Identifying Māori English and Pākehā English from Suprasegmental Cues: A Study Based on Speech Resynthesis

A thesis
submitted in partial fulfilment
of the requirements for the Degree
of
Master of Arts
in Linguistics in the
University of Canterbury
by
Anita Szakay

University of Canterbury
2007

In loving memory of Ratahi Waigth.

Takoto mai rā e te whaea,
Me hoki koe ki ō tūpuna,
Ki Hawaiki Nui,
Ki Hawaiki Roa,
Ki Hawaiki Pāmaomao.
Moe mai rā, moe mai rā.
Arohanui!

Abstract

This thesis investigates the suprasegmental properties of Māori English and Pākehā English, the two main ethnolects of New Zealand English. Firstly, in a Production Experiment the speech of 36 New Zealenders is acoustically analysed. Using the Pairwise Variability Index (PVI) to measure syllabic rhythm, the study reveals that the two ethnic varieties display differing rhythmic patterns, with Māori English being significantly more syllable-timed than Pākeha English. It is also shown that, overall, Māori speakers use a higher percentage of High Rising Terminals than Pākehā speakers. The results relating to pitch suggest that Māori English pitch is becoming higher over time, with young Māori speakers producing a significantly higher mean pitch than young Pākehā speakers.

Secondly, a Perception Experiment using 107 listeners is carried out to investigate the role of suprasegmental information in the identification of Māori English and Pākehā English. The ability of listeners to identify the two dialects based on prosodic cues only is tested in seven different speech conditions. The various conditions aim to isolate the precise suprasegmental features participants may use to identify speaker ethnicity. The results reveal that listeners are aware of the differing rhythmic properties of Māori English and Pākehā English, and are capable of tuning into the rhythmic characteristics of a speaker to use it as a cue in dialect identification, with some level of accuracy. The perceptual relevance of other prosodic cues is also discussed and the results indicate that, based on certain stereotypes, Māori English speech is assumed to be low-pitched, monotonous, hesitant and slow in pace. It is also shown that listeners who have had greater exposure to Māori English perform significantly better in a dialect identification task than those who are not integrated into Māori social networks, proving that the linguistic experience of the listener is a key indicator of his or her performance in ethnic dialect identification.

Table of Contents

List of Figures

List of Tables

Acknowledgements

Firstly, I would like to acknowledge all 36 speakers and 107 listeners who dedicated their time to participate in this experiment. Without you this thesis simply would not exist.

A very special thanks must be extended to my wonderful supervisor Dr. Jen Hay. Your infinite wisdom and enthusiasm helped keep me on track, and your 'gentle prodding' ensured that I was never able to lose focus. I cannot be grateful enough for your contribution to this thesis.

I would also like to thank my assistant supervisor Dr. Alex D'Arcy. Your encouragement and insight has been invaluable. The knowledge you have been able to pass on was *like* tremendously helpful.

To the University of Canterbury Linguistics Department, thank you for providing a stimulating and fun environment in which to study. I feel privileged to have known such a fantastic group of people.

I would also like to express my appreciation to all members of our weekly ONZE meetings for their useful comments and for their willingness to be 'guinea pigs' for my experiments.

I am also thankful to the University of Canterbury for providing a Masters scholarship that immensely helped with the cost of this research.

I am grateful for the useful feedback I received from audiences at the following conferences: 3rd International Postgraduate Linguistic Conference, 10th New Zealand Language and Society Conference, and the 11th Australasian Speech Science and Technology Conference.

To Gabriella Motoi and Ben Eves, I am thankful that you took the time to help in the construction of my experiment. Thanks for lending your voices and your support throughout this thesis.

Special thanks to Joseph Ramsay for your support. Without you this would not have been possible.

To my dear friend Emma Parnell, I would not have survived this year without you and our coffee breaks, so a big thank you must also go your way for making me laugh, keeping me sane, and generally being a great friend.

Lastly, a very special thanks to my family for their unwavering love and support from afar. Nélkületek soha nem jutottam volna el idáig. Köszönöm hogy mindvégig hittetek bennem. Borzasztóan hiányoztok!

Part I

Introduction & Background

Chapter I

Introduction

Most New Zealanders would agree that there is a distinctive Māori variety of English. Linguists, on the other hand, have been trying to identify its unique characteristics since the 1960s with not much success. Only in the last fifteen years have linguists begun to confirm some core features which differentiate the English of Māori and Pākehā speakers (Bell 2000). Most of these are phonetic and discourse features, although some linguists have researched prosodic features as well (Holmes and Ainsworth 1996, Holmes and Ainsworth 1997, Warren 1998). It seems that some of the most salient features of Māori English are in fact prosodic, namely rhythm and intonation.

The first aim of the present study is to establish whether Māori English and Pākehā English differ in their rhythmic patterns, as has been suggested by previous research. Other prosodic features are also acoustically analysed in an attempt to identify to what extent the two major ethnic dialects of New Zealand English differ from one another in suprasegmental cues.

Secondly, a perception experiment is carried out investigating the role of suprasegmental information in the identification of Māori English and Pākehā English by naive listeners of each of these varieties. The ability of listeners to identify the two dialects based on prosodic cues only is tested in seven different speech conditions. The various conditions aim to isolate the precise suprasegmental features listeners may use to identify the ethnicity of a speaker. Participants' exposure to the Māori English variety is also measured to investigate whether listeners' linguistic background has an effect on their accuracy at a dialect identification task.

Finally, we investigate the perceptual relevance of the prosodic parameters that were identified in the acoustic analysis. The main objective of this investigation, therefore, is to conduct a contrastive examination of the prosodic structure of Māori English and Pākehā English in order to establish to what extent the two dialects differ prosodically and whether naive listeners are able to tune into fine prosodic details to facilitate ethnic judgements.

The overall thesis structure is as follows:

- Part I of this thesis is intended to present important background information. Chapter 2 provides a general introduction on suprasegmentals, with special emphasis on linguistic rhythm. Features of Māori English will also be discussed, both segmental and suprasegmental. The chapter concludes with an exploration of previous studies that attempted to separate segmental information from suprasegmentals in various dialect identification experiments.

- Part II presents the production experiment. It first details the methodology and the acoustic analysis used to investigate prosodic features of Māori English and Pākehā English, then describes and discusses the results of the experiment. The results relating to rhythm are consistent with previous research indicating a significant difference between the two dialects. Some unexpected results concerning pitch are also discussed.

- Part III forms the major part of this thesis. It consists of four chapters, all of which deal with the perception experiment. Chapter 4 describes the methodology used to set up the dialect identification task, and gives comprehensive information on how the passages were re-synthesised and how each of the seven different speech conditions were created. Chapter 5 presents detailed results from the statistical analyses for each condition. These results are then discussed and summarised in Chapter 6. Finally, Chapter 7 gives a general summary of the thesis, and presents the important conclusions and suggestions for future research.

Chapter II

Background

The present chapter aims to provide background information about the various elements incorporated in the title of this thesis. Thus, first various aspects of suprasegmental features will be discussed with special attention to rhythm, followed by a section on Māori and Pākehā English. Then previous dialect identification experiments will be mentioned, with special emphasis on studies that made use of various techniques from speech science in an attempt to isolate and compare different suprasegmental cues that listeners might rely on when identifying the dialect of a speaker.

2.1 Suprasegmental Features of Speech

The present study is concerned with suprasegmental or prosodic features, which are the aspects of speech production that involve more than a single segment, that is, more than just a single vowel or consonant. Prosody can refer to, for example, intonation, tone of voice, emphasis, pauses, and changes in speech rate. Suprasegmental features of speech may carry grammatical and affective information but can also be used in different contexts to convey social information. The three most often mentioned phonetic dimensions of prosody are intonation, rhythmic features and voice quality.

Intonation is manifested primarily by pitch dynamics, which is an auditory property. The acoustic correlate of pitch is the fundamental frequency (F0) of the sound wave, which can be instrumentally measured and manipulated. Numerous studies have attempted to characterise the role of intonation in the perception of a dialect by separating segmental information from intonation (see Section 2.3.1.). To eliminate pitch dynamics from the speech signal, passages can be monotonised by flattening the intonation. This makes it possible to investigate whether naive listeners rely more on segments or intonation when identifying a dialect. Another technique used to isolate segmental information is to high-pass filter the speech signal. A high-pass filter eliminates low frequency information but retains vowels and consonants, which are generally contained in the higher frequencies.

The opposite of high-pass filtering is low-pass filtering, which retains frequencies between 0Hz and the cut-off frequency. This is generally around 400Hz in experiments that aim to eliminate segmental information from the speech signal but retain both intonation and rhythm. As linguistic rhythm is one of the focal points in this thesis, Section 2.1.1 discusses it in more detail. Low-pass filtering arguably also retains voice quality information in the signal.

Descriptions of voice quality have traditionally consisted of qualitative terms such as *warm*, *shrill*, *creaky*, *shimmery*, *breathy*, *hoarse*, *jittery* and *rough*. The acoustic and articulatory correlates of these terms have not been well defined. However, it seems possible to examine *jitter* by measuring F0 perturbation, *shimmer* by amplitude perturbation and *hoarse*-ness by harmonics-to-noise ratio (Walton and Orlikoff 1994). The present study does not investigate features of voice quality, however, it will be argued that speaker voice quality might be extremely relevant in listeners' ethnic judgements in the New Zealand context.

2.1.1 Linguistic Rhythm

Languages are traditionally divided into two main categories based on their rhythm. *Stress-timed languages*, such as English and Dutch, are said to have regular intervals between main stresses, while in *syllable-timed languages*, for example Spanish and Italian, it is syllables that occur at regular intervals. The difference between the two groups of languages was noted as early as 1940 by Lloyd James, who used the metaphor 'morse-code rhythm' for the first group of languages and 'machine-gun rhythm' for the second. Pike (1946) also strongly expressed this dichotomy, renaming the first group 'stress-timed' and the second group 'syllable-timed' languages. Abercrombie (1967) was the first to claim that the difference in rhythm was based on the isochrony of either syllables or interstress intervals.[1]

More recently, however, the strict categorical distinction between stress- and syllable-timing has been contradicted by various empirical studies. Research has not provided support from duration measurements for isochronous timing (Laver 1994), thus isochrony has been described as a tendency rather than an absolute (Beckman 1992) . Dauer (1987) proposed that all languages are more or less stress-based, while Miller (1984) suggests that languages are best described along a continuum of syllable- and stress-timing.

Below is a schematic representation of stress- and syllable-timing, where each block represents one syllable. In syllable-timed languages the duration of syllables is near equal, while in stress-timed languages the durational variability of syllables is much greater.

Syllable-timing:

Stress-timing:

There have been various proposals for the instrumental measurement of rhythm that avoid language-dependant phonological concepts (e.g. syllables), and instead of calculating syllable- and interstress-durations, use purely phonetic characteristics of the speech signal. Ramus, Nespor and Mehler (1999) segmented speech into vocalic and consonantal segments and computed three acoustic correlates of rhythm:

a) $\%V$, the average proportion of vocalic intervals
b) ΔC, the average standard deviations of consonantal intervals and
c) ΔV, the average standard deviations of vocalic intervals.

Relying on these measurements, the authors plotted eight rhythmically different languages in a three-dimensional space and argued that their results support the notion of rhythm classes.

[1] Further work suggested a third category of mora-timed languages, including Japanese, Tamil and Māori itself (Bloch 1950, Han 1962, Ladefoged 1975). In mora-timed languages, successive morae are said to be near equal in duration. Morae are sub-units of syllables, thus these languages are more similar to syllable-timed languages.

In other studies, such as Low and Grabe (1995), Low, Grabe and Nolan (2000) and Grabe and Low (2002), the authors computed a Pairwise Variability Index (PVI), also based on the duration of vocalic and intervocalic segments. The PVI arguably provides a better control for speech rate variations than does the simple calculation of average standard deviation. The present study follows Grabe and Low (2002) in its method of measuring rhythm.

The Pairwise Variability Index

The vocalic Pairwise Variability Index is based on the relative difference in duration of successive vocalic segments and is normalised for local rate variations. A low PVI value shows less variation in vowel duration, and as such indicates a more syllable-timed language. Stress-timed languages, on the other hand, typically demonstrate shorter unstressed vowels alternating with longer vowels, resulting in a higher PVI.

Grabe and Low (2002) use a version of the PVI which is normalised for speech rate. The equation for the normalised PVI is given in Equation (2.1). The equation shows that the PVI is compiled by calculating the difference in duration between each pair of successive measurements, taking the absolute value of the difference and dividing it by the mean duration of the pair. The differences are then summed and multiplied by 100 because the normalisation produces fractional values. In Grabe and Low's (2002) study nPVI stands for the normalised vocalic PVI.

$$nPVI = 100 \times \left[\sum_{k=1}^{m-1} \left| \frac{d_k - d_{k+1}}{(d_k + d_{k+1})/2} \right| / (m-1) \right] \tag{2.1}$$

where m = *number of intervals* and d = *duration of the k^{th} interval*

Grabe and Low (2002) also calculate raw, non-normalised PVI values for intervocalic segments. They argue that it is unnecessary to apply the normalisation procedure to the consonantal segments because speech rate does not seem to affect the duration of consonants to the same extent as it does the duration of vowels. The raw intervocalic PVI is abbreviated rPVI in their study. In the present thesis, PVI refers to the normalised vocalic PVI, unless otherwise specified.

Grabe and Low (2002) report PVIs of 57.2 for British English and 52.3 for Singapore English, showing that Singapore English is a more syllable-timed variety of English. A selection of PVI values for various languages taken from their study is reported in Table 2.1. These values are based on measurements from reading passages. The subjects were asked to read the passage 'The North Wind and the Sun', or its corresponding translation in the relevant language. With the help of PVI values we can plot languages along a continuum of stress- and syllable-timing instead of categorically distinguishing between stress- and syllable-timed languages.

Table 2.1: PVI values for selected languages based on reading passages. Data taken from Grabe & Low (2002).

Language	PVI
Mandarin	27.0
Spanish	29.7
Luxembourg	37.7
Japanese	40.9
French	43.5
Polish	46.6
Rumanian	46.9
Welsh	48.2
Singapore English	52.3
Malay	53.6
British English	57.2
German	59.7
Dutch	65.5
Thai	65.8

2.2 Māori English

The number of fluent, native Māori language speakers has steadily decreased in New Zealand, and English has become the dominant language of almost all Māori people, although many are also familiar with the Māori language. In such circumstances, it is not surprising that a distinctive variety of Māori English has emerged to express ethnic identity and positive attitudes toward Māori culture (Holmes 2005).

The term New Zealand English (NZE) is not as well defined as it first might seem. King (1993) points out that what most sociolinguists would call NZE, is probably better labelled as Pākehā English, that is, the English spoken mainly by European New Zealanders. In fact, NZE covers many varieties, including both Pākehā English and Māori English. King also suggests that Māori English is not restricted to ethnically Māori speakers, but is also used by some Pākehā who either grew up or identify with Māori peer groups. It is also the case that not all ethnically Māori speak Māori English.

Previous research has suggested that the differences between Māori English and Pākehā English tend to be relative rather than absolute. There are many features that are shared by both dialects but where the frequency of forms in each variety differs. Since the 1990s, linguists working on Māori English have made numerous attempts to identify the core features that differentiate the two dialects, at least quantitatively if not qualitatively. Some of these studies concentrated on phonological features, such as the pronunciation of vowels and consonants (e.g., King 1993, Robertson 1994, Holmes 1996, Bell 2000), while others set out to identify possible prosodic differences (e.g., Bauer 1994, Holmes and Ainsworth 1996, Holmes and Ainsworth 1997, Warren 1998, Bell 2000). Some of the linguistic features of Māori English that have been so far identified are described in the following sections.

2.2.1 Segmental Features

Final /z/ devoicing

Devoicing of word final /z/ by young Māori children was noted as early as 1966 by Benton. More recent research confirmed that it has developed into and identifiable feature of Māori English, with Māori speakers using a higher percentage of the devoiced variant than non-Māori. Bell (2000) claims that devoicing of /z/ is part of stereotypical representations of Māori English, producing for example a lengthened [s:] in ‘boys’.

Initial /t/ non-aspiration

In Holmes and Ainsworth’s (1996) study there was a significant difference in the frequency with which Pākehā and Māori speakers used unaspirated word-initial /t/. The Māori participants used the non-aspirated variant more than seven times as often as did the Pākehā participants. The results in Bell (2000) confirm this tendency. It is worth noting that the Māori language itself traditionally has a relatively unaspirated /t/, although this seems to be changing under the influence of English (Bauer 1997).

Stopping, affrication and fronting of /θ/ and /ð/

Bell (2000) suggests that in the case of Māori English variants of /θ/ and /ð/, there is a confluence of Māori language influence and general vernacular usage, as shown in other varieties such as African American English. The Māori speaker in Bell’s study uses higher frequencies of the stopped, affricated and fronted variants than the Pākehā speaker. The fronted variants [f] and [v] in particular seem to be currently becoming more salient features of Māori English.

Fronting of /u/

In NZE /u/ seems to have a generally centralised realisation. However, Bell (2000) shows that his Māori speaker uses an extreme fronted version more than twice as often (38% of the time) than does the Pākehā speaker, and only produces 6% of the back token.

2.2.2 Suprasegmental Features

Syllable-timing

Although English in general is considered to be stress-timed, it has been suggested that NZE shows a tendency towards syllable-timing. Previous work on rhythm in New Zealand has also claimed that there may be a difference in timing patterns within the variety, with Māori speakers producing more syllable-timed speech than Pākehā speakers (Ainsworth 1993, Holmes and Ainsworth 1996, Holmes and Ainsworth 1997, Warren 1998). In fact, various patterns that signal a greater degree of syllable-timing were already noted in the speech of Māori children by Benton (1966). These include shortening of long vowels or diphthongs, devoicing and lengthening of consonants and use of full vowels for schwa in unstressed syllables.

To investigate the degree of syllable-timing in NZE, an auditory analysis of recordings from radio broadcasts was carried out by Ainsworth (1993) where she measured full-vowel vs. reduced-vowel ratio. The recordings were obtained from the BBC World News, two commercial New Zealand stations, the more conservative National Radio and the Māori news service, Mana News. Her results revealed that full vowels were retained nearly 60% of the time in the Māori sample, and only around 20% in the case of the recordings from the BBC and the two commercial New Zealand stations. Interestingly, the more prestigious National Radio sample retained about 40% full vowels. Ainsworth speculates that this has to do with speech rate, such that the slower, more deliberate style of National Radio results in an increase in full vowels, compared to the faster speech of the commercial radio stations.

A detailed acoustic analysis of these recordings by Warren (1998) confirmed that there is a rhythmic distinction between British English and NZE, and also one between Māori English and Pākehā English. Using a non-normalised version of the PVI, first put forward in Low and Grabe (1995), Warren demonstrated that Māori English is more syllable-timed than Pākehā English, which in turn is more syllable-timed than British English.

High Rising Terminals

The High Rising Terminal (HRT) is a salient rise in pitch at the end of non-interrogative intonation phrases and is a feature of many varieties of English. It has been extensively studied in NZE, and is reported to be used mainly by young, female speakers. A major study carried out by Britain (1992) showed evidence that this pattern is used in different proportions by Māori and Pākehā. His analysis indicated that Māori speakers use a significantly higher percentage of HRTs than Pākehā. The results also revealed that young Māori men use levels similar to women, while young Pākehā men are extremely low users of HRTs.

2.3 Previous Dialect Identification Experiments

Dialect identification experiments were carried out in the United States as early as the 1950s. Many of these studies involved playing recordings of African American and European American voices to listeners who were asked to identify the speaker's ethnicity (e.g., Dickens and Sawyer 1952, Stroud 1956). Such early studies revealed that Americans were able to correctly categorise the ethnicity of speakers much of the time. These experiments did not aim to determine what cues listeners used to distinguish speaker ethnicity.

New Zealand was, for a while, lagging behind America in investigations concerning different English accents, however, there have now been a number of relevant studies carried out, as Bayard (1995) points out. The first of these studies by Huygens and Vaughan (1983) investigated attitudes of Pākehā students toward Māori accents (as well as those of British and Dutch immigrants). Their study demonstrated that once speakers are identified as Māori based on their accent, they are rated as less educated, less intelligent, belonging to a lower social class and lazier than speakers classified as Pākehā. Interestingly, in Huygens and Vaughan's (1983) study, only 25% of participants correctly identified ethnically Māori speakers as such.

Robertson (1994) also investigated ethnic identification and attitudes in New Zealand. Her study confirmed the results from Huygens and Vaughan (1983), showing that once participants decided that a voice is Māori rather than Pākehā, the ratings on variables reflecting social power decline significantly. However, correct ethnic identification rates were higher in Robertson's study, with 55% of her participants accurately identifying Māori speakers. She also investigated the participants' social network and concluded that those listeners who are highly integrated into Māori society perform best at the dialect identification task.

2.3.1 Experiments with Suprasegmentals

As mentioned above, the earliest dialect identification experiments examined whether listeners can distinguish between dialects, without testing what cues listeners relied on. More recently, however, various speech science methods have been applied to the study of the perception of linguistic variation with the aim of isolating the precise prosodic features listeners use to identify the dialect of a speaker (Thomas and Reaser 2002). Many of these studies concentrated on differences between African American and European American speakers in the United States. Researchers have employed various techniques to manipulate the speech signal in order to identify the cues listeners rely on to distinguish between African American and European American speakers. These include playing passages forward, backward and time-compressed (Lass, Mertz and Kimmel 1978), low-pass filtered and high-pass filtered (Lass, Almerino, Jordan and Walsh 1980, Foreman 2000), and using resynthesized F0 levels (Hawkins 1992). In Lass et al.'s (1978) study ethnicity was correctly identified 71% of the time when the passages were played forward, 62% of the time when played backward and 65% of the time when played time-compressed. Accuracy levels in Lass et al. (1980) were 72% for unfiltered, 60% for low-pass filtered and 69% for high-pass filtered passages. This latter study concluded that formant information was more important for ethnic identification than F0 information. Hawkins (1992) resynthesized the /æ/ vowel at nine F0 levels using 4 male speakers, and concluded that lower F0 levels were associated with identification as African American. Foreman (2000) used low-pass filtering at 900Hz in order to get rid of some segmental information and make intonation itself more salient. She extracted 54 sentences from read scripts designed to bring out intonational patterns and determined that intonation could be a cue in identifying African American English, as ethnicity was identified with over 80% accuracy. Foreman (2000) also found that listeners with more exposure to both dialects were more accurate at the dialect identification task. This result is consistent with Robertson's (1994) study for NZE, but goes one step further in showing that degree of exposure is a relevant factor in a dialect identification task even when the speech signal has been manipulated.

In their experiment using Hyde County African American speakers, Thomas and Reaser (2002) applied three different treatments to their speech samples in an attempt to compare different cues. The first condition left the passages unmodified. The second treatment was monotonization, setting F0 at 120Hz for male speakers and 200Hz for female speakers. In the third condition low-pass filtering was applied at 300Hz. Their results suggested that vowel quality and intonation were the most useful cues in ethnic dialect identification. However, the authors noted that rhythm and voice quality could have also served as possible cues.

Experimental techniques to isolate certain prosodic cues are not exclusive to studies on the perception of African American English. Bezooijen and Gooskens (1999) describe two experiments that investigated the role of different linguistic levels in the identification of regional Dutch and British English varieties respectively. In the Dutch experiment they created three different conditions they named *prosodic*, *verbal* and *integral*. In the *prosodic* condition they applied low-pass filtering at 350Hz, while the *verbal* condition included monotonised speech. Unaltered passages were presented in the *integral* condition. Their results indicated that the identification of Dutch regional varieties is possible almost equally well with or without intonation, whereas identification is extremely difficult on the basis of just prosody. In the English experiment they introduced a fourth condition, namely *prosodic special*. In this version, passages with marked regional prosodic characteristics were presented to the listener. Not surprisingly, this last condition showed a small positive effect on the accuracy of dialect identification. The overall accuracy rates in the English experiment were also higher than the Dutch ones. Most English varieties were identifiable significantly above chance on the basis of just prosodic information, although the results indicated that verbal features of speech contained more cues to the geographic origin of a passage than prosodic features.

The prosodic contrast between varieties of Portuguese was investigated by Frota and Vigário (2001). Following the measures put forward in Ramus et al. (1999), the authors captured the rhythmic distinction between European Portuguese (EP) and Brazilian Portuguese (BP), showing that EP is clearly more stress-timed and BP is more syllable-timed. In a follow-up experiment Frota, Vigário and Martins (2002) investigated the perception of the prosodic difference between the two Portuguese varieties. To test whether EP and BP can be discriminated on the basis of prosodic information, the authors low-pass filtered the source sentences at 400Hz and then created two versions of each filtered sentence. In the first condition the pitch contour was left unaltered, while in the second condition F0 contour was made flat, taking the mean F0 value of the original contour. The results indicated that listeners could discriminate EP from BP based on low-pass filtered sentences where intonation is preserved. However, when intonation is eliminated from the low-pass filtered signal, the two varieties are no longer reliably differentiated. Thus Frota et al. (2002) conclude that intonation is a necessary cue in the identification of European and Brazilian Portuguese.

Leyden (2004) used a similar technique in her research on the relative importance of intonation and segmental structure to the mutual identifiability of Orkney and Shetland dialects of Scottish English. She created two conditions, one with normal, intelligible speech, and one where the passages were low-pass filtered at 300Hz. Within both of these conditions, two different intonation conditions were created, one with original speech melody and one with monotonised speech. In the perception experiment the manipulated speech segments were organised into four blocks in the following way:

1) LP-filtered, monotonised
2) LP-filtered, original intonation
3) original speech, monotonised
4) original speech, original intonation

Her results showed that low-pass filtered then monotonised speech contained no audible information that allows listeners to differentiate between Orkney and Shetland dialects. However, when the original intonation contour is preserved in the low-pass filtered signal, the two varieties were distinguished rather well. The differentiation between the two dialects was nearly maximal in the original but monotonous speech condition, which led Leyden to conclude that the two dialects differ distinctly with respect to their segmental information. She argues, however, that although the elimination technique proved suitable for demonstrating the role of intonation in distinguishing the two dialects, there is no direct comparison between segmental and prosodic cues. Therefore she designed a second experiment, which used speech segments with segmental and prosodic information in conflict, implementing the following transplantations: Orkney intonation contours were grafted onto the Shetland utterances, and Shetland contours superimposed on the Orkney utterances. This methodology aims to determine the relative contribution of segmental information and intonation contour. Just as for the first experiment, listeners were asked to give judgement scores on a ten-point scale based on the perceived regional origin of the speaker. The results from this second experiment suggested that the contribution of segments and intonation to the acceptability of the speech sample are roughly equal. Leyden notes that for Shetland listeners, segmental deviations contribute more to non-nativeness, while intonation was the stronger cue for Orcadians.

The sophisticated methodology of prosody transplantation was also applied to French and Maghrebian accented French[1] by de Mareüil, Brahimi and Gendrot (2004) in an attempt to examine the relative importance of prosody in the identification of a foreign accent. The authors ran two perception experiments: the first one concerned French listeners with an accent degree estimation task, the second one involved Algerian listeners with an origin identification task. The results revealed that a Maghrebian voice with a French prosody is judged as having a stronger accent degree than a French voice with a Maghrebian prosody. The results obtained from the Algerian condition confirm that the articulation of phonemes overrides prosody to account for Maghrebian accents in French.

Many of the above mentioned studies used low-pass filtering around 400Hz

[1] French spoken by Arabic and Berber natives

as a technique for deleting segmental information and preserving prosodic information. However, this methodology has been criticised by Ramus and Mehler (1999), who argue that filtering does not allow one to know which properties of the speech signal are eliminated and which ones are retained. They claim, for example, that listening to filtered speech makes it obvious that some segmental information is preserved and pitch does sometimes rise above 400Hz, especially for female voices. They also note that low-pass filtering does not make any distinction between rhythm and intonation, and much information would be gained by separating these two components of the speech signal. The authors go on to propose a new experimental paradigm using speech resynthesis to better assess the relative importance of the different elements of prosody. Their stimuli consisted of 20 sentences in Japanese and 20 sentences in English, all read by female native speakers. Four different kinds of transformation were applied to each sentence:
1) *saltanaj* condition
2) *sasasa* condition
3) *aaaa* condition
4) flat *sasasa* condition

In the *saltanaj* condition, all fricatives were replaced with the phoneme /s/, all stop consonants with /t/, all liquids with /l/, all nasals with /n/, all glides with /j/ and all vowels with /a/. Thus, all non-prosodic lexical and syntactic information was lost but global intonation, rhythm and broad phonotactics were preserved.

The *sasasa* condition consisted of replacing all consonants with /s/ and all vowels with /a/. Thus, in this condition only syllabic rhythm and intonation were preserved.

In the *aaaa* condition all phonemes were replaced by /a/, ensuring that the synthesized sentences sounded like one long /a/, varying continuously in pitch. Thus only the intonation of the original sentences was retained.

The final condition, *flat sasasa*, was identical to the *sasasa* condition, with the addition of monotonisation by setting the F0 values at a constant 230Hz. Thus, in this condition only syllabic rhythm was preserved.

Participants were all French native speakers and were told that they would be trained to recognise acoustically modified sentences from two languages, Sahatu and Moltec, which they were told are two real and exotic languages. What listeners were really doing, of course, was attempting to discriminate English from Japanese on the basis of various prosodic cues. The results reveal that the two languages were discriminable in all but the *aaaa* condition. The relatively high performance of listeners in the *flat sasasa* condition makes Ramus and Mehler (1999) conclude that syllabic rhythm is enough to allow for discrimination of English, a stress-timed language and Japanese, a mora-timed language.

One aim of the present thesis will be to investigate whether two rhythmically different varieties of New Zealand English could be identified on the basis of syllabic rhythm only. In a pilot study, Szakay (2006*a*) applied low-pass filtering at 400Hz to both Māori and Pākehā English in order to investigate whether listeners were able to differentiate the two main ethnic varieties from prosodic information only. This small-scale study used four male speakers in the experiment; two of them spoke Māori English and two spoke Pākehā English. These speakers were relatively polarised in terms of their accents, which explains the very high 94% mean accuracy rate in the unaltered speech condition. In the low-pass filtered condition listeners averaged an accuracy rate of 74%, indicating that New Zealanders can identify the ethnicity of Māori and Pākehā speakers based on suprasegmental cues only, with relatively high accuracy rates. The 20 participants in the study were each assigned a Māori Integration Index, similar to the ones used in Robertson (1994). The results were consistent with her study, indicating that highly integrated listeners perform significantly better in a dialect identification task, even when the signal is low-pass filtered. As low-pass filtering retains both rhythm and intonation in the speech signal, it will be the aim of the present study to investigate the relative importance of these two prosodic cues in ethnic identification in New Zealand.

Part II

Production Experiment

Chapter III

Production Experiment

The topic of the present chapter is the production experiment that was carried out as a precursor to the dialect identification experiment, which will be described in Part III of the thesis. A detailed acoustic analysis of the suprasegmental features of Māori English and Pākehā English was needed for two reasons. On the one hand, we were interested in investigating what the actual prosodic differences are between the two dialects, especially with regards to rhythm and pitch. On the other hand, the analysis was also needed so that we could later establish the perceptual relevance of any of these prosodic differences in a dialect identification experiment and to see if listeners are aware of the variation between Māori English and Pākehā English.

Chapter 3 has three main sections. First the methodology used in the production experiment will be detailed. This contains information about the speakers and the recordings, and explains how the acoustic analysis was carried out. The next section introduces the results of the analysis, indicating how and to what extent the two dialects differ in their prosodic features. The last section presents a discussion of these results.[1]

3.1 Methodology

3.1.1 Speakers and Recordings

In order to investigate prosodic differences in Pākehā English and Māori English and gather recordings for the stimulus tape in the perception experiment, 36 New Zealanders were recorded. The recordings were carried out in

[1] A preliminary version of this chapter has appeared as Szakay (2006*b*).

a quiet room in the participants' own home using a *Samba AV Digital Player and Recorder*, which produces files in *.wav* format. Unfortunately, this particular digital recorder does not capture any information in the speech signal above 4000Hz. Although this eliminates some linguistic information, the elements affected are primarily consonants such as sibilants, which are not the focus of the present study. Prosodic information is well represented below 4000Hz.

