) 1996*. Washington D. C.: Georgetown University Press.

Morley, J. (1999). Trends in speech/pronunciation: Instructional theory and practice. *TESOL Matters*, *9*(4).

Mortensen, A. (2008). *Refugees as "others" : social and cultural citizenship rights for refugees in New Zealand health services* (PhD Thesis). Massey University, Albany. Retrieved from http://hdl.handle.net/10179/631

Moyer, A. (2013). *Foreign accent: the phenomenon of non-native speech*. Cambridge: CUP.

Moyer, A. (2014). The social nature of L2 pronunciation. In J. Levis & A. Moyer (Eds.), *Social dynamics in second language accent* (pp. 11–29). Boston: De Gruyter Mouton.

Mullen, K. (1980). Rater reliability and oral proficiency evaluations. In J. Oller & K. Perkins (Eds.), *Research in language testing* (pp. 91–101). Rowley, Mass: Newbury House.

Munro, M. (1993). Productions of English vowels by native speakers of Arabic: Acoustic measurements and accentedness ratings. *Language and Speech*, *36*(1). https://doi.org/10.1177/002383099303600103

Munro, M. (1995). Nonsegmental factors in foreign accent: Ratings of filtered speech. *Studies in Second Language Acquisition*, *17*(1), 17–34. https://doi.org/10.1017/S0272263100013735

Munro, M. (1998). The effects of noise on the intelligibility of foreign-accented speech. *Studies in Second Language Acquisition*, *20*(2), 139–154.

Munro, M. (2008). Foreign accent and speech intelligibility. In J. Edwards & M. Zampini (Eds.), *Phonology and second language acquisition* (pp. 193–218). Amsterdam / Philadelphia: John Benjamins.

Munro, M. (2011). Intelligibility: Buzzword or buzzworthy? In J. Levis & K. LeVelle (Eds.), *Proceedings of the 2nd Annual Pronunciation in Second Language Learning and Teaching Conference* (pp. 7–16). Ames, Iowa: Iowa State University. Retrieved from http://pslltconference.com/index.php?option=com_content&view=article&id=18&Itemid=16

Munro, M., & Derwing, T. (1994). Evaluations of foreign accent in extemporaneous and read material. *Language Testing*, *11*(3). https://doi.org/10.1177/026553229401100302

Munro, M., & Derwing, T. (1995a). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, *45*(1), 73–97. https://doi.org/10.1111/ j.1467-1770.1995.tb00963.x

Munro, M., & Derwing, T. (1995b). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech*, *38*(3), 289–306. https://doi.org/10.1177/002383099503800305

Munro, M., & Derwing, T. (1999). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, *49*(supplement 1), 285–310. https://doi.org/10.1111/0023-8333.49.s1.8

Munro, M., & Derwing, T. (2001). Modeling perceptions of the accentedness and comprehensibility of L2 speech: The Role of Speaking Rate. *Studies in Second Language Acquisition*, *23*(4), 451–468.

Munro, M., & Derwing, T. (2006). The functional load principle in ESL pronunciation instruction: An exploratory study. *System*, *34*(4), 520–531. https://doi.org/10.1016/j.system.2006.09.004

Munro, M., & Derwing, T. (2008). Segmental acquisition in adult ESL learners: A longitudinal study of vowel production. *International Journal of Applied Linguistics*, *18*(3), 479–502. https://doi.org/10.1111/j.1467-9922.2008.00448.x

Munro, M., & Derwing, T. (2011). The foundations of accent and intelligibility in pronunciation research. *Language Teaching*, *44*(3), 316–327. https://doi.org/10.1017/S0261444811000103

Munro, M., & Derwing, T. (2015). A prospectus for pronunciation research in the 21st century: A point of view. *Journal of Second Language Pronunciation*, *1*(1), 11–42. https://doi.org/10.1075/jslp.1.1.01mun

Munro, M., Derwing, T., & Holtby, A. (2012). Evaluating individual variability in foreign accent comprehension. In J. Levis & K. LeVelle (Eds.), *Social Factors In Pronunciation Acquisition* (pp. 233–240). Ames, Iowa: Iowa State University. Retrieved from http://jlevis.public.iastate.edu/Proceedingsfrom3rdPSLLT%20updated.pdf

Munro, M., Derwing, T., & Morton, S. (2006). The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition*, *28*(1), 111–131. https://doi.org/10.1017/S0272263106060049

Munro, M., Derwing, T., & Thomson, R. (2015). Setting segmental priorities for English learners: Evidence from a longitudinal study. *International Review of Applied Linguistics in Language Teaching*, *53*(1), 39–60. https://doi.org/10.1515/iral-2015-0002,

Murphy, J. (2013). *Teaching Pronunciation*. Alexandria, VA: TESOL International Association.

Murphy, J. (2014a). Intelligible, comprehensible, non-native models in ESL/EFL pronunciation teaching. *System*, *42*, 258–269. https://doi.org/10.1016/j.system.2013.12.007

Murphy, J. (2014b). Teaching training programs provide adequate preparation in how to teach pronunciation. In L. Grant & D. Brinton (Eds.), *Pronunciation myths: applying second language research to classroom teaching* (pp. 188–224). Ann Arbor: University of Michigan Press.

Nagamine, T. (2002). An experimental study on the teachability and learnability of English intonational aspect: Acoustic analysis on f0 and native-speaker judgment task. *The Journal of Language and Linguistic Studies*, *1*(4), 362–387.

Nakamura, S. (2010). Analysis of relationship between duration characteristics and subjective evaluation of English speech by Japanese learners with regard to contrast of the stressed to the unstressed. *Journal of Pan-Pacific Association of Applied Linguistics*, *14*(1), 1–14.

Nakamura, S. (2011). Characteristics of contrast between the stressed and the unstressed in rhythm units observed in duration structure in English speech by Japanese learners. *Journal of Pan-Pacific Association of Applied Linguistics*, *15*(1), 177–189.

Nation, P. (1989). Improving speaking fluency. *System*, *17*(3), 377–384. https://doi.org/10.1016/0346-251X(89)90010-9

Nation, P., & Newton, J. (2009). *Teaching ESL/EFL listening and speaking*. New York: Routledge.

Nguyen, N. (2008). Interlanguage phonology and the pronunciation of English final consonant clusters by native speakers of Vietnamese. *Ohio Working Papers in Linguistics and Language Teaching*, *16*, 1–20.

Nguyen, T., & Ingram, J. (2005). Vietnamese acquisition of English word stress. *TESOL Quarterly*, *39*(2), 309–319. https://doi.org/10.2307/3588314

North, N. (1999). The employment of immigrants in New Zealand: Employers' perspectives and experiences. In C. Higgins & H. Shrewsbury (Eds.), *Global Trends and Local Issues: Proceedings of the seventh annual conference of the international employment relations association* (pp. 423–435). Lincoln University, Canterbury: Lincoln University.

North, N. (2007). *The employment of immigrants in New Zealand: the attitudes, policies, practices and experiences of employers*. Palmerston North N.Z.: Massey University. Retrieved from http://newsettlers.massey.ac.nz/publica-tions.php?tab=0&PHPSESSID=4d0d077790b9db56f4e8f85a0072d8fd#

NZQA. (2015). Qualifications - Search Results. Retrieved October 31, 2015, from http://www.nzqa.govt.nz/nzqf/search/results.do?q=&area=76682&searchSub-ject=Society+and+Cul-ture+%C2%BB+Language+and+Literature+%C2%BB+English+for+Speakers+of+Other+Languages&type=&lvl=&credit=&status=Current

O'Connor, J. D. (1973). *Phonetics*. Harmondsworth: Penguin.

O'Connor, R. (2014). *"Then came reality": lived experiences of refugee youth in their first 12 months in New Zealand* (p. 44). Hamilton, NZ: NZ Red Cross. Retrieved from https://www.redcross.org.nz/what-we-do/in-new-zealand/refugee-services/re-sources/

Odisho, E. (1988). *The Sound System of Modern Assyrian (Neo-Aramaic)*. Wiesbaden: Otto Harrassowitz Verlag.

Office of Ethnic Communities. (2012). *Accents - ethnic diversity in the workplace* [HTML5]. Retrieved from https://www.youtube.com/watch?v=XsB7p7M-3-Y&fea-ture=youtu.be

Ohata, K. (2004). Phonological differences between Japanese and English: Several poten-tially problematic areas of pronunciation for Japanese ESL/EFL learners. *Language Learning*, *22*, 29–41.

Oliver, P. (2000). *Employment for professional migrants to New Zealand: Barriers and oppor-tunities*. Wellington, N.Z.: Work and Income NZ, Centre for Operational Research and Evaluation.

O'Rourke, K., & Sinnott, R. (2006). The determinants of individual attitudes towards immi-gration. *European Journal of Political Economy*, *22*(4), 838–861. https://doi.org/10.1016/j.ejpoleco.2005.10.005

Osgood, C. (1964). Semantic differential technique in the comparative study of cultures. *American Anthropologist*, *66*(3), 171–200. https://doi.org/10.1525/aa.1964.66.3.02a00880

Pennington, M. (1996). *Phonology in English language teaching: An international approach*. London: Addison Wesley Longman Ltd.

Pennington, M. (1998). The teachability of phonology in adulthood: a re-examination. *IRAL*, *36*(4), 323–341. https://doi.org/10.1515/iral.1998.36.4.323

Pennington, M., & Richards, J. (1986). Pronunciation revisited. *TESOL Quarterly*, *20*, 207–26. https://doi.org/10.2307/3586541

Pickering, L. (2001). The role of tone choice in improving ITA communication in the class-room. *TESOL Quarterly*, *35*(2), 233–255. https://doi.org/10.2307/3587647

Pickering, L. (2004). The structure and function of intonational paragraphs in native and nonnative speaker instructional discourse. *English for Specific Purposes*, *23*(1), 19–43. https://doi.org/10.1016/S0889-4906(03)00020-6

Pilott, M. (1995). Is it Kiwi English they are learning? *TESOLANZ Journal*, *3*, 45–52.

Pinget, A.-F., Bosker, H., Quené, H., & de Jong, N. (2014). Native speakers' perceptions of fluency and accent in L2 speech. *Language Testing*, *31*(3), 349–365. https://doi.org/http://dx.doi.org.helicon.vuw.ac.nz/10.1177/0265532214526177

Pio, E. (2005). Knotted strands: Working lives of Indian women migrants in New Zealand. *Human Relations*, *58*(10), 1277–1299. https://doi.org/10.1177/0018726705059858

Pirt, G. (1990). Discourse intonation problems for non-native speakers. In M. Hewings (Ed.), *Papers in discourse intonation* (pp. 145–155). Birmingham: English Language Research, University of Birmingham.

Piske, T., MacKay, I. R. ., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, *29*(2), 191–215. https://doi.org/10.006/jpho.2001.0134

Portes, A. (2002). English-only triumphs, but the costs are high. *Contexts*, *1*(1), 10–15. https://doi.org/10.1525/ctx.2002.1.1.10

Préfontaine, Y. (2010). Differences in perceived fluency and utterance fluency across speech elicitation tasks: A pilot study. In *Papers from the Lancaster University Postgraduate Conference in Linguistics and Language Teaching* (pp. 134–154). Lancaster. Retrieved from http://www.ling.lancs.ac.uk/pgconference/v05/Prefontaine.pdf

Purcell, E., & Suter, R. (1980). Predictors of pronunciation accuracy: A reexamination. *Language Learning*, *30*(2), 271–287. https://doi.org/10.1111/j.1467-1770.1980.tb00319.x

Qualtrics. (2015). Qualtrics (Provo, UT). Retrieved from http://www.qualtrics.com in 2013

Quirk, R., & Svartvik, J. (1966). *Investigating linguistic acceptability*. The Hague: Mouton.

R Core Team. (2015a). *R: A language and environment for statistical   computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

R Core Team. (2015b). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Rajagopalan, K. (2010). The soft ideological underbelly of the notion of intelligibility in discussions about "World Englishes." *Applied Linguistics (Oxford)*, *31*(3), 465–470. https://doi.org/10.1093/applin/amq014

Rao Hill, S., & Tombs, A. (2011). The effect of accent of service employee on customer service evaluation. *Managing Service Quality: An International Journal*, *21*(6), 649–666. https://doi.org/10.1108/09604521111185637

Ray, G., & Zahn, C. (1999). Language attitudes and speech behavior. *Journal of Language and Social Psychology*, *18*(3), 310–319. https://doi.org/10.1177/0261927X99018003005

Reed, M., & Michaud, C. (2015). Intonation in research and practice: the importance of metacognition. In M. Reed & J. Levis (Eds.), *The handbook of English pronunciation* (pp. 454–470). Chichester, UK ; Malden, MA: Wiley-Blackwell.

Rey, A. (1977). Accent and employability: Language attitudes. *Language Sciences*, (47), 7–12.

Rice, K. (2007). Markedness in Phonology. In P. De Lacy (Ed.), *The Cambridge handbook of phonology* (pp. 79–97). Cambridge ; New York: Cambridge University Press.

Rich, M., Maeda, E., Cargile, A., & Rodriguez, J. (2010). "Oh, you speak English so well!": US American listeners' perceptions of" foreignness" among nonnative speakers. *Journal of Asian American Studies*, *13*(1), 59–79.