Of the 36 speakers, 24 identified themselves as Māori and 12 as Pākehā. All ranged between 18 and 65 years of age and were friends of the researcher. The Māori participants were recorded in the North Island so that their recordings could be used in the follow-up perception experiment in the South Island. This significantly decreases the likelihood that the voices would be recognised by participants in the perception experiment. this precaution was felt necessary given the relatively small size of the Māori community in the South Island.

Speakers were recorded reading a passage as well as telling a narrative. This resulted in 72 passages available for analysis and enabled comparison of potential prosodic differences according to style. The reading passage consisted of 6 sentences taken from the book titled *The Little Prince* (de Saint-Exupery 1943) (see Appendix C for the reading passage). Participants had a chance to study the passage before being recorded. To elicit more informal, spontaneous speech style passages, the speakers were recorded telling a narrative. They were asked to talk about rugby or other sports of their choice, as it is believed that most New Zealanders are keen on these topics and could easily and enthusiastically talk about them.

Previous research on Māori English has indicated that not all people of Māori descent speak with Māori accented English (King 1993). Similarly, not only ethnically Māori speak the dialect. It is also used by some Pākehā who are highly integrated into the Māori community. Impressionistically, not all ethnically Māori participants used in this research sounded equally 'Māori'. Nonetheless, they were all included and analysed in the production part of this study. It is the aim of the perception study presented in Part III to disentangle the various effects of different linguistic cues in order to establish what exactly it means to 'sound Māori'.

Table 3.1 shows the distribution of the 36 speaker sample, according to ethnicity, gender and age. The age range for the *younger* groups was 18–39, while for the *older* groups it was 40–60+.

Table 3.1: Distribution of the 36 speakers used in the production experiment according to ethnicity, gender and age.

	younger male	older male	younger female	older female
Māori	7	5	7	5
Pākehā	4	2	3	3

3.1.2 Māori Integration Index

To examine the extent to which each speaker has been exposed to Māori English, a Māori Integration Index (MII) score was calculated, similar to the social network strength scores used in Robertson (1994). This was based on background information gathered after each recording. See Appendix G for the complete background information sheet.

The primary aim of designing the MII was to use it in the follow-up perception experiment to investigate participants' social networks and level of integration into Māori society within the New Zealand context. Our hypothesis was that listeners who are more involved in Māori social practices would fare better in the dialect identification task, as they should have had more exposure to the dialect. We expected that the MII would also serve as a better indicator of accuracy in the task than participants' actual ethnicity alone. That is, some highly integrated Pākehā participants might be more accurate in identifying Māori English than some Māori listeners with a low integration index. A discussion of MII scores by listeners in the perception experiment will be presented in Section 4.5.1.

Although the MII was initially designed for the dialect identification experiment, it was decided to assign scores to the actual speakers as well. This way we could analyse whether there is a correlation between one's language behaviour - in this case the use of suprasegmentals - and one's social network, over and above one's ethnicity.

Points could be scored for own ethnicity, spouse's ethnicity, competence in the Māori language and general involvement in things Māori. Questions also related to how often the speaker watches Māori Television or listens to local iwi[1] radio stations or visits a marae[2]. Subjects were also asked about the ethnicity of the people they spend most of their time with, either at work or in their spare time. They also had a chance to state to what extent they perceive themselves to have been exposed to Māori English.

Two slightly different scales for Māori and Pākehā subjects were used, with a Pākehā participant being able to score slightly more points for the same answer than a Māori participant. In New Zealand culture, which is a predominantly Pākehā English environment, for a Pākehā participant it takes more of a conscious effort to be involved with the Māori community than for a Māori subject, who is often intrinsically involved. Based on this, Pākehā participants could potentially score an extra half point for each question compared to Māori participants. This excludes the question on the participant's own ethnicity where Pākehā subjects score zero. This way the maximum possible score was 16 for Māori and 17.5 for Pākehā subjects. Actual scores ranged between 6–13.5 for Māori and between 0–3.5 for Pākehā speakers. The histograms in Figure 3.1 illustrate the distribution of the Māori Integration Indices amongst Māori participants (left panel) and Pākehā participants (right panel).

Below is a detailed description of how the scores were assigned for each relevant question.

Question 3. Your ethnicity is:

	Māori	Pākehā	Other
by Māori	2	N/A	0
by Pākehā	N/A	0	0

[1] iwi = tribe

[2] marae = Māori meeting area, a quintessential gathering place for the Māori community

Question 4. If you have a partner, their ethnicity is:

	Māori	Pākehā	Other
by Māori	1.5	0	0
by Pākehā	2	0	0

Question 6. How well do you speak Te Reo Māori?

	0 (none)	1 (basic)	2	3	4	5 (fluent)
by Māori	0	0.5	1	1.5	2	2.5
by Pākehā	0	1	1.5	2	2.5	3

Question 7. How often do you listen to Māori radio stations? (eg. Tahu FM)?

	never	sometimes	often
by Māori	0	1	2
by Pākehā	0	1.5	2.5

Question 8. How often do you watch The Māori Television or other Māori TV programmes?

	never	sometimes	often
by Māori	0	1	2
by Pākehā	0	1.5	2.5

Question 9. Do you ever visit a marae?

	never	sometimes	often
by Māori	0	1	2
by Pākehā	0	1.5	2.5

Question 10. People you spend most of your time with (friends, colleagues etc...) are:

	Māori	Pākehā	Pasifika	Other
by Māori	2	0	0	0
by Pākehā	2.5	0	0	0

Question 11. In general, to what extent do you perceive yourself to have been exposed to Māori English?

	never	seldom	sometimes	often
by Māori	0	0	1	2
by Pākehā	0	0	1.5	2.5

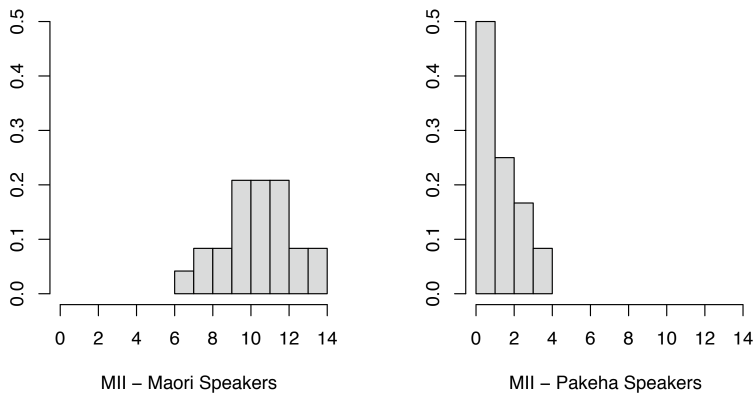


Figure 3.1: Histograms indicating the distribution of Māori Integration Indices for Māori speakers (left panel) and Pākehā speakers (right panel).

3.1.3 Phonetic Segmentation

All acoustic analysis and phonetic segmentation of the recordings was carried out using the Praat acoustic analysis software (Boersma and Weenink 2006). The first 3 sentences of each reading passage were used in the analysis, as well as 10–15 seconds of each narrative (depending on where the intonation phrase ended). Phonetic segmentation was done manually, using textgrids in Praat. Each passage was manually segmented into vowel, consonant and pause sequences, necessary for the analysis of syllabic rhythm. Both auditory and acoustic cues were used for the segmentation. Following Grabe and Low (2002), diphthongs as well as adjacent vowels were treated as one vocalic segment, whereas initial glides were marked as consonants.

3.1.4 Analysis of Rhythm

Altogether, 3281 vocalic segments were analysed and measured. Phrase final segments were included, while pauses and hesitations were excluded from the analysis. The normalised vocalic PVI values were calculated based on the difference in duration between each pair of vowels in successive syllables, taking the absolute value of the difference and dividing it by the mean duration of the pair. The output is multiplied by 100 as the normalisation produces fractional values. The equation is given in Equation (2.1). Consonantal PVI values were also computed for each passage.[1]

3.1.5 Analysis of Pitch

Pitch values were also calculated in Praat, with some of Praat's values hand-corrected to remove pitch tracker errors. Suspiciously low F0 values, typically a result of creaky voice, were excluded from the analysis. These were typically around 40Hz. Some very high pitched background noises and bleeps were also coded for and were not taken into consideration when computing pitch values. We calculated mean, minimum and maximum pitch, standard deviation of pitch as well as pitch range. Pitch range was simply computed as the difference between maximum and minimum pitch values over a passage. The obtained Hertz measurements were also converted to ERB (Equivalent Rectangular Bandwidth), which is arguably the most appropriate scale for the study of intonation (Daly and Warren 2001). The following formula was used to convert Hertz values to ERB:

- ERB = 16.7 log $(1 + f \,/\, 165.4)$, where f equals frequency in Hertz.

3.1.6 Analysis of High Rising Terminals

HRTs were only analysed in the spontaneous speech segments, as speakers did not produce a great number of HRTs in the reading passages. HRTs were manually labelled on a separate textgrid tier in Praat, from the valley to the peak of the phrase final pitch rise. A sample spectrogram is given in Figure 3.2 indicating how HRTs were labelled.

Both actual and potential occurrences were coded for, allowing for the computation of the percentage of HRTs used by each speaker. Of the 36 narrative passages, 16 did not contain any HRTs. For the speakers that did produce HRTs, we also calculated the mean duration of HRTs, and the mean pitch range of the HRTs. By simply dividing the pitch range by the duration of the HRT, an HRT ratio was calculated for each token, and the mean value

[1] I would like to thank Esther Grabe for making her PVI spreadsheet available online.

calculated for each speaker. The *absolute mean slope* in Hz / sec within each HRT was also measured by extracting the pitch contour for each HRT and using Praat's absolute mean slope function.

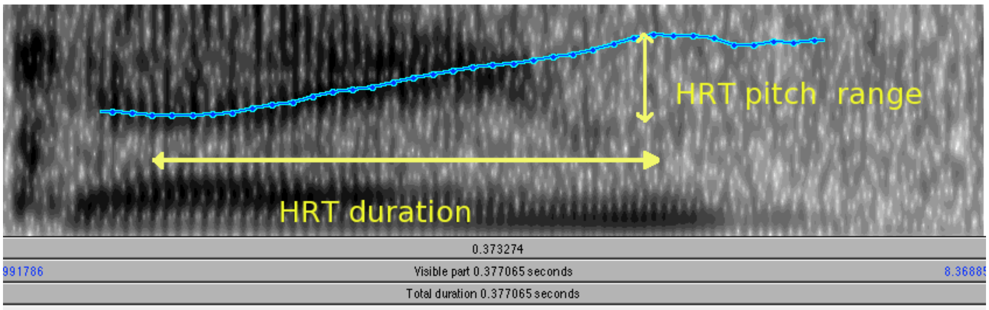


Figure 3.2: Sample spectrogram showing HRT duration and pitch range

3.2 Production Results

3.2.1 Rhythm

Rhythm and Speaker Ethnicity

The mean vocalic PVI values, calculated from both the reading passages and the narratives, show that the rhythm of Pākehā English is similar to that of British English, whereas Māori English is even more syllable-timed than Singapore English.[1] The mean PVI for Pākehā speakers was 58.7, and the mean PVI for Māori speakers was a low 47.3. There is a degree of variability within Pākehā and Māori English themselves, with PVI values ranging from 38.3 to 72.1 across the two varieties. However, the PVI for Māori English never exceeds 60 while the PVI for Pākehā English never drops below 50. A boxplot diagram illustrates these results in Figure 3.3.

[1] See Table 2.1 for PVI values for selected languages as reported by Grabe and Low (2002).

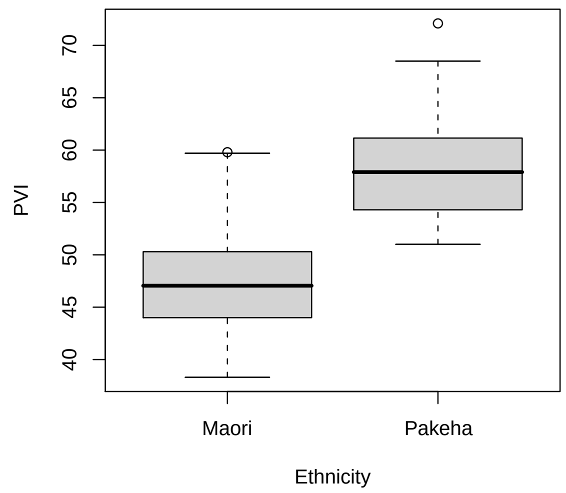


Figure 3.3: Rhythm and Ethnicity (Wilcoxon-test p<.0001)

There was no significant difference in consonantal PVI between the two varieties. The mean intervocalic PVI was 63.6 for Māori English and 64.7 for Pākehā English. Grabe and Low (2002) report an rPVI of 64.1 for British English and 68.2 for Singapore English. Thus, the values for the two New Zealand English varieties are very similar to the British values.

In order to investigate the effect of rhythm, a linear regression model was fit by hand.[1] The regression model showed significant effects of ethnicity (p<.0001) and age (p<.01). In other words, speaker rhythm significantly differs according to both the speaker's ethnicity and their age. The model predicts that Māori speakers are producing a lower PVI than Pākehā speakers, that is, Māori are significantly more syllable-timed. The coefficients for the linear model are shown in Table 3.2 and the corresponding anova table is given in Table 3.3.

[1] All statistical analyses used in this study were carried out using the 'R' statistical package (R-Development-Core-Team 2004).

Table 3.2: Coefficient table for model of PVI

Variable	Coef	S.E.	t	P
Intercept	43.25	1.57	27.39	0.000e+00
ethnicity=Pākehā	11.47	1.29	8.85	5.420e-13
age	0.12	0.04	2.90	5.002e-03

Table 3.3: Anova table for model of PVI

Factor	d.f.	Partial SS	MS	F	P
ethnicity	1	2106.0	2106.0	78.46	<.0001
age	1	225.7	225.7	8.41	0.005
REGRESSION	2	2313.4	1156.7	43.09	<.0001
ERROR	69	1852.1	26.8		

Rhythm and Speaker Age

As mentioned above, the linear regression model indicated that speaker rhythm is affected not only by ethnicity but also by speaker age. Older speakers of both Māori and Pākehā English produced significantly higher PVIs than younger speakers, which seems to indicate that New Zealand English as a whole is becoming more syllable timed. This trend is illustrated in Figure 3.4, which plots the model predictions for different age groups, with the coefficients set to the level for Māori speakers in the left panel and for Pākehā speakers in the right panel. Note, that the background information sheet (see Appendix G) asked speakers to circle an age group they belong to, rather than asking for their actual age. These age groups were used in the statistical analysis of the data, with 18 and 19 year old speakers being grouped under the label '10'.

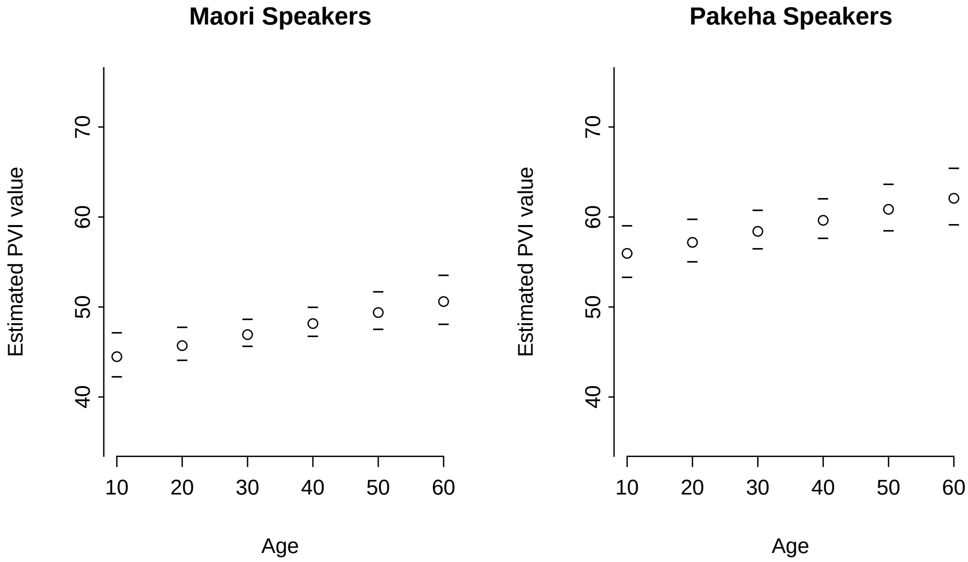


Figure 3.4: Model predictions of PVI for different age groups, with the coefficients set to Māori speakers (left panel) and Pākehā speakers (right panel)

Rhythm and Māori Integration Index

The MII was not included in the linear model due to its colinearity with ethnicity. However, there is a significant correlation between PVI values and MII (Spearman's rho = .65, p<.0001). Speakers who are more integrated into Māori society, that is have a higher MII, produced more syllable-timed speech than speakers with no or low integration. Figure 3.5 demonstrates this strong correlation.

It is difficult to separate the effects of MII and ethnicity. None of the Pākehā speakers in this study scored higher on the MII than any Māori participant, as discussed in Section 3.1.2. MII scores ranged between 6–13.5 for Māori and 0–3.5 for Pākehā. It would be interesting to see PVI values for highly integrated Pākehā to see whether the MII is doing more than just distinguishing between the two ethnic groups on the basis of rhythmic production.[1] Our initial

[1] The results from the perception experiment in Chapter 5 will demonstrate that MII is clearly doing more than just a binary distinction between Māori and Pākehā participants.

hypothesis is that Pākehā with a higher MII score would be more syllable-timed than Pākehā with low MII scores. This remains a topic for further research.

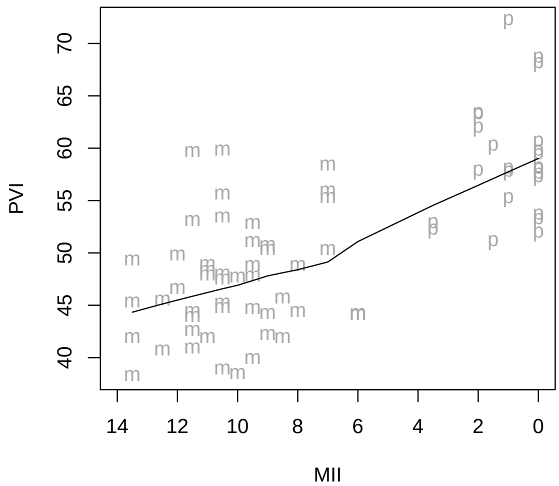


Figure 3.5: Rhythm and Māori Integration Index. m = Māori participant, p = Pākehā participant. The line represents a non-parametric scatterplot smoother fit through the data (Cleveland 1979).

Rhythm and Passage Type

Although Grabe and Low (2002) only report PVI values for read passages, the current method enabled the examination of the effect of style on rhythm. The mean PVI values for the two different passage types are reported in Table 3.4.

Table 3.4: Mean PVI values according to passage type and ethnicity

	reading passage	spontaneous speech
Māori English	46.4	48.1
Pākehā English	57.2	60.2

For both Māori and Pākehā participants there was a tendency for the reading passages to be more syllable-timed than the narratives, however this difference only reached the level of significance in case of the Pākehā participants (Wilcoxon paired test, p<.05). This is shown in Figure 3.6. Note however, that passage type did not show significance in the linear regression model reported above.

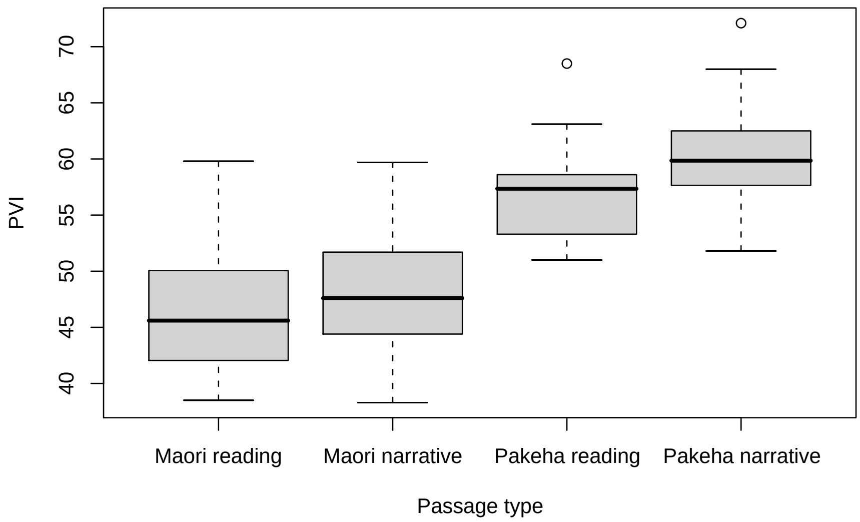


Figure 3.6: Rhythm and Passage Type and Ethnicity

Rhythm and Speaker Gender

Gender did not have a significant effect on rhythm. While Māori females tended to be somewhat less syllable-timed than Māori males, this trend is far from significant. There was virtually no rhythmic difference between Pākehā males and females. The mean PVI values by gender and ethnicity are shown in Table 3.5 below.

Table 3.5: Mean PVI values according to gender and ethnicity

	female speakers	male speakers
Māori English	47.8	46.7
Pākehā English	58.7	58.7

Rhythm and Speech Rate

A significant correlation was found between the PVI values of Pākehā speakers and speech rate (Spearman's rho = -.45, p = 0.026). Shown in Figure 3.7, as the speech rate of Pākehā speakers increases, their PVI values decrease, that is, the faster they speak, the more syllable-timed they become. However, speech rate does not affect the rhythm of Māori English speakers. This will be discussed in more detail in Section 3.3.1.

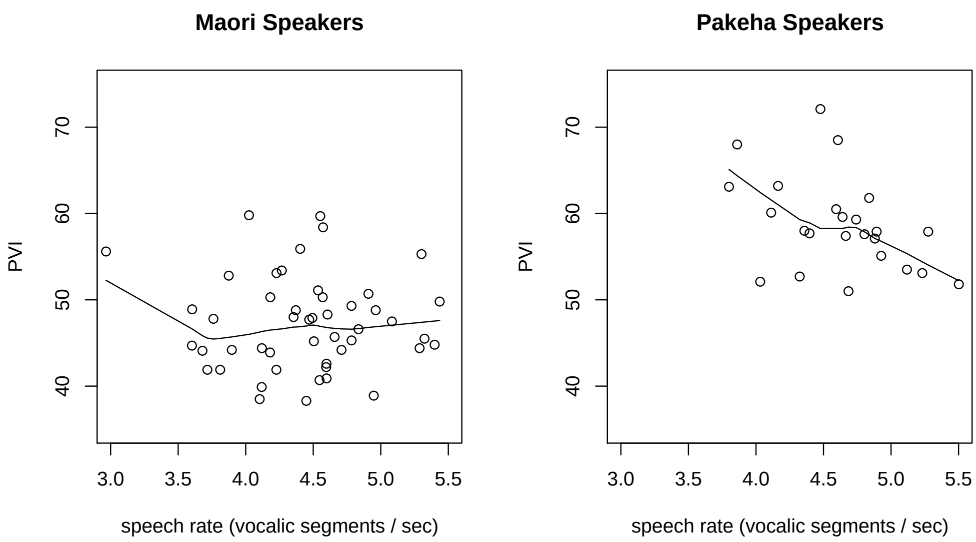


Figure 3.7: The correlation between rhythm and speech rate. Pākehā speakers: p <.05, rho = -.45. Māori speakers: not significant.

3.2.2 Pitch

In another linear regression model we investigated possible differences in mean pitch. The results of this model are shown below, with the coefficients in Table 3.6 and the anova table in Table 3.7.

Predictably, gender had a highly significant effect on mean pitch (p<.0001), with men producing significantly lower mean pitch values than women. However, over and above this gender effect there was also a significant effect of ethnicity (p<.01). Māori participants had a significantly higher mean pitch than Pākehā participants. This is most obvious for Māori male participants, whose mean pitch was 128.4Hz as opposed to their Pākehā counterparts with a mean pitch of 109.5Hz. A boxplot diagram in Figure 3.8 displays the male mean pitch values. Female mean pitch values were 196.7Hz for Māori and 189.5Hz for Pākehā speakers.

Table 3.6: Coefficient table for model of mean pitch

Variable	Coef	S.E.	t	P
Intercept	217.13	7.67	28.30	0.00000
gender=male	-72.87	4.82	-15.11	0.00000
ethnicity=Pākehā	-36.93	12.87	-2.86	0.00552
age	-5.50	1.97	-2.78	0.00696
ethnicity=Pākehā * age	7.28	3.62	2.00	0.04887

Table 3.7: Anova table for model of mean pitch

Factor	d.f.	Partial SS	MS	F	P
gender	1	94057.1	94057.1	228.54	<.0001
ethnicity	2	4436.2	2218.1	5.39	0.0068
age	2	3340.3	1670.1	4.06	0.0217
ethnicity * age	1	1656.5	1656.5	4.03	0.0489
REGRESSION	4	100022.8	25005.7	60.76	<.0001
ERROR	67	27574.1	411.5		

The linear regression model also indicated that age interacting with ethnicity had a significant effect on mean pitch ($p<.05$). This interaction is shown in Figure 3.9, which illustrates that, assuming apparent time, Māori mean pitch is getting higher over time whereas Pākehā mean pitch, if anything, is getting lower.

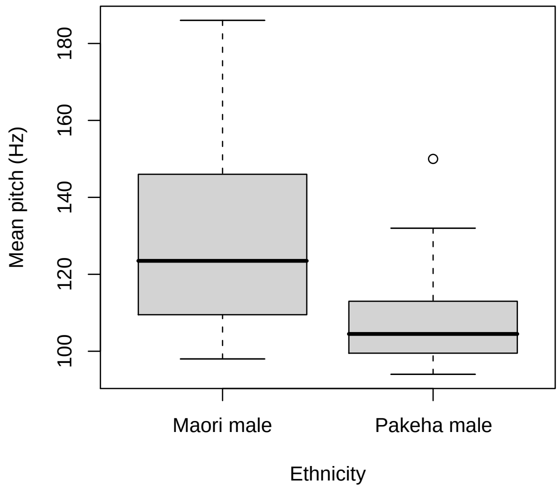


Figure 3.8: Difference in mean pitch between Māori and Pākehā males (Wilcoxon-test p<.001)

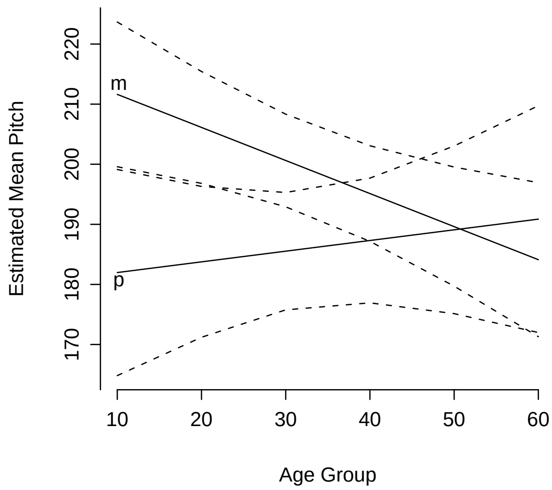


Figure 3.9: Model effects showing estimated mean pitch by age group and ethnicity. Dashed lines indicate 95% confidence intervals.

Although there were highly significant differences in pitch range and the standard deviation of pitch between the sexes (confirming the results of Daly and Warren (2001)), there was no such difference between Māori English and Pākehā English. Conversion of Hertz measurements to ERB did not result in significant differences either. Table 3.8 displays the mean values for various pitch measurements. None of these show significant differences between the two ethnic dialects other than the mean pitch values mentioned above. It is tempting to assume that the reason behind a significantly higher Māori mean pitch might be that Māori speakers have been shown to use a higher percentage of HRTs. However, as the following section will show, in this study Pākehā males actually produced a higher percentage of HRTs than their Māori counterparts. Moreover, if HRTs were affecting pitch in this way, we would also expect a significant effect of SD of pitch, which we don't get. Thus, the higher mean pitch values are clearly not a consequence of a higher percentage of HRTs.

Table 3.8: Mean values for pitch measurements (Hz) according to ethnicity and gender.

	Māori		Pākehā	
	female	male	female	male
mean pitch	196.8	128.4	189.6	109.5
minimum pitch	101.9	82.7	94.6	74.9
maximum pitch	307.4	199.7	308.8	172.1
pitch range	205.5	117.0	214.2	97.2
SD of pitch	33.8	16.9	33.0	14.2

3.2.3 High Rising Terminals

The results of the analysis relating to HRTs reveal that on average Māori speakers do use a higher percentage of HRTs. The effects on the percentage of HRTs used by the speaker were investigated in another linear regression model. It showed a significant interaction between speaker ethnicity and speaker gender. The coefficients for the linear model are shown in Table 3.9 and the corresponding anova table is given in Table 3.10.

The model predictions, illustrated in Figure 3.10, reveal that Māori females use the highest percentage of HRTs of all our groups. However, the Pākehā male speakers recorded in this study are also frequent users of HRTs. In fact, they produced a higher percentage of HRTs than did the Māori males as well as the Pākehā females.[1] It is possible that the topic of the narratives had an influence on the percentage of HRTs used by the speaker. Pākehā females might not be as excited about rugby as Pākehā men, which could explain why Pākehā females were not producing more HRTs than the men, as previous research had indicated they would (Britain 1992). At the same time, it is also possible that HRT use by Pākehā men is undergoing change, and they might be using HRTs with a higher frequency in general than they were before.

Table 3.9: Coefficient table for model of percentage of HRTs.

Variable	Coef	S.E.	t	P
Intercept	46.84	7.85	5.96	1.208e-06
ethnicity=Pākehā	39.89	13.60	-2.93	6.170e-03
gender=male	-28.29	11.10	-2.54	1.586e-02
ethnicity=p * gender=m	55.32	19.23	2.87	7.123e-03

Table 3.10: Anova table for model of percentage of HRTs.

Factor	d.f.	Partial SS	MS	F	P
ethnicity	2	7317.0	3658.5	4.94	0.0135
gender	2	6994.1	3497.0	4.72	0.0159
ethnicity * gender	1	6120.0	6120.0	8.27	0.0071
REGRESSION	3	8191.1	2730.3	3.69	0.0218
ERROR	32	23689.2	740.2		

[1] This high percentage of Pākehā male HRTs will have a significant effect on accuracy levels in the perception experiment.

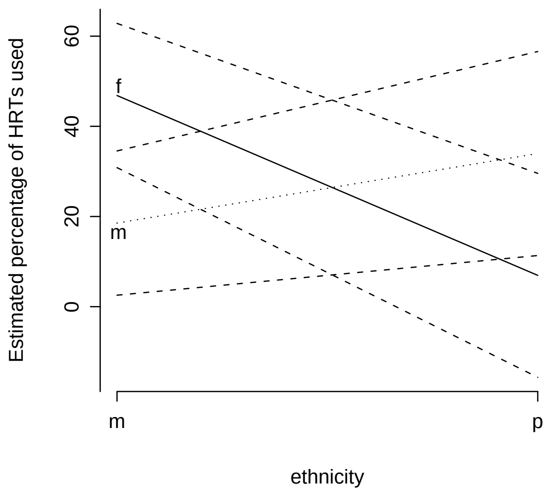


Figure 3.10: Model effects showing estimated percentage of HRTs used by gender and ethnicity

Not all of the 36 narrative passages analysed in this study contained HRTs. The number of speakers using HRTs is shown in Table 3.11.

Table 3.11: Number of speakers producing HRTs (narratives only).

	total N of speakers	producing HRTs
Māori female	12	10 (83%)
Māori male	12	5 (42%)
Pākehā female	6	2 (33%)
Pākehā male	6	3 (50%)

While speakers with 0% HRTs were included in the above mentioned statistical model, they were naturally ignored for the analysis of actual HRT properties. We investigated the effects on the mean duration, pitch range, ratio and absolute mean slope of HRTs for the 20 speakers who did produce some HRTs. Among these, HRT ratio and pitch range showed a significant difference between male and female speakers. However, none of them showed a significant effect of ethnicity.

3.2.4 Speech Rate and Pauses

We also investigated possible differences in speech rate, the number of pauses and the mean pause duration between the two dialects. None of these showed significant results. Speech rate was calculated as number of vocalic segments per second. Māori participants showed a slightly lower speech rate, with a mean of 4.44 V/sec, as opposed to the mean of 4.62 V/sec for Pākehā speakers. Māori participants also paused slightly more often, with the mean length of these pauses being longer than for Pākehā participants. The number of pauses and pause lengths were calculated based on actual pauses between intonation phrases, excluding hesitations such as *hmm* or *er*. All of these trends are far from significant. The mean values are given on Table 3.12.

Table 3.12: Mean values for speech rate, number of pauses and mean pause duration.

	Māori speakers	Pākehā speakers
Speech rate	4.44 V/sec	4.62 V/sec
Number of pauses	3.15	2.75
Mean pause duration	0.22 sec	0.21 sec

As mentioned above in Section 3.2.1, there was a significant correlation between speech rate and rhythm in the case of Pākehā speech, where the rhythm of Pākehā speakers becomes more syllable-timed as the speech rate increases. Another correlation was found between the number of pauses and the mean pause duration of the speaker in both dialects (Spearman's rho = .25, p = 0.029). The results indicate that speakers who pause more often, also have longer mean pause lengths. This positive correlation is illustrated in Figure 3.11.

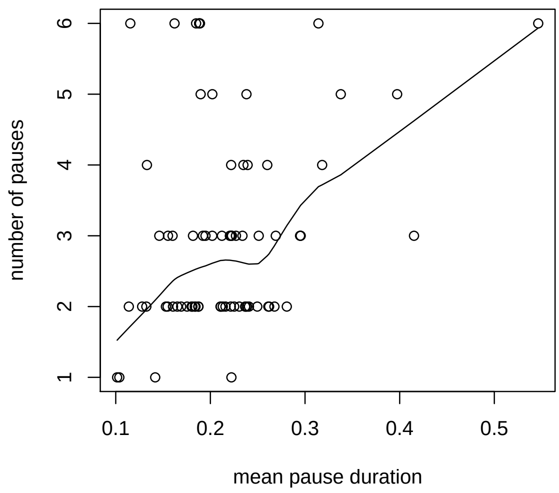


Figure 3.11: Correlation between number of pauses and mean pause duration (Spearman's rho = .25, p = 0.029).

3.3 Discussion of Production Results

3.3.1 Rhythm

The results of the production experiment clearly show that Māori English is a significantly more syllable-timed dialect than Pākehā English. They also demonstrate that a within-variety difference in rhythmic properties can exist, as it does within New Zealand English. Table 3.13 repeats the PVI values shown in Table 2.1, including the values for Māori English and Pākehā English from this study to demonstrate where these two dialects of New Zealand English fit along the continuum from syllable-timing to stress-timing.

Grabe and Low (2002) report only one value for British English. It would be interesting to see what degree of variability is present among various dialects of British English regarding rhythm. Note that if we were to assign a single PVI value for New Zealand English by calculating the mean value across Māori and Pākehā English, we would get a PVI of 53.0, still lower than that of Singapore English. However, while Pākehā speech may be moving in the direction of syllable-timing, it is still the case that only *young* Pākehā New

Zealanders are more syllable-timed than the British speakers analysed by Grabe and Low (2002).[1]

Table 3.13: PVI values for selected languages. Data taken from Grabe & Low (2002) with PVI included for Māori English and Pākehā English. Based on reading passages. Values in brackets indicate mean PVI for narratives.

Language	PVI
Mandarin	27.0
Spanish	29.7
Luxembourg	37.7
Japanese	40.9
French	43.5
Māori English	**46.4** (47.3)
Polish	46.6
Rumanian	46.9
Welsh	48.2
Singapore English	52.3
Malay	53.6
British English	57.2
Pākehā English	**57.2** (58.7)
German	59.7
Dutch	65.5
Thai	65.8

The results relating to rhythm and age are of great interest as they seem to describe a change in progress toward increased syllable-timing in both Māori English and Pākehā English. Such a change in timing patterns over time has not been documented in New Zealand English before. The fact that languages and language varieties can experience a rhythmic shift is not unknown. Eek and Help (1987), for example, describe Estonian as a syllable-timed language that has undergone a Great Rhythm Shift in its history from a rhythmically more complex type. Crystal (1996) also predicts that more and more English dialects that are less stress-timed are likely to emerge as the language comes into contact with various other languages that are traditionally categorised as syllable-timed. The question arises whether we

[1] Based on the reading passages, the mean PVI for young Pākehā is 55.6 as opposed to the British English mean PVI 57.2.

can tell *when* these New Zealand dialects - and specifically Māori English - started to become more syllable-timed. With the help of the ONZE corpus (Gordon, Hay and Maclagan forthcoming) it might be possible to answer this question by analysing the speech of people born early on in the last century.

It is interesting to note that the acoustic analysis carried out by Warren (1998) suggested that not only is Māori English more syllable-timed than Pākehā English, but Pākehā English in turn is more syllable-timed than British English, as mentioned in Section 2.2.2. The results from the present study confirm the latter observation only in the case of *young* Pākehā speakers. When the mean PVI is calculated across all age groups, we do not find a significant difference between the rhythmic characteristics of British English and Pākehā English. The results from the two studies are not directly comparable, however, as they used different methods. Based on newsreader samples, Warren (1998) calculated syllabic - and also foot - PVIs, whereas the current study uses vocalic PVI values based on reading passages and spontaneous speech fragments. It is likely that the discrepancy between the results of two studies relating to Pākehā English and British English is caused by a combination of various effects, such as speaker age, different genre of passages and a different type of PVI measurement.

It has been demonstrated in numerous sociolinguistic studies that speakers use a more standardised variety of their language when reading a passage than when speaking spontaneously (e.g. Labov (1972)). Based on this, it would be expected that the PVI values would be higher for reading passages than for narratives. In fact, the opposite trend emerged from the present data. Pākehā participants produced significantly more syllable-timed speech for the reading passages. One explanation might be that being stress-timed is no longer necessarily associated with the standardised variety and that syllable-timing - as the incoming form - is gaining prestige. However, it is also possible that a lower PVI is just an inherent characteristic of this particular reading passage. See Appendix C for the reading passage.

The correlation between the speaker's MII and PVI confirms the suggestion by Holmes (2005) that syllable-timing appears to be a sociolinguistically sensitive feature of Māori English, which varies in degree according to the background of the speaker and the relative importance of signalling one's Māori identity. Milroy (1987) also shows in her Belfast study that the strongest vernacular speakers were generally those whose neighbourhood network ties were the strongest, a pattern complicated by the interaction of other social variables such as age and gender. Similarly, in our case, those speakers who are highly integrated into Māori networks are the ones who exhibit the most syllable-timed rhythmic patterns.

Although the vocalic PVI is assumed to be normalised for speech rate, our results nevertheless indicated a correlation between PVI and speech rate in the case of Pākehā speakers. Faster speech caused Pākehā English to become more syllable-timed, while Māori English rhythm was not affected by speech rate. Rather than being socially and ethnically motivated, it is possible that this difference is conditioned by linguistic factors. Dellwo (2004) demonstrates that stress-timed languages tend to vary in rhythm as a function of speech rate, while the rhythm of syllable-timed languages seems to be unaffected by changes in speech rate. The fact that we find the same difference between Pākehā English and Māori English indicates that Māori English clearly patterns with syllable-timed languages even in this respect. Moreover, our results show that such a difference with regard to rhythm and speech rate can exist not only between different languages but also between different varieties of the same language and *within* different dialects of the same (national) variety.

3.3.2 Pitch

Differences between women and men have received the greatest amount of attention in the literature on pitch, as Biemans (2000) points out. Gender differences relating to pitch can obviously be linked to differences in physiology, the build of the larynx and the vocal folds. The present study showed that on top of the unsurprising gender effect, there was also a significant effect of ethnicity on mean pitch. Māori participants had a significantly higher mean pitch, which was most obvious for male speakers. This difference between the two male groups cannot be explained by anatomy. Rather, it has to be socially constructed and as such indicates ethnicity-related speaker identity.

Studies in the past have shown that pitch can be influenced by culture. Majewski, Hollien and Zalewski (1972) showed that the pitch of Polish men was lower than the pitch of American men, and Loveday (1981) described differences between Japanese and British use of pitch. Biemans (2000) draws from several studies on pitch to show average values for both males and females and gives a range between 100–125Hz for men, which indicates just how high the mean pitch of Māori males, at 128.4Hz, is.

This difference in the use of mean pitch between Māori and Pākehā speakers seems to be a recent phenomenon. As Figure 3.9 has shown, the ethnicity effect is interacting with age. Māori mean pitch is getting higher over time, whereas Pākehā pitch is lowering, resulting in a highly significant difference between young Māori and Pākehā speakers. The results of Walker (2006) also suggested that male pitch in New Zealand English might be lowering over time. Her subjects did not include Māori participants, so the conclusion from her study can be taken to confirm that Pākehā male pitch might indeed be lowering, while Māori pitch is becoming higher, as the present study indicates.

Gussenhoven (2002) claims that high pitch is socially associated with submissiveness, politeness, vulnerability and femininity, while low pitched voices suggest authority, assertiveness, masculinity. It remains to be determined whether these perceptual characteristics bear any explanatory value in the New Zealand context. What is certain is that these two dialects within New Zealand English are becoming increasingly distinct, at least with respect to pitch.

Overall, while both Māori English and Pākehā English are shifting increasingly toward the syllable-timed end of the continuum, the PVI values are significantly differentiated, thus the two ethnolects display distinct rhythmic qualities. Moreover, the two varieties are diverging in terms of mean pitch, thus further widening the gulf between the main varieties of New Zealand English. Whether naive listeners are aware of these prosodic changes will be one of the topics investigated in Part III of this thesis.

Part III

Perception Experiment

Introduction to Part III

The production experiment described in Chapter 3 served as a precursor to the dialect identification experiment, which is the topic of Part III of this thesis. The main objective of the perception task was to investigate whether listeners can correctly identify a speaker's ethnicity and to find out what cues listeners use for ethnic identification. Can they identify Māori English and Pākehā English based solely on prosodic features, such as rhythm and intonation, or do they also need segmental cues, such as the pronunciation of particular vowels and consonants? The question also arose whether listeners are aware of the dialectal differences shown by the results of the production experiment, described in Section 3.2. If they are conscious of the contrasting rhythmic properties and the different usage of pitch between the two dialects, could they accurately use them as cues in the perception task?

In order to isolate the precise features listeners use to distinguish between ME and PE, it was necessary to eliminate segmental information from the speech signal, as well as to separate rhythm and intonation from each other. Seven different conditions were created for use in the perception experiment, each of them retaining different linguistic cues available for the listeners.

Part III of the thesis is divided into four chapters. Chapter 4 describes the methodology used to set up the experiment. It includes information about the participants, as well as the passages chosen to be used in the perception task. The methodology chapter also gives a detailed explanation of how the seven conditions used on the stimulus tape were created. Chapter 5 describes the results from the dialect identification task, giving information about each condition in separate sections. The discussion of these results is presented in Chapter 6, while the conclusion and summary of the main points are found in Chapter 7.

Chapter IV

Methodology for Perception Experiment

4.1 Passages

As described in Section 3.1.1, 36 speakers participated in the production experiment. Both a reading passage and spontaneous speech were recorded for each speaker, but only the narratives would be used in the perception task. There were two main reasons behind this decision. Firstly, the narratives provided more interesting and more natural intonation patterns than the reading passages. Secondly, the experiment would be too lengthy and tiring for the participants if both passage types were included in the task.[1]

Because of the time constraints on the perception experiment, not all 36 speakers could be included in the task. As mentioned in Section 3.1.1, not all Māori English speakers sounded equally Māori and not all Pākehā speakers sounded equally Pākehā. Our aim was to choose 20 speakers for the perception experiment who were easily identifiable as Māori or Pākehā in the normal, unmodified speech condition. We argue that there would be not much point in trying to identify a speaker's ethnicity from suprasegmental cues alone, if they were unlikely to be correctly identified based on unaltered speech. In order to decide which speakers to include in the experiment, a test session was run among linguists in which they rated each speaker based on how Māori or Pākehā-sounding they found them to be. Although it was our aim to choose speakers who were identified with high accuracy during the test run, two speakers were nonetheless included who had been misidentified by most of the linguists. The inclusion of these two speakers (*m18*, a younger Māori female and *m11*, an older Māori male) was felt necessary to

[1] Note however, that one clear advantage of using the reading passages over the narratives would be that the content of the passages could be controlled for.

balance the numbers in each group. We also felt it would be interesting to see whether some participants with a high Māori Integration Index (for MII see Section 3.1.2) would fair better in identifying these two speakers than our test group had, the members of which all scored fairly low on the MII.

The distribution of the selected speakers according to ethnicity, gender and age is shown in Table 4.1. The age range for the *younger* groups was 18–30, while for the *older* groups it was 40–60+.

Table 4.1: Distribution of the 20 speakers used in the perception task according to ethnicity, gender and age.

	younger male	older male	younger female	older female
Māori	4	2	2	2
Pākehā	4	2	2	2

The reason why there were twice as many younger male speakers chosen to be included in the experiment is twofold. On the one hand, it has been previously argued that the features of Māori English are most salient for younger male speakers (King 1999). Thus, we were more confident that this group would yield positive results in a dialect identification task. On the other hand, we simply had more younger male speakers recorded than any of the other groups.

4.2 Creating the Conditions

In order to investigate exactly what cues listeners tune in to when identifying an accent, various test conditions were created. Our main aim was to be able to disentangle the various effects of different linguistic cues in order to establish what exactly it means to "sound" Māori or Pākehā. Each condition contained different linguistic cues available for the listener. Based on these we would try to determine whether rhythm or intonation – alone or combined – could serve as a sufficient cue for identifying the ethnicity of the speaker.

The first six conditions provided only suprasegmental information, either separately or in various combinations. In Conditions 1 to 5 we used speech re-synthesis, while in Condition 6 we applied low-pass filtering to eliminate segmental information. The final test condition presented the unmodified version of the passages. The following is a list of the seven test conditions that were used in the perception experiment:

(1) Duration-normalised Rhythm Only at Mean Pitch
(2) Non-duration-normalised Rhythm Only at Mean Pitch
(3) Non-duration-normalised Rhythm Only at Own Pitch
(4) Intonation Only
(5) Rhythm and Intonation Together
(6) Low-pass Filtered at 400Hz
(7) Unmodified

The subsections below provide a thorough explanation on how these different test conditions were created.

4.2.1 Condition One - Duration-normalised Rhythm Only at Mean Pitch

The purpose of the first condition was to investigate whether participants could identify the two dialects if the one and only available cue they had was rhythm. To create a rhythm only condition, all segmental information was eliminated and the intonation was flattened. All passages had been manually segmented into vowel and consonant sequences and marked up in a Praat textgrid, as described in section 3.1.3. A Praat script was written that replaced each consonant and each pause by silence, while vowels were replaced by a tone complex, also created in Praat. This technique ensured that what listeners would hear as rhythm in the perception task, would closely correspond to the vocalic PVI values assigned previously to each speaker in the production experiment. The sound replacing vowels was produced from a tone complex as a sum of a number of cosine waves with equidistant frequencies at a sampling frequency of 8000Hz. It was created at the mean pitch across all speakers according to gender. This was 118Hz for male speakers and 188Hz for female speakers. Gender information was kept in the signal so it could be investigated whether rhythm is interpreted differently in the case of male and female voices.

Note that previous research used a different method for creating a rhythm only condition. For example, Leyden (2004) first low-pass filtered the samples to obtain a rhythm and intonation condition, then the speech was monotonised by changing the pitch contour into a flat line. However, after such transformation the acoustic signal becomes extremely degraded.[1] Gooskens and van Bezooijen (2002) omitted this condition in their study for similar reasons.

Our initial aim was to create a *flat sasasa* condition by using the technique described in Ramus and Mehler (1999) (see Section 2.3.1). Following their methodology, we replaced all consonants with /s/ and all vowels with /a/. However, the resultant stimulus material did not resemble human speech. Instead, it created the impression that two different sounds were being played simultaneously: one Morse-code-like vowel in the foreground with a hissing noise in the background. The /sa/ segments were clearly not being perceived as the onset and nucleus belonging to the same syllable. Thus, the decision was made to abandon the *flat sasasa* condition and replace all consonants with silence. This way it was easier to relate the stimulus material to human speech, which was essential as we hoped that listeners would be able to associate it with either Māori English or Pākehā English.

To make sure that participants would not base their judgements on whether the speech was fast or slow but on the actual rhythm alone, it was necessary to normalise the passages for speech rate in this condition. Consequently, duration normalisation was performed, which speeds up slow speakers and slows down fast speakers. The normalisation process was carried out the following way. Mean segment duration of vowels was calculated respectively for each individual speaker as well as across all speakers. A scaling factor was defined for each speaker by dividing the mean vowel duration across all speakers by the mean vowel duration of the individual speaker. Then, the duration of each actual vowel in the passage was multiplied by the scaling factor. The same procedure was then carried out for all consonants and all pauses. The resulting normalised duration was used when replacing the original segments by silence in the case of consonants and pauses, and by the tone complex in the case of vowels. The normalisation formula is given in

[1] Indeed, low-pass filtered then monotonised speech reminds one of Darth Vader's heavy breathing.

Equation (4.1).

$$nD_{current} = \frac{meanD_{all-speakers}}{meanD_{current-speaker}} \times D_{current} \tag{4.1}$$

where D = *segment duration* and nD = *normalised segment duration*

Figure 4.1 demonstrates a sample spectrogram from the duration normalised, flat rhythm only condition. When compared with the corresponding unmodified version shown in Figure 4.2[1], it can be observed how the segments marked as consonants were replaced by silence. The straight line of the pitch tracker in Figure 4.1 illustrates that the intonation is flattened. In this sample, the pitch is kept constant at 118Hz, as for all male speakers. Upon closer inspection, it can also be seen how the duration of the vocalic segments is slightly different from the corresponding unaltered speech sample in Figure 4.2. This results from the duration normalisation procedure.

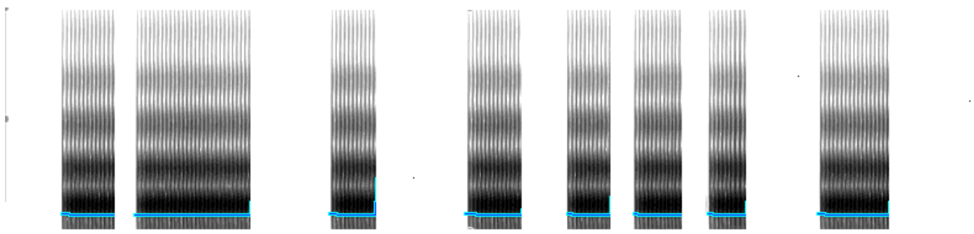


Figure 4.1: Sample spectrogram from Condition One

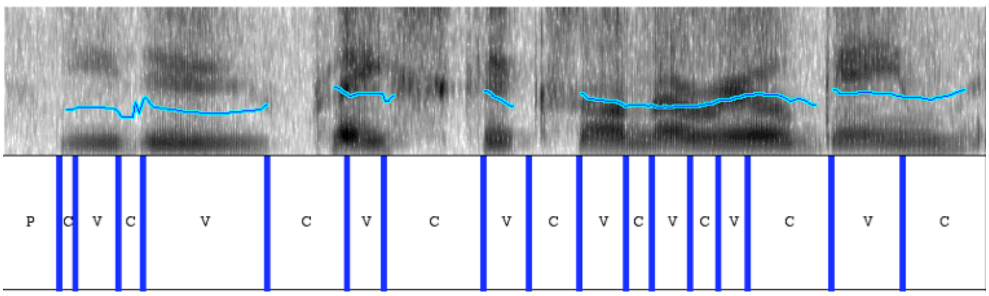


Figure 4.2: Sample spectrogram from Condition Seven (unmodified speech)

[1] All spectrograms shown in this section represent the same speech fragment from one male speaker.

4.2.2 Condition Two - Non-duration-normalised Rhythm Only at Mean Pitch

The second condition was essentially the same as the first one, except duration normalisation was not applied. As in Condition One, consonants and pauses were replaced by silence while vowels were replaced by a tone complex at the mean pitch across all speakers (118Hz for male and 188Hz for female speakers). The differentiating factor between these two conditions is that all segments retained their original duration. This enabled participants to rely on speech rate as well as rhythm when identifying a dialect.

Figure 4.3 shows a sample spectrogram from Condition Two. The pitch is again kept constant at 118Hz, but because duration normalisation was not used, the vocalic segments differ slightly in length in comparison to those in Figure 4.1, from Condition One.

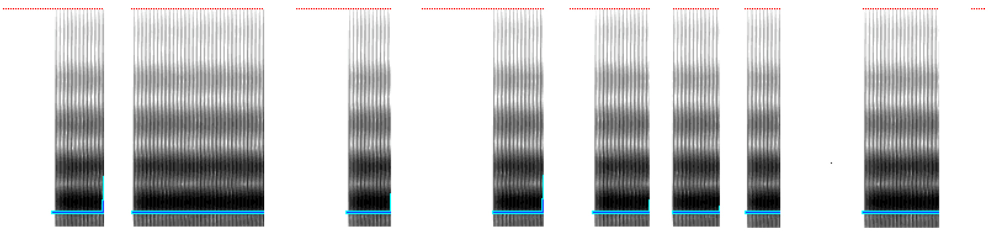


Figure 4.3: Sample spectrogram from Condition Two

4.2.3 Condition Three - Non-duration-normalised Rhythm Only at Own Pitch

The third condition was again flat rhythm only, except this time the speakers' own mean pitch was used instead of the mean pitch across all speakers. The tone complex replacing each vocalic segment was created at the mean pitch of that particular passage. This enabled participants to now use mean pitch as a cue in the dialect identification task, in addition to rhythm and speech rate.

When calculating mean pitch, some unusual tokens were excluded from the analysis. All "bad pitch" was marked up in a textgrid. These included very high pitched bleepy noises as well as suspiciously low, glottalised tokens. Mean pitch was calculated based on the remaining tokens. For a more detailed description on how pitch was analysed see Section 3.1.5.

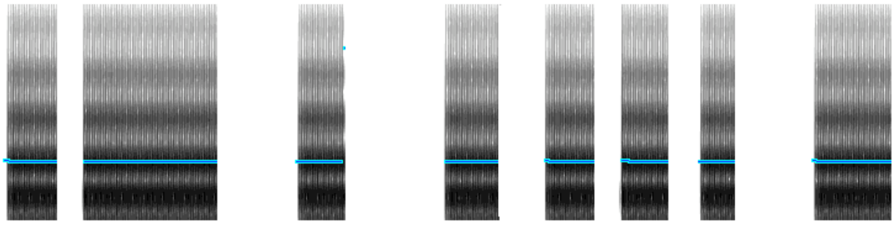


Figure 4.4: Sample spectrogram from Condition Three

The sample spectrogram for this condition is shown in Figure 4.4. It is essentially the same as the spectrogram for Condition Two, except the pitch is now higher than the average mean pitch used previously. It is held constant at 169Hz, which is the mean value for this particular speaker.

4.2.4 Condition Four - Intonation Only

The purpose of the fourth condition was to test whether participants could distinguish between Māori English and Pākehā English if they had only pitch dynamics to rely on. This intonation only condition was created by using the *Pitch Tier* in Praat. First the *Pitch Tier* was extracted from the original sound file, then it was degraded to a *Point Process*. Finally, the *Point Process* was converted into a *Sound Object* by humming it. Praat's humming procedure involves running the sound through a sequence of second-order filters that represent five formants.

In the hummed version the pitch contour is linearly interpolated between two adjacent points, even across original pauses. This has the undesirable effect that the end of an intonation phrase cannot be recognised. To avoid this, each segment originally marked as a pause in the textgrid was replaced by silence in the hummed version. This allowed HRTs to be recognised, making it possible to investigate whether participants use HRTs as a cue in this condition.

As with the previous condition, some unusual tokens were excluded when analysing pitch. High, bleepy noised pitch was removed in this condition but the glottalised low pitch tokens were retained. As these are normally the result of creaky voice, their inclusion created an intonation plus voice quality condition in a way, and the glottalised tokens could be used as a cue for ethnic identification by the participants.

Figure 4.5 illustrates the sample spectrogram for this condition. Notice how the pitch contour is interpolated between adjacent points. This has the result of all originally voiceless segments obtaining pitch movements. Compare this spectrogram with the one from the unaltered Condition Seven (Figure 4.2), where the voiceless segments are clearly indicated by breaks in the pitch movement. This sample section of the spectrogram starts off with a pause (marked as P in the textgrid for Condition Seven), then the pitch contour is interpolated across originally voiceless segments.

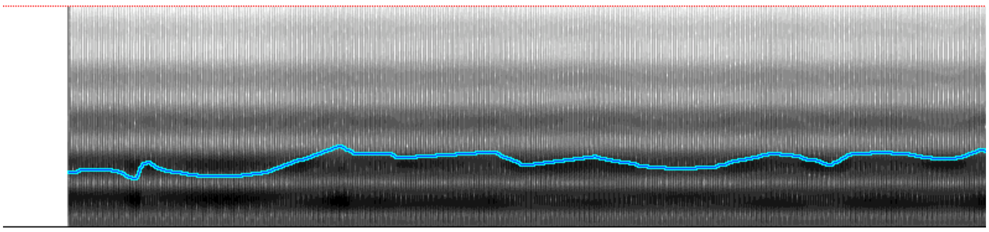


Figure 4.5: Sample spectrogram from Condition Four

4.2.5 Condition Five - Rhythm and Intonation

In Condition Five our aim was to find out how accurately participants would identify the two dialects if they could hear both rhythm and intonation together. A hummed version of each sound file was created, as for Condition Four. Each original vocalic segment was then replaced by the corresponding hummed segment, while consonants and pauses were replaced by silence. This condition is similar to Condition Three, with the exception that all vowels retain their original pitch movements instead of being replaced by a monotonous tone complex.

Compare the spectrogram for this condition, shown in Figure 4.6, with the spectrogram taken from Condition Three (Figure 4.4). The only difference between the two is that the pitch dynamics during vocalic segments are now clearly visible, as opposed to Condition Three, where intonation was flattened.

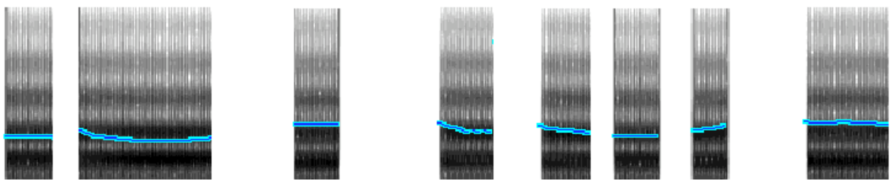


Figure 4.6: Sample spectrogram from Condition Five

4.2.6 Condition Six - Low-pass Filtered at 400Hz

Several previous studies have used low-pass filtered speech for dialect identification tasks (see Section 2.3.1). It is claimed that low-pass filtering at 400Hz eliminates segmental information, which is mainly contained in the higher frequencies. At the same time low-pass filtering is said to retain both syllabic rhythm and intonation, as pitch rarely rises higher than 400Hz. However, listening to filtered speech makes it obvious that it is not only rhythm and intonation that are preserved. In addition to rhythm and intonation, low-pass filtering seems to also retain the voice quality of the speech. This makes Condition Six somewhat different from Condition Five, and accordingly both were used in the perception task. If participants' accuracy differs in the two conditions, it would suggest that they are using voice quality as a cue in dialect identification.[1]

Low-pass filtering at 400Hz was carried out using Praat with a smoothing of 50Hz. As the filtering produces a muffled sound, the amplitude was multiplied by four to increase loudness. This ensured that Condition Six would not be noticeably quieter than the other conditions. Figure 4.7 shows how all spectral information above 400Hz is eliminated from the speech signal.

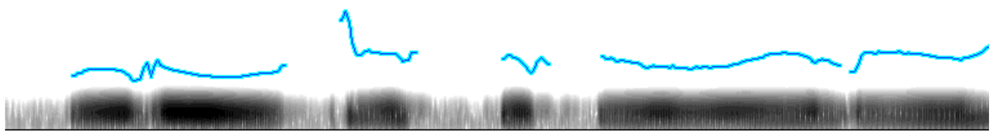


Figure 4.7: Sample spectrogram from Condition Six

4.2.7 Condition Seven - Unaltered speech

The final condition contained the original passages and acted as a control condition. Although this is the unaltered condition, two changes were nonetheless carried out. To make sure that participants would not base their judgements on non-standard syntax rather then on the actual phonetic properties of the speech, two words were cut from two of the original passages. The word *more* was deleted from the phrase *more faster* (speaker *m02*), while *don't* was deleted from the phrase *you don't see*

[1] And possibly some limited segmental information, such as F1, for example.

nothing (speaker *m28*). A test run was carried out among a group of linguists to make sure that the altered phrases still sounded natural.[1]

A sample spectrogram with its corresponding textgrid from the unmodified speech condition is shown in Figure 4.2.

4.3 Stimulus Tape

The seven speech conditions were organised into seven blocks with each of the twenty passages – one from each speaker – randomised within each block. The various test conditions were played in an ascending order, starting with Condition One. That is, the presentation of material was so arranged that the amount of linguistic information that was made available to the listeners increased from one block to the next one. This prevented listeners from transferring information gathered from one block over to the next block.

Most of the seven blocks were preceded by an example passage to familiarise listeners with the following condition. It was felt necessary to avoid recording a New Zealand English speaker (either Pākehā or Māori) for the example passage, as their perceived ethnicity might influence the listener's answers in the condition proper. We decided to use a non-English speaker for introducing the conditions as we felt that listening to a foreign language example would be unlikely to prime things Māori or Pākehā in the participants' mind. This allows us to be more confident that the results were not skewed. A young female Romanian speaker was recorded for this purpose. To keep with the theme, she was asked to talk about *rugby* in her native language. At the beginning of a new condition, first the unaltered example passage was presented, followed by the corresponding modified speech. This demonstrated to listeners how the modified version of the speech relates to normal human speech. The foreign language example passage was omitted before Conditions Two and Three, as they were impressionistically very similar to Condition One and a new introduction was not necessary.

The mean duration of the passages used in the task was 13.4 seconds, with the actual lengths ranging between 10–15 seconds. The interstimulus interval was 1.5 seconds, which was measured from the offset to the onset of the number introducing the next passage. Another pause of 1.5 seconds was played prior to each stimulus

[1] These words were present for the other conditions, as Condition Seven was dealt with after the speech resynthesis had been completed. They were also taken into account when PVI values were calculated in the production experiment.

onset. The numbers were read out by a young male Pākehā speaker, while the different condition blocks were introduced by another young male Pākehā speaker. As the task was quite lengthy, a two minute break was held after Condition Four. Thus, Part One consisted of 80 passages in four conditions and Part Two consisted of 60 passages in the remaining three conditions. All in all, each experimental session lasted about 55 minutes, including instruction time and the filling out of a background information sheet (see Appendix G) by the participant.

The manipulated speech fragments were converted from digital to analogue and then recorded onto a cassette tape. Most of the subjects were tested individually in their own home or in the author's office. The stimuli were presented over headphones at a comfortable listening level. At certain times, however, it was necessary to speed up the experimental procedure by using small groups of maximum 5 people instead of individual sessions. In these cases the material was presented in a quiet room through speakers instead of headphones.

4.4 Answer Sheets

Participants were issued with response sheets on which they were asked to indicate, for each passage, whether they thought a particular speaker was Māori or Pākehā. They responded by circling a number on a four-point scale ranging from 1 *'very Pākehā sounding'* to 4 *'very Māori sounding'*, as shown in Figure 4.4. Participants were provided with information relating to each speaker's age (*'young'* or *'old'*) and gender; they had only to decide on the speaker's perceived ethnicity. See Appendix F for a complete sample response sheet.