Rifkin, B., & Roberts, F. (1995). Error gravity: A critical review of research design. *Language Learning*, *45*(3), 511. https://doi.org/10.1111/j.1467-1770.1995.tb00450.x

Riggenbach, H. (1991). Toward an Understanding of Fluency: A Microanalysis of Nonnative Speaker Conversations. *Discourse Processes*, *14*(4), 423–441. https://doi.org/10.1080/01638539109544795

Robinson, C., & Schumacker, R. (2009). Interaction effects: Centering, variance inflation factor, and interpretation issues. *Multiple Linear Regression Viewpoints*, *35*(1). https://doi.org/http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.620.5853&rep=rep1&type=pdf

Romova, Z., Smith, J., & Neville-Barton, P. (2008). Can I change the way I speak? An exploration into pronunciation and fluency after three years of tertiary EAL study. *Prospect*, *23*(3), 12–23.

Rose, R. (1998). *The communicative value of filled pauses in spontaneous speech* (M.A.). University of Birmingham, Birmingham UK. Retrieved from http://www.roselab.sci.waseda.ac.jp/resources/file/madissertation.pdf

Rose, R. (2008). Filled pauses in language teaching: Why and how. In *EPICS III* (Vol. 29, pp. 47–64). Seville, Spain: Gunma Prefectural Women's University. Retrieved from http://www.roselab.sci.waseda.ac.jp/resources/file/epics3talk.pdf

Rose, R. (2013). Crosslinguistic corpus of hesitation phenomena: a corpus for investigating first and second language speech performance. In *Interspeech 2013* (pp. 992–996). Lyon, France: ISCA. Retrieved from http://www.roselab.sci.waseda.ac.jp/resources/file/2013_interspeech_rose_cchp_final.pdf

Rose, R. (2015). *Temporal variables in first and second language speech and perception of fluency*. Glasgow. Retrieved from http://www.roselab.sci.waseda.ac.jp/resources/file/2015_ichps_rose_slides.pdf

Rossiter, M. (2009). Perceptions of L2 fluency by native and non-native speakers of English. *The Canadian Modern Language Review / La Revue Canadienne Des Langues Vivantes*, *65*(3), 395–412. https://doi.org/10.3138/cmlr.65.3.395

Royal Society of New Zealand. (2013). *Languages in Aotearoa New Zealand* (p. 8). Wellington, N.Z. Retrieved from http://assets.royalsociety.org.nz/media/Languages-in-Aotearoa-New-Zealand.pdf

Rubin, D. (2012). The power of prejudice in accent perception: Reverse linguistic stereotyping and its impact on listener judgements and decisions. In J. Levis & K. LeVelle (Eds.), *Social factors in pronunciation acquisition* (pp. 11–17). Ames, Iowa: Iowa State University. Retrieved from http://jlevis.public.iastate.edu/Proceedings-from3rdPSLLT%20updated.pdf#page=16

Rummel, R. (2002). Understanding factor analysis. Retrieved September 30, 2015, from http://www.hawaii.edu/powerkills/UFA.HTM

Ryan, E. (1983). Social psychological mechanisms underlying native speaker evaluations of non-native speech. *Studies in Second Language Acquisition*, *5*(2), 148–159. https://doi.org/10.1017/S0272263100004824

Saito, K. (2011). Examining the role of explicit phonetic instruction in native-like and comprehensible pronunciation development: An instructed SLA approach to L2 phonology. *Language Awareness*, *20*(1), 45–59. https://doi.org/10.1080/09658416.2010.540326

Saito, K., Trofimovich, P., & Isaacs, T. (2015). Second language speech production: Investigating linguistic correlates of comprehensibility and accentedness for learners at different ability levels. *Applied Psycholinguistics*, *FirstView*, 1–24. https://doi.org/10.1017/S0142716414000502

Sato, C. (1985). Task variation in interlanguage phonology. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 181–196). Rowley, Mass: Newbury House.

Sato, C. (1991). Sociolinguistic variation and language attitudes in Hawaii. In J. Cheshire (Ed.), *English around the world:Sociolinguistic Perspectives* (pp. 647–663). Cambridge: CUP. Retrieved from http://dx.doi.org/10.1017/CBO9780511611889.044

Scarcella, R., & Oxford, R. (1994). Second language pronunciation: state of the art in instruction. *System*, *22*(2), 221–230. https://doi.org/10.1016/0346-251X(94)90058-2

Schairer, K. (1992). Native speaker reaction to non-native speech. *The Modern Language Journal*, *76*(3), 309–319. https://doi.org/10.1111/j.1540-4781.1992.tb07001.x

Schellekens, P. (2001). *English language as a barrier to employment, education and training* (No. 4RP 210/98). London: Department of Education and Skills. Retrieved from https://nationalstrategies.standards.dcsf.gov.uk/publications/eOrderingDownload/4RP21098.pdf

Schmidt, C. (2010). Systemic discrimination as a barrier for immigrant teachers. *Diaspora, Indigenous, and Minority Education*, *4*(4), 235–252. https://doi.org/10.1080/15595692.2010.513246

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*(2), 129. https://doi.org/10.1093/applin/11.2.129

Scoop Media. (2014). Auckland Airport aiming to become top Chinese holiday market. Retrieved July 12, 2015, from http://www.scoop.co.nz/stories/BU1405/S00595/auckland-airport-aiming-to-become-top-chinese-holiday-market.htm

Scott, K. (1999). *The impact of accent, noise, and linguistic predictability on the intelligibility of non-native speakers of English.* University of Florida, Gainesville, Florida. Retrieved from http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA370565

Segalowitz, N. (2010). *Cognitive bases of second language fluency*. New York;London: Routledge.

Segrest Purkiss, S., Perrewé, P., Gillespie, T., Mayes, B., & Ferris, G. (2006). Implicit sources of bias in employment interview judgments and decisions. *Organizational Behavior and Human Decision Processes*, *101*(2), 152–167. https://doi.org/10.1016/j.obhdp.2006.06.005

Selinker, L. (1972). Interlanguage. *IRAL*, *10*, 209–232. https://doi.org/10.1515/iral.1972.10.1-4.209

Selinker, L., & Lamendella, J. (1981). Updating the interlanguage hypothesis. *Studies in Second Language Acquisition*, *3*(2), 201–220. https://doi.org/10.1017/S0272263100004186

Seyedabadi, S., Fatemi, A., & Pishghadam, R. (2015). Towards better teaching of pronunciation: review of literature in the area. *Mediterranean Journal of Social Sciences*, *6*(4 S1), 76.

Shee, N. H. (2008). A descriptive grammar of Geba Karen. *Chiang Mai, Thailand: Payap University: MA Thesis*. Retrieved from http://ic.payap.ac.th/graduate/linguistics/theses/Naw_Hsar_Shee_Thesis.pdf

Shields, M., & Price, S. (2001). Language fluency and immigrant employment prospects: Evidence from Britain's ethnic minorities. *Applied Economics Letters*, *8*(11), 741–745. https://doi.org/10.1080/13504850010038678

Shih, J. (2002). "...Yeah, I could hire this one, but I know it"s gonna be a problem': How race, nativity and gender affect employers' perceptions of the manageability of job seekers. *Ethnic and Racial Studies*, *25*(1), 99–119. https://doi.org/10.1080/01419870120112076

Shrout, P., & Fleiss, J. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420–428. https://doi.org/10.1037/0033-2909.86.2.420

Sicola, L. (2014). Pronunciation, perception, and power: Helping non-native English speakers find their voice in the workplace. *Voice and Speech Review*, *8*(2), 207–211. https://doi.org/10.1080/23268263.2014.906960

Sicola, L., & Darcy. (2015). Integrating pronunciation into the language classroom. In M. Reed & J. Levis (Eds.), *The handbook of English pronunciation* (pp. 471–487). Chichester, UK ; Malden, MA: Wiley-Blackwell.

Singer, M. (1988). The effect of applicant ethnicity on selection decisions: Are results generalisable from résumé to interview research? *Australian Journal of Psychology*, *40*(4), 423–431. https://doi.org/10.1080/00049538808260061

Singer, M., & Eder, G. (1989). Effects of ethnicity, accent, and job status on selection decisions. *International Journal of Psychology*, *24*(1), 13. https://doi.org/10.1080/00207594.1989.10600029

Smit, U. (2002). The interaction of motivation and achievement in advanced EFL pronunciation learners. *IRAL*, *40*(2), 89–116. https://doi.org/10.1515/iral.2002.009

Smith, J. (2012). Teaching pronunciation with multiple models. *New Zealand Studies in Applied Linguistics*, *17*(2), 107–115.

Smith, J., & Beckmann, B. (2010). Noticing-reformulation tasks as a stimulus towards continued autonomous phonological development. *New Zealand Studies in Applied Linguistics*, *16*(1), 36.

Smith, J., & Brown, A. (2011). Exploring the role of pronunciation in New Zealand's brain waste. *Communication Journal of New Zealand*, *12*(1). Retrieved from http://search.ebscohost.com/login.aspx?direct=true&profile=ehost&scope=site&authtype=crawler&jrnl=11754486&AN=67154129&h=8PutHj0NYwUgWi4FMi7YSJhL032m9sYiY2I2U0cKhD3i4EyZYzKx2aAhidwFa-fIlBkoRf8V4jUYEijUJkag09w%3D%3D&crl=c

Smith, L., & Nelson, C. (1985). International intelligibility of English: Directions and resources. *World Englishes*, *4*(3), 333–342. https://doi.org/10.1111/j.1467-971X.1985.tb00423.x

Sobrun-Maharaj, A., Abramson, S., Hoque, E., & Rossen, F. (2008). *Survey of key informants for a study of migrant and refugee youth settlement and social inclusion in New Zealand*. Auckland, N.Z.: University of Auckland.

Southwood, M. H., & Flege, J. (1999). Scaling foreign accent: Direct magnitude estimation versus interval scaling. *Clinical Linguistics & Phonetics*, *13*(5), 335–349. https://doi.org/10.1080/026992099299013

Speer, S., Warren, P., & Schafer, A. (2011). Situationally independent prosodic phrasing. *Laboratory Phonology*, *2*(1), 35–98. https://doi.org/10.1515/labphon.2011.002

Spoonley, P., & Bedford, R. (2012). *Welcome to our world? Immigration and the reshaping of New Zealand*. Auckland, N.Z.: Dunmore Pub.

Spoonley, P., Gendall, P., & Trlin, A. (2007). *Welcome to our world: The attitudes of New Zealanders to immigrants and immigration* (Research report No. 14) (p. 81). Palmerston North, N.Z.: Massey University.

Stanfield, R. (1986a). *Asian Language Notes No.1. Some likely areas of difficulty for Asian learners of English. Burmese, Chinese, Khmer, Vietnamese:* (3rd ed.). Canberra: Australian Govt. Pub. Service.

Stanfield, R. (1986b). *Asian Language Notes No.2. Some likely areas of difficulty for Asian learners of English.Tetum, Portuguese, Japanese* (3rd ed.). Canberra: Commonwealth Schools Commission Australia.

Statistics New Zealand. (2007). Classification - ethnicity, complete. Retrieved March 29, 2015, from http://www.stats.govt.nz/methods/classifications-and-standards/classification-related-stats-standards/ethnicity.aspx

Statistics New Zealand. (2015a). Estimated resident population (ERP), national population by ethnic group, age, and sex, 30 June 1996, 2001, 2006, and 2013. Retrieved February 5, 2015, from http://nzdotstat.stats.govt.nz/wbos/Index.aspx?DataSetCode=TABLECODE7511

Statistics New Zealand. (2015b). Ethnic groups in New Zealand. Retrieved May 8, 2016, from http://www.stats.govt.nz/Census/2013-census/profile-and-summary-reports/quickstats-culture-identity/ethnic-groups-NZ.aspx

Statistics New Zealand. (2015c). International travel and migration: November 2015. Retrieved January 12, 2016, from http://www.stats.govt.nz/browse_for_stats/population/Migration/IntTravelAndMigration_HOTPNov15/Commentary.aspx

Statistics New Zealand. (2015d). *Subnational ethnic population projections: 2013(base)–2038*. Retrieved from http://www.stats.govt.nz/browse_for_stats/population/estimates_and_projections/SubnationalEthnicPopulationProjections_MR2013base.aspx

Strange, W. (1992). Learning non-native phoneme contrasts: Interactions among subject, stimulus and task variables. In Y. 'ichi Tohkura, E. Vatikiotis-Bateson, & Y. Sagisaka (Eds.), *Speech perception, production and linguistic structure* (pp. 197–219). Amsterdam;Washington;Oxford: IOS Press.

Suenobu, M., Kanzaki, K., & Yamane, S. (1992). An experimental study in the intelligibility of Japanese English. *International Review of Applied Linguistics in Language Teaching*, *30*(2), 146–156.

Swan, M. (2001). German Speakers. In M. Swan & B. Smith (Eds.), *Learner English: a teacher's guide to interference and other problems* (2nd ed, pp. 37–51). Cambridge ; New York: CUP.

Swan, M., & Smith, B. (1987). *Learner English: A teacher's guide to interference and other problems* (1st ed.). Cambridge ; New York: CUP.

Swan, M., & Smith, B. (Eds.). (2001). *Learner English: A teacher's guide to interference and other problems* (2nd ed.). Cambridge ; New York: CUP.

Tajima, K., Port, R., & Dalby, J. (1997). Effects of temporal correction on intelligibility of foreign-accented English. *Journal of Phonetics*, *25*(1), 1–24. https://doi.org/10.1006/jpho.1996.0031

Takavoli, P., & Skehan, P. (2005). Strategic planning, task structure and performance testing. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 239–273). Amsterdam: John Benjamins.

Tan, L. (2009). Foreign students mastering Nu Zillund tongue-twisters. *jobs2me.net*. Auckland, N.Z. Retrieved from http://www.jobs2me.net/bbs/Topic.aspx?BoardID=9&TopicID=861

Tarone, E. (1979). Interlanguage as chameleon. *Language Learning*, *29*(1), 181–191. https://doi.org/10.1111/j.1467-1770.1979.tb01058.x

Taylor, D. (1981). Non-native speakers and the rhythm of English. *International Review of Applied Linguistics in Language Teaching*, *XIX*(3), 219–226. https://doi.org/10.1515/iral.1981.19.1-4.219

Temperley, S. (1987). Linking and deletion in final consonant clusters. In J. Morley (Ed.), *Current perspectives on pronunciation: Practices anchored in theory.* (pp. 65–82). Washington D. C.: TESOL.

Tertiary Education Commission. (2015). Our current funds. Retrieved March 3, 2016, from http://www.tec.govt.nz/Funding/Our-current-funds/

Thomas, E. (2014). Accent choices and language contact. In J. Levis & A. Moyer (Eds.), *Social dynamics in second language accent* (pp. 119–144). Boston: De Gruyter Mouton.

Thompson, I. (1991). Foreign accents revisited: The English pronunciation of Russian immigrants. *Language Learning*, *41*(2), 177–204. https://doi.org/10.1111/j.1467-1770.1991.tb00683.x

Thomson, R. (2014). Accent reduction and pronunciation instruction are the same thing. In L. Grant & D. Brinton (Eds.), *Pronunciation myths: Applying second language research to classroom teaching* (pp. 160–187). Ann Arbor: University of Michigan Press.

Thornbury, S. (1993). Having a good jaw: voice-setting phonology. *ELT Journal*, *47*(2), 126–131. https://doi.org/10.1093/elt/47.2.132

Tiffen, B. (1992). A study of the intelligibility of Nigerian English. In A. van Essen & E. Burkart, *Homage to W.R.Lee: Essays in English as a foreign or second language* (pp. 255–259). Berlin: Foris.

Timming, A. (2016). The effect of foreign accent on employability: A study of the aural dimensions of aesthetic labour in customer-facing and non-customer-facing jobs. *Work, Employment & Society*. https://doi.org/10.1177/0950017016630260

Tombs, A., Rao Hill, S., Lee, N., & Lings, I. (2014). The effect of service employees' accent on customer reactions. *European Journal of Marketing*. https://doi.org/10.1108/EJM-03-2013-0115

Trofimovich, P., & Baker, W. (2006). Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition*, *28*(1), 1–30. https://doi.org/DOI:10.1017/S0272263106060013

Trofimovich, P., & Isaacs, T. (2012). Disentangling accent from comprehensibility. *Bilingualism: Language and Cognition*, *15*(4), 905–916. https://doi.org/10.1017/S1366728912000168

UCLA Language Materials Project. (2014). Amharic. Retrieved July 5, 2015, from http://lmp.ucla.edu/Profile.aspx?menu=004&LangID=7

Ushioda, E. (2008). Motivation and good language learners. In C. Griffiths, *Lessons from good language learners* (p. 19). Cambrdidge: CUP.

Valles, B. (2015). *The impact of accented English on speech comprehension* (Ph.D.). The University of Texas at El Paso, Texas, USA. Retrieved from http://search.proquest.com/docview/1700786798

Van den Doel, W. (2006). *How friendly are the natives? An evaluation of native-speaker judgements of foreign-accented British and American English*. Utrecht: LOT (Netherlands Graduate School of Linguistics). Retrieved from http://igitur-archive.library.uu.nl/dissertations/2006-1018-200515/UUindex.html

Van Els, T., & De Bot, K. (1987). The role of intonation in foreign accent. *The Modern Language Journal*, *72*, 147–155. https://doi.org/10.2307/327199

Vance, A. (2015). New Zealand is to take an extra 600 Syrian refugees. Retrieved March 2, 2016, from http://www.stuff.co.nz/national/politics/71820234/New-Zealand-is-to-take-an-extra-600-Syrian-refugees

Varonis, E., & Gass, S. (1982). The comprehensibility of non-native speech. *Studies in Second Language Acquisition*, *4*(2), 114–136. https://doi.org/10.1017/S027226310000437X

Vasilescu, I., Candea, M., & Adda-Decker, M. (2005). Perceptual salience of language-specific acoustic differences in autonomous fillers across eight languages. In *Proceedings of Interspeech*. Lisbonne, France: pp.n.a. Retrieved from http://hal.upmc.fr/hal-00875151

Vertovec, S. (2007). Super-diversity and its implications. *Ethnic and Racial Studies*, *30*(6), 1024–1054. https://doi.org/10.1080/01419870701599465

Ward, C., & Masgoret, A. (2008). Attitudes toward immigrants, immigration, and multiculturalism in New Zealand: A social psychological analysis. *International Migration Review*, *42*(1), 227–248. https://doi.org/10.1111/j.1747-7379.2007.00119.x

Ward, C., & Masgoret, A.-M. (2007). Immigrant entry into the workforce: A research note from New Zealand. *International Journal of Intercultural Relations*, *31*(4), 525–530. https://doi.org/10.1016/j.ijintrel.2007.03.001

Ward, C., Masgoret, A.-M., & Vauclair, M. (2011). *Attitudes towards immigrants and immigrant experiences: Predictive models based on regional characteristics*. Wellington, N.Z.: Department of Labour, NZ Government. Retrieved from http://dol.govt.nz/publications/research/attitudes-towards-immigrants-experiences-regional/attitudes-towards-migrants.pdf

Warren, P. (1998). Timing patterns in New Zealand English rhythm. *Te Reo*, *41*, 80–93.

Warren, P. (2012). *Introducing psycholinguistics*. Cambridge: CUP.

Warren, P., & Britain, D. (2000). Intonation and prosody in New Zealand English. In A. Bell & K. Kuiper (Eds.), *New Zealand English* (pp. 146–172). Wellington, N.Z.: Victoria University Press.

Warren, P., Elgort, I., & Crabbe, D. (2009). Comprehensibility and prosody ratings for pronunciation software development. *Language, Learning & Technology*, *13*(3), 87–102.

Watanabe, M., & Rose, R. (2012). Pausology and hesitation phenomena in second language acquisition. In *The Routledge encyclopedia of second language acquisition.* (pp. 480–483). London: Routledge. Retrieved from http://www.roselab.sci.waseda.ac.jp/resources/file/watanabe_rose_pausology_and_hesitation_phenomena_in_sla.pdf

Watanabe, Y. (2008). New Zealand attitudes towards foreign-accented English. *Te Reo*, *51*, 99–127.

Watkins, J. (2001). Burmese. *Journal of the International Phonetic Association*, *31*(2). https://doi.org/10.1017/S0025100301002122

Watts, N. (1981). The attitudes of New Zealanders to speakers with foreign accents. *Rostra*, (16), 3–5.

Watts, N., & Trlin, A. (2000). Diversity as a productive resource: Employment of immigrants from non-English-speaking backgrounds in New Zealand. *Social Policy Journal of New Zealand*, *15*, 87–101.

Watts, N., White, C., & Trlin, A. (2001). Immigrant and provider perceptions of ESOL learning environments. *TESOLANZ Journal*, *9*, 85–97.

Watts, N., White, C., & Trlin, A. (2002). Roles and responsibilities in English language provision for adult NESB immigrants and refugees. In R. Prasad & N. Van der Walt (Eds.), *Vibrant voices and visions for ethnic New Zealand : declaration* (p. 51- 77). Wellington, N.Z.: New Zealand Federation of Ethnic Councils.

Wennerstrom, A. (1994). Intonational meaning in English discourse: A study of non-native speakers. *Applied Linguistics*, *15*(4), 399–420. https://doi.org/10.1093/applin/15.4.399

Wennerstrom, A. (1998). Intonation as cohesion in academic discourse: A study of Chinese speakers of English. *Studies in Second Language Acquisition*, *20*(1), 1–25.

Werker, J., Gilbert, J., Humphrey, K., & Tees, R. (1981). Developmental aspects of cross-language speech perception. *Child Development*, *52*(1), 349–355.

Werker, J., & Tees, R. (1984). Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America*, *75*(6), 1866–1878. https://doi.org/10.1121/1.390988

Whorf, B. (1964). A linguistic consideration of thinking in primitive communities. In D. Hymes (Ed.), *Language in Culture and Society* (pp. 129–141). New York: Harper & Row.

Wiese, R. (1984). Language production in foreign and native languages: Same or different? In H. Dechert, D. Möhle, & M. Raupach (Eds.), *Second language productions.* (pp. 11–25). Tübingen: G. Narr.

Williams-van Klinken, C. (2003). *Tetun Language Course*. Peace Corps East Timor. Retrieved from http://www.tetundit.tl/Publications/PC%20Tetun%20-%20minus%20cover.pdf

Win, T. T. (1998). *Burmese-English accent: Description, causes, and consequences* (Ph.D.). Northern Illinois University, Illinois. Retrieved from http://search.proquest.com.helicon.vuw.ac.nz/pqdtglobal/docview/304442112/abstract/65DA844DD7B04D83PQ/34?accountid=14782

Witton-Davies, G. (2011). Fluency and pronunciation: The missing link? In *Proceedings of the 2011 International Conference and workshop on TEFL and applied linguistics* (pp. 183–193). National Taiwan University: Crane Publishing. Retrieved from http://www.forex.ntu.edu.tw/en/files/writing/3744_9b8376b3.pdf

Wolff, H. (1964). Intelligibility and inter-ethnic attitudes. In D. Hymes (Ed.), *Language in Culture and Society* (pp. 440–445). New York: Harper & Row.

Wood, D. (2001). In search of fluency: What is it and how can we teach it? *Canadian Modern Language Review*, *57*(4), 573. https://doi.org/10.3138/cmlr.57.4.573

Wray, A., & Perkins, M. (2000). The functions of formulaic language: an integrated model. *Language and Communication*, *20*(1), 1–28. https://doi.org/10.1016/S0271-5309(99)00015-4

Wurm, L., & Fisicaro, S. (2014). What residualizing predictors in regression analyses does (and what it does not do). *Journal of Memory and Language*, *72*, 37–48. https://doi.org/10.1016/j.jml.2013.12.003

Xu, D. (2006). *Job-hunting experiences of immigrant women in canada: A study of professional women from the People's Republic of China* (PCERII Working Paper Series No. WP01-06). Edmonton: University of Alberta. Retrieved from http://pcerii.metropolis.net

Yavas, M. (2011). *Applied English Phonology*. Oxford;Maldon MA: John Wiley & Sons.

Zetterholm, E. (2014). Final stops or not? The importance of final consonants for an intelligible accent. In J. Levis & S. McCrocklin (Eds.), *Pronunciation and Assessment* (pp. 22–29). Ames, Iowa: Iowa State University. Retrieved from http://jlevis.public.iastate.edu/cv_publish.html

Zielinski, B. (2006). The intelligibility cocktail: An interaction between speaker and listener ingredients. *Prospect*, *21*(1), 22–45.

Zielinski, B. (2008). The listener: No longer the silent partner in reduced intelligibility. *System*, *36*(1), 69–84. https://doi.org/10.1016/j.system.2007.11.004

Zielinski, B., & Yates, L. (2014). Pronunciation instruction is not appropriate for beginning-level learners. In L. Grant & D. Brinton (Eds.), *Pronunciation myths: applying second language research to classroom teaching* (pp. 56–79). Ann Arbor: University of Michigan Press.

Zimmer, K., & Orgun, O. (1992). Turkish. *Journal of the International Phonetic Association*, *22*(1–2), 43–45. https://doi.org/10.1017/S0025100300004588

11 APPENDICES

List of Appendices

11.1 APPENDIX A: FORMS AND INVITATIONS

11.1.1 Appendix A1 Migrant Information consent and questionnaire

VICTORIA UNIVERSITY OF WELLINGTON
Te Whare Wananga o te Upoko o te Ika a Maui

School of Linguistics and Applied Language Studies

PO Box 600, Wellington 6140

Researcher: Marty Pilott   marty.pilott@vuw.ac.nz
Supervisors: Paul Warren   paul.warren@vuw.ac.nz   04 463 5631
              Sasha Calhoun   sasha.calhoun@vuw.ac.nz   04 463 9537

Research: *Migrant Pronunciation*
INFORMATION
I am studying pronunciation for a doctorate at Victoria University. The university has agreed
    that I am allowed to ask people to help me with this survey, so I would like to ask for
    your help by recording an interview with you. At first, I will use the interviews for a test,
    to see if the survey is working properly.
I want to find out what kind of pronunciation employers like to hear, so ESOL teachers will
    know more about how to teach pronunciation to their students.
I am going to make computer recordings of different migrants speaking. The recordings will
    be heard on the internet, but not by everyone. I will give a password to some employers
    so only they can listen.
 The employers will then complete a form on the internet (a **survey**) and tell me what they
    think about the pronunciation. The recordings will be **anonymous** which means that no-
    one will know who is speaking. Your name will not be in the survey. I will keep all the
    recordings safe at Victoria University with a password so they cannot be used for any
    other reason.
You can ask me any questions you like before you agree to help with this survey.
When you sign the consent (agreement) form, you agree that I can record an interview at
    Victoria University and use it for the survey. After the survey starts, I will not be able to
    take your interview out, because it will be anonymous.
You can ask for a report of the survey. I will send it to you when it is finished.
Marty Pilott

VICTORIA UNIVERSITY OF WELLINGTON
Te Whare Wananga o te Upoko o te Ika a Maui

School of Linguistics and Applied Language Studies
PO Box 600, Wellington 6140

Researcher: Marty Pilott   marty.pilott@vuw.ac.nz
Supervisors: Paul Warren   paul.warren@vuw.ac.nz   04 463 5631
                Sasha Calhoun   sasha.calhoun@vuw.ac.nz   04 463 9537

<u>Research</u>:   *Migrant Pronunciation*
CONSENT TO HELP WITH SURVEY
NAME: _____

I understand the information sheet.