Table 4.2: Sample taken from the instruction sheet.

			very Pākehā sounding	somewhat Pākehā sounding	somewhat Māori sounding	very Māori sounding
#1	young	female	1	2	3	4

Participants were required to always circle a scale position, even if they felt they had to guess. It should be noted that, since the scale has no midpoint, the subjects were forced to make a decision, no matter how tentative. Although a four point scale was presented on the answer sheet, the statistical analysis presented in Chapter 5 will treat the perceived ethnicity of the speaker as a strictly binary variable: a score of 1 or 2 = Pākehā, a score of 3 or 4 = Māori.

As noted above, speaker gender information was provided on the answer sheet, as previous research has shown that speaker gender can be correlated with ethnic identifications, suggesting that listeners rely on different cues for the ethnic identification of females and males (Thomas and Lass 2005). Moreover, having the gender information available avoids participants using the two different pitches in the early conditions as markers of ethnicity. As the signal in the synthesised conditions did not contain information about speaker age, this was also provided on the answer sheet. It is therefore possible to investigate whether the various suprasegmental cues are interpreted differently for younger and older voices.

4.5 Participants

Altogether, 107 subjects took part in the perception experiment, all born and raised in New Zealand. 55 of them marked themselves as Pākehā and 52 claimed to be of Māori descent. The distribution of the participants according to age, gender and ethnicity is shown in Table 4.3 .

Table 4.3: Distribution of the 107 participants in the perception task according to ethnicity, gender and age.

	10's	20's	30's	40's	50's	60's
Māori female	5	8	5	2	2	2
Māori male	5	11	5	3	2	2
Pākehā female	9	12	2	4	3	3
Pākehā male	6	3	4	5	2	2

The dialect identification experiment was carried out in the South Island of New Zealand. As previously mentioned in Section 3.1.1, the Māori speakers used in the task were recorded in the North Island so that their recordings could be used in the perception experiment. It decreased the likelihood that the voices used in the task would be recognised by the listeners. This was necessary given the relatively small size of the Māori community in South Island and in Christchurch in particular.

Listeners were not paid for their participation, but they did received home-made savoury snacks and cakes.

4.5.1 Māori Integration Index

Previous research (e.g., Clopper 2004) has indicated that people with more exposure to a particular dialect are more accurate in identifying that dialect in a perception experiment. In order to test this hypothesis in the New Zealand context, it was necessary to devise a procedure for characterising social network structure which reflects local social practice. A Māori Integration Index (MII) was designed to measure participants' level of involvement in Māoridom. The MII was constructed form responses to eight questions presented on the background information sheet (see Appendix G.) Section 3.1.2 provided a detailed description of how the actual scores were assigned, and also discusses the distribution of MII assigned to the speakers used in the production experiment.

The histograms in Figure 4.8 illustrate the distribution of the Māori Integration Indices amongst Māori participants (left panel) and Pākehā participants (right panel). The average score for Māori participants was 11.95, with the median at 13, while Pākehā averaged a score of 3.3 with their median being at 2.5.

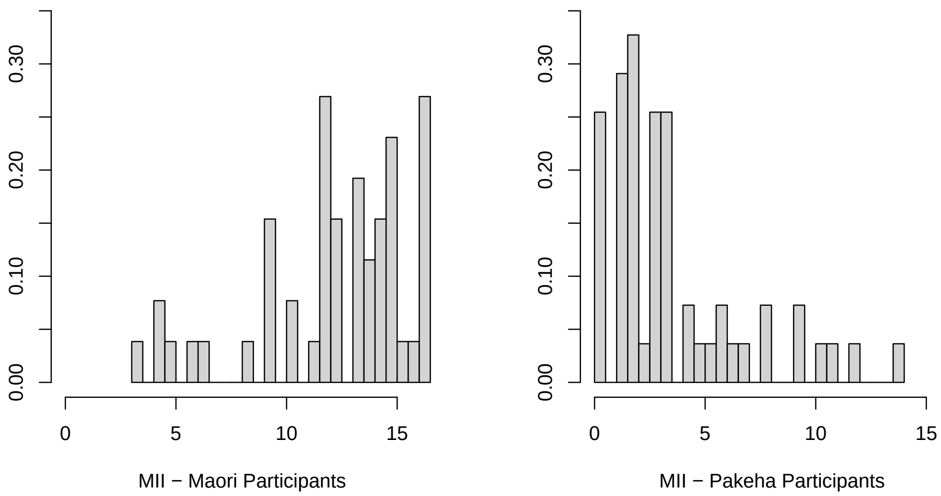


Figure 4.8: Histograms indicating the distribution of Māori Integration Indices for Māori participants (left panel) and Pākehā participants (right panel).

As previously discussed in Section 3.2.1, none of the Pākehā speakers scored higher on the MII than any Māori speaker, making it difficult to tell whether the MII was doing more than just separating the two ethnic groups. In the perception experiment, however, the distribution of MII scores amongst participants was very different, with the desired result of some Pākehā scoring higher than some Māori subjects. That is, some Pākehā listeners were more integrated into Māori society than some low-scoring Māori participants. This provides a good testing ground to examine whether the MII is a better indicator of participants' accuracy in the dialect identification task than the social variable of *ethnicity* alone.

Chapter V

Perception Results

This chapter provides a description of the results from the perception experiment. The results from each of the seven conditions are presented in separate sections, starting with Condition Seven, unaltered speech, working back through to Condition One, which is discussed last. Two logistic regression models were fit by hand for each condition. The first one tested the effects on participants' actual responses, that is, whether they perceived the speaker to be Māori or Pākehā. This model will be referred to as *the model of perceived ethnicity.* The aim of the second regression model we fit in each condition was to check what affected the accuracy of the answers, that is, whether participants were able to correctly identify the speaker's ethnicity. This model will be referred to as *the model of accuracy.* The two models are described in separate subsections for each condition. Below in Table 5.1 is a list of the 22 variables that were used in the analysis of the results and had the potential to show up in any one of the statistical models.

Table 5.1: Variables used in the analysis of the perception results.

External Factors	Rhythm & Pitch	HRTs & Speech Rate
speaker ethnicity	PVI	percentage of HRTs
speaker gender	mean pitch	mean HRT duration
speaker age	maximum pitch	mean HRT pitch range
speaker MII	minimum pitch	mean slope of HRT
participant ethnicity	pitch range	HRT ratio
participant gender	SD of pitch	number of pauses
participant age		mean pause duration
participant MII (PMII)		speech rate

Below is a list of various things to be noted about the variables and the statistical analysis:

- Speaker age and participant age are used in categories grouped by decades, rather than showing actual age. That is, the possible values are 10, 20, 30, 40, 50 and 60.

- Measurements for standard deviation of pitch were converted to logSD of pitch, in order to compress the values and eliminate outliers.

- Not all of the 20 speakers used in the perception experiment produced HRTs. In fact, 3 Māori and 5 Pākehā speakers did not use any. Phonetic properties of HRTs, such as mean duration, range, slope and ratio, were analysed based on the 12 speakers who did produce some percentage of HRTs. These values were left blank for the speakers who did not have HRTs to be measured.

- As some of the variables correlate with, or are predictable by, other variables, variance inflation factors were carefully checked for each statistical model to measure multi-colinearity of the independent variables. All combinations of predictors used in our fit were tolerated by the regression models. For example, some models allowed the presence of both participant ethnicity and participant MII as predictors.

5.1 Condition Seven - Unaltered Speech

Although Condition Seven was presented last during the dialect identification experiment, the results from this condition will be described first. This condition retained all segmental and suprasegmental information, so it is not surprising that participants' accuracy was highest in this condition. The overall accuracy was a very high 92%. Table 5.2 shows the percentage of correct responses by participant and speaker ethnicity. Pākehā speakers were easier to identify for both Māori and Pākehā participants. However, Māori participants were more successful at identifying Māori speakers than were Pākehā participants.

Table 5.2: Condition Seven - Percentage of correct responses by participant and speaker ethnicity.

	Māori speakers	Pākehā speakers	Total
Māori participants	93%	95%	94%
Pākehā participants	86%	95%	91%

5.1.1 Perceived Ethnicity - Condition 7

In the first of the two logistic regression models we investigated the effects of the perceived ethnicity, that is whether participants identified a speaker as Māori or Pākehā. The variables that showed significant effects were:

(a) PVI interacting with participant MII
(b) % of HRTs used interacting with speaker ethnicity
(c) minimum pitch
(d) standard deviation of pitch interacting with speaker gender
(e) mean slope of HRTs

The coefficient table for this model is presented in Table 5.3, while the corresponding anova table is shown in Table 5.4.

Table 5.3: Condition Seven - Coefficient table for model of perceived ethnicity.

Variable	Coef	S.E.	Wald Z	P
Intercept	-31.75	7.53	-4.21	0.0000
PVI	0.04	0.04	1.12	0.2607
PMII	-0.58	0.18	-3.24	0.0012
log SD of pitch	27.41	5.61	4.88	0.0000
speaker gender=male	32.50	7.57	4.29	0.0000
% of HRTs	-0.13	0.01	-7.95	0.0000
speaker ethnicity=Pākehā	-3.70	1.14	-3.25	0.0011
minimum pitch	0.02	0.00	3.21	0.0013
mean slope of HRTs	-0.01	0.00	-6.28	0.0000
PVI * PMII	0.01	0.00	3.10	0.0019
log SD of pitch * speaker gender=male	-24.71	5.38	-4.59	0.0000
% of HRTs * speaker ethnicity=Pākehā	0.13	0.02	6.08	0.0000

Table 5.4: Condition Seven - Anova table for model of perceived ethnicity.

Factor	Chi-Square	d.f.	P
PVI	15.53	2	0.0004
PMII	11.51	2	0.0032
log SD of pitch	26.25	2	<.0001
speaker gender	24.26	2	<.0001
% of HRTs	65.21	2	<.0001
speaker ethnicity	64.82	2	<.0001
minimum pitch	10.31	1	0.0013
mean slope of HRTs	39.46	1	<.0001
PVI * PMII	9.64	1	0.0019
log SD of pitch * speaker gender	21.03	1	<.0001
% of HRTs * speaker ethnicity	36.97	1	<.0001
TOTAL INTERACTION	57.81	3	<.0001
TOTAL	385.18	11	<.0001

The model shows a significant effect of PVI interacting with the participant's MII (PMII). A 3D diagram illustrating this interaction can be seen in Figure 5.1. All figures in this chapter show model predictions, holding all other significant effects constant. The y-axis shows the log odds of a Pākehā response, where higher values indicate that participants are more likely to perceive the speaker as Pākehā and lower values indicate that they are more likely to perceive the speaker as Māori. The back panel of the 3D diagram shows the responses of participants with a high MII. Highly integrated participants perceive speakers with high PVI values (i.e. stress-timing) as very Pākehā, and speakers with low PVI values (i.e. syllable-timing) as very Māori. The responses of participants with a low MII are shown in the front panel of the diagram. They use PVI as a cue for ethnic identification in a similar fashion to those with high PMII's, with stress-timed speakers being perceived as more Pākehā, and syllable-timed speakers as more Māori. However, the trend is not as pronounced as it is for highly integrated participants, who appear very confident that syllable-timing is an indicator of a Māori speaker and stress-timing is and indicator of a Pākehā speaker.

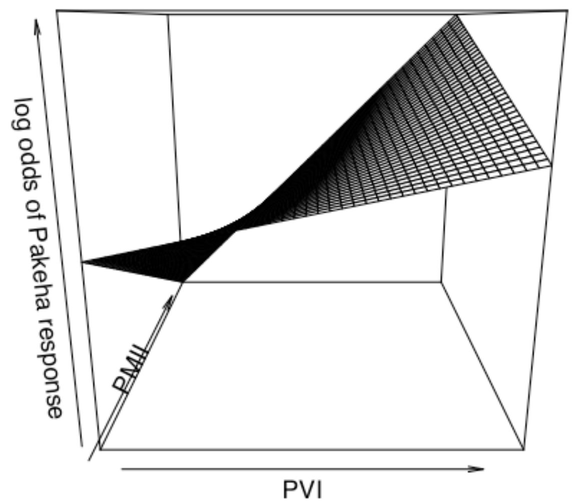


Figure 5.1: Condition Seven - Model effects showing log odds of Pākehā response by PVI and participant MII.

Perceived ethnicity is affected by another significant interaction, which occurs between the percentage of HRTs and the speaker's actual ethnicity. This is shown in Figure 5.2. The percentage of HRTs seems to be a very important cue for dialect identification in the case of Māori speakers. If a Māori speaker has a low percentage of HRTs, then they are more likely to be identified as Pākehā. However, if they use a lot of HRTs then they are considered to be Māori sounding. Listeners do not make use of the percentage of HRTs in the case of Pākehā speakers. As the relatively flat line around zero indicates on the graph, the percentage of HRTs used by a Pākehā speaker does not seem to influence participants' responses in either direction. This presumably indicates that HRT use is a secondary cue, unlikely to influence responses alone, but increasing the certainty of the listener if a speaker has other characteristics that make them sound Māori.

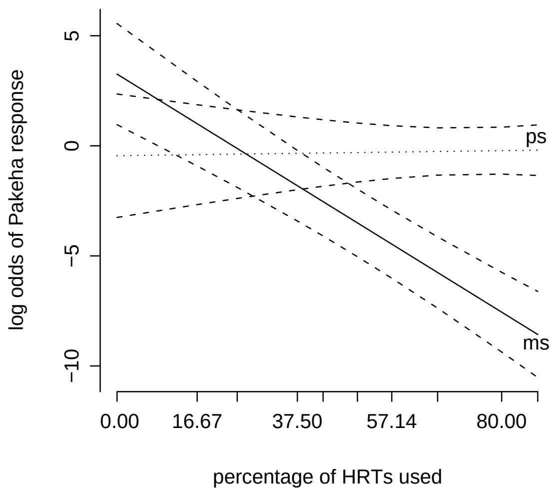


Figure 5.2: Condition Seven - Model effects showing log odds of Pākehā response by percentage of HRTs used and speaker ethnicity.

Minimum pitch also had a significant effect on perceived ethnicity. Speakers whose minimum pitch is very low are identified as Māori, whereas a high minimum pitch seems to suggest that the speaker is Pākehā. See Figure 5.3. This is true for both Māori and Pākehā participants, as there was no interaction with participant ethnicity in the model. This is an interesting result in that it differs from the production results, which indicated that there is no significant difference between the two ethnolects in terms of minimum pitch. In fact, as shown in Section 3.2.2, Māori speakers have a significantly higher mean pitch than Pākehā speakers.

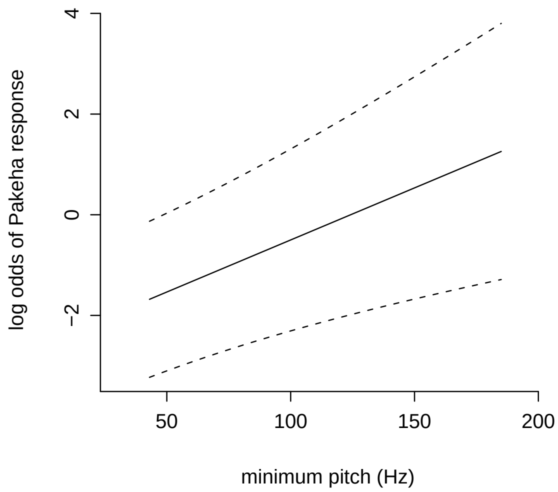


Figure 5.3: Condition Seven - Model effects showing log odds of Pākehā response by minimum pitch.

Minimum pitch is not the only cue relating to pitch that is used for ethnic identification in this condition. The standard deviation of pitch interacting with speaker gender also had a significant effect on perceived ethnicity. This is illustrated in Figure 5.4. Overall, the higher the standard deviation of pitch, the more likely the speaker will be identified as Pākehā. However, if the standard deviation of pitch is low, participants are inclined to say the speaker is Māori. This trend is more pronounced for female speakers than male speakers, that is, monotonous female speakers are considered to be very Māori sounding.

The last variable that showed a significant effect on perceived ethnicity in Condition Seven is the mean slope of HRTs. Speakers whose HRTs demonstrate a steeper rise in pitch are considered to be more Māori sounding.[1] See Figure 5.5.

[1] This is based on the 12 speakers that produced HRTs in the perception experiment. Values were left blank for the 8 speakers who did not have HRTs.

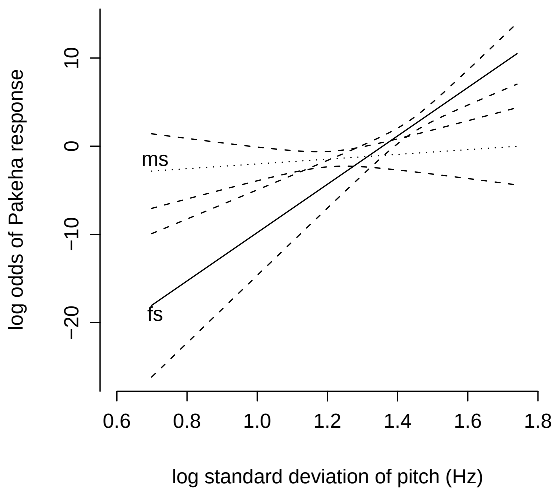


Figure 5.4: Condition Seven - Model effects showing log odds of Pākehā response by standard deviation of pitch and speaker gender.

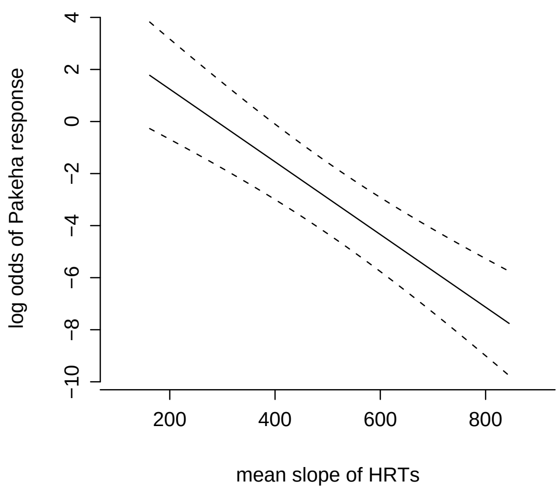


Figure 5.5: Condition Seven - Model effects showing log odds of Pākehā response by mean slope of HRTs.

5.1.2 *Accuracy - Condition 7*

In the second regression model we investigated the effects on the accuracy of participants' responses. Our aim was to find out which variables help listeners correctly identify a speaker's ethnicity and which ones make it more difficult. The following list shows the variables that had a significant effect on accuracy in our model:

(a) participant MII interacting with participant gender
(b) PVI
(c) percentage of HRTs used interacting with speaker gender
(d) standard deviation of pitch interacting with speaker ethnicity
(e) mean pitch interacting with speaker ethnicity

The coefficient table for this model is shown in Table 5.5, and the anova table is presented in Table 5.6.

Table 5.5: Condition Seven - Coefficient table for model of accuracy.

Variable	Coef	S.E.	Wald Z	P
Intercept	-4.03	2.43	-1.66	0.0968
PVI	-0.12	0.02	-4.44	0.0000
log SD of pitch	13.28	1.57	8.42	0.0000
speaker ethnicity=Pākehā	6.59	1.34	4.90	0.0000
% of HRTs	0.00	0.00	0.47	0.6389
speaker gender=male	3.66	0.95	3.84	0.0001
mean pitch	-0.04	0.00	-5.40	0.0000
participant gender=male	-0.61	0.26	-2.27	0.0234
PMII	0.04	0.02	2.19	0.0287
log SD of pitch * speaker ethnicity=Pākehā	-8.87	1.07	-8.24	0.0000
% of HRTs * speaker gender=male	-0.04	0.01	-3.81	0.0001
speaker ethnicity=Pākehā * mean pitch	0.04	0.00	5.37	0.0000
participant gender=male * PMII	0.08	0.03	2.46	0.0141

Table 5.6: Condition Seven - Anova table for model of accuracy.

Factor	Chi-Square	d.f.	P
PVI	19.68	1	<.0001
log SD of pitch	80.37	2	<.0001
speaker ethnicity	85.52	3	<.0001
% of HRTs	22.11	2	<.0001
speaker gender	16.10	2	0.0003
mean pitch	36.09	2	<.0001
participant gender	6.40	2	0.0408
PMII	28.59	2	<.0001
log SD of pitch * speaker ethnicity	67.91	1	<.0001
% of HRTs * speaker gender	14.49	1	0.0001
speaker ethnicity * mean pitch	28.88	1	<.0001
participant gender * PMII	6.03	1	0.0141
TOTAL INTERACTION	75.91	4	<.0001
TOTAL	166.51	12	<.0001

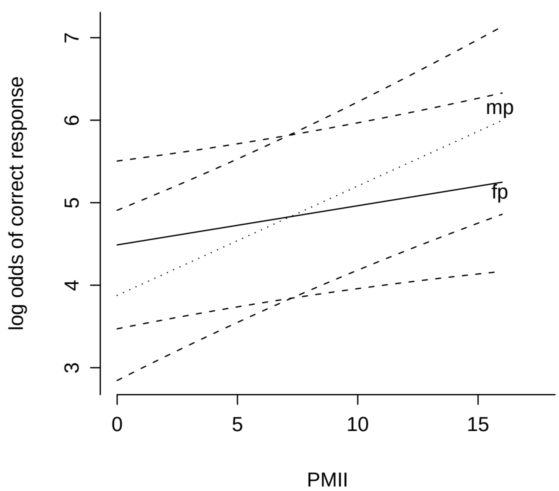


Figure 5.6: Condition Seven - Model effects showing log odds of correct response by participant MII and participant gender.

The model clearly indicates that people with a high MII are better at correctly identifying speaker ethnicity. Figure 5.6 illustrates the interaction between participant MII and participant gender. The y-axis now shows the log odds of a correct response. The higher the value, the more likely the answer is to be correct. As can be seen from the graph, highly integrated people are more accurate at the task than are those with a low MII. Participant gender also plays a role. Highly integrated male listeners are slightly more accurate than highly integrated females. However, the opposite is true for people with a low integration index, where females are doing slightly better than their male counterparts.

Speaker rhythm also affects participants' accuracy. As Figure 5.7 shows, listeners are more accurate at identifying speakers with a low PVI value. That is, people find it easier to identify the ethnicity of a syllable-timed speaker as Māori, while they make more mistakes when trying to identify the ethnicity of a more stress-timed speaker as Pākehā.

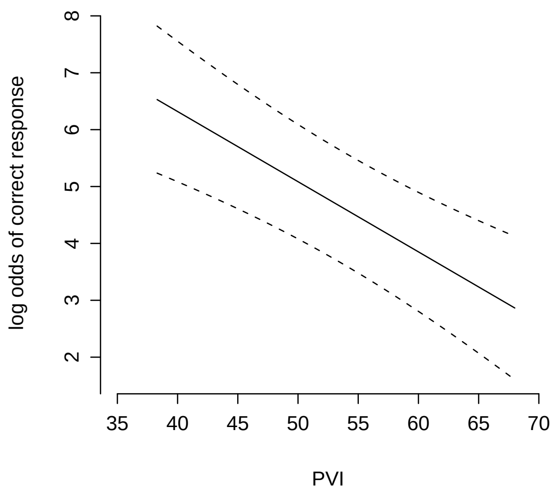


Figure 5.7: Condition Seven - Model effects showing log odds of correct response by PVI.

The percentage of HRTs used by a speaker also has a significant effect on whether listeners can correctly identify their ethnicity, but this effect interacts with speaker gender. This interaction is shown in Figure 5.8. If a speaker is female, her percentage of HRTs does not appear to affect the accuracy of participants in identifying her as Māori or Pk̄ehā. However, if a male speaker produces many HRTs, his ethnicity is more likely to be misidentified. Recall that in this data set, Pākehā males produced proportionally more HRTs than did Māori males (see Section 3.2.3). This might explain the low accuracy ratings for male speakers with a high percentage of HRTs, because listeners expect them to be Māori, but, in fact, many of them are Pākehā.

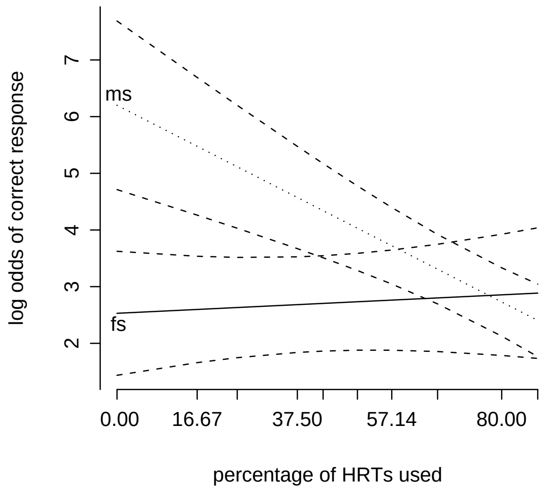


Figure 5.8: Condition Seven - Model effects showing log odds of correct response by percentage of HRTs used and speaker gender.

Figure 5.9 illustrates the interaction of standard deviation of pitch and speaker ethnicity in this model. Participants are more accurate at identifying speaker ethnicity if the speaker's standard deviation of pitch is high. A high standard deviation of pitch might make distinctive intonational patterns and HRTs more perceivable for the listener, thus facilitating correct identification of speaker ethnicity. Monotonous Māori speakers are least accurately identified. As the model of perceived ethnicity showed, participants are inclined to mark speakers with a low standard deviation of pitch as Māori. However, the production experiment indicated that this is not the case, as there is no significant difference between Māori and Pākehā SD of pitch.

Mean pitch interacting with speaker ethnicity also showed a significant effect on accuracy. Participants are more likely to misidentify Māori speakers who have a high mean pitch, indicating that listeners expect Māori speakers to have a low mean pitch. In the case of Pākehā speakers, mean pitch does not influence the accuracy of responses.

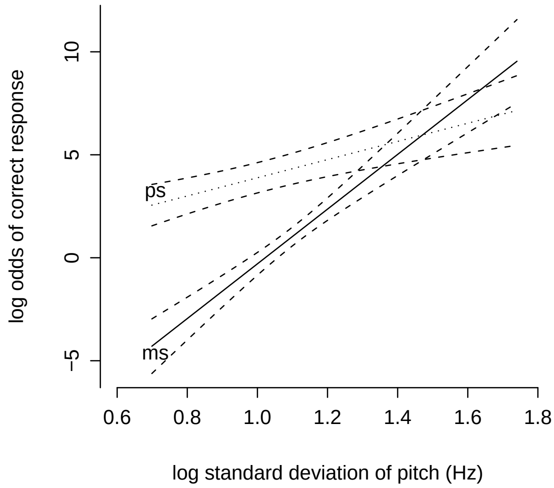


Figure 5.9: Condition Seven - Model effects showing log odds of correct response by standard deviation of pitch and speaker ethnicity.

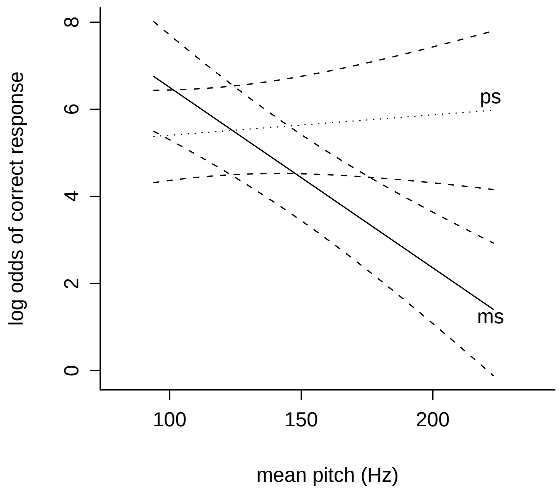


Figure 5.10: Condition Seven - Model effects showing log odds of correct response by mean pitch and speaker ethnicity.

5.1.3 Summary for Condition Seven

Below is a summary of all the variables that showed a significant effect in our statistical models in Condition Seven. The effect of *increasing values* of each variable on the likelihood of Pākehā response is shown in Table 5.7, while Table 5.8 summarises the effect on the accuracy of responses.

Table 5.7: Condition Seven - Summary of significant variables in model of perceived ethnicity.

increasing value of variable	likelihood of Pākehā response
PVI	increases, more so for high MII participants
% of HRTs	decreases for Māori speakers
minimum pitch	increases
SD of pitch	increases, more so for female speakers
mean slope of HRTs	decreases

Table 5.8: Condition Seven - Summary of significant variables in model of accuracy.

increasing value of variable	likelihood of correct response
participant MII	increases, steeper for male participants
PVI	decreases
% of HRTs	decreases for male speakers
SD of pitch	increases, steeper for Māori speakers
mean pitch	decreases for Māori speakers

5.2 *Condition Six - Low-pass Filtered at 400Hz*

Low-pass filtering at 400Hz eliminates most segmental information from the speech signal but preserves rhythm and intonation, and arguably voice quality as well. As described in Section 2.2.1, there are noted differences in segmental features between Māori and Pākehā English. The fact that these segmental cues are unavailable for participants in this condition to facilitate dialect identification results in lower accuracy ratings when compared to the unaltered speech condition. In contrast to Condition Seven, in this condition Māori speakers were easier to identify than were Pākehā speakers overall (cf. Table 5.2). Māori participants in particular achieved a high 75% accuracy when identifying fellow Māori English speakers. The average percentages of correct responses for this condition are given in Table 5.9, showing both speaker and participant ethnicity.

Table 5.9: Condition Six - Percentage of correct responses by participant and speaker ethnicity.

	Māori speakers	Pākehā speakers	Total
Māori participants	75%	62%	69%
Pākehā participants	66%	60%	63%

5.2.1 *Perceived Ethnicity - Condition 6*

Just as for Condition Seven, the first of the two logistic regression models we fit for this condition investigated the effects of perceived ethnicity. The variables in this model showing a significant effect on whether a speaker is identified as Māori or Pākehā are listed below:

(a) PVI interacting with participant MII
(b) % of HRTs interacting with participant ethnicity & participant gender
(c) minimum pitch
(d) speech rate
(e) number of pauses

The coefficient table for this model is shown in Table 5.10, and the anova table is presented in Table 5.11.

Table 5.10: Condition Six - Coefficient table for model of perceived ethnicity.

Variable	Coef	S.E.	Wald Z	P
Intercept	-5.42	0.85	-6.36	0.0000
% of HRTs	-0.00	0.00	-0.03	0.9759
participant ethnicity=Pākehā	0.89	0.24	3.74	0.0002
participant gender=male	0.31	0.19	1.58	0.1142
PVI	0.04	0.01	4.05	0.0001
PMII	-0.28	0.06	-4.57	0.0000
minimum pitch	0.01	0.00	6.31	0.0000
speech rate	0.52	0.11	4.44	0.0000
number of pauses	-0.12	0.03	-3.91	0.0001
% of HRTs * participant ethnicity=Pākehā	-0.02	0.00	-4.93	0.0000
% of HRTs * participant gender=male	-0.00	0.00	-0.09	0.9322
participant ethnicity=Pākehā * part gender=male	-0.96	0.27	-3.48	0.0005
PVI * PMII	0.00	0.00	4.50	0.0000
% of HRTs * part ethn=Pākehā * part gender=male	0.02	0.00	3.53	0.0004

The model indicates that participants do use rhythm as a cue for dialect identification when low-pass filtering is applied. Highly integrated people use it very well, marking syllable-timed speakers as Māori and stress-timed speakers as Pākehā sounding. This can be seen on the back panel of the diagram in Figure 5.11. A speaker with a high PVI value is very likely to be marked as Pākehā, while a speaker with a low PVI value is very likely to be identified as Māori by participants who have a high MII. The front panel shows the answers of participants who have a low integration index. These listeners also use rhythm to decide whether a speaker is Māori or Pākehā, although not quite as well as highly integrated participants do. Nevertheless, for these listeners syllable-timing still triggers a Māori response while stress-timing suggests that the speaker is much more likely to be Pākehā.

In addition to rhythm, low-pass filtering also preserves pitch dynamics in the speech signal. Participants in this condition use minimum pitch to help them decide whether a speaker is Pākehā or Māori. Speakers with a low minimum pitch are perceived as Māori, while a high minimum pitch prompts a Pākehā response. As there is no statistical interaction with participant ethnicity, the model estimates that all participants use minimum pitch in this fashion. See Figure 5.12.