I agree to go to Victoria University to record an interview, and this interview will be used by
    Marty Pilott to study pronunciation for his doctorate.

I understand that the recorded interview will be kept at Victoria University of Wellington. I
    also understand that the interview will be kept with a password, and that only the peo-
    ple named above will be able to see any information about me. No-one else who listens
    to the interview will know who is speaking.

I understand that because the interviews are anonymous, I cannot change my mind after
    the survey starts.

I understand that I can ask for some of the results from the study to be sent to me later.

Please tick here if you wish to receive a copy of a short summary of the study. ☐
Email or postal address to which the report can be sent:

I know that I can ask questions before starting this survey.
I agree to help with the survey
Signed

VICTORIA UNIVERSITY OF WELLINGTON
Te Whare Wananga o te Upoko o te Ika a Maui

School of Linguistics and Applied Language Studies

PO Box 600, Wellington 6140

Researcher: Marty Pilott   marty.pilott@vuw.ac.nz
Supervisors: Paul Warren   paul.warren@vuw.ac.nz   04 463 5631
                Sasha Calhoun   sasha.calhoun@vuw.ac.nz   04 463 9537

<u>Research</u>:   *Migrant Pronunciation*

INFORMATION ABOUT YOU
(This information will be anonymous, confidential, and cannot be traced to you)
What kind(s) of work are you looking for – choose one or more.

Caregiving	☐
Cleaning	☐
Driving	☐
Hairdressing	☐
Handling goods (e.g. stacking)	☐
Manual labouring	☐
Retail	☐

Other (Please state what) ……………………………………………..

How long have you been in New Zealand?
Less than three years
Three years or more

What is your native language?
 ………………………………..
Please provide contact details if I need to call / email you:
 …………………………………………………………………………………
 ………………………………………………………………………………….

11.1.2 Appendix A2 Employer Survey Intro Consent Instructions

INTRODUCTION

Acceptability of Migrant Pronunciation Survey

 School of Linguistics and Applied Language Studies   PO Box 600, Wellington 6140
Researcher: Marty Pilott   marty.pilott@vuw.ac.nz
Supervisors: Paul Warren   paul.warren@vuw.ac.nz   04 463 5631
Sasha Calhoun   sasha.calhoun@vuw.ac.nz   04 463 9537
Angela Joe   angela.joe@vuw.ac.nz   04 463 5607

Research: Migrant Pronunciation

Approved by Human Ethics Committee, Victoria University of Wellington

INFORMATION - Please read!

My name is Marty Pilott and I am undertaking research for my PhD to study the way employers respond to different types of migrant pronunciation. This project has been given ethical approval by the Victoria University of Wellington Human Ethics Committee and information gathered will be kept and used as part of my PhD.

The online survey that you will take part in involves the following tasks:

Questionnaire: I start with a few questions about you because it's important to know whether different types of employers tend to give different responses. You will be asked to complete a brief questionnaire that will tell me about your background (type of work you offer, first language, etc.)

Survey: You will be asked to listen to samples of migrant speech online, and comment on them by responding to survey questions.

The survey is done online in your own time. The entire session should take around twenty minutes. You can stop part way through and continue later (though this must be on the same computer) and the survey will not be submitted until you reach the end. I will analyse the data you provide, and may use it in contributions to research publications, which will be publicly available. Your identity will remain confidential, and no personal information will be stored with the experiment data. If the audio samples do not display or load, you may need to have Adobe Flash Player installed on your computer.

If you have any questions or concerns you are invited to contact me or my supervisors (contacts given above). Please email me if you wish to receive a short summary of the survey.

CONSENT FORM

School of Linguistics and Applied Language Studies
PO Box 600, Wellington 6140
(See p.1 for contact details)
Research: Migrant Pronunciation

CONSENT TO PARTICIPATE IN RESEARCH

By clicking on the Continue button below, you indicate that you understand the procedure of this research as set out in the INFORMATION on the previous page, and you know that:
You can seek more information before starting this survey.
The responses you provide in the survey will be stored in the School of Linguistics and Applied Language Studies at Victoria University of Wellington.
Your data will remain confidential, and no personal information that may be used to identify you uniquely will be stored with the data.
As the responses are anonymous, it will not be possible for me to withdraw any responses provided.
If you request, details of the purpose of the research will be discussed with you once you have completed your participation, and you may ask for a summary of results from the study to be sent to you at a later date.
This survey will only be submitted when you reach the final page. By clicking on the CONTINUE button below you are agreeing to take part in this research.

11.2 Appendix B: Employer Survey

11.2.1 Appendix B1 Text for Migrant Interview Script

The aim of the Interview Script is to elicit fluent unscripted speech samples from migrants. These will be recorded and used as samples for employers to rate.

Content: Employment interview, general questions

Length: Two minutes, to allow for selection of optimum portion

Format: New question posed at time close to 30s, 1 min, 1 min 30 of interviewee talking time so that all interviewees answer the same four questions. If the interviewee cannot answer, or gives a very brief answer, then an alternative is offered.

Presentation: Interview questions in written rubric. Participants will be given the questions beforehand to think about them.

Text spoken to interviewee:

Imagine that this is an interview for a job that you want to do. What is the job? OK. Remember for each question, give me an answer and then add a bit more information.

1. *Tell me about a job you have done. What were your responsibilities, and what did you do each day?*
2. *Do you like working by yourself or with other people?*
3. *What makes you a good worker?*
4. *How do you think other people would describe you?*

11.2.2 Appendix B2 SURVEY Employer Information

EQ1 Please state the kind(s) of work you offer – select all that apply.
1. Caregiving
2. Cleaning
3. Driving
4. Hairdressing
5. Handling goods (e.g. stacking)
6. Labouring
7. Retail
8. Other (Please state what) _____

EQ2 How long have you been in New Zealand?
1. I was born here
2. Less than 2 years
3. 2 years or more

EQ3 What is your native language?
1. English
2. Other _____

EQ4 Which of the following best describes your view about migrant employees using their own language?
1. They can use their own language whenever they like.

2. They must use English with co-workers and customers.
3. They must use English at work when it is appropriate
4. They must use English while actually on the job.
5. They must use English at all times during work and breaks.
6. They should learn English and use it all the time

EQ5 Have you ever employed people who are not native speakers of English?
1. Yes
2. No

EQ6 [If 5=Yes] Have any of your non-native speakers had pronunciation problems which affected their work?
1. Yes
2. No

EQ7 [If 5 and 6 = Yes] Please briefly describe the problem(s).
[Open response]

11.2.3 Appendix B3 SURVEY Part A Intro and questions

SURVEY PART A INTRODUCTION

INSTRUCTIONS

In this survey I am dealing with migrants who are applying for positions which do not require an advanced level of English or high qualifications. Typical jobs are cleaning, labouring, hairdressing, care-giving, driving and retail.   Your responses   I want to find out what you, as an employer, think about the pronunciation of migrants who might apply for a job with your business. You will hear recordings of mock interviews with a range of migrants. I would like you to imagine that they would be suitable for a job you are offering, so you are only thinking about their pronunciation.

 What do we mean by "Pronunciation"?   Pronunciation means the way someone speaks, including their accent and how well you can understand them. As I'm asking for your impressions, you don't need any knowledge of the subject to answer this survey.

Can I go back to listen again?   Not after you have heard the sample and pressed "Continue". I would like you to listen to each sample once only, and the browser's "back" button will close the survey. How long will it take?   The survey (5 samples) usually takes no more than 20 minutes.

ASI SURVEY PART A: GENERAL FEATURES MIGRANT SPEECH SAMPLES

Imagine that the migrant you are about to listen to is applying for a job at your workplace and that their experience and qualifications make them suitable for the job. Please listen to the sample once and focus only on the way they speak.
Each migrant was asked these questions:
1. Tell me about a job you have done. What were your responsibilities, and what did you do each day?
2. Do you like working by yourself or with other people?
3. What makes you a good worker?

4. How do you think other people would describe you?

The above questions and personal identification have been erased from the recordings to save time; what you hear are their answers. After hearing each sample, you will be asked questions on the following:

1. Whether the migrant's pronunciation would be acceptable to you for employment purposes.
2. Whether you think their pronunciation would be acceptable to their supervisor, other employees or customers.
3. Your impressions of their speech - what it shows of their attitude to work, their friendliness and politeness.
4. What you thought of their accent - whether you could understand them, and whether they were easy to follow or not.

PART A: QUESTIONS FOR EACH SAMPLE

[Note: all questions except the first used a 9-point Likert scale. The default setting was 0, which triggered an error message if not changed. Scales were labelled only at the extremes.]

A0 [NUMBER] Click on the Play button to listen once, then click Continue when you have heard enough (at least one minute please).

A. QUESTIONS SAMPLE A013

Which country do you think the migrant comes from? Enter your answer in the box:

[Free response]

Use the slider to enter a score of 1 (minimum) to 9 (maximum).

[1= Not Acceptable 9=Very Acceptable]

AQ1. Overall Acceptability

AQ1i How acceptable is the migrant's pronunciation for a job with your company?

AQ2. Specific Acceptability

How acceptable is the migrant's pronunciation for communication with the following people?

AQ2i Their supervisor?

AQ2ii Other employees?

AQ2iii Customers?

AQ3. Impressions

From the way this person speaks, how would you place them on the following scales?

AQ3i Work   [1= Lazy 9=Hard worker]

AQ3ii Friendliness   [1= Unfriendly 9=Very Friendly]

AQ3iii Politeness   [1= Impolite 9=Very Polite]

AQ4. Accent and Understanding   What is this migrant's accent like?

AQ4i How much of the sample did you understand?

[1= Nothing 9=All of it]

AQ4ii How easy was it to follow what they were saying?

[1= Very difficult 9=Very easy]

AQ4iii How strong was their accent?

[1= Very STRONGLY accented 9=Very LITTLE accent]

11.2.4 Appendix B4 SURVEY Part B Intro and questions

SURVEY PART B INTRODUCTION

Imagine that the migrant you are about to listen to is applying for a job at your workplace and that their experiences and qualifications make them suitable for the job. Please focus only on the way they speak.
Click on the Speaker icon then answer the questions. You will be able to hear the sample once only.   Each interviewee was asked these questions:

1. Tell me about a job you have done. What were your responsibilities, and what did you do each day?
2. Do you like working by yourself or with other people?
3. What makes you a good worker?
4. How do you think other people would describe you?

The questions have been erased from the recordings so you don't have to listen to them every time. After hearing each sample, you will be asked questions on specific pronunciation features.

PART B: QUESTIONS FOR EACH SAMPLE

[Note: all questions except the first used a 9-point Likert scale. The default setting was 0, which triggered an error message if not changed. Scales were labelled only at the extremes.]

B0 [Number]   Click on the Play button to listen once, then click Continue when you have heard enough (at least one minute please).

B. QUESTIONS SAMPLE B002

Which country do you think the migrant comes from? Enter your answer in the box:
[Free response]

Use the slider to enter a score of 1 (minimum) to 9 (maximum).
BQ1. Overall Acceptability
     BQ1i How acceptable is the migrant's pronunciation for a job with your company?
    [1= Not Acceptable 9=Very Acceptable]
BQ2. Specific Features   How would you rate these specific features of their pronunciation?
BQ2i Individual sounds
     Vowels – e.g. first pronounced correctly, or as farst or feerst
     Consonants – e.g. job pronounced correctly or as chob or jop?
    [1= Many sounds wrong 9=Generally correct]
BQ2ii Stress on words   For example, eleMENtary (correct stress) not eLEMentary (incorrect stress)
    [1= Many wrong stresses 9=Generally correct]
BQ2iii Intonation   How the voice rises and falls in a sentence, e.g.   My ↑last job was quite ↑inte↓resting, or ↑Where are you ↑go↓ing?
    [1= Intonation often wrong 9=Generally correct]
    BQ2iv Added or missing sounds   For example, saying ba' instead of back (missing), or sikis instead of six (added).
[1= Many added/missing sounds 9=Generally correct]
BQ3. General Impression
Finally, how would you describe the migrant's speech?
    BQ3i   **[1= Slow 9=Fast]**
    BQ3ii   **[1= Monotonous 9=Interesting]**
    BQ3iii **[1= Choppy 9=Smooth]**
    BQ3iv   **[1= Hesitant 9=Fluent]**

11.2.5 Appendix B5 Languages, Categories and Speakers

Rationale for coding languages

Number	Region	Short name	Includes
1	Pacific	Pacific	Polynesian, Tetum, Fiji
2	South East Asian	SEAsian	Vietnamese, Cambodian, Thai, Filipino, Myanmarese
3	Chinese	Chinese	
4	Japanese	Japanese	
5	Korean	Korean	
6	South Asian	SAsian	Indian, Sri Lankan, Bengali
7	African	African	Non-Arabic: Somali, Amharic
8	Middle Eastern	MidEastern	Arabic, Assyrian, Farsi, Turkish, Israeli
9	Eastern European	EEuropean	Russian; Not Spanish or Portuguese
10	Western European	WEuropean	European, Greek, Italian
11	South American	SAmerican	Spanish and Portuguese
12	Could not tell	Other	18 observations outs of 424
13	Other	Other	South African
14	Asian	Asian	No further specification given
2 choices: 1st taken, except "Asian" if stated alone but placed in other category if specified.			
"maybe" etc. ignored.			
Spain/South America = South America			

11.3 APPENDIX C: BASE RATING

11.3.1 Appendix C1 Feature notes for raters

Overall:
9: The odd mistake, not like a native speaker, but otherwise competent and unremarkable.
6-8: Competent range. Some errors but in general the feature is not a problem.
5: This feature is not severe but could be seen as a problem
2-4: This feature is clearly a problem. Definite departure from native speaker norms
1: Needs remedial work. Feature impacts heavily on other areas.
NOTE:
Names of features used in the final version of this thesis have been added in square brackets.

QUESTIONS	Comments
Country	Not crucial but it will be interesting to see if TESOL teachers are better than inexperienced raters. [Omitted]
AQ4i How much of the sample did you understand? *[Intelligibility]*	Our understanding is likely to be better than inexperienced raters, so base this on the *immediate impression*, not listening again. 1 = Frequently unintelligible 5 = Significant areas of unintelligibility 9 = Understood everything
AQ4ii How easy was it to follow what they were saying? *[Comprehensibility]*	Mark down if not immediately intelligible. 1 = Listener must struggle to follow 5 = Careful listening needed for parts of the sample 9 = No effort required to understand
AQ4iii How strong was their accent? *[Accentedness]*	Try to be objective – do not base on *familiarity* with the accent or on *intelligibility*. All of these speakers have an accent. Accent is based on a range of regular variations (prosodic and segmental) from native speaker norms. Your reference accents can be any locally accepted (e.g. including British or US) but not a mixture (BUT acknowledge developing fluency – e.g. partial adoption of native pronunciation.) 1 = Most aspects of speech are very deviant from NS norms 5 = Many approximations to NS norms, other aspects clearly accented. 9 = Accent is barely noticeable.
BQ2i Individual sounds [Segmentals]	

QUESTIONS	Comments
BQ2ia Consonants	Look for non-English versions, e.g. /x/ for /h/ or retroflex versions as in Indian speech; allophonic variation failure, e.g. unaspirated initial /t/ or "try" [traɪ] with full [t] and [r] instead of [tʃɹaɪ]; conflation (e.g. /r/ with /l/). Conflation can cause two errors if it leads to unintelligibility or difficulty. Treat rhotic pronunciation or flapped /r/ as a minor fault. Individually these may be normal in locally accepted dialects, so consider whether a native speaker would view the type of pronunciation as significant in the context of the speaker's accent. 1 = Consonants errors would cause significant communication problems 5 = Some major or frequent minor errors 9 = No noticeable errors
BQ2ib Vowels	Listen to vowels in context. A form within one accent may sound wrong in another. Vowels cannot be absolute so must be judged by existence of distinctions, e.g. *bat / bet* are correctly NZ [bɛt] / [bet] or RP [bæt] / [bɛt] but must not be identical and not diphthongs (e.g. typical Chinese [beɪt]). Pay attention to loss of length distinctions such as between /ʌ/ and /ɑ:/ which may conflate the sounds. Failure to reduce vowels: deal with under BQ2iia *Use of word stress.* 1 = Vowel errors would cause significant communication problems 5 = Some major or frequent minor errors 9 = No noticeable errors
BQ2ii Stress on multisyllabic words [Prosody: word stress]	
BQ2iia Use of word stress	Lack of vowel reduction suggests lack of word stress. English word stress tends to involve lengthening, volume and pitch, and lack of pitch variation also produces a non-English pattern. Also note secondary stress e.g. [2]multisy[1]llabic. 1 = No evidence of word stress; even stresses throughout 5 = Use of word stress erratic, weakly evident 9 = Speaker consistently places a stress on multisyllabic words, even if it is wrong.
BQ2iib Word stress placement	Mark down all multisyllabic words which use non-native stress. Note that some words (e.g. *fourteen*) are predictably variable. Listen for phrase groups with alternative meanings e.g. *toy factory.* 1 = Numerous wrongly placed word stresses 5 = Occasional wrongly placed word stresses 9 = Consistent correct use.
BQ2iii Prosody	
BQ2iiia Standard intonation pattern	Use of high rising terminals can only be interpreted as developing fluency if other pronunciation features suggest this too. Mark down constant use of the same pattern or little pitch variation. Look for wrong stress on function words and pronouns ("so I bought IT"). Try repeating sentences as you would say them. 1 = Speaker does not control standard intonation patterns 5 = Evidence of accurate intonation but also some clearly wrong 9 = Most intonation patterns similar to a NS norm
BQ2iiib Contrastive intonation	Listen closely for new information (stressed) and old information (destressed) and mark down inability to do this. Mark up use of contrastive stress to convey additional information ("but THAT wasn't the problem"). 1 = Speaker does not use contrastive intonation 5 = Some use of contrastive intonation 9 = New and important content words regularly stressed / known destressed.

QUESTIONS	Comments
BQ2iiic Coarticulation	Coarticulation includes *sandhi* (e.g. youʷare), blending and assimilation found in fluent speech (e.g. "pasttense"). Lack of coarticulation, or frequent glottal stops and gaps between words, will also show up in "choppiness". Try repeating phrases as you would say them to see if some modification at the boundaries would be expected. 1 = Words are all distinct and no linkage is made 5 = Frequent smooth transitions from word to word; some use of coarticulation 9 = Regular use of coarticulatory features.
BQ2iiid Pauses and Phrasing	Mark down awkward pausing or pauses of excessive length (which will show up in slowness as well). Mark up pauses which support a sense of narrative. 1 = Pauses & phrasing lacking, excessive or wrongly placed 5 = Pauses & phrasing often used appropriately, but often not. 9 = Regular correct use of pauses & phrasing
BQ2iv Added or missing sounds	
BQ2iva Added or missing final sounds	Treat unreleased final consonants or replacements by glottal stops as missing where clarity is compromised. Treat missing suffixes –s and –ed, or wrongly pronounced –ed as in [ɑːskɛd], as minor errors. Deal with replacements (e.g. /fæs/ instead of /fæt/) under BQ2ia *Consonants*. 1 = Regularly misses/adds a range of final sounds 5 = Noticeable number of missed/added final sounds. 9 = No noticeable added/missed final sounds.
BQ2ivb Other added or missing sounds	Include [səˈpred] or [ɛˈsprɛd] for spread, [ˈsɪkɪs] for six, and sounds missing from clusters. Treat wrongly pronounced –ed as in [ɑːskɛd], as a minor error. 1 = Regularly misses/adds a range of non-final sounds 5 = Noticeable number of missed/added non-final sounds. 9 = No noticeable added/missed non-final sounds.
BQ3. General Impression "extensive use of restructurings and self-corrections may result in incoherent utterances" (Albrechtsen et al., 1980, p. 386)	
Slow-Fast 1-9 *[Speed]*	Treat 5 as neutral. Only start marking down or up where the speed is a noticeable factor. 1 = Painfully slow 5 = Speed is not a noticeable problem 9 = Rapid speed probably interferes with comprehensibility
Monotonous-Interesting 1-9 *[Variation]*	A monotonous voice will also show up in poor intonation, perhaps slow, and little pitch range. 1 = Deadly dull 5 = Adequate 9 = A vocal performer (good use of intonation and contrastive stress)
Choppy-Smooth 1-9 *[Smoothness]*	Choppiness may show up in lack of fluency, poor intonation and low coarticulation. 1 = Words.sound.unnaturally.separated; or syllables are all equally and forcefully stressed (as in some Japanese speech). 5 = Clearly a non-native speaker but there is evidence of flow. 9 = Very good use of coarticulation, intonation and rhythm.

QUESTIONS	Comments
Hesitant-Fluent 1-9 *[Fluidity]*	May repeat other features. A hesitant speaker may have bursts of fluency but often has to pause to think. Halting speech causes irritation (Fayer & Krasinski, 1987). 1 = Frequent long breaks to process what to say. May use a lot of fillers (er, um) or repetition. 5 = Clearly a non-native speaker, not rapid, but does not have to stop often to think. 9 = Can produce reasonably fast speech with no preparation and little pausing.
BQ4. Interference 1=Significant, 9=Minimal	*Interference* here means "features which are sufficiently noticeable as to influence a listener's judgement of the speaker's pronunciation". Lexis and grammar were areas of low rater agreement. This is an additional area in which to judge features which might interfere with a listener's perception of pronunciation. Move down from scores of 6-9 only when you think interference is likely.
BQ4i Lexis	1 = Use of many words which are wrong, badly produced or completely unintelligible causing serious distraction or irritation 5 = Some words would cause distraction or irritation, e.g. regular use of "customent" for "customer" 9 = Words are generally used in a way which not cause problems.
BQ4ii Grammar	Do not mark down poor grammar unless it affects comprehensibility. Untrained raters view grammar as less important than teachers do. (Hinofotis & Bailey, 1981, p. 125). Some features (e.g. final –s) belong to both missing consonants and grammar. 1 = Grammar faults are a serious impediment to communication 5 = There are some faults which could cause distraction or lack of comprehension. 9 = Grammar may have faults (e.g. all in present tense) but native speakers would generally ignore them.
BQ4iii Voice quality	Ratings here may be reflected in intonation. VQ can be affected by tenseness (English is generally lax) and energy or articulation (e.g. staccato effect in Arabic speakers – Swan and Smith (1987, p. 143)). Also take account of voice clarity, noting overpronouncing and mumbling (Isaacs, 2008) . Look for *non-native* quality. 1 = Voice quality causes irritation and disrupts message 5 = Some concerns about voice quality 9 = Voice quality does not cause any problem
BQ4iv Intrusive noise	Nonverbal noises only. Note frequency and intensity of clicks, tutting noises, and fillers such as aah or mm. Ignore recording noises such interference with microphone. 1 = Frequent loud intrusive noises which would strongly irritate a listener. 5 = Intrusive noises are evident and could be distracting 9 = Not noticeable

11.3.2 Appendix C2 Rater Response Table

Sample number: 0		1	2	3	4	5	6	7	8	9	

A. QUESTIONS

Which country do you think the migrant comes from?
Write the number in the correct column

		1	2	3	4	5	6	7	8	9	
AQ4i How much of the sample did you understand?	Nothing										All of it
AQ4ii How easy was it to follow what they were saying?	Very difficult										Very easy
AQ4iii How strong was their accent?	Very STRONGLY accented										Very LITTLE accent
B. QUESTIONS											
BQ2ia Consonants	Many consonants wrong										Generally correct
BQ2ib Vowels	Many vowels wrong										Generally correct
BQ2ii Stress on multisyllabic words	Stress not evident										Stress always evident
BQ2iib Word stress placement	Often wrong										Generally correct
BQ2iii Prosody											
BQ2iiia intonation - general	Intonation often wrong										Generally correct
BQ2iiib Contrastive intonation	Contrast often wrong										Generally correct
BQ2iiic Coarticulation	Coarticulation often wrong										Generally correct
BQ2iiid Pauses and Phrasing	Pauses/Phrasing often wrong										Generally correct
BQ2iv Added or missing sounds											
BQ2iva Added or missing finals	Many added/missing finals										Generally correct
BQ2ivb Other added or missing sounds	Many added/missing sounds										Generally correct
BQ3. General Impression											
BQ3i	Slow										Fast
BQ3ii	Monotonous										Interesting
BQ3iii	Choppy										Smooth
BQ3iv	Hesitant										Fluent
BQ4. Interference											
Note where the following non-phonological features are likely to cause interference.											
BQ4i Lexis	Minimal										Significant
BQ4ii Grammar	Minimal										Significant
BQ4iii Voice quality	Minimal										Significant
BQ4iv Intrusive noise	Minimal										Significant