Table 5.11: Condition Six - Anova table for model of perceived ethnicity.

Factor	Chi-Square	d.f.	P
% of HRTs	50.38	4	<.0001
participant ethnicity	26.43	4	<.0001
participant gender	28.90	4	<.0001
PVI	142.74	2	<.0001
PMII	20.92	2	<.0001
minimum pitch	39.76	1	<.0001
speech rate	19.74	1	<.0001
number of pauses	15.31	1	1e-04
% of HRTs * participant ethnicity	24.29	2	<.0001
% of HRTs * participant gender	24.13	2	<.0001
participant ethnicity * participant gender	14.35	2	8e-04
PVI * PMII	20.27	1	<.0001
% of HRTs * part ethnicity * part gender	12.48	1	4e-04
TOTAL INTERACTION	59.63	5	<.0001
TOTAL	279.5	13	<.0001

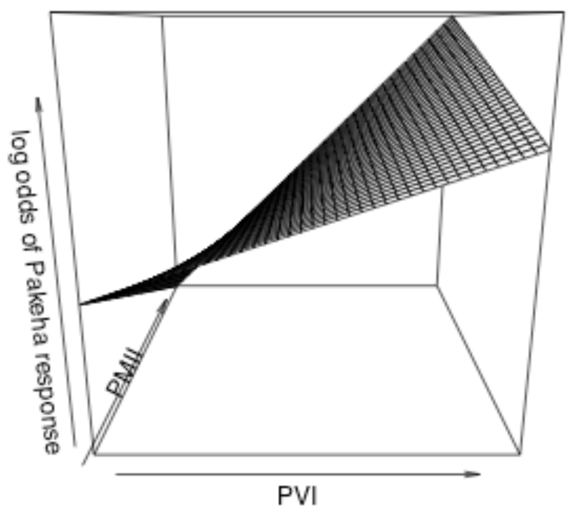


Figure 5.11: Condition Six - Model effects showing log odds of Pākehā response by PVI and participant MII.

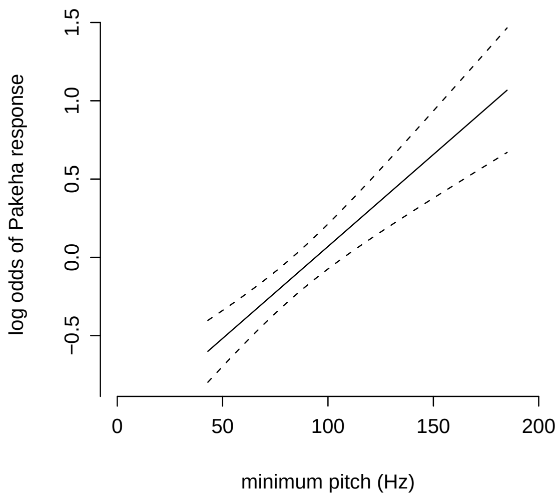


Figure 5.12: Condition Six - Model effects showing log odds of Pākehā response by minimum pitch.

As pitch movement is retained by low-pass filtering, HRTs are easily perceivable in this condition. However, not all participants seem to use the percentage of HRTs present to identify speaker ethnicity. Whether a speaker uses many HRTs or none at all does not seem to have an effect on the responses of Māori participants in this condition. This is illustrated in the left panel of Figure 5.13 by the flat lines for both Māori male and female participants. The behaviour of Pākehā males parallels that of Māori participants, as they do not seem to use HRTs as a cue either. Pākehā females, however, do tune in to the percentage of HRTs used by a speaker in this condition. For them a speaker who uses many HRTs sounds very Māori, while the lack of HRTs indicates that the speaker is more likely to be Pākehā. See the right panel of Figure 5.13 for this gender difference.

Syllabic rhythm is not the only temporal feature of the speech signal retained by low-pass filtering at 400Hz. Speech rate and the number of pauses in the utterance also had a significant effect on perceived ethnicity in this condition. The model estimates that all participants, regardless of their own ethnicity, identified fast speakers as more Pākehā and slow speakers as more Māori. This is shown in Figure 5.14. Also, the more often a speaker pauses, the more likely they are perceived as Māori, as can be seen in Figure 5.15.

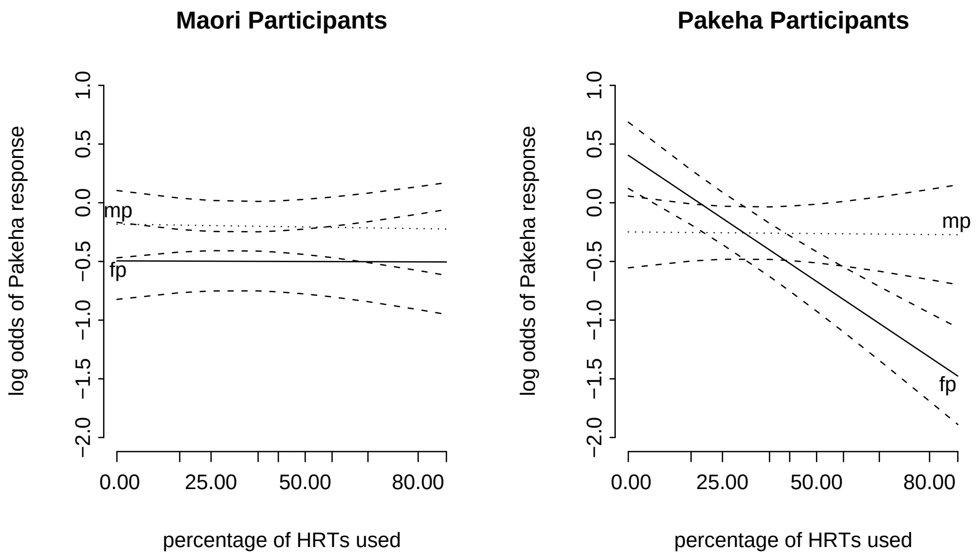


Figure 5.13: Condition Six - Model effects showing log odds of Pākehā response by percentage of HRTs used and participant gender. Coefficients set to Māori participants in left panel. Coefficients set to Pākehā participants in right panel.

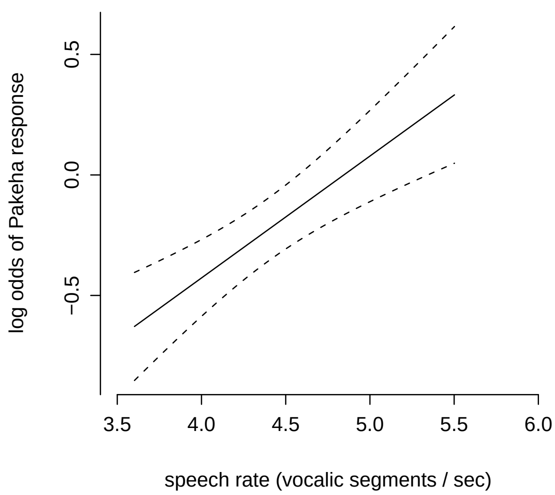


Figure 5.14: Condition Six - Model effects showing log odds of Pākehā response by speech rate.

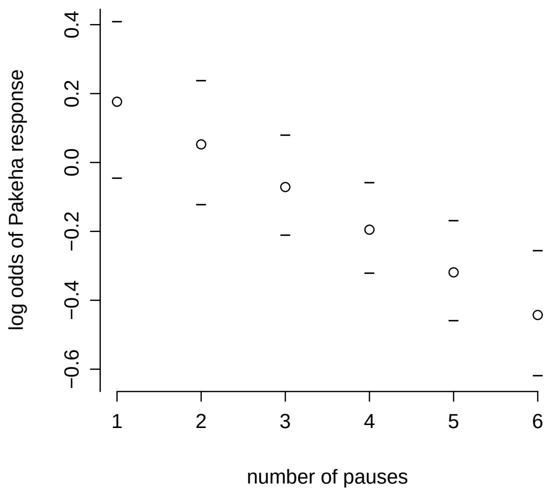


Figure 5.15: Condition Six - Model effects showing log odds of Pākehā response by number of pauses.

5.2.2 Accuracy - Condition 6

In the second regression model for this condition we investigated the effects on the correctness of participants' responses to examine which variables help listeners accurately identify a speaker's ethnicity when low-pass filtering is applied. The following is a list of the variables that had a significant effect on accuracy in our model:

(a) participant MII interacting with participant gender
(b) PVI
(c) % of HRTs interacting with speaker ethnicity and participant ethnicity
(d) number of pauses interacting with speaker ethnicity
(e) minimum pitch
(f) speech rate

The coefficient table for this model is shown in Table 5.12, and the anova table is presented in Table 5.13.

Table 5.12: Condition Six - Coefficient table for model of accuracy.

Variable	Coef	S.E.	Wald Z	P
Intercept	6.96	1.17	5.92	0.0000
participant gender=male	-0.61	0.16	-3.77	0.0002
PMII	0.03	0.01	2.33	0.0199
PVI	-0.07	0.01	-5.16	0.0000
% of HRTs	0.00	0.00	1.61	0.1072
speaker ethnicity=Pākehā	1.25	0.45	2.75	0.0060
participant ethnicity=Pākehā	-0.71	0.25	-2.80	0.0051
number of pauses	0.01	0.05	0.28	0.7833
speech rate	-0.46	0.15	-2.92	0.0035
minimum pitch	-0.01	0.00	-5.50	0.0000
participant gender=male * PMII	0.05	0.01	2.99	0.0028
% of HRTs * speaker ethn=Pākehā	-0.01	0.00	-1.90	0.0580
% of HRTs * part ethn=Pākehā	0.02	0.00	4.34	0.0000
speaker ethn=Pākehā * part ethn=Pākehā	1.37	0.28	4.83	0.0000
speaker ethn=Pākehā * number of pauses	-0.20	0.07	-2.88	0.0040
% of HRTs * speaker ethn=Pākehā * part ethn=Pākehā	-0.02	0.00	-4.59	0.0000

The model shows that the level of integration into Māori society significantly affects a participant's accuracy at the task. Listeners with a high integration index are more likely to correctly identify a speaker's ethnicity. There is also a gender difference amongst people with a low integration index, where females are slightly more correct than males. This gender effect disappears as the MII increases. Figure 5.16 shows this interaction between participants' MII and gender.

Similarly to Condition Seven, rhythm also had a significant effect on whether a speaker's ethnicity was correctly identified in Condition Six. Participants were more accurate at determining speaker ethnicity if the speaker was syllable-timed, that is, had a lower PVI value. This is illustrated in Figure 5.17.

Table 5.13: Condition Six - Anova table for model of accuracy.

Factor	Chi-Square	d.f.	P
participant gender	14.19	2	0.0008
PMII	28.79	2	<.0001
PVI	26.61	1	<.0001
% of HRTs	62.38	4	<.0001
speaker ethnicity	72.13	5	<.0001
participant ethnicity	29.28	4	<.0001
number of pauses	21.25	2	<.0001
speech rate	8.53	1	0.0035
minimum pitch	30.30	1	<.0001
participant gender * PMII	8.97	1	0.0028
speaker ethn * number of pauses	8.28	1	0.0040
% of HRTs * speaker ethn * part ethn	21.09	1	<.0001
TOTAL INTERACTION	78.69	6	<.0001
TOTAL	153.10	15	<.0001

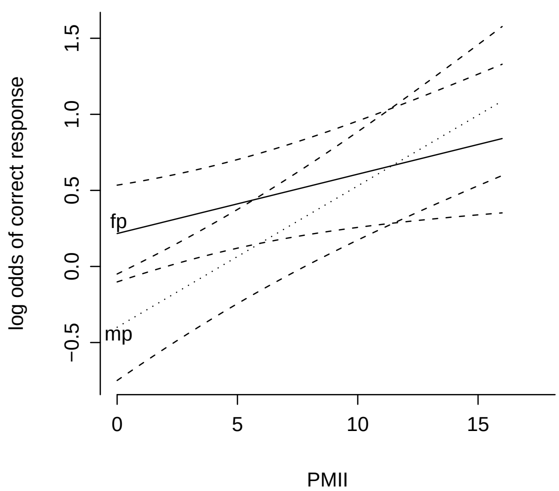


Figure 5.16: Condition Six - Model effects showing log odds of correct response by participant MII and participant gender.

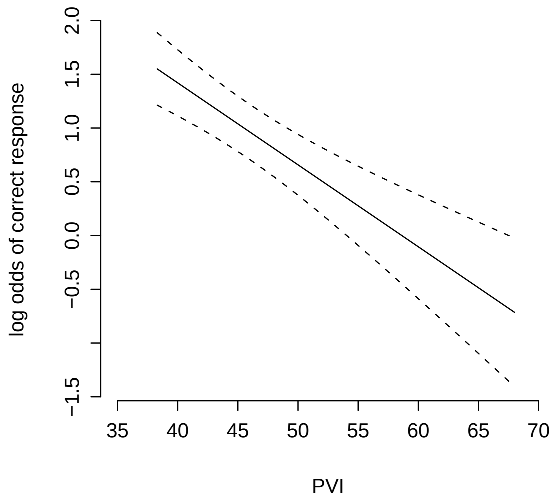


Figure 5.17: Condition Six - Model effects showing log odds of correct response by PVI.

The percentage of HRTs used only affected the accuracy of Pākehā participants. This is not surprising as the previous model of perceived ethnicity indicated that Māori participants were not using HRTs as a cue in this low-pass filtered condition. Pākehā participants, on the other hand, were most accurate when a Māori speaker used a high percentage of HRTs or when a Pākehā speaker did not use them at all. This three-way interaction between speaker ethnicity, the percentage of HRTs used and participant ethnicity is shown in Figure 5.18. The coefficients are set to Māori participants in the left panel, and to Pākehā participants in the right panel.

Participants' responses were more likely to be incorrect if the speaker's minimum pitch was high. As mentioned above, the model of perceived ethnicity in this condition illustrated that participants identified a higher minimum pitch with Pākehā speakers. However, the production experiment showed no significant differences between Māori and Pākehā speakers. This could explain the inaccuracy of participants' responses when it comes to high minimum pitch.

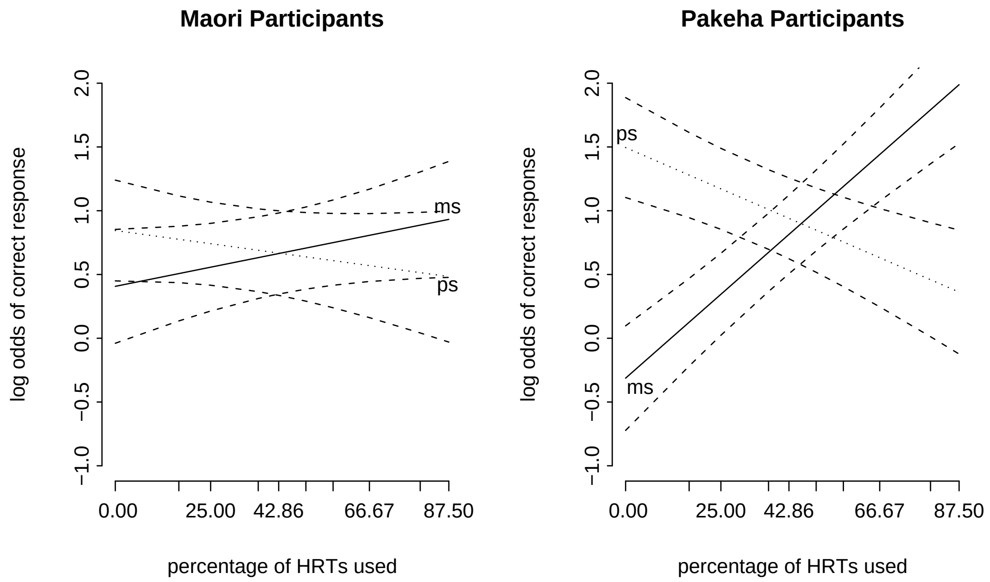


Figure 5.18: Condition Six - Model effects showing log odds of correct response by percentage of HRTs used and speaker ethnicity. Coefficients set to Māori participants in left panel and to Pākehā participants in right panel.

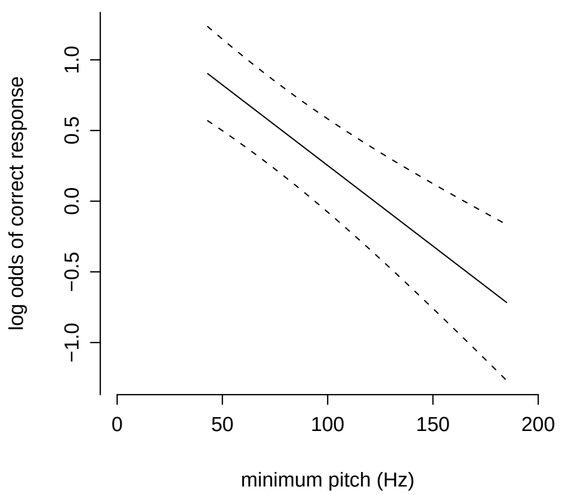


Figure 5.19: Condition Six - Model effects showing log odds of correct response by minimum pitch.

As mentioned above, speech rate and the number of pauses influenced participants' perception of the speaker's ethnicity in this condition. They also had a significant effect on the accuracy of the responses. Figure 5.20 shows that slow speakers were more accurately identified than fast speakers in this condition. Recall that speakers with a high speech rate were identified as more Pākehā. However, the results from the production experiment illustrated that there is no significant difference in speech rate between Māori and Pākehā English, as described in Section 3.2.4. Thus, the lower accuracy rate in the case of fast speakers is probably due to fast Māori speakers being misidentified as Pākehā. It may also be the case that a slower speech rate enables people to more accurately tune in to other cues, thereby increasing the accuracy of the responses.

The model of perceived ethnicity showed that speakers who pause more often during an utterance are more likely to be identified as Māori. The number of pauses only had an effect on the correctness of responses in the case of Pākehā speakers. Whether a Māori speaker uses fewer or more pauses does not affect participants' accuracy. On the other hand, the more often a Pākehā speaker pauses, the more likely they will be incorrectly identified as Māori. This interaction between the number of pauses and speaker ethnicity is illustrated in Figure 5.21.

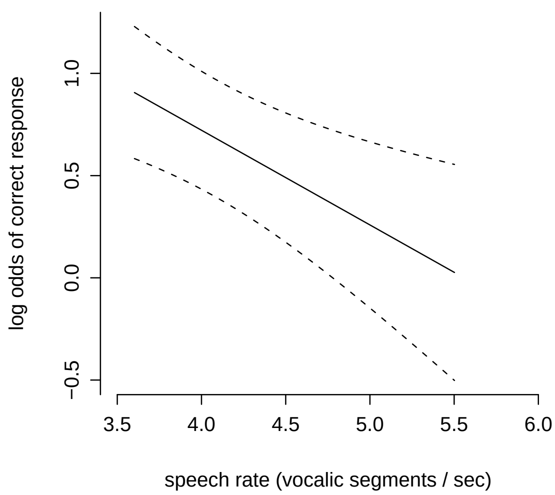


Figure 5.20: Condition Six - Model effects showing log odds of correct response by speech rate.

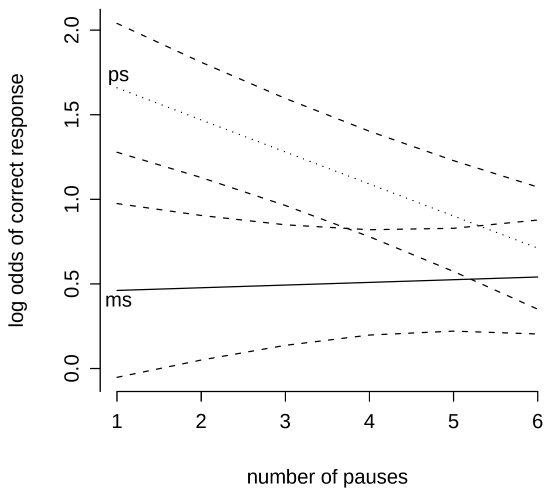


Figure 5.21: Condition Six - Model effects showing log odds of correct response by number of pauses and speaker ethnicity.

5.2.3 Summary for Condition Six

The tables below give a summary of all the variables that showed a significant effect in our statistical models in Condition Six. The effect of *increasing values* of each variable on the likelihood of Pākehā response is shown in Table 5.14, and Table 5.15 summarises the effect on the accuracy of responses.

Table 5.14: Condition Six - Summary of significant variables in model of perceived ethnicity.

increasing value of variable	likelihood of Pākehā response
PVI	increases, more so for high MII participants
minimum pitch	increases
% of HRTs	decreases for Pākehā female participants
speech rate	increases
number of pauses	decreases

Table 5.15: Condition Six - Summary of significant variables in model of accuracy.

increasing value of variable	likelihood of correct response
participant MII	increases, steeper for male participants
PVI	decreases
minimum pitch	decreases
% of HRTs	only affects Pākehā participants:
	increases for Māori speakers
	decreases for Pākehā speakers
speech rate	increases
number of pauses	decreases

5.3 Condition Five - Rhythm and Intonation

The aim of Condition Five was to investigate how well participants could identify a dialect if the only available cues they had were syllabic rhythm and intonation. In this condition each vowel was replaced by its corresponding hummed version that retained its pitch dynamics, while each consonant was replaced by silence. For more information on how the condition was created see Section 4.2.5. Accuracy ratings in this condition were significantly lower than in Condition Six, where low-pass filtering seems to retain slightly more information in the speech signal than just rhythm and intonation alone. Again the highest accuracy ratings were achieved by Māori participants, who managed to correctly identify other Māori speakers 59% of the time. Pākehā participants also performed above chance, although slightly worse than the Māori listeners. For the percentages of correct responses shown by both speaker and participant ethnicity see Table 5.16.

Table 5.16: Condition Five - Percentage of correct responses by participant and speaker ethnicity.

	Māori speakers	Pākehā speakers	Total
Māori participants	59%	56%	58%
Pākehā participants	53%	53%	53%

5.3.1 *Perceived Ethnicity - Condition 5*

The first of the two logistic models we fit in this condition examined the effects of participants' perception of a speaker's ethnicity. The following list shows the variables that had a significant effect on whether a speaker is identified as Māori or Pākehā when only rhythm and intonation are kept in the signal.

(a) PVI interacting with participant MII
(b) PVI interacting with the % of HRTs
(c) standard deviation of pitch

The coefficient table for this model is given in Table 5.17 and the anova table can be seen in Table 5.18.

Table 5.17: Condition Five - Coefficient table for model of perceived ethnicity.

Variable	Coef	S.E.	Wald Z	P
Intercept	-1.18	0.67	-1.77	0.0774
PMII	-0.21	0.05	-3.83	0.0001
PVI	0.01	0.01	0.98	0.3275
log SD of pitch	0.75	0.19	3.90	0.0001
% of HRTs	0.02	0.01	1.84	0.0654
PMII * PVI	0.00	0.00	3.65	0.0003
PVI * % of HRTs	-0.00	0.00	-2.84	0.0045

Table 5.18: Condition Five - Anova table for model of perceived ethnicity.

Factor	Chi-Square	d.f.	P
PMII	15.58	2	0.0004
PVI	34.39	3	<.0001
log SD of pitch	15.24	1	0.0001
% of HRTs	62.93	2	<.0001
PMII * PVI	13.31	1	0.0003
PVI * % of HRTs	8.06	1	0.0045
TOTAL INTERACTION	21.09	2	<.0001
TOTAL	97.20	6	<.0001

Rhythm appears in the model in two significant interactions. The first one, illustrated in Figure 5.22, shows the combined effects that a speaker's PVI and a participant's MII have on perceived ethnicity. The back panel of this 3D diagram shows the probability of a Pākehā response in the case of participants who have a high MII. These listeners use a speaker's rhythmic characteristics very well as a cue to facilitate dialect identification in this condition. If the speaker is stress-timed, that is, has a higher PVI value, they will be perceived as very Pākehā sounding. If the speaker is more syllable-timed, which is indicated by a lower PVI value, then they will be identified as Māori by these highly integrated participants. However, people with a low MII do not rely much on rhythm to decide whether a speaker is Māori or Pākehā. If anything, they are using it the wrong way, with stress-timed speakers marked as slightly more Māori sounding than syllable-timed speakers.

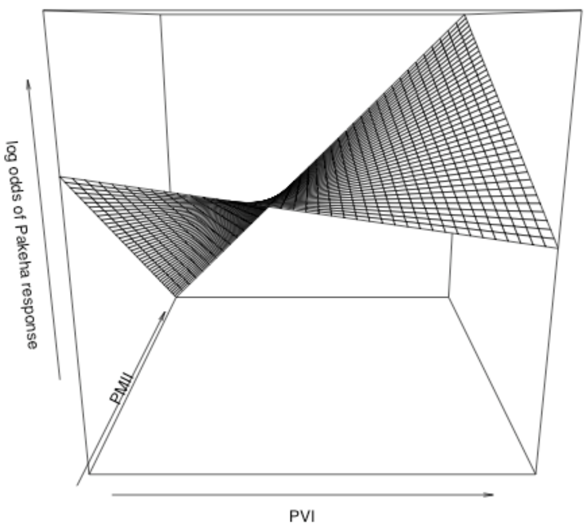
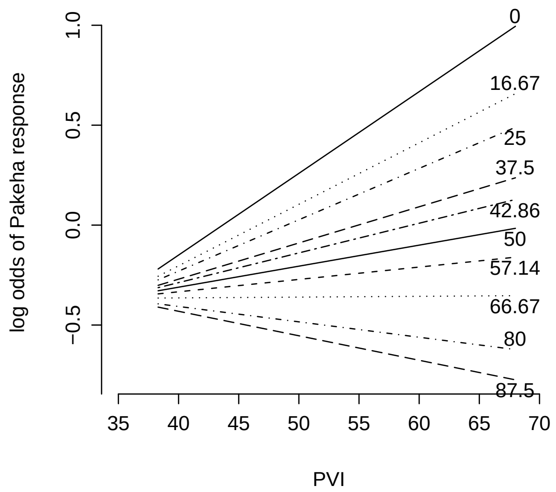


Figure 5.22: Condition Five - Model effects showing log odds of Pākehā response by PVI and participant MII.

The second interaction in which PVI shows up, is the one with the percentage of HRTs used by the speaker. If a speaker is very syllable-timed, they are identifed as Māori. In this case, it does not matter what percentage of HRTs the speakers uses. However, if the speaker is more stress-timed, then the perceived ethnicity is in fact determined by the frequency of HRTs. This operates in the following way. If the stress-timed speaker does not use HRTs, they will be identified as very Pākehā. However, if the stress-timed speaker does use a lot of HRTs, they will be perceived as Māori, despite having a high PVI value. That is, the frequency of HRTs seems to be more important a cue than rhythm itself, *if* the speaker is stress-timed. On the other hand, rhythm is more important than HRTs *if* the speaker is very syllable-timed. The interaction between rhythm and the percentage of HRT is shown in Figure 5.23



.

Figure 5.23: Condition Five - Model effects showing log odds of Pākehā response by PVI and percentage of HRTs used.

Although consonants were replaced by silence in this condition, vowels preserved their original pitch movements. This not only made HRTs easily recognisable but also the standard deviation of pitch, which also had a significant effect on the perceived ethnicity of the speaker. If a speaker's pitch is 'all over the place', that is, they have a high standard deviation of pitch, they are more likely to be perceived as Pākehā. Speakers who have a low standard deviation of pitch, and as such sound more monotonous, are considered to be more Māori sounding. See Figure 5.24.

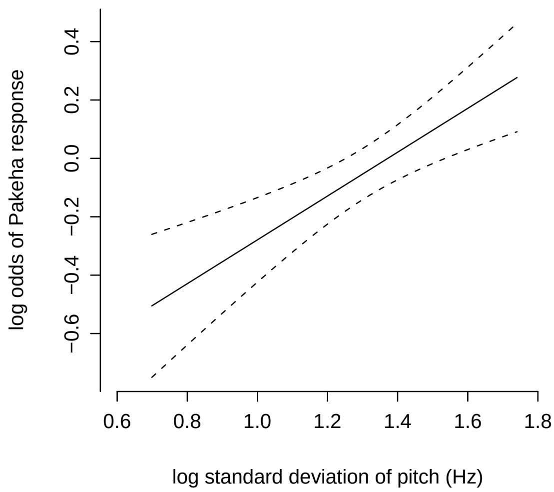


Figure 5.24: Condition Five - Model effects showing log odds of Pākehā response by standard deviation of pitch.

5.3.2 Accuracy - Condition 5

In order to investigate the effects of the accuracy of responses in this rhythm and intonation condition, a second regression model was fit by hand. The following variables had a significant effect in this model:

(a) participant MII
(b) PVI
(c) % of HRTs used interacting with speaker ethnicity
(d) mean pause duration

The coefficient table for this model is given in Table 5.19 and the anova table can be found in Table 5.20.

Table 5.19: Condition Five - Coefficient table for model of accuracy.

Variable	Coef	S.E.	Wald Z	P
Intercept	-2.97	0.78	-3.82	0.0001
PMII	0.02	0.00	3.61	0.0003
% of HRTs	-0.00	0.00	-0.85	0.3969
speaker ethnicity=Pākehā	-0.29	0.25	-1.17	0.2422
PVI	0.05	0.01	3.43	0.0006
mean pause duration	2.97	0.84	3.54	0.0004
% of HRTs * speaker ethnicity=Pākehā	-0.01	0.00	-4.83	0.0000

Table 5.20: Condition Five - Anova table for model of accuracy.

Factor	Chi-Square	d.f.	P
PMII	13.07	1	3e-04
% of HRTs	51.73	2	<.0001
speaker ethnicity	44.91	2	<.0001
PVI	11.79	1	6e-04
mean pause duration	12.54	1	4e-04
% of HRTs * speaker ethnicity	23.30	1	<.0001
TOTAL	67.71	6	<.0001

Participants' MII is a very good indicator of their accuracy at identifying Māori English and Pākehā English from the non-segmental cues of rhythm and intonation together. As Figure 5.25 shows, the higher the participants' MII, the more accurate they are at the task. Notice that the variable of *participant ethnicity* did not make it into our model, suggesting that the MII assigned to each listener is a better indicator of one's accuracy than their ethnicity alone.

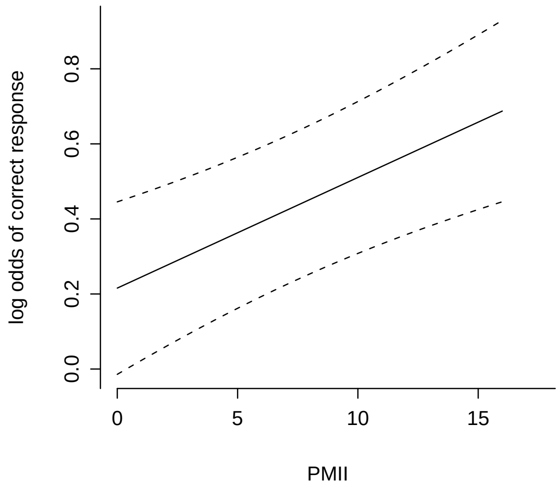


Figure 5.25: Condition Five - Model effects showing log odds of correct response by participant MII.

The timing patterns of a speaker also had a significant effect on whether their ethnicity was correctly identified. In this condition those speakers who had a more stress-timed rhythm were more accurately identified than the speakers with syllable-timed speech. This trend can be seen in Figure 5.26.

Our model of accuracy also shows a significant interaction between the percentage of HRTs used by the speaker and their ethnicity. The percentage of HRTs used by a Māori speaker does not influence participants' accuracy in either direction. However, if a Pākehā speaker uses many HRTs, participants will incorrectly identify them as Māori.

The last variable showing a significant effect on accuracy in this condition was the mean duration of pauses used by the speaker. Listeners were most accurate at identifying the ethnicity of speakers whose pauses were the longest.