11.4 APPENDIX D: STATISTICAL MODELS

11.4.1 Appendix D1 LanguageGuess

```
clmm_LgeActual_LgeGuess.R
Reference level SEAsian

Acceptable ~ LgeActual + Factor1 + Factor2 + Factor3 + (1| Employer) +(1|
Speaker)

 link  threshold nobs logLik  AIC      niter         max.grad cond.H
 logit flexible   424  -797.39 1636.78 2438(15160) 7.79e-04 6.2e+02
Random effects:
 Groups    Name          Variance Std.Dev.
 Employer (Intercept) 2.6755   1.6357
 Speaker  (Intercept) 0.8998   0.9486
Number of groups:  Employer 95,  Speaker 40
Coefficients:
                    Estimate Std. Error z value Pr(>|z|)
LgeActualAfrican      -0.3520      0.7083  -0.497 0.619247
LgeActualChinese      -0.7844      0.9329  -0.841 0.400464
LgeActualJapanese     -0.3399      0.8328  -0.408 0.683162
LgeActualMidEastern   -0.8872      0.6752  -1.314 0.188884
LgeActualPacific      -0.2506      0.8621  -0.291 0.771324
LgeActualSAmerican    -0.1346      0.6908  -0.195 0.845474
LgeActualSAsian       -0.1988      0.7790  -0.255 0.798540
LgeActualWEuropean     0.9155      0.7856   1.165 0.243902
Factor1                0.6746      0.2205   3.059 0.002219 **
Factor2                0.8771      0.2510   3.495 0.000475 ***
Factor3                0.7324      0.2497   2.934 0.003351 **
Threshold coefficients:
    Estimate Std. Error z value
1|2  -4.5226     0.5391  -8.389
2|3  -3.2610     0.4966  -6.567
3|4  -2.3900     0.4766  -5.014
4|5  -1.5840     0.4655  -3.403
5|6  -0.4409     0.4602  -0.958
6|7   0.7784     0.4627   1.682
7|8   2.4677     0.4807   5.133
8|9   4.0447     0.5194   7.787


anova(fm0,fm1)
fm0 Acceptable ~ Factor1 + Factor2 + Factor3 + (1 | Employer) + (1 |
Speaker)
fm1 Acceptable ~ LgeActual + Factor1 + Factor2 + Factor3 + (1 | Employer) +
(1 | Speaker)
    link: threshold:
fm0 logit flexible
fm1 logit flexible

    no.par    AIC  logLik LR.stat df Pr(>Chisq)
fm0    13 1626.1 -800.05
fm1    21 1636.8 -797.39  5.3175  8     0.7232
```

Language Guess

```
clmm_LgeActual_LgeGuess

Reference level SEAsian

Acceptable ~ LgeGuess + Factor1 + Factor2 + Factor3 + (1| Employer) + (1|
Speaker)
 link  threshold nobs logLik  AIC      niter       max.grad cond.H
 logit flexible  424  -792.25 1634.50 3291(19529) 3.03e-04 3.4e+02


Random effects:
 Groups    Name         Variance Std.Dev.
 Employer (Intercept) 2.822    1.680
 Speaker  (Intercept) 1.135    1.065
Number of groups:  Employer 95,  Speaker 40
```

Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)	
LgeGuessAfrican	-0.10327	0.39854	-0.259	0.79555	
LgeGuessAsian	-0.09475	0.69721	-0.136	0.89190	
LgeGuessChinese	0.09314	0.43127	0.216	0.82901	
LgeGuessEEuropean	0.38602	0.46117	0.837	0.40257	
LgeGuessJapanese	0.36228	0.74198	0.488	0.62536	
LgeGuessKorean	**1.80833**	**0.81093**	**2.230**	**0.02575**	*****
LgeGuessMidEastern	-0.47244	0.47742	-0.990	0.32239	
LgeGuessOther	-0.22997	0.47747	-0.482	0.63005	
LgeGuessPacific	-0.80376	0.63591	-1.264	0.20625	
LgeGuessSAmerican	0.86466	0.59818	1.445	0.14832	
LgeGuessSAsian	0.11960	0.46479	0.257	0.79692	
LgeGuessWEuropean	0.60183	0.50635	1.189	0.23462	
Factor1	0.67145	0.20476	3.279	0.00104	**
Factor2	0.88550	0.21723	4.076	4.58e-05	***
Factor3	0.67137	0.22326	3.007	0.00264	**

Threshold coefficients:

	Estimate	Std. Error	z value
1|2	-4.3605	0.4688	-9.301
2|3	-3.0620	0.4206	-7.280
3|4	-2.1672	0.3976	-5.450
4|5	-1.3299	0.3833	-3.469
5|6	-0.1480	0.3767	-0.393
6|7	1.1005	0.3829	2.874
7|8	2.8314	0.4122	6.869
8|9	4.4438	0.4655	9.546

anova(fm0,fm2)
```
Likelihood ratio tests of cumulative link models:
fm0 Acceptable ~ Factor1 + Factor2 + Factor3 + (1 | Employer) + (1 |
Speaker)
fm2 Acceptable ~ LgeGuess + Factor1 + Factor2 + Factor3 + (1 | Employer) +
(1 | Speaker)
    link: threshold:
fm0 logit flexible
fm2 logit flexible
```

	no.par	AIC	logLik	LR.stat	df	Pr(>Chisq)
fm0	13	1626.1	-800.05			
fm2	25	1634.5	-792.25	15.59	12	0.2107

Conclusions

The second lowest comparative rating in this and other models is for Asian and the highest for Korean. Given that Employer Language Guess is rarely accurate, was this rating the result of (1) favourable assumptions about "Korean" and unfavourable about "Asian", so that the employer ratings were subsequently skewed; or (2) actual differences in speaker competence, so that the employer ratings were not skewed but their language guesses may have reflected attitudes about relative ability?

Language guess "Korean" was made for speakers 07, 08 (x3), 15, 37, 48 and 50. None were actually Korean (in fact, there were none amongst the samples). Speakers 07, 37 and 50 were South East Asian, and 08 Chinese, while 15 was South American and 48 was African.

Language guess "Asian" was made for speakers 05, 07, 08, 15, 19, 23, 25, 32 and 42. All were in fact Asian but of course this is a huge category including many unrelated languages. 05 and 32 were Middle Eastern; 07, 19, 23, and 42 were South East Asian; 08 was Chinese; 15 and 25 were South Asian.

The baseline rating for these speakers was checked to see if there was any correlation between guesses of "Korean" or "Asian" and speaker competence.

BASELINE	Intelligibility	Comprehensibility	Accentedness
Korean	6.7	6.6	4.5
Without 07	7.4	6.8	5.2
Asian	6.6	5.6	4.1
Without 07	7.0	5.9	4.5
Overall	7.0	6.2	4.8

When the outlier Speaker 07 (who also features in the lowest, category, Asian) is removed from the baseline rating of those judged by Employers to be "Korean", their ratings are all well above the average for all speakers (105%, 109% and 108% respectively). Those judged to be "Asian" are all lower: 94%, 90% and 85% respectively).

Even if Speaker 07 is also removed from the "Asian" group for consistency, "Korean" still exceeds "Asian" by 105%, 115% and 1.15% respectively.

We can therefore conclude that the correlation between an Employer guess of "Korean" and higher acceptability, and a guess of "Asian" and lower acceptability, is likely to be a result of the actual pronunciation competence of the speakers, and not to a skewing of the Employer ratings of those guessed to belong to those groups.

11.4.2 Appendix D2: Sex of Employer and Migrant with Interaction

E_S_Gender_Interaction.R

```
Cumulative Link Mixed Model fitted with the Laplace approximation
formula: Acceptable ~ Gender + Factor1 + Factor2 + Factor3 + (1 | Employer)
+ (1 | Gender:Speaker)
 link  threshold nobs logLik  AIC     niter      max.grad cond.H
 logit flexible  424  -799.50 1627.00 1300(9813) 9.81e-04 2.0e+02
Random effects:
 Groups          Name        Variance Std.Dev.
 Employer        (Intercept) 2.684    1.638
 Gender:Speaker (Intercept) 1.033    1.016
Number of groups:  Employer 95,  Gender:Speaker 40
Coefficients:
        Estimate Std. Error z value Pr(>|z|)
GenderM  -0.4194     0.3994  -1.050  0.29362
Factor1   0.6348     0.1963   3.233  0.00122 **
Factor2   0.8698     0.2054   4.234  2.3e-05 ***
Factor3   0.6308     0.2188   2.883  0.00394 **

Threshold coefficients:
    Estimate Std. Error z value
1|2  -4.5651     0.4599  -9.926
2|3  -3.3002     0.4089  -8.072
3|4  -2.4297     0.3846  -6.317
4|5  -1.6243     0.3703  -4.387
5|6  -0.4784     0.3606  -1.327
6|7   0.7445     0.3612   2.061
7|8   2.4318     0.3833   6.344
8|9   4.0020     0.4336   9.230
Standard model with interaction
Cumulative Link Mixed Model fitted with the Laplace approximation
formula: Acceptable ~ Gender * EmpGender + Factor1 + Factor2 + Factor3 +
(1 | Employer) + (1 | Gender:Speaker)
 link  threshold nobs logLik  AIC     niter       max.grad cond.H
 logit flexible  424  -798.88 1629.76 1749(13065) 2.25e-04 3.3e+02
Random effects:
 Groups          Name        Variance Std.Dev.
 Employer        (Intercept) 2.664    1.632
 Gender:Speaker (Intercept) 1.022    1.011
Number of groups:  Employer 95,  Gender:Speaker 40
Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
GenderM            -0.6691     0.4642  -1.442  0.14943
EmpGenderM         -0.4069     0.4637  -0.878  0.38020
Factor1             0.6268     0.1957   3.202  0.00136 **
Factor2             0.8749     0.2048   4.272 1.94e-05 ***
Factor3             0.6322     0.2180   2.900  0.00373 **
GenderM:EmpGenderM  0.4424     0.4210   1.051  0.29334
Threshold coefficients:
    Estimate Std. Error z value
```

```
1|2  -4.7885     0.5290  -9.052
2|3  -3.5258     0.4856  -7.260
3|4  -2.6570     0.4649  -5.716
4|5  -1.8524     0.4524  -4.095
5|6  -0.7068     0.4436  -1.593
6|7   0.5176     0.4429   1.169
7|8   2.2073     0.4593   4.806
8|9   3.7778     0.5011   7.538
```

Standard model without interaction

```
Cumulative Link Mixed Model fitted with the Laplace approximation
formula: Acceptable ~ Gender + EmpGender + Factor1 + Factor2 + Factor3 +
(1 | Employer) + (1 | Gender:Speaker)
 link  threshold nobs logLik  AIC     niter        max.grad cond.H
 logit flexible  424  -799.43 1628.86 1853(15127) 1.55e-02 NaN
```
Random effects:
```
 Groups          Name         Variance Std.Dev.
 Employer        (Intercept) 2.677     1.636
 Gender:Speaker (Intercept) 1.031     1.015
Number of groups:  Employer 95,  Gender:Speaker 40
```
Coefficients:
```
          Estimate Std. Error z value Pr(>|z|)
GenderM     -0.4184       NA     NA      NA
EmpGenderM  -0.1448       NA     NA      NA
Factor1      0.6343       NA     NA      NA
Factor2      0.8691       NA     NA      NA
Factor3      0.6300       NA     NA      NA
```
Threshold coefficients:
```
    Estimate Std. Error z value
1|2  -4.6463       NA     NA
2|3  -3.3819       NA     NA
3|4  -2.5110       NA     NA
4|5  -1.7053       NA     NA
5|6  -0.5594       NA     NA
6|7   0.6635       NA     NA
7|8   2.3506       NA     NA
8|9   3.9206       NA     NA
Warning message:
In summary.clmm(fm6, CORR = TRUE) :
  Variance-covariance matrix of the parameters is not defined
ANOVA
Likelihood ratio tests of cumulative link models:
fm5 Acceptable ~ Factor1 + Factor2 + Factor3 + (1 | Employer) + (1 |
Speaker)
fm6 Acceptable ~ SGender + Factor1 + Factor2 + Factor3 + (1 | Employer) +
(1 | SGender:Speaker)
    no.par   AIC  logLik LR.stat df Pr(>Chisq)
fm5    13 1626.1 -800.05
fm6    14 1627.0 -799.50  1.0939  1     0.2956
```

11.4.3 Appendix D3: clmm Acceptability + 3 factors (the 3-Factor Model)

clmm_Acceptable_V4_Survey_13Vars_All.R

```
Acceptable ~ Factor1 + Factor2 + Factor3 + (1 | Employer) +  (1 | Speaker)
 link   threshold nobs logLik  AIC      niter        max.grad cond.H
 logit flexible   424  -800.05 1626.09 1270(10247) 9.43e-04 1.0e+02
```
Random effects:
```
 Groups    Name         Variance Std.Dev.
 Employer (Intercept) 2.686     1.639
 Speaker  (Intercept) 1.069     1.034
Number of groups:  Employer 95,   Speaker 40
```
Coefficients:
```
         Estimate Std. Error z value Pr(>|z|)
```
Factor1   0.6626   0.1971   3.361 0.000776 ***
Factor2   0.8651   0.2078   4.162 3.15e-05 ***
Factor3   0.6831   0.2161   3.161 0.001573 **
Threshold coefficients:
```
    Estimate Std. Error z value
1|2  -4.3128     0.3857 -11.180
2|3  -3.0491     0.3263  -9.343
3|4  -2.1790     0.2978  -7.318
4|5  -1.3726     0.2801  -4.899
5|6  -0.2259     0.2696  -0.838
6|7   0.9967     0.2742   3.635
7|8   2.6836     0.3074   8.731
8|9   4.2542     0.3705  11.483


R Script:
library(ordinal)

emp = read.table("EmployerswithRaterM.txt", header = TRUE, check.names =
TRUE, sep = "\t")
emp$Acceptable<-ordered(emp$Acceptable)
fm1<-clmm(as.factor(Acceptable) ~ Factor1 + Factor2 + Factor3 + (1|Em-
ployer) + (1|Speaker), emp)
summary (fm1)
```

11.4.4 Appendix D4: Principal Component Analysis *workpca.R*

Variables factor map (PCA)
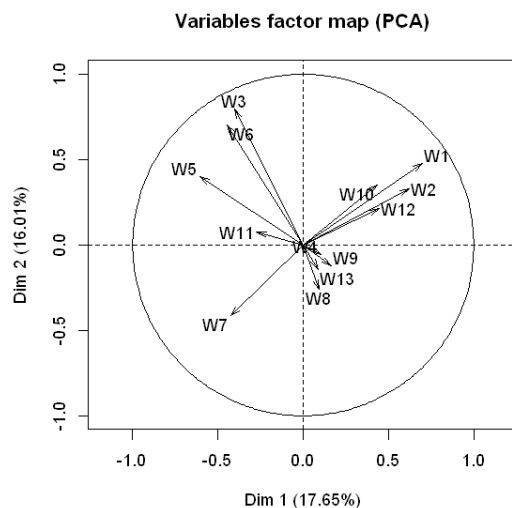
```
       eigenvalue percentage of variance cumulative percentage of
       variance
comp 1    2.2943095               17.648535                     17.64853
comp 2    2.0816274               16.012519                     33.66105
comp 3    1.3218775               10.168288                     43.82934
comp 4    1.1637691                8.952070                     52.78141
comp 5    1.0771322                8.285633                     61.06704
comp 6    1.0512915                8.086857                     69.15390
comp 7    1.0304027                7.926175                     77.08008
comp 8    0.9471730                7.285946                     84.36602
comp 9    0.6568814                5.052934                     89.41896
comp 10   0.6403381                4.925678                     94.34463
comp 11   0.3054896                2.349920                     96.69455
comp 12   0.2732531                2.101947                     98.79650
comp 13   0.1564548                1.203499                    100.00000
```

Result:

C1   W1 (Caregiving), W2 (Cleaning), W10 (Nursing), W12 (Kitchen)
C2   W3 (Driving), W5 (Handling goods), W6 (Labouring, removals), W11 (Office)
C3   W4 (Hairdressing), W8 (Hospitality, customer), W9 (Security), W13 (Other)
C4   W7 (Retail)

11.4.5 Appendix D5: Grammar and Lexis as predictors of Acceptability

Accept~Grammar_Lexis.R

Summary of Results

Grammar added:

When Grammar is introduced as a predictive variable, only Prosody and Fluency remains as predictors, Grammar is also a predictor ($p < 0.05$), but the p value of Segmentals becomes too high. Grammar is highly correlated with the 3 Factors. ANOVA shows a better fit than the original model.

Lexis added:

Factors 1 and 2 (Prosody and Fluency) remain as predictors. Lexis is strongly identified as a predictor ($p < 0.01$), but the p value of Factor 3 (Segmentals) becomes too high. Lexis is highly correlated with the 3 Factors. ANOVA shows a better fit.

Grammar + Lexis

Lexis, Prosody and Factor 2 are all predictors at with similar estimates. The p values of Grammar and Segmental become too high. Grammar is correlated with Lexis and Factors 2 and 3; Lexis is less strongly correlated with the 3 Factors. Normally the Factors show little correlation, but in this analysis all Factors show a higher correlation. ANOVA shows a better fit ($p < 01$).

Grammar*Lexis

Only Prosody and Fluency remain as predictors yet, ANOVA reports that it is a better fit than the original model.

Acceptability with Grammar

```
fm6<-clmm(Acceptable ~ Grammar + Prosody + Fluency + Segmental + (1|Em-
ployer) + (1|Speaker), emp)

summary(fm6,CORR=TRUE)
Cumulative Link Mixed Model fitted with the Laplace approximation
formula: Acceptable ~ Grammar + Prosody + Fluency + Segmental + (1 | Em-
ployer) +  1 | Speaker)
 link  threshold nobs logLik  AIC     niter       max.grad cond.H
 logit flexible  424  -797.77 1623.54 1357(11453) 3.91e-04 5.8e+03
Random effects:
 Groups   Name         Variance Std.Dev.
 Employer (Intercept) 2.6827   1.6379
 Speaker  (Intercept) 0.9081   0.9529
Number of groups:  Employer 95,  Speaker 40
Coefficients:
       Estimate Std. Error z value Pr(>|z|)
```

Grammar   0.3334   0.1527   2.183   0.02903 *
Prosody   0.5337   0.1944   2.745   0.00605 **
Fluency   0.5655   0.2363   2.393   0.01670 *
```
Segmental 0.3603       0.2497    1.443   0.14904


Threshold coefficients:
    Estimate Std. Error z value
1|2  -2.2490     0.9947   -2.261
2|3  -0.9820     0.9808   -1.001
3|4  -0.1074     0.9789   -0.110
4|5   0.7036     0.9805    0.718
5|6   1.8530     0.9856    1.880
6|7   3.0728     0.9936    3.093
7|8   4.7500     1.0101    4.703
8|9   6.3138     1.0355    6.097
```
Correlation of Coefficients:
```
        Grammar Prosody Fluency  ST1   ST2
Prosody
```
-0.2874
```
Fluency
```
-0.5568   0.1952
```
Segmental
```
-0.5746   0.1949 0.3077 0.0418

anova (fm5,fm6)

```
Likelihood ratio tests of cumulative link models:

    formula:
fm5 Acceptable ~ Prosody + Fluency + Segmental + (1 | Employer) + (1 |
Speaker)
fm6 Acceptable ~ Grammar + Prosody + Fluency + Segmental + (1 | Employer) +
(1 | Speaker)
    link: threshold:
fm5 logit flexible
fm6 logit flexible


    no.par   AIC  logLik LR.stat df Pr(>Chisq)
fm5    13 1626.1 -800.05
```
fm6   14 1623.5 -797.77 4.5545 1   0.03283 *

Acceptability with Lexis

```
fm7<-clmm(Acceptable ~ Lexis + Prosody + Fluency + Segmental + (1|Employer)
+ (1|Speaker), emp)
summary(fm7,CORR=TRUE)
Cumulative Link Mixed Model fitted with the Laplace approximation
formula: Acceptable ~ Lexis + Prosody + Fluency + Segmental + (1 | Em-
ployer) +  (1 | Speaker)

 link  threshold nobs logLik  AIC     niter      max.grad cond.H
 logit flexible  424  -796.21 1620.43 1382(8092) 1.27e-04 7.0e+03

Random effects:
 Groups   Name        Variance Std.Dev.
 Employer (Intercept) 2.6821   1.6377
 Speaker  (Intercept) 0.8061   0.8978
Number of groups:  Employer 95,  Speaker 40

Coefficients:
        Estimate Std. Error z value Pr(>|z|)
Lexis     0.4073    0.1415   2.879  0.00399 **
Prosody   0.4585    0.1910   2.401  0.01634 *
Fluency   0.5507    0.2143   2.570  0.01018 *
Segmental 0.3309    0.2285   1.448  0.14756


Threshold coefficients:
    Estimate Std. Error z value
1|2 -1.5889     0.9866   -1.610
2|3 -0.3037     0.9801   -0.310
3|4  0.5766     0.9814    0.588
4|5  1.3849     0.9831    1.409
5|6  2.5318     0.9892    2.559
6|7  3.7530     1.0004    3.751
7|8  5.4283     1.0208    5.318
8|9  6.9798     1.0452    6.678
Correlation of Coefficients:
        LexisR  Prosody Fluency  ST1  ST2
Prosody -0.3480
Fluency -0.4746  0.2058
Segmntl -0.5091  0.2112  0.2283
```

Acceptability ~ Grammar + Lexis

```
Cumulative Link Mixed Model fitted with the Laplace approximation
formula: Acceptable ~ Grammar + Lexis + Prosody + Fluency + Segmental +
    (1 | Employer) + (1 | Speaker)
data:    emp
 link  threshold nobs logLik  AIC     niter     max.grad cond.H
 logit flexible  424  -795.75 1621.50 1415(9924) 4.95e-04 1.6e+04

Random effects:
 Groups   Name         Variance Std.Dev.
 Employer (Intercept) 2.6815   1.6375
 Speaker  (Intercept) 0.7772   0.8816
Number of groups:  Employer 95,  Speaker 40
```

Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)	
GrammarR	0.1612	0.1664	0.969	0.3328	
Lexis	**0.3297**	**0.1606**	**2.053**	**0.0401**	*****
Prosody	**0.4351**	**0.1901**	**2.289**	**0.0221**	*****
Fluency	**0.4660**	**0.2285**	**2.039**	**0.0414**	*****
Segmntl	0.2421	0.2433	0.995	0.3197	

Threshold coefficients:

	Estimate	Std. Error	z value
1|2	-1.1115	1.0910	-1.019
2|3	0.1717	1.0853	0.158
3|4	1.0528	1.0878	0.968
4|5	1.8632	1.0914	1.707
5|6	3.0115	1.0990	2.740
6|7	4.2314	1.1101	3.812
7|8	5.9035	1.1293	5.227
8|9	7.4546	1.1530	6.466

Correlation of Coefficients:

	Grammar	Lexis	Prosody	Fluency
Lexis	**-0.4922**			
Prosody	-0.1228	**-0.2400**		
Fluency	**-0.3764**	-0.1969	**0.2360**	
Segmntl	**-0.3735**	**-0.2267**	**0.2408**	0.3370

Acceptability ~ Interaction Grammar*Lexis

```
fm8<-clmm(Acceptable ~ Grammar*Lexis + Prosody + Fluency + Segmental +
(1|Employer) + (1|Speaker), emp)
Cumulative Link Mixed Model fitted with the Laplace approximation
formula: Acceptable ~ Grammar * Lexis + Prosody + Fluency + Segmental +  (1
| Employer) + (1 | Speaker)

 link  threshold nobs logLik  AIC     niter      max.grad cond.H
 logit flexible  424  -795.73 1623.46 1660(11927) 1.00e-04 1.5e+06

Random effects:
 Groups   Name         Variance Std.Dev.
 Employer (Intercept) 2.6847   1.6385
 Speaker  (Intercept) 0.7765   0.8812
Number of groups:  Employer 95,  Speaker 40

Coefficients:
```

```
              Estimate Std. Error z value Pr(>|z|)
Grammar       0.100274  0.350764  0.286   0.7750
Lexis         0.272601  0.330743  0.824   0.4098
Prosody       0.433396  0.190232  2.278   0.0227 *
Fluency       0.470019  0.229353  2.049   0.0404 *
Segmntl       0.257021  0.254798  1.009   0.3131
Grammar:Lexis 0.009861  0.049997  0.197   0.8436
Threshold coefficients:
    Estimate Std. Error z value
1|2  -1.4372    1.9794  -0.726
2|3  -0.1559    1.9837  -0.079
3|4   0.7245    1.9879   0.364
4|5   1.5347    1.9903   0.771
5|6   2.6828    1.9946   1.345
6|7   3.9032    1.9983   1.953
7|8   5.5767    2.0018   2.786
8|9   7.1297    2.0072   3.552
```

anova (fm5,fm7)

```
Likelihood ratio tests of cumulative link models:
     formula:
fm5 Acceptable ~ Prosody + Fluency + Segmental + (1 | Employer) + (1 |
Speaker)
fm7 Acceptable ~ Lexis + Prosody + Fluency + Segmental + (1 | Employer) +
(1 | Speaker)
    link: threshold:
fm5 logit flexible
fm7 logit flexible


    no.par    AIC  logLik LR.stat df Pr(>Chisq)
fm5     13 1626.1 -800.05
fm7     14 1620.4 -796.21  7.6682  1    0.00562 **
```

11.4.6 Appendix D6: Qualities as IVs run against standard model (Survey A)

```
Qualities_Centered.R
Semantic differential qualities centred and run with the three factors as
IVs.
Cumulative Link Mixed Model fitted with the Laplace approximation

Acceptable ~ cFriendliness + cWorkhard + cPoliteness + Factor1 +
    Factor2 + Factor3 + (1 | Employer) + (1 | Speaker)
data:    empA

 link  threshold nobs logLik  AIC    niter      max.grad cond.H
 logit flexible  209  -366.86 765.73 1661(6579) 2.03e-04 1.2e+02
Random effects:
 Groups    Name          Variance Std.Dev.
 Employer (Intercept) 1.3849   1.1768
  Speaker  (Intercept) 0.5702    0.7551
Number of groups:  Employer 45,  Speaker 40
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
cFriendliness  0.2107     0.1822   1.157  0.24739
cWorkhard      0.7384     0.1801   4.101 4.12e-05 ***
cPoliteness    0.2277     0.1837   1.240  0.21512
Factor1        0.4864     0.1972   2.466  0.01365 *
Factor2        0.6459     0.2017   3.202  0.00137 **
```

```
Factor3          0.6990      0.2192   3.189  0.00143 **
```
Threshold coefficients:
```
    Estimate Std. Error z value
1|2  -4.6598      0.5150   -9.049
2|3  -3.3279      0.4116   -8.086
3|4  -2.4093      0.3550   -6.786
4|5  -1.6948      0.3214   -5.273
5|6  -0.2655      0.2893   -0.918
6|7   1.1476      0.3023    3.797
7|8   2.7756      0.3682    7.538
8|9   4.3589      0.4744    9.188
```

Correlation of Coefficients:
```
           cFriend  cWork  cPolite  Factor1
cWorkhard  -0.3037
cPoliteness -0.5666 -0.2599
Factor1     0.0068 -0.0060  0.0621
Factor2     0.0313  0.0434 -0.0346   0.1233
Factor3    -0.0131  0.0966 -0.0289   0.1019
```