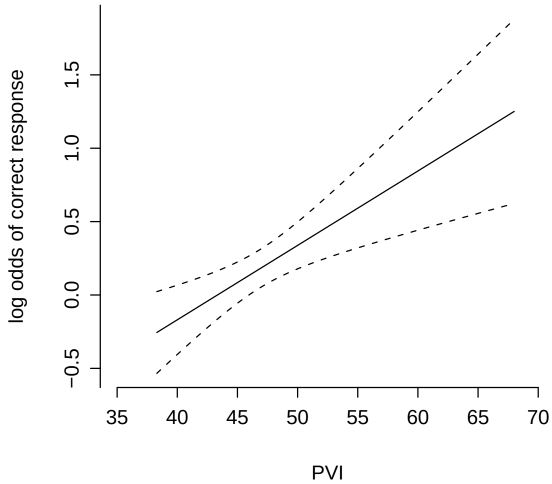


Figure 5.26: Condition Five - Model effects showing log odds of correct response by PVI.

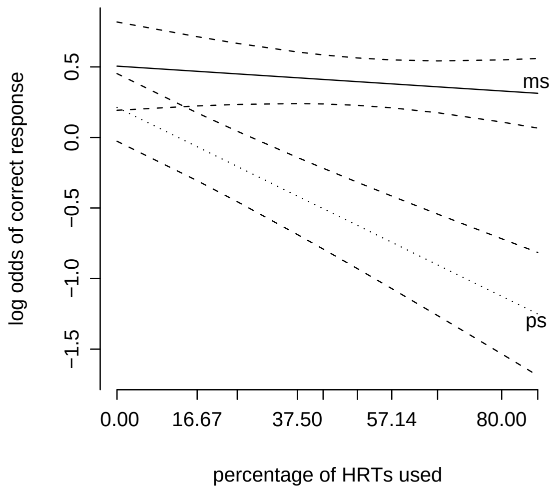


Figure 5.27: Condition Five - Model effects showing log odds of correct response by percentage of HRTs used and speaker ethnicity.

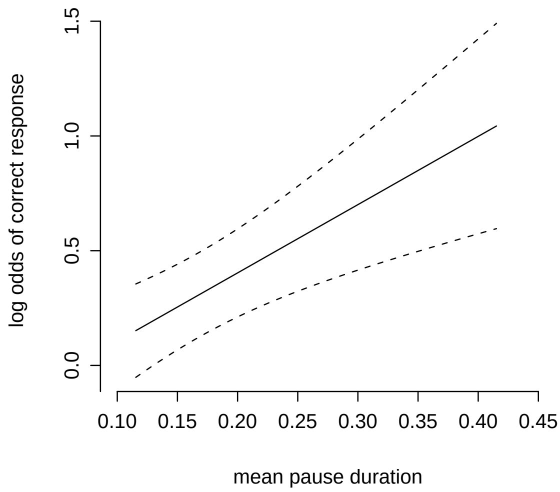


Figure 5.28: Condition Five - Model effects showing log odds of correct response by mean pause duration.

5.3.3 Summary for Condition Five

The tables below give a summary of all the variables that showed a significant effect in our statistical models in Condition Five. The effect of *increasing values* of each variable on the likelihood of Pākehā response is shown in Table 5.21, and Table 5.22 summarises the effect on the accuracy of responses.

Table 5.21: Condition Five - Summary of significant variables in model of perceived ethnicity.

increasing value of variable	likelihood of Pākehā response
PVI	increases for high MII participants
% of HRTs	decreases for stress-timed speakers
SD of pitch	increases

Table 5.22: Condition Five - Summary of significant variables in model of accuracy.

increasing value of variable	likelihood of correct response
participant MII	increases
PVI	increases
% of HRTs	decreases for Pākehā speakers
mean pause duration	increases

5.4 Condition Four - Intonation Only

Condition Four was designed to investigate whether participants could identify Māori English and Pākehā English from intonation only. A hummed version of each utterance was created with the pitch contour linearly interpolated between adjacent points. Original pauses were then replaced by silence, marking the end of each intonation phrase. This way HRTs were retained in the signal and could potentially be used as a cue by participants in the dialect identification task. Section 4.2.4 describes in more detail how this condition was created.

Overall, participants' accuracy was lower in Condition Four than in Condition Five, where rhythm as well as intonation were preserved in the speech signal. Somewhat surprisingly though, Māori participants were more accurate at identifying fellow Māori speakers from intonation alone than they were in Condition Five, where they had both rhythm and intonation to rely on. They achieved a relatively high 64% accuracy rate when identifying Māori speakers (cf. 59% in Condition Five)compared to a low 48% when identifying Pākehā speakers. Pākehā participants performed below chance, with an average 47% accuracy. The percentages of correct responses for this condition are shown in Table 5.23, indicating both participant and speaker ethnicity.

Table 5.23: Condition Four - Percentage of correct responses by participant and speaker ethnicity.

	Māori speakers	Pākehā speakers	Total
Māori participants	64%	48%	56%
Pākehā participants	45%	49%	47%

5.4.1 *Perceived Ethnicity - Condition 4*

As in the previous conditions, the first one of the two regression models we fit investigated the effects of perceived ethnicity. The variables that had a significant effect on participants' perception about a speaker's ethnicity were:

(a) standard deviation of pitch
(b) % of HRTs used interacting with participant ethnicity

The coefficient table for this model is given in Table 5.24 and the anova table can be found in Table 5.25.

Table 5.24: Condition Four - Coefficient table for model of perceived ethnicity.

Variable	Coef	S.E.	Wald Z	P
Intercept	-1.40	0.25	-5.52	0.0000
log SD of pitch	1.13	0.19	5.92	0.0000
participant ethnicity=Pākehā	0.00	0.12	0.01	0.9948
% of HRTs	-0.01	0.00	-6.06	0.0000
participant ethnicity=Pākehā * % of HRTs	0.01	0.00	4.32	0.0000

Table 5.25: Condition Four - Anova table for model of perceived ethnicity.

Factor	Chi-Square	d.f.	P
log SD of pitch	35.02	1	<.0001
participant ethnicity	38.14	2	<.0001
% of HRTs	36.67	2	<.0001
participant ethnicity * % of HRTs	18.67	1	<.0001
TOTAL	78.15	4	<.0001

This condition preserved all pitch dynamics in the speech signal, including the standard deviation of pitch. This had a significant effect on whether a speaker is perceived to be Māori or Pākehā sounding. As can be seen in Figure 5.29, the higher a speaker's standard deviation of pitch, the more likely they are to be identified as Pākehā. On the other hand, more monotonous speakers with a flatter intonational patterns are considered to be Māori sounding. Again, there is no statistical interaction with participant ethnicity, thus the model predicts this trend to be true for both Māori and Pākehā participants.

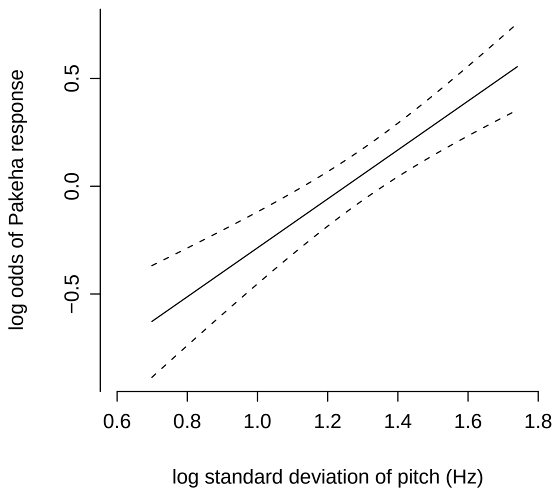


Figure 5.29: Condition Four - Model effects showing log odds of Pākehā response by standard deviation of pitch.

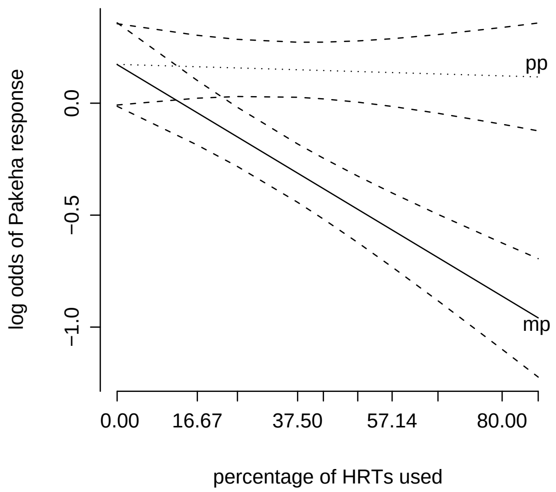


Figure 5.30: Condition Four - Model effects showing log odds of Pākehā response by standard deviation of pitch and participant ethnicity.

Unlike the standard deviation of pitch, HRTs were only used by Māori participants as a cue in this condition. For Māori listeners, the more HRTs a speaker uses, the more likely the speaker is to be identified as Māori sounding. Pākehā participants, on the other hand, do not make use of HRTs as a cue for dialect identification in the intonation alone condition: whether a speaker uses many HRTs or not does not influence Pākehā participants' responses in either direction. This is illustrated in Figure 5.30 by the flat line marking Pākehā responses.

5.4.2 Accuracy - Condition 4

In the second regression model we investigated the effects on the accuracy of participants' responses. Our aim was to find out which variables help listeners correctly identify a speaker's ethnicity and which ones make it more difficult. The following list shows the variables that had a significant effect on accuracy in our model:

(a) participant MII interacting with speaker ethnicity
(b) % of HRTs interacting with speaker ethnicity & participant ethnicity
(c) standard deviation of pitch interacting with speaker ethnicity

The coefficient table for this model is shown in Table 5.26, and the anova table is presented in Table 5.27.

Table 5.26: Condition Four - Coefficient table for model of accuracy.

Variable	Coef	S.E.	Wald Z	P
Intercept	0.63	0.58	1.08	0.2784
PMII	0.09	0.01	4.83	0.0000
speaker ethnicity=Pākehā	-1.91	0.69	-2.76	0.0058
log SD of pitch	-1.07	0.44	-2.42	0.0156
% of HRTs	0.00	0.00	1.89	0.0591
participant ethnicity=Pākehā	0.15	0.26	0.57	0.5667
PMII * speaker ethnicity=Pākehā	-0.08	0.02	-3.00	0.0027
speaker ethnicity=Pākehā * log SD of pitch	2.19	0.49	4.45	0.0000
speaker ethnicity=Pākehā * % of HRTs	-0.02	0.00	-5.01	0.0000
% of HRTs * participant ethnicity=Pākehā	-0.00	0.00	-0.89	0.3756
speaker ethn=Pākehā * participant ethn=P	-0.44	0.35	-1.26	0.2070
% of HRTs * speaker ethn=P * part ethn=P	0.02	0.00	3.38	0.0007

Table 5.27: Condition Four - Anova table for model of accuracy.

Factor	Chi-Square	d.f.	P
PMII	23.69	2	<.0001
speaker ethnicity	75.96	6	<.0001
log SD of pitch	32.37	2	<.0001
% of HRTs	32.26	4	<.0001
participant ethnicity	15.87	4	0.0032
PMII * speaker ethnicity	8.98	1	0.0027
speaker ethnicity * log SD of pitch	19.79	1	<.0001
speaker ethnicity * % of HRTs	25.14	2	<.0001
% of HRTs * participant ethnicity	15.64	2	0.0004
speaker ethnicity * % of HRTs * participant ethnicity	11.43	1	0.0007
TOTAL INTERACTION	65.70	6	<.0001
TOTAL	121.02	11	<.0001

The percentages of correct responses in Table 5.23 indicated that Māori participants were much more accurate at identifying Māori speakers from intonation alone than were Pākehā participants. As our logistic regression model of accuracy reveals, however, the difference is not between ethnically Māori and Pākehā listeners but between people with a high and low MII. Figure 5.31 shows the interaction between participant MII and speaker ethnicity. The more integrated to Māori society the listener is, the more accurate they are at identifying Māori speakers from intonation alone. However, having a high MII does not seem to facilitate one's ability to correctly identify Pākehā speakers in this condition, as the flat line relating to Pākehā speakers in the graph indicates.

Speaker ethnicity also features in another significant interaction in our model of accuracy. This three way interaction between percentage of HRTs used, speaker ethnicity and participant ethnicity is illustrated in Figure 5.32. Coefficients are set to Māori participants in the left panel and to Pākehā participants in the right panel. Pākehā participants' accuracy is not influenced by what percentage of HRTs the speaker uses in this condition. However, Māori participants incorrectly identify those Pākehā speakers who use a lot of HRTs as Māori. This inaccuracy stems from the fact that Māori participants perceive a high percentage of HRTs as very Māori sounding, as our model of perceived ethnicity for this condition has shown.

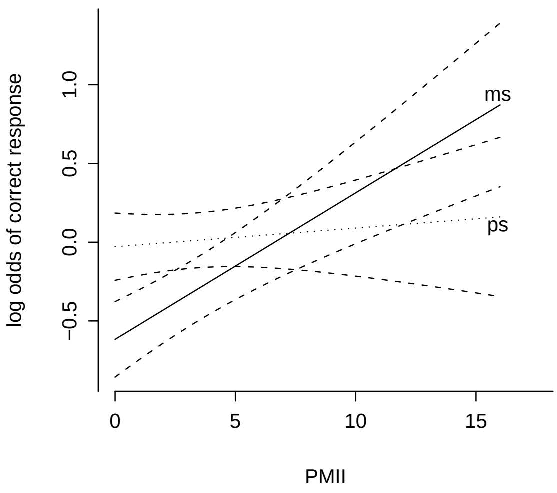


Figure 5.31: Condition Four - Model effects showing log odds of correct response by participant MII and speaker ethnicity.

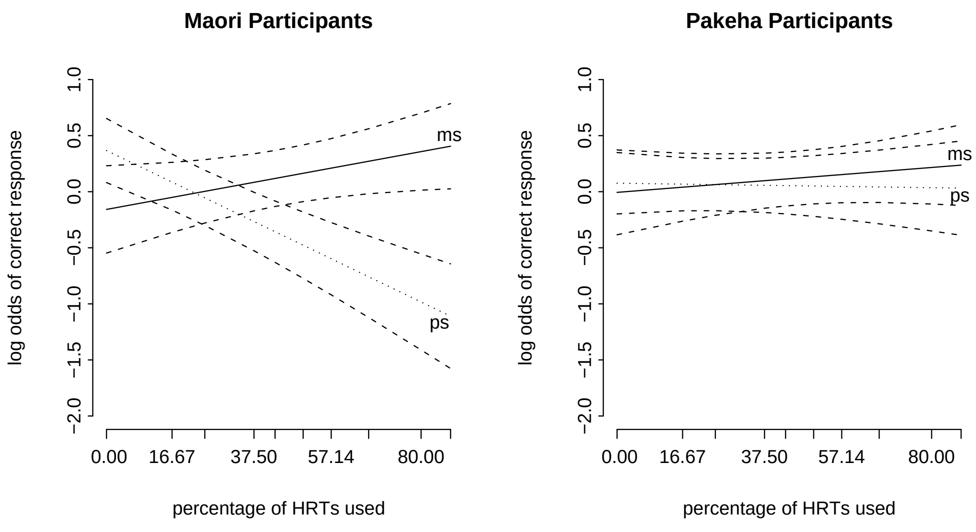


Figure 5.32: Condition Four - Model effects showing log odds of correct response by percentage of HRTs used and speaker ethnicity. Coefficients set to Māori participants in left panel and to Pākehā participants in right panel.

The last significant interaction in this model once again includes the variable of speaker ethnicity, this time interacting with the standard deviation of pitch used by the speaker. If a Pākehā speaker has a high standard deviation of pitch, they are correctly identified as Pākehā. Māori speakers, on the other hand, are expected to have a more monotonous intonational pattern. Thus, those Māori speakers who do have a high standard deviation of pitch are incorrectly identified in this condition.

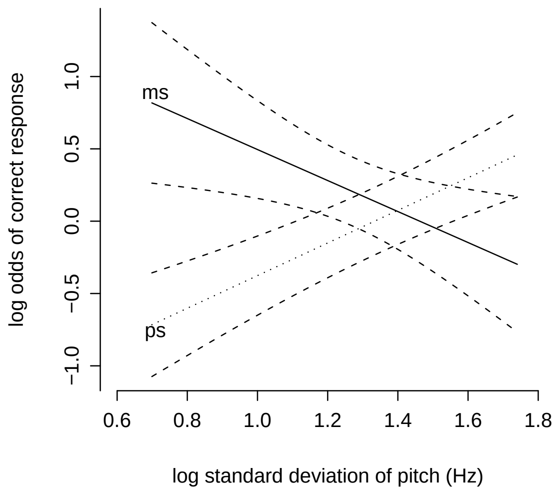


Figure 5.33: Condition Four - Model effects showing log odds of correct response by standard deviation of pitch and speaker ethnicity.

5.4.3 Summary for Condition Four

The tables below give a summary of all the variables that showed a significant effect in our statistical models in Condition Four. The effect of *increasing values* of each variable on the likelihood of Pākehā response is shown in Table 5.28, and Table 5.29 summarises the effect on the accuracy of responses.

Table 5.28: Condition Four - Summary of significant variables in model of perceived ethnicity.

increasing value of variable	likelihood of Pākehā response
% of HRTs	decreases for Māori participants
SD of pitch	increases

Table 5.29: Condition Four - Summary of significant variables in model of accuracy.

increasing value of variable	likelihood of correct response
participant MII	increases for Māori speakers
% of HRTs	only affects Māori participants:
	increases for Māori speakers
	decreases for Pākehā speakers
SD of pitch	increases for Pākehā speakers
	decreases for Māori speakers

5.5 *Condition Three - Non-duration-normalised Rhythm Only at Own Pitch*

Condition Three was created to examine whether listeners could distinguish the two ethnic varieties if the speech signal contained syllabic rhythm but no intonation. Consonants were replaced by silence, while the pitch dynamics of vowels were flattened at the speaker's own mean pitch. As this condition was not duration normalised, speech rate is also available as a cue for ethnic identification in addition to rhythm and mean pitch. See Section 4.2.3 for a more detailed description of how this condition was created.

In this condition the overall accuracy rate was an equal 51% for both Māori and Pākehā participants. However, both groups performed better at identifying their own dialect, with Māori listeners reaching 53% when identifying other Māori speakers and Pākehā participants averaging 52% at marking Pākehā speakers. For the percentages of correct responses shown by both speaker and participant ethnicity see Table 5.30.

Table 5.30: Condition Three - Percentage of correct responses by participant and speaker ethnicity.

	Māori speakers	Pākehā speakers	Total
Māori participants	53%	49%	51%
Pākehā participants	49%	52%	51%

5.5.1 Perceived Ethnicity - Condition 3

In the first of the two logistic regression models, we investigated the effects of the perceived ethnicity, that is whether participants identified a speaker as Māori or Pākehā. The variables that showed significant effects are listed below:

(a) PVI interacting with participant gender
(b) mean pitch interacting with participant ethnicity
(c) speech rate

The coefficient table for this model is presented in Table 5.31, while the corresponding anova table is shown in Table 5.32

The regression model indicates that rhythm is used by listeners in this condition to facilitate dialect identification. However, females and males are attending to the rhythmic properties of the speaker quite differently. Female participants tend to use PVI in the correct way, marking stress-timed speakers as more Pākehā sounding and syllable-timed speakers as more Māori sounding. Male participants, on the other hand, use PVI in the wrong direction. For them, a low PVI indicates a Pākehā speaker while a a high PVI is a sign for a Māori speaker. This interaction between PVI and participant gender is illustrated in Figure 5.34.

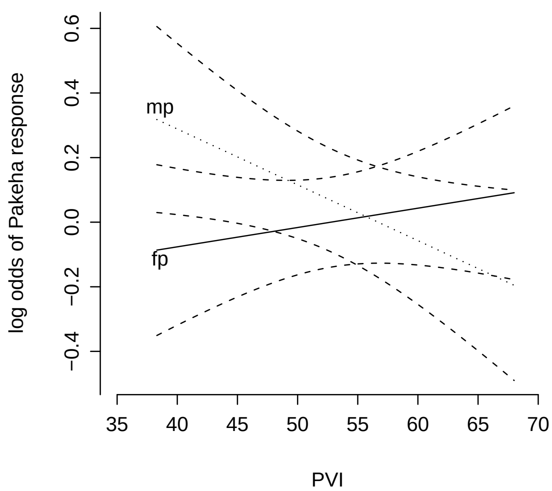


Figure 5.34: Condition Three - Model effects showing log odds of Pākehā response by PVI and participant gender.

Table 5.31: Condition Three - Coefficient table for model of perceived ethnicity.

Variable	Coef	S.E.	Wald Z	P
Intercept	-1.77	0.61	-2.88	0.0039
mean pitch	0.00	0.00	0.26	0.7912
participant ethnicity=Pākehā	-0.61	0.33	-1.84	0.0665
speech rate	0.29	0.08	3.28	0.0011
PVI	0.00	0.00	0.77	0.4441
participant gender=male	1.29	0.60	2.13	0.0332
mean pitch * participant ethnicity=P	0.00	0.00	2.32	0.0202
PVI * participant gender=male	-0.02	0.01	-2.04	0.0411

Table 5.32: Condition Three - Anova table for model of perceived ethnicity.

Factor	Chi-Square	d.f.	P
mean pitch	12.74	2	0.0017
participant ethnicity	7.72	2	0.0210
speech rate	10.74	1	0.0011
PVI	4.88	2	0.0872
participant gender	4.74	2	0.0936
mean pitch * participant ethnicity	5.39	1	0.0202
PVI * participant gender	4.17	1	0.0411
TOTAL INTERACTION	9.37	2	0.0092
TOTAL	30.22	7	0.0001

A somewhat more significant indicator of perceived ethnicity in this condition is the mean pitch of the speaker. It is a particularly good indicator of perceived ethnicity by Pākehā participants. They identify a speaker with a low mean pitch as more Māori sounding, while a high mean pitch suggests that the speaker is Pākehā. At the same time, Māori listeners do not make much use of the speaker's mean pitch as a cue for ethnic identification. Whether a speaker has a high or low mean pitch does not influence their responses in either direction. Figure 5.35 shows the interaction between a speaker's mean pitch and the participant's ethnicity.

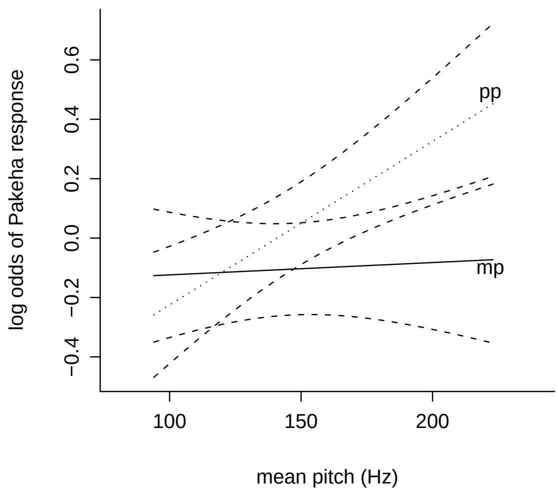


Figure 5.35: Condition Three - Model effects showing log odds of Pākehā response by mean pitch and participant ethnicity.

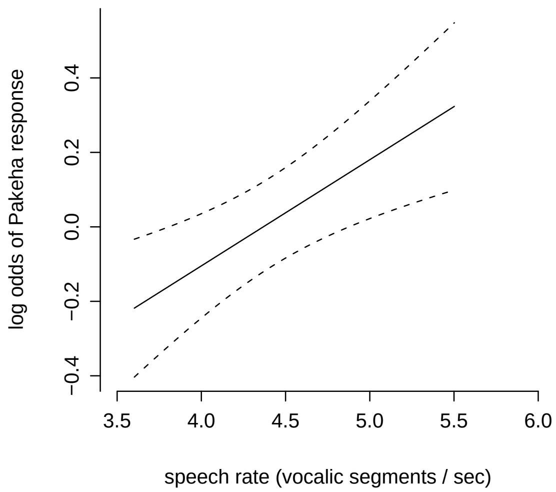


Figure 5.36: Condition Three - Model effects showing log odds of Pākehā response by speech rate.

The most significant predictor of perceived ethnicity in this condition is speech rate. Whether the speech is fast or slow has an effect on listeners' perception of the speakers ethnicity. All participants agree that faster speakers are more Pākehā sounding, while a lower speech rate suggests that the speaker is Māori. This is shown in Figure 5.36.

5.5.2 *Accuracy - Condition 3*

The effects of the participants' accuracy in this condition were investigated in our second regression model. The following variables reached statistical significance:

(a) participant MII interacting with speaker ethnicity
(b) participant gender
(c) mean pitch interacting with speaker ethnicity

The coefficient table for this model is presented in Table 5.33, while the corresponding anova table is shown in Table 5.34

Table 5.33: Condition Three - Coefficient table for model of accuracy.

Variable	Coef	S.E.	Wald Z	P
Intercept	0.48	0.28	1.67	0.0946
PMII	0.04	0.01	3.81	0.0001
speaker ethnicity=Pākehā	-0.68	0.36	-1.87	0.0612
participant gender=male	-0.17	0.08	-1.99	0.0471
mean pitch	-0.00	0.00	-2.56	0.0104
PMII * speaker ethnicity=Pākehā	-0.05	0.01	-3.18	0.0015
speaker ethnicity=Pākehā * mean pitch	0.00	0.00	3.09	0.0020

Table 5.34: Condition Three - Anova table for model of accuracy.

Factor	Chi-Square	d.f.	P
PMII	14.98	2	0.0006
speaker ethnicity	19.56	3	0.0002
participant gender	3.94	1	0.0471
mean pitch	9.63	2	0.0081
PMII * speaker ethnicity	10.12	1	0.0015
mean pitch * speaker ethnicity	9.52	1	0.0020
TOTAL INTERACTION	19.43	2	0.0001
TOTAL	27.24	6	0.0001

As in all other conditions, the model of accuracy again demonstrates that a participant's level of integration into Māoridom is a highly significant predictor of their accuracy at distinguishing between Māori English and Pākehā English. In particular, a higher integration index facilitates correct identification of Māori speakers in the present condition. The interaction between participants' MII and the speaker's ethnicity is illustrated in Figure 5.37.

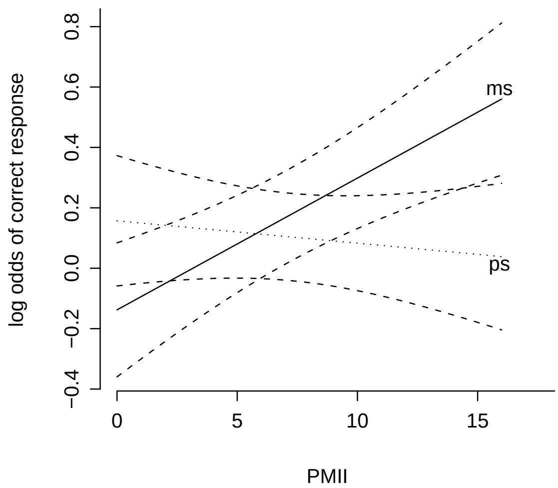


Figure 5.37: Condition Three - Model effects showing log odds of correct response by participant MII and speaker ethnicity.

Regardless of integration index, females performed better at identifying the ethnicity of the speaker in this condition, where only syllabic rhythm, mean pitch and speech rate were available cues. The higher accuracy rate of female participants is shown in Figure 5.38.

As mentioned above, the mean pitch of a speaker played a significant role in participants' perception of the speaker's dialect. Interacting with the variable of speaker ethnicity, it also has a significant effect in our model of accuracy. Figure 5.39 illustrates that Māori speakers with a high mean pitch are incorrectly identified, while the ones that do have a low mean pitch are correctly marked as Māori. The opposite trend emerges for Pākehā speakers. If they have a high mean pitch, they are accurately identified as Pākehā. However, a low mean pitch results in inaccurate responses, because it is perceived to be a characteristic of Māori English.

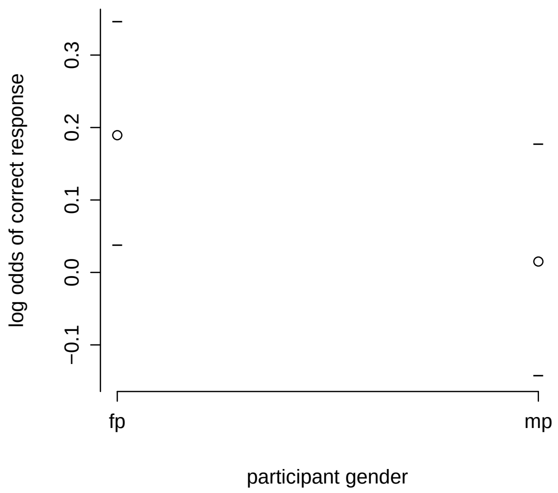


Figure 5.38: Condition Three - Model effects showing log odds of correct response by participant gender.

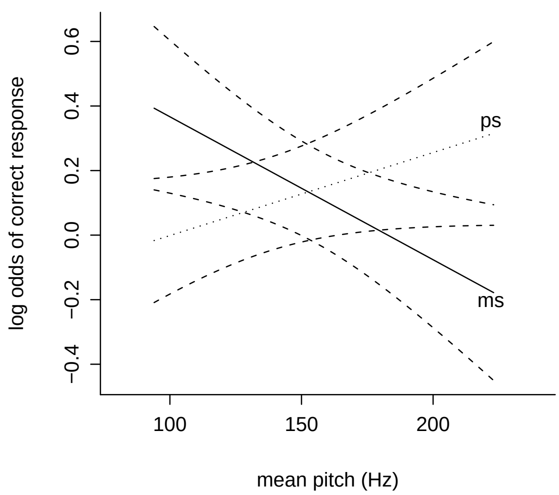


Figure 5.39: Condition Three - Model effects showing log odds of correct response by mean pitch and speaker ethnicity.

5.5.3 *Summary for Condition Three*

The tables below give a summary of all the variables that showed a significant effect in our statistical models in Condition Three. The effect of *increasing values* of each variable on the likelihood of Pākehā response is shown in Table 5.35, and Table 5.36 summarises the effect on the accuracy of responses[1].

Table 5.35: Condition Three - Summary of significant variables in model of perceived ethnicity.

increasing value of variable	likelihood of Pākehā response
PVI	increases for female participants
	decreases for male participants
mean pitch	increases for Pākehā participants
speech rate	increases

Table 5.36: Condition Three - Summary of significant variables in model of accuracy.

increasing value of variable	likelihood of correct response
participant MII	increases for Māori speakers
mean pitch	increases for Pākehā speakers
	decreases for Māori speakers
participant gender	higher for female participants

5.6 *Condition Two - Non-duration-normalised Rhythm Only at Mean Pitch*

Condition Two is essentially the same as Condition Three, except this time the speaker's own mean pitch is not retained in the speech signal. Instead, the sine wave complex replacing vowels were created at the mean pitch across all speakers according to gender. This was 118Hz for male speakers and 188Hz for female speakers. Consult Section 4.2.2 for a more detailed description about the creation of this condition.

[1] The variable of participant gender is of course not a continuous but a binary variable.

Overall, Māori participants performed at a higher level of accuracy than Pākehā participants. Māori listeners correctly identified fellow Māori speakers 56% of the time,and Pākehā speakers 49% of the time. Pākehā participants only achieved a 48% accuracy rate in total. Table 5.37 contains all percentages of correct responses shown by both participant and speaker ethnicity.

Table 5.37: Condition Two - Percentage of correct responses by participant and speaker ethnicity.

	Māori speakers	Pākehā speakers	Total
Māori participants	56%	49%	53%
Pākehā participants	49%	48%	48%

5.6.1 Perceived Ethnicity - Condition 2

The first of the two logistic models we fit in this condition examined the effects of participants' perception of a speaker's ethnicity. The following list shows the variables that had a significant effect on whether a speaker is identified as Māori or Pākehā in Condition Two.

(a) PVI interacting with participant MII
(b) speaker gender interacting with participant MII
(c) speech rate
(d) number of pauses interacting with participant ethnicity

Unlike in Condition Three, the mean pitch of the speaker does not seem to be an important factor in this condition.

The coefficient table for this model is given in Table 5.38 and the anova table can be seen in Table 5.39.

As Condition Two did not use duration normalisation, speech rate was kept in the signal. The regression model shows that speech rate is again a good predictor of perceived ethnicity. There is no statistical interaction with participant ethnicity, thus the model predicts that all participants identify fast speakers as more Pākehā sounding and slow speakers as more Māori sounding. This significant trend is shown in Figure 5.40.

Table 5.38: Condition Two - Coefficient table for model of perceived ethnicity.

Variable	Coef	S.E.	Wald Z	P
Intercept	-1.58	0.76	-2.09	0.0368
PVI	-0.03	0.01	-3.88	0.0001
PMII	-0.17	0.06	-2.88	0.0039
speech rate	0.74	0.10	7.33	0.0000
number of pauses	0.11	0.04	2.91	0.0036
participant ethnicity=Pākehā	0.44	0.26	1.69	0.0909
speaker gender=male	0.01	0.15	0.09	0.9245
PVI * PMII	0.00	0.00	3.18	0.0015
number of pauses * part ethn=Pākehā	-0.12	0.05	-2.29	0.0222
PMII * speaker gender=male	-0.05	0.01	-3.08	0.0020

Table 5.39: Condition Two - Anova table for model of perceived ethnicity.

Factor	Chi-Square	d.f.	P
PVI	15.05	2	0.0005
PMII	22.95	3	<.0001
speech rate	53.67	1	<.0001
number of pauses	8.63	2	0.0134
participant ethnicity	5.39	2	0.0677
speaker gender	26.12	2	<.0001
PVI * PMII	10.09	1	0.0015
number of pauses * participant ethnicity	5.23	1	0.0222
PMII * speaker gender	9.51	1	0.0020
TOTAL INTERACTION	23.44	3	<.0001
TOTAL	104.36	9	<.0001

The logistic regression analysis of perceived ethnicity also reveals that some listeners were able to attend to a speaker's rhythmic properties in Condition Two. Rhythm interacting with the participant's MII shows up in our model as a significant predictor of whether participants identify a speaker as Māori or Pākehā. Figure 5.41 illustrates this interaction. The predicted responses of highly integrated listeners are shown on the back panel of the 3D diagram. They use rhythm in the correct way, marking speakers with a high PVI - that is stress-timing - as more Pākehā sounding, and speakers with a low PVI - that is syllable-timing - as very Māori sounding. However, those participants who scored low on the MII use rhythm in the opposite way. Their responses are shown on the front panel of the diagram. For them, a low PVI suggests that the speaker is Pākehā, while a high PVI indicates that the speaker is more Māori sounding. It is plausible that participants with a low integration index are not actually tuning in to the speaker's rhythmic properties but instead they are relying on other cues, such as speech rate. Recall that speech rate showed a significant correlation with PVI in the production experiment, as described in Section 3.2.1. Pākehā participants' rhythm becomes significantly more syllable timed the faster they speak. This correlation was illustrated in Figure 3.7. As mentioned above, a faster speech rate triggers a more Pākehā response in this condition. Thus, it is possible that participants with a low MII will identify some speakers with a low PVI as more Pākehā sounding simply because of their faster speech rate, ignoring - or not being able to rely on - the actual rhythmic characteristics of the speaker.

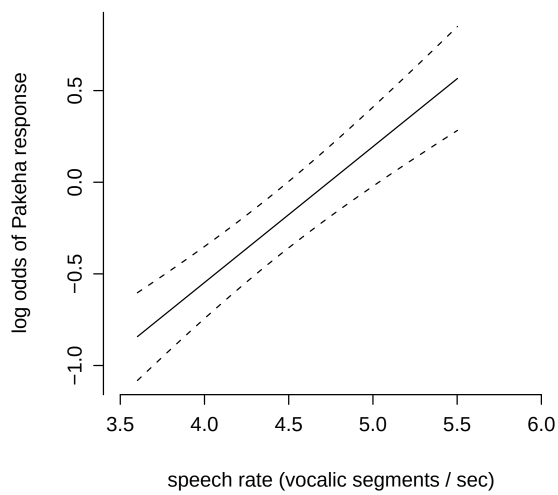


Figure 5.40: Condition Two - Model effects showing log odds of Pākehā response by speech rate.

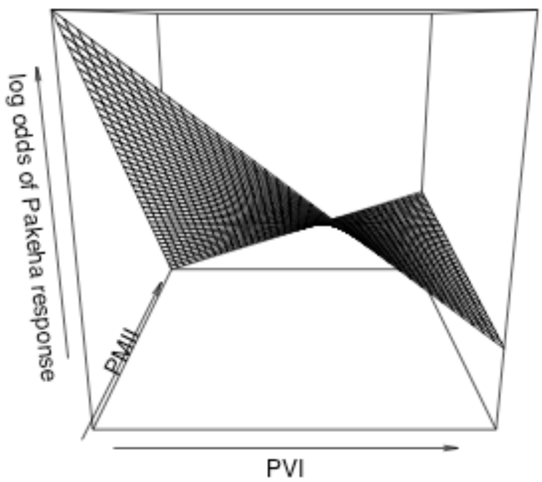


Figure 5.41: Condition Two - Model effects showing log odds of Pākehā response by PVI and participant MII.

Participant MII also features in another interaction in our statistical model. Highly integrated participants seem to make use of speaker gender as a cue in distinguishing the two dialects, whereby male speakers are perceived to be more Māori sounding than female speakers. This speaker gender effect is not present for lowly integrated participants. Recall that in this condition two different fundamental frequency values were used for male and female speakers. The sine waves replacing male vowels were created at 118Hz, and those replacing female vowels at 188Hz. It is possible that the highly integrated participants were tuning in to this pitch difference, actually perceiving a low pitch as more Māori sounding than a high pitch. This could explain why males are identified as more Māori sounding in this condition. The interaction between participant MII and speaker gender is demonstrated in Figure 5.42.

The number of pauses used by a speaker also features in our statistical model, in interaction with participant ethnicity. Pākehā participants' perception of a speaker's ethnicity is not influenced by the number of pauses used. However, Māori participants perceive a low number of pauses as more Māori sounding and a high number of pauses as more Pākehā sounding in this condition. See Figure 5.43 for the interaction between the number of pauses used by a speaker and participant ethnicity.

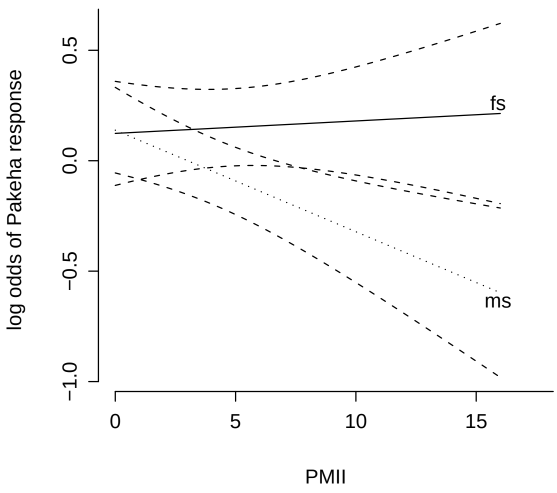


Figure 5.42: Condition Two - Model effects showing log odds of Pākehā response by participant MII and speaker gender.

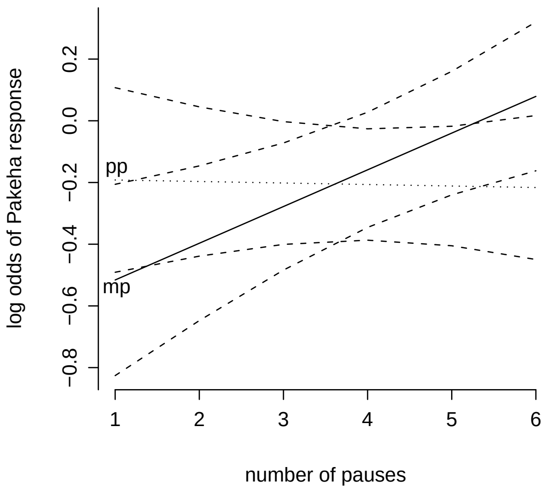


Figure 5.43: Condition Two - Model effects showing log odds of Pākehā response by number of pauses and participant ethnicity.

5.6.2 Accuracy - Condition 2

In the second regression model for Condition Two we investigated the effects on the accuracy of participants' responses to examine which variables, if any, help listeners correctly identify a speaker's ethnicity. The following is a list of the variables that had a significant effect on accuracy in our model:

(a) participant MII interacting with speaker ethnicity
(b) speech rate interacting with participant gender
(c) number of pauses interacting with speaker gender
(d) speaker age

The coefficient table for this model is shown in Table 5.40, and the anova table is presented in Table 5.41.

Table 5.40: Condition Two - Coefficient table for model of accuracy.

Variable	Coef	S.E.	Wald Z	P
Intercept	-0.00	0.70	-0.01	0.9922
speaker ethnicity=Pākehā	0.08	0.15	0.55	0.5791
PMII	0.04	0.01	3.86	0.0001
speaker age	-0.11	0.03	-3.63	0.0003
number of pauses	-0.08	0.04	-2.05	0.0404
speaker gender=male	-0.93	0.25	-3.68	0.0002
speech rate	0.14	0.13	1.08	0.2810
participant gender=male	1.75	0.79	2.21	0.0272
speaker ethnicity=Pākehā * PMII	-0.03	0.01	-2.07	0.0387
number of pauses * speaker gender=male	0.17	0.05	3.09	0.0020
speech rate * participant gender=male	-0.42	0.17	-2.40	0.0163

Again the regression analysis of accuracy in Condition Two shows the effect of highly integrated listeners being able to perform better at the dialect identification task. Our model shows a significant effect of participant MII interacting with speaker ethnicity, which is illustrated in Figure 5.44. A high MII facilitates correct identification of both dialects, but in particular, it helps Māori speakers be identified more accurately.

Table 5.41: Condition Two - Anova table for model of accuracy.

Factor	Chi-Square	d.f.	P
speaker ethnicity	7.23	2	0.0270
PMII	15.75	2	0.0004
speaker age	13.16	1	0.0003
number of pauses	9.54	2	0.0085
speaker gender	14.38	2	0.0008
speech rate	6.03	2	0.0490
participant gender	8.37	2	0.0153
speaker ethnicity * PMII	4.27	1	0.0387
number of pauses * speaker gender	9.53	1	0.0020
speech rate * participant gender	5.77	1	0.0163
TOTAL INTERACTION	19.76	3	0.0002
TOTAL	52.27	10	<.0001

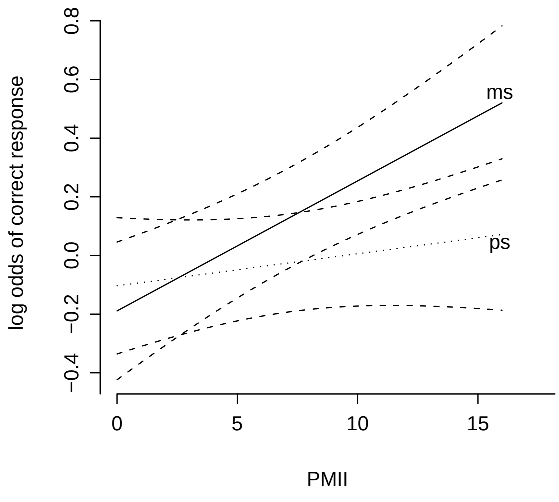


Figure 5.44: Condition Two - Model effects showing log odds of correct response by participant MII and speaker ethnicity.

Speech rate played a role in our model of perceived ethnicity above. Interacting with participant gender it is also a significant predictor of listeners' accuracy at the task. The faster the speaker, the more likely they will be incorrectly identified by male participants. In contrast, the accuracy of responses by females is less influenced by the speech rate of the speaker. If anything, female accuracy increases with the speaker's speech rate. This gender difference in accuracy is not present for slow speakers. See Figure 5.45 for the interaction between speech rate and participant gender.

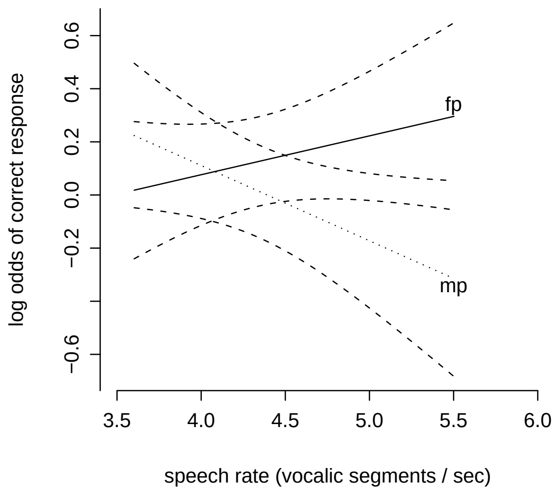


Figure 5.45: Condition Two - Model effects showing log odds of correct response by speech rate and participant gender.

Another significant interaction is between the number of pauses used and speaker gender. Male speakers are more likely to be misidentified if they have a low number of pauses. In contrast, the ethnicity of females is most accurately identified when they use few pauses in their speech. The responses become less accurate as the number of female pauses increases. This interaction between the number of pauses used by the speaker and his or her ethnicity is shown in Figure 5.46.

The final variable that showed a significant effect on accuracy in Condition Two is speaker age. Participants were most accurate when it came to identifying young speakers. As Figure 5.47 illustrates, the estimated accuracy rate drops as speaker age increases. Note that whether a speaker belonged to the *older* or *younger* age group was indicated on the participant's answer sheet (see Appendix F). It is also worth noting that of all 14 regression models fit for the perception experiment -

including those in Condition One, not yet discussed -, this is the only one where speaker age shows significance.

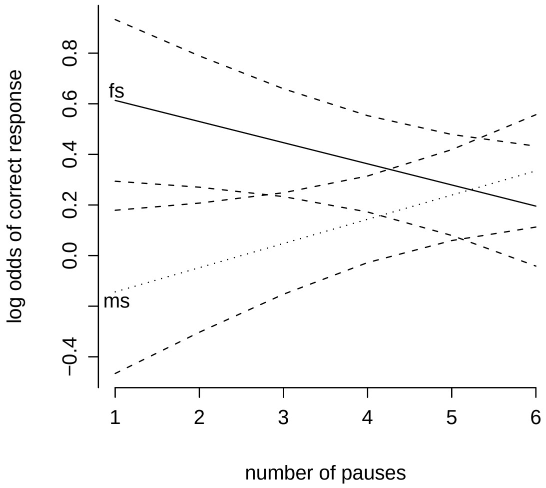


Figure 5.46: Condition Two - Model effects showing log odds of correct response by number of pauses and speaker gender.

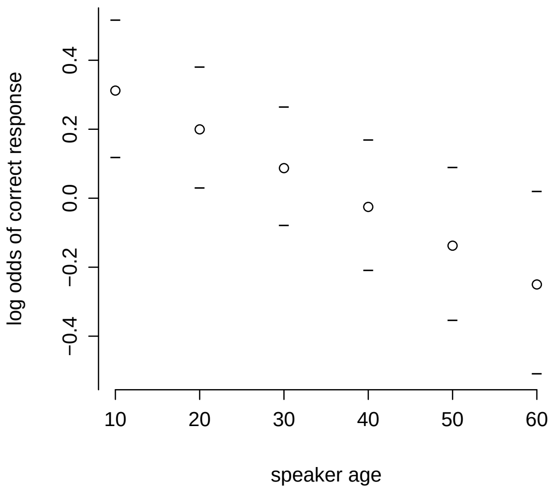


Figure 5.47: Condition Two - Model effects showing log odds of correct response by speaker age.

5.6.3 *Summary for Condition Two*

The tables below give a summary of all the variables that showed a significant effect in our statistical models in Condition Two. The effect of *increasing values* of each variable on the likelihood of Pākehā response is shown in Table 5.42, and Table 5.43 summarises the effect on the accuracy of responses.

Table 5.42: Condition Two - Summary of significant variables in model of perceived ethnicity.

increasing value of variable	likelihood of Pākehā response
PVI	increases for high MII participants
	decreases for low MII participants
participant MII	decreases for male speakers
number of pauses	increases for Māori participants
speech rate	increases

Table 5.43: Condition Two - Summary of significant variables in model of accuracy.

increasing value of variable	likelihood of correct response
participant MII	increases for Māori speakers
speech rate	increases for female participants
	decreases for male participants
number of pauses	increases for female speakers
	decreases for male speakers
speaker age	decreases

5.7 *Condition One - Duration-normalised Rhythm Only at Mean Pitch*

The previous condition examined whether naive listeners could distinguish between Māori English and Pākehā English if intonation and the speaker's own mean pitch are both eliminated from the speech signal, thus only preserving syllabic rhythm. However, apart from rhythm, Condition Two also retained the speech rate of the speakers. Condition One was therefore created to investigate how participants would fair at the dialect identification task if they did not have speech rate as a cue to rely on. To eliminate the difference between fast speakers and slow speakers, a duration-normalisation procedure was carried out. The duration of vowels, consonants and pauses independently underwent their own time-normalisation process. Section 4.2.1 explains how this procedure was applied. Again, consonants were replaced by silence, while vowels were replaced by sine waves created at the mean pitch across all speakers according to gender (118Hz for men, 188Hz for women). Conditions One to Three all created the impression of listening to a machine gun or morse-code, something very unlike natural human speech.

In Condition One participants averaged an accuracy rate just below 50%. However, Māori participants managed to correctly identify other Māori speakers 54% of the time. Table 5.44 below presents the percentage of correct responses shown by speaker ethnicity and participant ethnicity.

Table 5.44: Condition One - Percentage of correct responses by participant and speaker ethnicity.

	Māori speakers	Pākehā speakers	Total
Māori participants	54%	44%	49%
Pākehā participants	49%	47%	48%

5.7.1 Perceived Ethnicity - Condition 1

The first of the two logistic models we fit in this condition examined the effects of participants' perception of a speaker's ethnicity. The following list shows the variables that had a significant effect on whether a speaker is identified as Māori or Pākehā when only syllabic rhythm is kept in the signal.

(a) PVI interacting with the % of HRTs
(b) mean pause duration interacting with participant MII
(c) number of pauses interacting with participant ethnicity
(d) speaker gender interacting with participant gender and ethnicity

The coefficient table for this model is given in Table 5.45 and the anova table can be seen in Table 5.46.

Table 5.45: Condition One - Coefficient table for model of perceived ethnicity.

Variable	Coef	S.E.	Wald Z	P
Intercept	-0.71	0.65	-1.09	0.2742
PMII	-0.06	0.02	-2.47	0.0136
mean pause duration	-1.32	1.08	-1.23	0.2200
PVI	0.01	0.00	1.96	0.0504
% of HRTs	0.03	0.01	2.60	0.0092
number of pauses	0.01	0.04	0.29	0.7691
participant ethnicity=Pākehā	-0.50	0.31	-1.62	0.1062
speaker gender=male	-0.06	0.19	-0.34	0.7370
participant gender=male	-0.16	0.19	-0.85	0.3955
PMII * mean pause duration	0.28	0.10	2.72	0.0065
PVI * % of HRTs	-0.00	0.00	-3.10	0.0019
number of pauses * participant ethnicity=Pākehā	0.10	0.05	1.96	0.0498
speaker gender=male * participant gender=male	0.28	0.25	1.09	0.2757
participant ethn=Pākehā * speaker gender=male	0.32	0.24	1.31	0.1891
participant ethn=Pākehā * participant gender=male	0.89	0.28	3.15	0.0016
part ethn=Pākehā * spk gend=male * part gend=male	-1.41	0.36	-3.88	0.0001

The logistic regression analysis of perceived ethnicity in Condition One shows an unexpected result. The percentage of HRTs used by the speaker turns out to be one of the best predictors of whether participants identify the speaker as Māori or Pākehā. As pitch dynamics were completely eliminated from the speech signal by a monotone in this condition, this is surprising. Nevertheless, the percentage of HRTs used interacting with the speaker's rhythmic properties has a significant effect in our model of perceived ethnicity. The interaction between PVI and the percentage of HRTs is illustrated in Figure 5.48[1]. This diagram shows that in the case of stress-timed speakers, participants use the percentage of HRTs as a cue to decide whether the speaker sounds Māori or Pākehā. The more HRTs are used, the more Māori sounding the stress-timed speaker is. This effect is not present for the more syllable-timed speakers.

[1] This diagram is similar to the one shown in Condition Five (Figure 5.23), except that one does not have the cross-over effect.

Table 5.46: Condition One - Anova table for model of perceived ethnicity.

Factor	Chi-Square	d.f	P
PMII	7.46	2	0.0240
mean pause duration	8.61	2	0.0135
PVI	9.82	2	0.0074
% of HRTs	24.45	2	<.0001
number of pauses	8.38	2	0.0152
participant ethnicity	22.34	5	0.0005
speaker gender	22.81	4	0.0001
participant gender	20.59	4	0.0004
PMII * mean pause duration	7.42	1	0.0065
PVI * % of HRTs	9.62	1	0.0019
number of pauses * participant ethnicity	3.85	1	0.0498
part ethnicity * speaker gender * part gender	15.04	1	0.0001
TOTAL INTERACTION	42.10	7	<.0001
TOTAL	66.75	15	<.0001

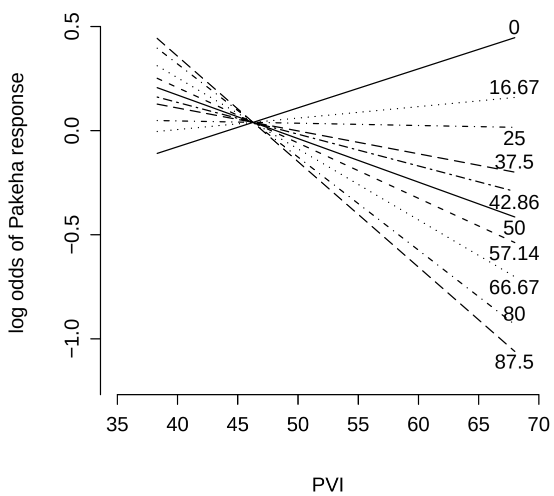


Figure 5.48: Condition One - Model effects showing log odds of Pākehā response by PVI and percentage of HRTs used.

As the difference between fast and slow speakers was diminished in Condition One, participants had to tune into cues other than speech rate, and pauses in the speech signal prove to be useful cues. The second of the four significant interactions in our model of perceived ethnicity is between the number of pauses used by the speaker and the ethnicity of the participant, which is shown in Figure 5.49. Māori participants' responses are not influenced by the number of pauses used by the speaker, but Pākehā participants are more likely to identify the speaker as Māori if the speaker does not pause very often during the passage.

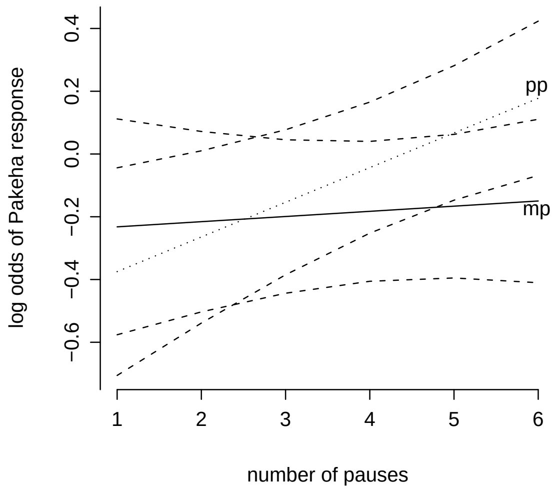


Figure 5.49: Condition One - Model effects showing log odds of Pākehā response by number of pauses and participant ethnicity.

Another aspect of pauses also shows a significant effect in this model. The mean duration of pauses used by the speaker interacts with the participant's MII. Although the length of pauses were time-normalised in this condition, participants seem to make use of the mean pause duration as a cue for dialect identification. Figure 5.50 shows the effects of the interaction between mean pause duration and participant MII. The back panel demonstrates the estimated responses of highly integrated participants. They perceive a long mean pause duration to be an indicator of a Pākehā speaker, while short pauses suggest to them that the speaker is Māori. Participants with a low integration index, however, do not make much use of mean pause duration for their decision about the speaker ethnicity. If anything, they seem to think that longer pauses are an indicator of a more Māori sounding speaker. This can be seen on the front panel of the diagram. Section 3.2.4 indicated that there was no significant difference in production between the mean pause duration of the two ethnolects.

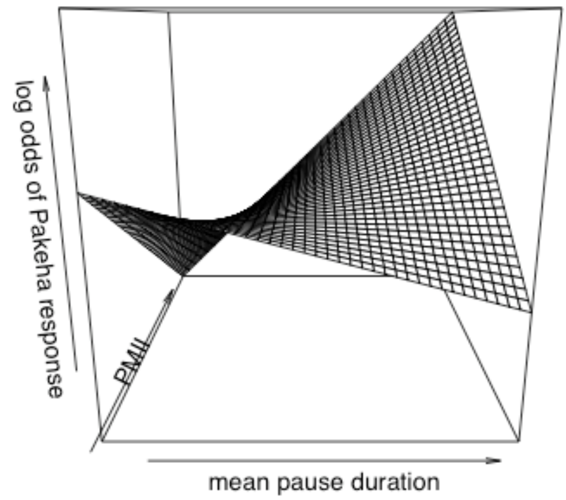


Figure 5.50: Condition One - Model effects showing log odds of Pākehā response by mean pause duration and participant MII.

The last significant interaction in our model is a three-way interaction between speaker gender, participant ethnicity and participant gender. This is shown in Figure 5.51. Pākehā male participants are using speaker gender as a cue for ethnic dialect identification. They mark female speakers as more Pākehā sounding and male speakers as more Māori sounding. However, Māori participants as well as Pākehā female participants do not rely on speaker gender to decide on the ethnicity of the speaker. It is likely that Pākehā male participants are tuning into the different fundamental frequency values used for the sine waves replacing female and male vowels in this condition. The higher frequency at 188Hz used for females might suggest to the participants that the speaker is Pākehā, while the lower frequency at 118Hz used for male vowels may trigger a more Māori response. This is similar to what we find in the model of perceived ethnicity in Condition Two, where highly integrated participants were using speaker gender to distinguish between the two dialects.

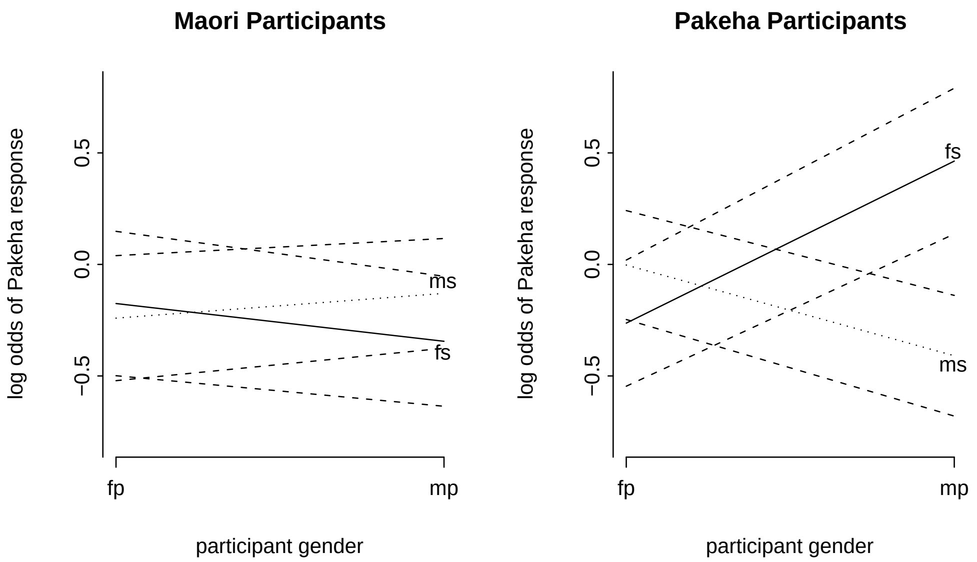


Figure 5.51: Condition One - Model effects showing log odds of Pākehā response by participant gender and speaker gender. Coefficients set to Māori participants in left panel and to Pākehā participants in right panel.

5.7.2 Accuracy - Condition 1

In the second regression model we investigated the effects on the accuracy of participants' responses. Our aim was to find out which variables help listeners correctly identify a speaker's ethnicity. The following list shows the variables that had a significant effect on accuracy in our model:

(a) participant MII interacting with speaker gender
(b) number of pauses interacting with speaker ethnicity & participant ethnicity
(c) percentage of HRTs used

The coefficient table for this model is shown in Table 5.47, and the anova table is presented in Table 5.48.

The logistic regression analysis of the accuracy of responses indicates that participants' level of integration into Māori culture is a relevant factor in this condition, just as it was in all other conditions. Participant MII is in a significant interaction with speaker gender, as illustrated in Figure 5.52. The higher their MII, the more accurate participants become at identifying male speakers. However, a high MII does not facilitate the correct identification of female speakers.

Table 5.47: Condition One - Coefficient table for model of accuracy.

Variable	Coef	S.E.	Wald Z	P
Intercept	0.08	0.36	0.22	0.8227
% of HRTs	-0.00	0.00	-4.28	0.0000
PMII	-0.00	0.01	-0.45	0.6531
speaker gender=male	-0.35	0.15	-2.28	0.0228
speaker ethnicity=Pākehā	-0.17	0.38	-0.45	0.6493
participant ethnicity=Pākehā	0.61	0.45	1.36	0.1733
number of pauses	0.09	0.06	1.35	0.1777
PMII * speaker gender=male	0.03	0.01	2.05	0.0405
spk ethn=Pākehā * part ethnicity=Pākehā	-0.75	0.52	-1.44	0.1511
spk ethn=Pākehā * number of pauses	-0.07	0.08	-0.84	0.4035
participant ethnicity=Pākehā * n of pauses	-0.15	0.09	-1.69	0.0912
spk ethn=Pkh * part ethn=Pkh * n of pauses	0.25	0.11	2.16	0.0304

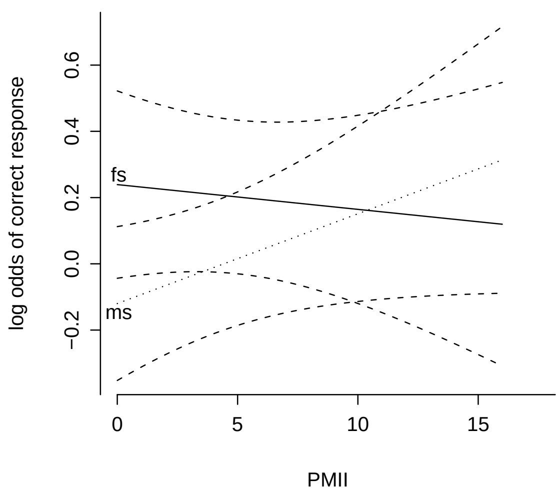


Figure 5.52: Condition One - Model effects showing log odds of correct response by participant MII and speaker gender.

Table 5.48: Condition One - Anova table for model of accuracy.

Factor	Chi-Square	d.f.	P
% of HRTs	18.35	1	<.0001
PMII	5.16	2	0.0756
speaker gender	5.32	2	0.0701
speaker ethnicity	19.65	4	0.0006
participant ethnicity	7.98	4	0.0921
number of pauses	8.36	4	0.0793
PMII * speaker gender	4.20	1	0.0405
speaker ethnicity * participant ethnicity	7.52	2	0.0233
speaker ethnicity * number of pauses	5.69	2	0.0582
participant ethnicity * number of pauses	4.69	2	0.0960
speaker ethn * part ethn * number of pauses	4.69	1	0.0304
TOTAL INTERACTION	11.35	5	0.0449
TOTAL	40.12	11	<.0001

The number of pauses used by the speaker also affected listeners' accuracy in this condition, although differently so for Māori and Pākehā participants. The three way interaction between the number of pauses, speaker ethnicity and participant ethnicity is shown in Figure 5.53. Coefficients are set to Māori participants in the left panel and to Pākehā participants in the right panel. Māori listeners become more accurate at identifying Māori speakers as the number of pauses used by the speaker increases. In contrast, Pākehā participants become less accurate at identifying Māori speakers as the number of pauses gets higher. This is a consequence of the results shown in our model of perceived ethnicity for this condition, where Pākehā participants identified a high number of pauses with Pākehā speakers. And as expected, Pākehā participants' accuracy rate at identifying a speaker as Pākehā increases with the number of pauses used.