CLMM Model with Workhard only
```
Acceptable ~ cWorkhard + Factor1 + Factor2 + Factor3 + (1 | Employer) +
(1 | Speaker)
data:    empA

 link  threshold nobs logLik  AIC     niter       max.grad cond.H
 logit flexible  209  -370.22 768.44 1085(5139) 3.07e-04 4.9e+01
```
Random effects:
```
 Groups    Name          Variance Std.Dev.
 Employer (Intercept) 1.4541    1.2059
 Speaker  (Intercept) 0.5879    0.7667
Number of groups: Employer 45,  Speaker 40
```
Coefficients:
```
          Estimate Std. Error z value Pr(>|z|)
```
cWorkhard   1.0413   0.1417   7.349 1.99e-13 ***
```
Factor1    0.4607      0.1973   2.335 0.019542 *
Factor2    0.6640      0.2035   3.263 0.001104 **
Factor3    0.7386      0.2205   3.349 0.000811 ***
```
Threshold coefficients:
```
    Estimate Std. Error z value
1|2  -4.6085      0.5120   -9.001
2|3  -3.2945      0.4117   -8.003
3|4  -2.3914      0.3569   -6.700
4|5  -1.6841      0.3240   -5.198
5|6  -0.2742      0.2921   -0.939
6|7   1.1287      0.3041    3.711
7|8   2.7530      0.3678    7.486
8|9   4.3001      0.4696    9.157
```
Correlation of Coefficients:
```
cWkhard Fac1    Fac2
Factor1  0.0324
Factor2  0.0443  0.1224
Factor3  0.0857  0.1075
```

ANOVA with 3-factor model

```
Likelihood ratio tests of cumulative link models:
fm0 Acceptable ~ Factor1 + Factor2 + Factor3 + (1 | Employer) + (1 |
Speaker)
fm2 Acceptable ~ cWorkhard + Factor1 + Factor2 + Factor3 + (1 | Employer) +
(1 | Speaker)
```

```
    link: threshold:
fm0 logit flexible
fm2 logit flexible

    no.par   AIC  logLik LR.stat df Pr(>Chisq)
fm0    13 834.00 -404.00
fm2    14 768.44 -370.22  67.559  1  < 2.2e-16 ***

Model with Workhard is different
```

11.4.7 Appendix D7: Residence

Residence.R

Hypothesis: That Residence is a predictor of Acceptability.

Conclusion: P values do not support the hypothesis.

P values in ANOVA comparing model including Residence with model without Residence do not support the hypothesis.

```
Cumulative Link Mixed Model fitted with the Laplace approximation
formula: Acceptable ~ Residence + Factor1 + Factor2 + Factor3 + (Resi-
dence:1 | Employer) + (1 | Speaker)
link  threshold nobs logLik  AIC     niter      max.grad cond.H
 logit flexible  424  -799.59 1627.17 1325(8219) 2.20e-04 5.9e+02

Random effects:
 Groups   Name         Variance Std.Dev.
 Employer (Intercept) 2.650    1.628
 Speaker  (Intercept) 1.077    1.038
Number of groups:  Employer 95,  Speaker 40

Coefficients:
         Estimate Std. Error z value Pr(>|z|)
Residence  -0.4036     0.4196  -0.962 0.336099
Factor1     0.6643     0.1976   3.362 0.000774 ***
Factor2     0.8650     0.2083   4.153 3.29e-05 ***
Factor3     0.6832     0.2166   3.154 0.001609 **

Threshold coefficients:
    Estimate Std. Error z value
1|2  -4.8407     0.6811  -7.108
2|3  -3.5772     0.6460  -5.537
3|4  -2.7074     0.6299  -4.298
4|5  -1.9014     0.6201  -3.067
5|6  -0.7536     0.6116  -1.232
6|7   0.4703     0.6096   0.771
7|8   2.1560     0.6225   3.463
8|9   3.7259     0.6539   5.698

ANOVA
fm3 Acceptable ~ Factor1 + Factor2 + Factor3 + (1 | Employer) + (1 |
Speaker)
```

```
fm4 Acceptable ~ Residence + Factor1 + Factor2 + Factor3 + (Residence:1 |
Employer) + (1 | Speaker)
    link: threshold:
fm3 logit flexible
fm4 logit flexible
    no.par   AIC  logLik LR.stat df Pr(>Chisq)
fm3     13 1626.1 -800.05
fm4     14 1627.2 -799.59  0.9232  1     0.3366
```

11.4.8 Appendix D8: Employee Use of English

```
UseEnglish.R
Cumulative Link Mixed Model fitted with the Laplace approximation

formula: Acceptable ~ UseEnglish + Factor1 + Factor2 + Factor3 + (1 |
Employer) + (1 | Speaker)
 link  threshold nobs logLik  AIC      niter        max.grad cond.H
 logit flexible  424  -797.76 1631.52 2236(16953) 7.10e-04 2.8e+03
```

Random effects:
```
 Groups    Name          Variance Std.Dev.
 Employer (Intercept) 2.505    1.583
 Speaker  (Intercept) 1.048    1.024
Number of groups:  Employer 95,  Speaker 40
```
Coefficients:
```
                        Estimate Std. Error z value Pr(>|z|)
UseEnglishAlways          0.5219     1.1338   0.460 0.645302
UseEnglishAppropriate    0.6022     1.1741   0.513 0.608026
UseEnglishCustomers      0.6934     1.0916   0.635 0.525296
UseEnglishWork           0.7545     1.4043   0.537 0.591081
UseEnglishWork&Breaks   -1.0194     1.3141  -0.776 0.437898
Factor1                  0.6651     0.1956   3.400 0.000675 ***
Factor2                  0.8537     0.2063   4.138 3.51e-05 ***
Factor3                  0.6741     0.2146   3.142 0.001681 **
```
Threshold coefficients:
```
    Estimate Std. Error z value
1|2  -3.7896     1.1003  -3.444
2|3  -2.5220     1.0872  -2.320
3|4  -1.6513     1.0810  -1.528
4|5  -0.8436     1.0773  -0.783
5|6   0.3006     1.0753   0.280
6|7   1.5164     1.0762   1.409
7|8   3.1989     1.0865   2.944
8|9   4.7676     1.1098   4.296
```

```
        ANOVA
Likelihood ratio tests of cumulative link models:
link:
fm0 Acceptable ~ Factor1 + Factor2 + Factor3 + (1 | Employer) + (1 |
Speaker)            logit
fm1 Acceptable ~ UseEnglish + Factor1 + Factor2 + Factor3 + (1 | Employer)
+ (1 | Speaker) logit
    threshold:
fm0 flexible
fm1 flexible

    no.par    AIC   logLik LR.stat df Pr(>Chisq)
fm0     13 1626.1 -800.05
fm1     18 1631.5 -797.76   4.571  5     0.4704
```

11.4.9 Appendix D9: Marty Factor Analysis Minus Speed

```
Data: ExpertRatings_Martyonly.txt
Factor Analysis with myself as the only Experienced Rater
factanalMarty_13Factors.R
Call:
factanal(x = er[5:17], factors = 3, scores = "regression")

Uniquenesses:
    Consonants          Vowels    StressMarked    StressPlace     Intonation
         0.44            0.38            0.57           0.48           0.00
     Contrast    Coarticulation       Pausing    AddMissFinal    AddMissOther
         0.31            0.36            0.16           0.25           0.35
   Monotonity      Choppiness      Hesitancy
         0.23            0.40            0.26

Loadings:
               Factor1 Factor2 Factor3
StressMarked    0.53
StressPlace     0.67
Intonation      0.98
Contrast        0.76
Coarticulation          0.71
Pausing         0.42    0.82
Variety         0.61    0.62
Smoothness              0.70
Fluidity                0.78
Consonants                      0.66
Vowels          0.44            0.61
AddMissFinal                    0.77
AddMissOther                    0.80


               Factor1 Factor2 Factor3
SS loadings       3.21    3.18    2.42
Proportion Var    0.25    0.24    0.19
Cumulative Var    0.25    0.49    0.68


Test of the hypothesis that 3 factors are sufficient.
The chi square statistic is 58.89 on 42 degrees of freedom.
p<0.05
```

11.5 Appendix E Pronunciation problems – qualitative responses

Have any of your non-native speakers had pronunciation problems which affected their work?

Please briefly describe the problem(s).

- Communication is the key to all successful organisations. Often our staff misunderstand, or we can't understand them. / Often some guests complain that cannot understand some staff members.
- Customers found it hard to understand
- Speed/efficiency in workload when not understanding instructions/information - i.e. language barriers between customers and/or co-workers. /
- Staff was on reception and sometimes it was difficult getting correct name or details clearly from phone calls.
- In our operation we have communications between our patrols and the local monitoring company. If the patrolman is say from 6 the comms operator finds it harder to understand what they are saying.
- Their pronunciation very hard to catch.
- Hard to understand on the telephone, but okay face to face
- When people are having problems with their English pronunciations, it will make others hard to understand what they are saying and this can cause conflict amongst colleagues or with customers.
- Telephone conversations where person ringing in cannot fully understand the reply. Customers at checkout not understanding response from staff member. co-workers not understanding verbal response
- Some accents are too strong - funny case - ice cream in South African accent can sound like arse cream!
- Nothing major- just communicating with customers in a fast paced environment is sometimes difficult when there are words misunderstood.
- Customers unable to understand them, so service was diminished due to them not being able comprehend correctly
- Elderly deaf and dementia residents can have difficulty understanding the accents
- Special instructions relating to a job not clearly understood, a lot of communication is via radio telephone so it is important what is said is clear and concise
- Difficult to understand what they mean, very strong accent
- Difficult pronunciation and hard to understand by customers and colleagues
- Not understanding written instructions
- Understanding direction and Health and Safety information
- Sometimes it is required to clarify a question/request as you can see they don't fully understand, so this is only a problem as it can be time consuming under some circumstances.
- misunderstandings & customers not being able to understand some phrases
- Some of my clients are not always supportive of staff who have some difficulty with English

- The level of English spoken by this person caused problems in communicating tasks and standards required in their department. This led to performance issues.
- Not being able to speak English and make their wishes known
- Difficulty in understanding what they are trying to pronounce. i.e. the sound they make is not what I expect the word they are trying to say to sound like especially with v's and w's
- When dealing with customers or other staff they can be misunderstood or not understand the communication they are receiving.
- Clarity of understanding what they are saying or receiving instructions.
- At times it was very difficult for elderly and deaf clients to understand and when clients have short term memory loss they cannot concentrate on what is being said.
- Difficulty particularly on the phone. Miscommunication. Saying yes when they clearly do not understand. The elderly have difficulty as some speak too softly with pronounced accents.
- I would not call it a problem, just as they learn the language and become efficient. Sometimes customers cannot understand them, we work with the employee & make sure they are not put in a situation.
- Some workers get misunderstood due to pronunciation problems. In one case a customer was asking for polonies and the worker directed the customer to Checkouts where the Manager's name is [similar sounding].
- no major problems, just takes a little longer to understand each other... often demonstrating things is useful and quite comical for both parties at time(senses of humour always good)
- Too shy, because of their communication barriers. Patrons not happy when cannot understand answers to basic questions.
- When they answer the phone sometimes our customers are unable to understand them. Many times they do not follow instructions as they have not understood them
- misunderstand instructions
- An engineer who was very good, but due to very strong 8i accent customers would get frustrated as they couldn't understand him on the phone.
- Difficulty in following specific work instructions due to differences with accents and pronunciation; 40% of the managers are immigrants themselves.
- One worker gets ignored by other workers fairly often because his accent is so bad we can't understand what he is saying. We smile and nod.
- Problems understanding the words used. Most commonly between two people who both have English as a second Language.
- Could not communicate with others apart from using primary words, used body language
- Misunderstanding with co-workers due to differences in languages and cultures, difficulty in understanding and responding to customers in English, difficulty in carrying out their job
- Problems understanding the words that sound similar ie letter and ladder
- Pronunciation issues - dialect hard for colleagues to understand (the person is an IT helpdesk worker). Especially difficult over the phone.

- Failure to understand instructions / - Failure to be able to communicate problems / - Customers can't understand employee / - Co-workers cannot understand.
- Understanding written & verbal instruction
- Understanding English directives and instruction
- Difficulty in staff being understood by residents (clients), family members and colleagues and especially by phone resulting in poor/misinformation & sometimes lack of trust.
- Need to be clearly understood. Using street names address suburbs - nationwide. Pronunciation needs to be accurate. Security language can change meanings of English words.
- Customer Service is where they find it difficult to communicate or to have a clear understanding on what the customer is inquiring.
- Customers not being able to understand.
- Problems when ordering supplies for jobs. I now tell them to text me.
- we had both Irish with broad accents and 2n people very accented English
- Minor clarification required.
- The problem in our industry is that we have clients that are very hard of hearing so with strong accents and/or poor pronunciation of words it makes employees that much harder to understand.
- Being understood by customers both in person, and (especially) via telephone
- Trying to get the customer to understand clearly what they are saying when answering a question
- Guests having to say 'pardon' all the time and getting them to repeat themselves
- Unclear pronunciation or word choice. Accent heavily impacted on speech. Lack of confidence also affected speech which was quiet and disjointed.
- New Zealand customers are sometimes confused by the message they are saying and get frustrated through lack of understanding them
- Using a telephone and or Radio communications - difficult to understand.
- Inability to interact with customers effectively while doing their job. Hard to understand.
- Using the public address appears to accentuate an accent and it is often difficult to understand what is being said.
- Particularly on the telephone, clients have difficulty understanding their accent.
- Unable to understand something simple by incorrect pronunciation