As mentioned above, it is peculiar that the percentage of HRTs showed a significant effect on perceived ethnicity in this flat, monotonous condition. Nonetheless, it proves to be a good predictor of participants' accuracy as well. The accuracy rate drops as the percentage of HRTs used by the speaker increases. That is, listeners were most accurate when no HRTs were used and least accurate when the speaker used a high percentage of HRTs.

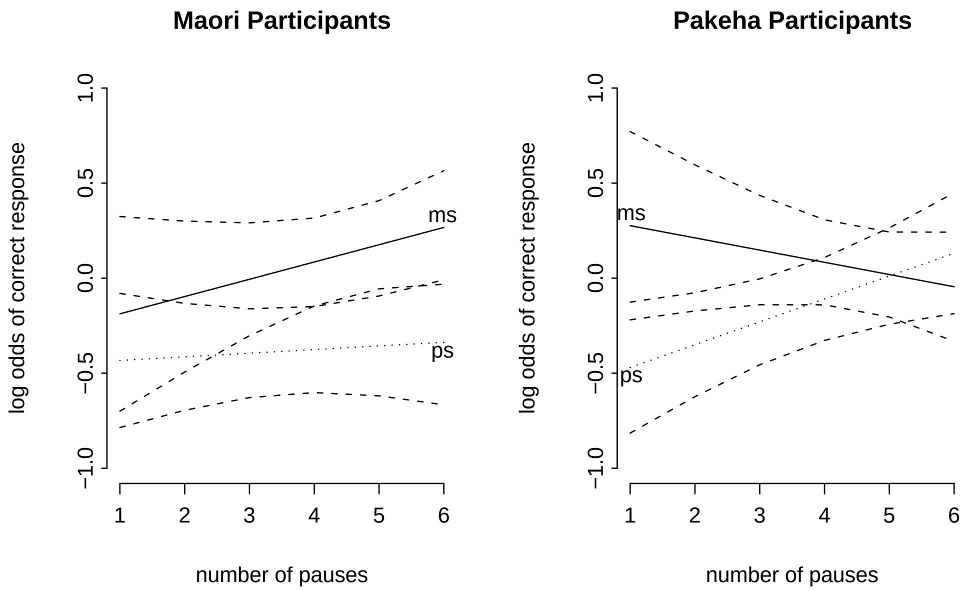


Figure 5.53: Condition One - Model effects showing log odds of correct response by number of pauses and speaker ethnicity. Coefficients set to Māori participants in left panel and to Pākehā participants in right panel.

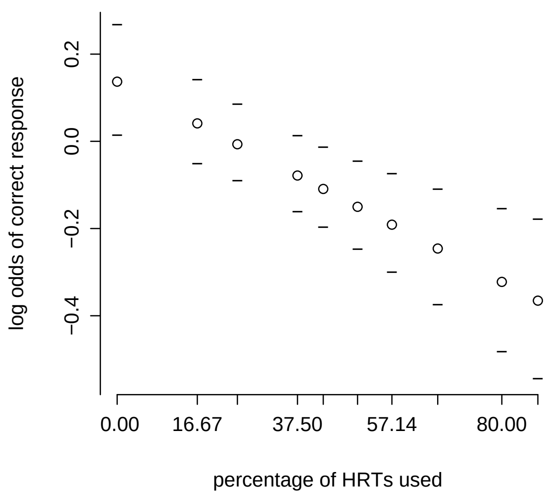


Figure 5.54: Condition One - Model effects showing log odds of correct response by percentage of HRTs used.

5.7.3 *Summary for Condition One*

The tables below give a summary of all the variables that showed a significant effect in our statistical models in Condition One. The effect of *increasing values* of each variable on the likelihood of Pākehā response is shown in Table 5.49, and Table 5.50 summarises the effect on the accuracy of responses.

Table 5.49: Condition One - Summary of significant variables in model of perceived ethnicity.

increasing value of variable	likelihood of Pākehā response
% of HRTs	decreases for syllable-timed speakers
mean pause duration	increases for high MII participants
number of pauses	increases for Pākehā participants
speaker gender	only affects Pākehā male participants:
	higher for female speakers
	lower for male speakers

Table 5.50: Condition One - Summary of significant variables in model of accuracy.

increasing value of variable	likelihood of correct response
participant MII	increases for male speakers
number of pauses	affects Māori participants:
	increases for Māori speakers
	affects Pākehā participants:
	increases for Pākehā speakers
	decreases for Māori speakers
% of HRTs	decreases

Chapter VI

Discussion of Results

Compared to previous studies, the results of the present study show a great increase in New Zealanders' accuracy at correctly identifying Māori speakers, with an average of 89% correct responses in the unaltered speech condition. Previous research has indicated that listeners' attempts at classifying Māori and Pākehā ethnicity based on a speaker's accent have often been inaccurate in the past. Holmes, Murachver and Bayard (2001) note that only 25% of Huygens and Vaughan's (1983) and 55% of Robertson's (1994) participants managed to correctly identify Māori speakers as Māori, while their own participants' identification rates reached 74%. Possible reasons behind this considerable increase over the last two decades at correctly identifying Māori speakers will be discussed in Section 6.3.1.[1]

Not surprisingly, Condition Seven had the highest percentage of correct responses, as it retained all segmental and non-segmental cues in the speech signal. As expected, the more cues that were available for the listener, the better they performed at the forced-choice dialect categorisation task. The average accuracy of participants in each condition is shown in Figure 6.1, which plots the mean accuracy of all 107 participants. As can be seen from the graph, both Māori and Pākehā listeners show a steep rise in accuracy between Conditions Five and Six, which indicates that low-pass filtering retains certain cues that facilitate dialect identification for both ethnic groups. The greatest increase in accuracy levels is between Conditions Six and Seven, showing that providing segmental cues in addition to the suprasegmental information in the speech signal is extremely helpful for listeners.

[1] Note, however, that the 10 Māori speakers used in the perception experiment were chosen from the 24 Māori speakers who had been recorded earlier, exactly on the basis of relatively high accuracy rates during the test run. This would no doubt also contribute to a higher percentage of correct responses in this study.

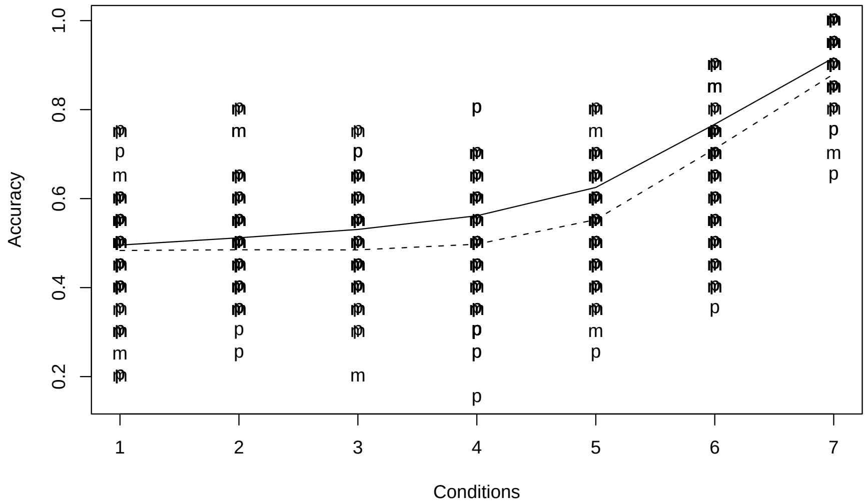


Figure 6.1: Summary of participants' mean accuracy in each condition. m = Māori participant, p = Pākehā participant. The lines represent non-parametric scatterplot smoothers fit through the data (Cleveland 1979). The dashed line indicates Pākehā participants and the solid line indicates Māori participants.

A summary of the percentage of correct responses in all conditions is given in Table 6.1, showing both participant and speaker ethnicity.[1] On average, Māori participants performed better in each condition. They are particularly good at identifying other Māori speakers and perform above chance when doing so, even in the most degraded listening conditions. It is worth noting that the intonation alone Condition Four shows a striking difference between the accuracy of Māori and Pākehā participants when identifying Māori speakers. Pākehā participants only achieved 45%, while Māori participants managed to correctly identify other Māori speakers 64% of the time, showing they can use intonation very well to identify fellow Māori speakers.

[1] For simplicity's sake, Figure 6.1 plots the accuracy of ethnically Māori and Pākehā listeners, but the reader is reminded that the real difference is between participants with a higher MII versus participants with a lower index. Converting the continuous variable of MII into a binary one by creating two groups, one with MII scores greater than 7.5 and one with MII scores lower than 7.5, creates a similar looking graph.

Table 6.1: Summary of percentage of correct responses in all conditions by participant and speaker ethnicity.

Conditions	Participants	Māori speakers	Pākehā speakers	Total
Condition Seven	Māori	93%	95%	94%
	Pākehā	86%	95%	91%
Condition Six	Māori	75%	62%	69%
	Pākehā	66%	60%	63%
Condition Five	Māori	59%	56%	58%
	Pākehā	53%	53%	53%
Condition Four	Māori	64%	48%	56%
	Pākehā	45%	49%	47%
Condition Three	Māori	53%	49%	51%
	Pākehā	49%	52%	51%
Condition Two	Māori	56%	49%	53%
	Pākehā	49%	48%	48%
Condition One	Māori	54%	44%	49%
	Pākehā	49%	47%	48%

Both Māori and Pākehā listeners performed above chance in Conditions Five to Seven, the rhythm plus intonation, the low-pass filtered and the unaltered speech conditions. In the lower conditions, however, participants did not always reach chance level. Despite the relatively poor overall levels of performance in Conditions Four to One, the results of the regression analyses suggest that listeners were able to attend to some of the relevant properties that distinguish the two ethnic varieties of New Zealand English, even when they could only rely on very few prosodic cues. Table 6.2 shows all the variables that had a significant effect in our models of *perceived ethnicity*, and whether or not they were interacting with another variable. In turn, Table 6.3 indicates the significant variables in our models of *accuracy* for all conditions. The ordering of variables in Table 6.2 is loosely based on their relevance in the different conditions. Variables that are most relevant in Condition Seven are listed at the top, while variables significant in the lower conditions are listed toward the bottom of the table. The same order is maintained in table 6.3 for easy comparison.

Table 6.2: Summary of significant variables in models of perceived ethnicity. Open bullets indicate an interaction with another variable, while solid bullets mark the lack of interaction. A shaded cell indicates that the particular phonetic cue is theoretically not perceivable in the corresponding condition.

	Condition 7	Condition 6	Condition 5	Condition 4	Condition 3	Condition 2	Condition 1
speaker ethnicity	○						
mean slope of HRT	●						
minimum pitch	●	●					
percentage of HRTs	○	○	○	○			○
PVI	○	○	○		○	○	○
SD of pitch	○		●	●			
participant MII	○		○			○	○
speaker gender	○					○	○
participant ethnicity		○		○	○	○	○
speech rate		●			●	●	
number of pauses		●				○	○
mean pitch					○		
participant gender					○		○
mean pause duration							○
speaker age							
participant age							
speaker MII							
maximum pitch							
pitch range							
mean HRT pitch range							
mean HRT ratio							
mean HRT duration							

Table 6.3: Summary of significant variables in models of accuracy. Open bullets indicate an interaction with another variable, while solid bullets mark the lack of interaction. A shaded cell indicates that the particular phonetic cue is theoretically not perceivable in the corresponding condition.

	Condition 7	Condition 6	Condition 5	Condition 4	Condition 3	Condition 2	Condition 1
speaker ethnicity	○	○	○	○	○	○	○
mean slope of HRT							
minimum pitch		●					
percentage of HRTs	○	○	○	○			●
PVI	●	●	●				
SD of pitch	○			○			
participant MII	○	○	●	●	○	○	○
speaker gender	○					○	○
participant ethnicity		○		○			○
speech rate		●				○	
number of pauses		○				○	●
mean pitch	○				○		
participant gender	○	○			●	○	
mean pause duration			●				
speaker age						●	
participant age							
speaker MII							
maximum pitch							
pitch range							
mean HRT pitch range							
mean HRT ratio							
mean HRT duration							

6.1 Discussion of Models of Perceived Ethnicity

6.1.1 Rhythm

The results of the perception experiment clearly indicate that participants are able to attend to the rhythmic characteristics of a speaker and use them as a cue for distinguishing between the two ethnic dialects of New Zealand English. Previous research has suggested that there might be a difference in timing patterns between Māori English and Pākehā English, however, no other study has shown that naive listeners are in fact aware of the variation and are capable of tuning into this rhythmic difference to help identify a speaker's ethnicity. Table 6.2 shows that PVI is a significant predictor of perceived ethnicity in *all* conditions, except in Condition Four, where intonation only was retained.[1] The presence of PVI in all other six conditions suggests that listeners can and do make use of a speaker's rhythmic properties even in degraded listening conditions to facilitate dialect identification.

However, not all participants are equally good at using PVI the correct way. The interaction of speaker rhythm and participants' MII in Conditions Seven, Six, Five and Two shows that highly integrated listeners are predictably better at interpreting PVI values to correctly identify a speaker's ethnicity. These highly integrated listeners always perceive a stressed-timed speaker as more Pākehā sounding, and a more syllable-timed speaker as more Māori sounding. This is in line with the results of the production experiment described in Chapter 3, which demonstrated that Māori English speakers are in fact significantly more syllable-timed than Pākehā speakers. Listeners with a low MII, on the other hand, only manage to use PVI in the right direction in the unaltered speech and the low-pass filtered speech conditions. In the rhythm and intonation only condition they are not able to rely on rhythm, while in Condition Two they use it in an opposite direction. This difference in the perception of syllabic rhythm between highly and lowly integrated listeners indicates that greater exposure to a dialect not only facilitates listeners' ability to recognise and identify segmental characteristics of particular vowels and consonants but also the prosodic features of the particular dialect.

The MII was not the only variable PVI interacted with in our models of perceived ethnicity. Participant gender in Condition Three also had a significant effect on how rhythm was interpreted. In this flat, rhythm only condition females were

[1] The lack of PVI as a significant predictor of perceived ethnicity in Condition Four suggests that the separation of rhythm and intonation was successful in this condition by completely eliminating rhythm from the speech signal.

able to use PVI to identify stress-timed speakers as more Pākehā sounding and syllable-timed speakers as more Māori sounding. Men, however, tended to mark syllable-timed speakers as more Pākehā.

The logistic regression analysis of perceived ethnicity in Condition Five also reveals the relative importance of rhythm in comparison with the intonational patterns of the speaker. It showed a significant interaction between speaker rhythm and the percentage of HRTs used by the speaker. The results indicate that if a speaker is very syllable-timed, they will be identified as Māori sounding, no matter what percentage of HRTs they are using. For example, as Figure 5.23 demonstrated, the model predicts that a speaker with a low PVI will tend to be marked as Māori sounding even if they do not use HRTs at all. That is, rhythm is more important a cue for ethnic dialect identification if the speaker is syllable-timed. In the case of more stress-timed speakers, however, the deciding factor of whether a speaker will be perceived as Māori or Pākehā is in fact their use of HRTs. Even the most stress-timed of speakers will be perceived as Māori sounding if they frequently make use of HRTs. A similar pattern emerged from our analysis in Condition One, where PVI also interacted this way with the percentage of HRTs used by the speaker. This is a very peculiar result, as Condition One did not retain pitch movement in the speech signal, thus HRTs should not be perceivable at all. Consequently, it makes sense to question whether in addition to a rising pitch characteristic, HRTs may also demonstrate some inherent durational properties which would be available as a cue in Condition One. For example, the length of HRTs might be inherently longer than that of non-rising terminal contours. However, post-hoc statistical analysis reveals no significant difference in duration between rising and non-rising terminals. Another avenue for research could be to investigate the relative durational difference between the phrase-final terminal and the immediately preceding syllables, where HRTs and non-rising terminals might display different ratios. An explanation remains to be found how the percentage of HRTs used by the speaker could be a significant predictor of perceived ethnicity in a duration-normalised, rhythm only condition.

6.1.2 High Rising Terminals

As discussed in the previous section, rhythm seems to be a reliable cue for ethnic dialect identification in the New Zealand context. However, the use of HRTs is just as useful a cue for the listener. Previous research has shown that Māori English speakers are likely to use a higher percentage of HRTs than Pākehā speakers (e.g., Britain (1992), Warren and Britain (2000)). The results from the production experiment in this study also indicated that overall Māori speakers use significantly more HRTs.[1] The results described in Chapter 5 reveal that listeners' perception is in line with the findings of these studies, in that participants generally expect Māori speakers to use more HRTs in their speech than do Pākehā speakers.

The percentage of HRTs used by the speaker is a variable that showed a significant effect in all conditions where intonation was retained in the speech signal, that is Conditions Four to Seven.[2] In Condition Seven, where all information, including segmental features, was preserved in the speech signal, the percentage of HRTs interacting with speaker ethnicity had a significant effect on perceived ethnicity. Only the percentage of HRTs used by Māori speakers influenced participants' responses. The more HRTs a Māori speaker uses, the more Māori sounding they will be marked by the listener. However, participants do not make use of HRTs in the case of Pākehā speakers. This suggests that the segmental features of Pākehā English are more important for ethnic dialect identification and as such override the frequency of the HRTs used by the speaker. The regression model in Condition Seven also featured another aspect of HRTs, namely the absolute mean slope. A high absolute mean slope indicates a steeper rise in pitch during the high rising terminal. The steeper this rise, the more Māori sounding the speaker is perceived to be. The results from the production experiment did not show any significant difference in absolute mean slope of HRTs between Māori English and Pākehā English (Section 3.2.3). It is, however, plausible that such a difference might exist, as Stanton (2006) indicated in her small-scale study, which compared HRTs in Pākehā, Māori and Pasifika English. Her results suggest that Māori HRTs do in fact show a steeper rise in pitch than Pākehā and Pasifika ones. A more extensive study is needed to confirm these results. Condition Seven is the only condition in our study where the mean slope of HRTs shows up in the statistical model as a significant predictor of perceived ethnicity.

In Condition Six, low-pass filtered condition, where segmental information is no longer available, only Pākehā females are tuning into the percentage of HRTs. The direction of their responses is the same as in all other conditions, that is the more HRTs a speaker uses, the more Māori sounding they are perceived to be. Pākehā

[1] Note, however, that Pākehā men in this study used proportionally more HRTs than Māori men.

[2] In addition, as mentioned in Section 6.1.1, it also featured in our model of perceived ethnicity in Condition One, where intonation is not in fact theoretically perceivable.

men and Māori participants are tuning in to cues other than HRTs when the low-pass filter is applied (e.g., rhythm, speech rate, minimum pitch and the number of pauses).[1]

Condition Five demonstrates that HRTs in some instances serve as more useful an indicator of perceived ethnicity than rhythm itself. As discussed above in Section 6.1.1, in the rhythm and intonation condition a stress-timed speaker's perceived ethnicity will depend on how often they use HRTs. Again, more HRTs are an indicator of a more Māori sounding speaker. In the case of syllable-timed speakers, however, rhythm is more important a factor for listeners than the percentage of HRTs. In Condition Five the statistical model predicts that all participants use these variables the same way, regardless of their ethnicity. In contrast, only Māori participants rely on HRTs in the intonation only Condition Four. Again, a greater number of HRTs suggests that the speaker is more Māori sounding. Pākehā listeners do not make use of HRTs in this condition to distinguish between the two dialects. Instead, they rely on the speaker's standard deviation of pitch.

6.1.3 Pitch

The pitch characteristics of the speaker – in one form or another – proved to be a significant cue in all seven conditions of the perception experiment.

The standard deviation of pitch was found to be a very important cue for listeners. It is a significant predictor of a speaker's perceived ethnicity in all conditions where pitch dynamics were preserved in the speech signal, with the exception of the low-pass filtered condition[2]. The overall pattern is very clear: the higher one's standard deviation of pitch, the more likely they will be identified as Pākehā sounding. In the unaltered speech condition the SD of pitch interacts with speaker gender. If a female speaker has a low SD of pitch, that is, sounds monotonous, she will be considered very Māori sounding. Pākehā females on the other hand are expected to have a high SD of pitch.

These results are fascinating, especially in light of the production data. The results from the production experiment demonstrated that there is in fact no significant difference in the standard deviation of pitch between Māori English and Pākehā

[1] It is also possible that they are tuning into the voice quality of the speech, which is arguably preserved by low-pass filtering but was not investigated in this study.

[2] Note, however, that pitch nonetheless features in the statistical model in Condition Six, in the form of minimum pitch.

English (Section 3.2.2). Yet, Māori speakers are assumed to be speaking with a flat intonation, while Pākehā speakers are expected to be very "up and down"-y when it comes to pitch movement. None of the statistical models showed an interaction between the SD of pitch of a speaker and participant ethnicity, which indicates that not only Pākehā listeners perceive Māori speakers to be more monotonous sounding, but Māori people themselves believe so as well.

The fact that there is a significant gap between what speakers actually produce and what listeners expect to be hearing shows people's misconceptions about the difference between the two ethnic dialects in terms of pitch, and seems to suggest that stereotypes are at work. There is some possibility that these stereotypes stem from historical facts, reflecting an earlier stage of Māori English as a diverging dialect of New Zealand English. It is possible then, that listeners simply have not caught up with the changes yet.

A similar trend emerged with regard to minimum pitch. In Conditions Seven and Six, in addition to the standard deviation of pitch, participants also tuned in to the minimum pitch used by the speaker. Again, the production experiment demonstrated that there is no significant difference between the minimum pitch used by Māori and Pākehā speakers. Listeners, nonetheless, expect Māori speakers to have a lower minimum pitch than Pākehā speakers.

The divergence between production and perception is even better illustrated in the case of the speaker's mean pitch. Section 3.2.2 of the production experiment showed the somewhat unexpected result of Māori English mean pitch getting higher over time, while Pākehā mean pitch, if anything, is getting lower. Listeners' perception about the difference in mean pitch between Māori and Pākehā English seems to be at odds with these results. Participants do rely on the mean pitch of a speaker in Conditions Three, Two and One when many of the other cues have been eliminated from the speech signal.[1] However, they perceive a lower mean pitch to be a characteristic of a Māori speaker and a higher mean pitch to indicate a Pākehā speaker. This is completely different from what is happening in production, where Māori speakers in fact use a significantly higher mean pitch. As this seems to be a change over time, it is plausible that listeners are not yet consciously aware of this new feature of Māori English. When people are overtly asked whether they think Māori speakers have a higher or lower mean pitch than

[1] assuming that in Conditions Two and One the way they are using speaker gender as a cue simply reflects the different mean pitch values used for creating the signal

Pākehā speakers, they tend to reply 'lower pitch' without hesitation. This might be the result of certain physical stereotypes held about Māori being big and bulky. However, when they are asked to imitate a Māori speaker, they almost always use a higher pitch in doing so. This suggests that subconsciously they might be aware of the ongoing change in Māori English pitch and, if such is the case, then we might expect that with time, perception results regarding mean pitch will adjust as listeners become actively aware of this new prosodic feature of the ethnolect.

6.1.4 Speech Rate and Pauses

Pitch is not the only characteristic that demonstrates possible stereotypes associated with Māori English. Paralinguistic features, such as speech rate and number of pauses used by a speaker, also suggest that stereotypes are at work. Speech rate features in three of the seven conditions as a significant predictor of listeners' perception of a speaker's ethnicity. Even though the results from the production experiment indicated that there is no significant difference between the speech rate of Māori English and Pākehā English, listeners nonetheless identify slower speakers as Māori and faster speakers as Pākehā sounding. The logistic regression models in Conditions Six, Three and Two[1] all predict that fast speakers tend to be identified as Pākehā. Speech rate does not interact with participant ethnicity in these models, suggesting that both Māori and Pākehā participants perceive a lower speech rate as a characteristic of Māori English. Speech rate has been shown relevant in ethnic identification in the past. Bayard (1995), for example, found similar results in his study, where slower speakers were judged to be of Māori or Pasifika origin as opposed to being Pākehā.

Roach (1982) claims that speakers of stress-timed languages perceive syllable-timed speech as spoken faster than stress-timed speech. Based on this, we could expect that the stress-timed Pākehā speakers would perceive the more syllable-timed speech of Māori English as faster sounding than their own variety. However, it seems that this does not apply in the New Zealand context, where certain attitudes and stereotypes associated with Māori English might run too deep. These stereotypes will be discussed more in Section 6.3.

It is not only slow speech rate that is perceived to be characteristic of Māori

[1] Speech rate was not found to be a significant variable in Condition One, which seems to suggest that the duration normalisation procedure of vocalic and consonantal segments was successful.

English; a high number of pauses and hesitations is also assumed to be a feature of the variety. The regression model in Condition Six indicated that the more often a speaker pauses, the more likely they will be identified as Māori sounding. However, things get a little more complicated in the rhythm only conditions. In Condition Two, Māori participants perceive a lot of pauses to be an indicator that the speaker is Pākehā, while in Condition One it is Pākehā participants who think that a high number of pauses means a Pākehā speaker.

In the lack of many other available cues, participants seem to rely on the mean duration of pauses in Condition One. Participants with a high MII appear to perceive speakers with a long mean pause duration as Pākehā sounding. This is surprising, as the length of pauses were duration normalised in this condition, so theoretically listeners should not be using it as a cue. It seemed plausible to investigate whether the mean pause duration correlated with the standard deviation of pause durations, which could be perceivable in this condition. Mean pause duration positively correlates with the number of pauses, however, we found no such correlation between mean pause length and the SD of pause duration. It remains to be explained how such cues as mean pause duration (and in fact HRTs) could be relevant factors in this duration normalised, rhythm only condition.

6.1.5 Māori Integration Index and Participant Ethnicity

The linguistic experience of the participant has repeatedly been shown to affect the perception of dialect variation (e.g., Preston 1986, Tamasi 2003, Clopper 2004). Most of these studies were concerned with regional variation in the United States and demonstrated a correlation between the mobility of listeners and their dialect perception. Linguistic experience does not only play a role in the perception of geographical dialects but also in the perception of ethnic varieties. In New Zealand, Robertson (1994) and Szakay (2006*a*) illustrated that linguistic experience influences one's performance in an ethnic dialect categorisation task. The results of the present study are consistent with their research, showing that participants who are highly integrated into Māori society behave differently in a dialect identification task from those participants with a low integration index. Our logistic regression analyses reveal that participants' MII is a significant variable in the models of perceived ethnicity in five of the seven conditions.[1]

[1] The other two conditions feature the variable of participant ethnicity instead of MII.

Figure 6.2 summarises the interaction between speaker rhythm and participant MII in the relevant conditions. The y-axis of these 3D diagrams shows the likelihood of a Pākehā response. A higher value indicates that the speaker is more likely to be identified as Pākehā, while a low value signals a Māori response. The x-axis demonstrates the continuum from syllable-timing to stress-timing. The predicted responses of highly integrated participants are shown in the back panel, while the front panel of the diagram displays the perceived ethnicity by listeners with a low MII.

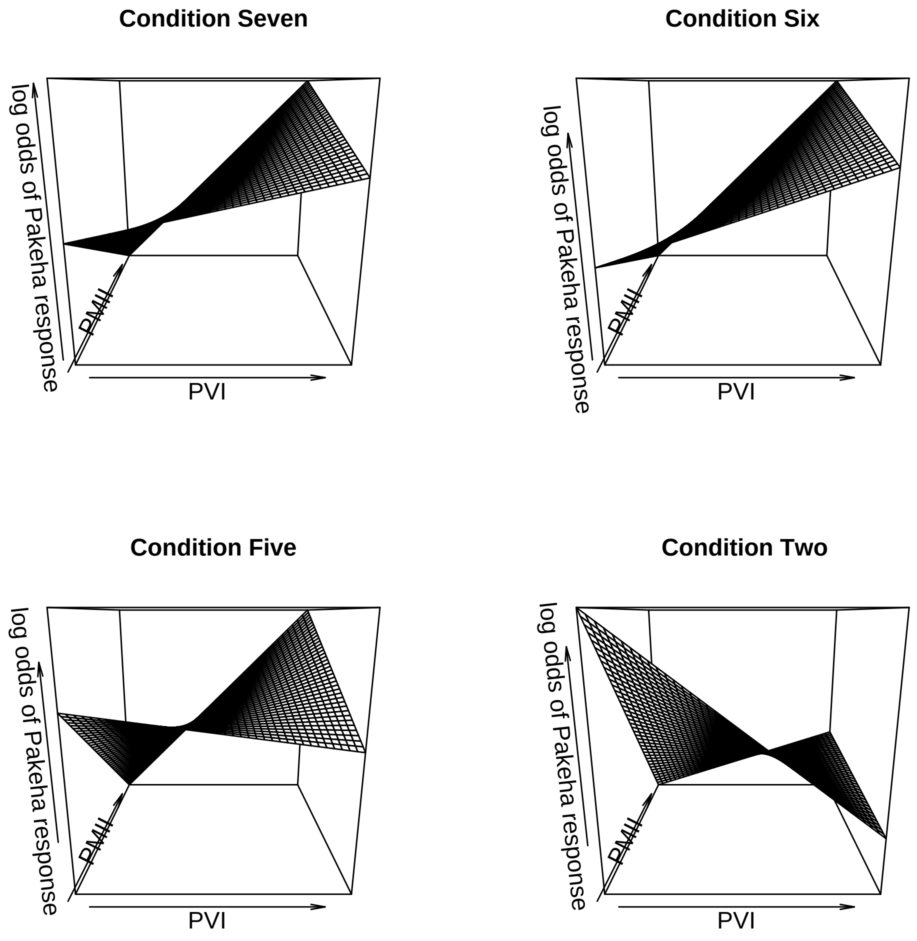


Figure 6.2: Summary of the interaction between participant MII and speaker rhythm in the models of perceived ethnicity in Conditions Seven, Six, Five and Two.

In Conditions Seven, Six, Five and Two, MII interacts with PVI in a way that shows highly integrated listeners being consistently better at interpreting speaker rhythm. These participants are able to rely on the rhythmic characteristics of a speaker and identify stress-timed speech as more Pākehā sounding and syllable-timed speech as more Māori sounding, even in the most degraded listening conditions. Participants with a low integration index, however, tend to only use rhythm this way in the unaltered speech and low-pass filtered conditions, when many other cues are available as well. In the lower conditions they are not able to rely on rhythm to identify a speaker's ethnicity. This confirms the hypothesis that more exposure to a dialect facilitates the identification of not only segmental but also prosodic features as belonging to the particular dialect.

MII did not show significance in our models of perceived ethnicity in Conditions Four and Three. However, the variable of participant ethnicity did, demonstrating a difference in performance between Māori and Pākehā listeners in relation to the non-segmental cues they used for dialect identification. In the intonation only condition, Māori participants were able to tune into the percentage of HRTs used by a speaker, while Pākehā participants relied on other cues (such as SD of pitch). In Condition Three, the difference in performance is again based on participant ethnicity, where Pākehā listeners use the speaker's mean pitch as a cue. For them, the higher the mean pitch of the speaker, the more likely they will be identified as Pākehā. Māori listeners on the other hand, do not rely on mean pitch in this condition. This illustrates that listeners with different linguistic experience and different backgrounds will make use of, and rely on, different linguistic cues when attempting to identify a speaker's dialect.

6.1.6 Gender

Gender has been found to interact with dialect variation in numerous ways (Labov 1990). In speech production, women tend to lead phonological changes in progress, while in a stable sociolinguistic situation, men tend to use more non-standard forms. Perceptual research has indicated that there might also be a gender effect in speech perception, with men and women interpreting linguistic cues differently (Hay, Nolan and Drager 2006). In the present study, gender did not play a very important role in our models of perceived ethnicity. Participant gender only showed up as a significant predictor of perceived ethnicity in two conditions. In Condition Three, female and male participants were attending differently to a speaker's PVI. Women were able to tune into a speaker's rhythmic characteristics and identify syllable-timed speakers as slightly more Māori sounding and stress-timed speakers as slightly more Pākehā sounding. Men did not manage to use rhythm in this fashion. In Condition One, Pākehā male participants were behaving differently from Māori participants as well as from Pākehā female participants. They relied on speaker gender – arguably used as a cue for mean pitch – when identifying speaker ethnicity. Speaker gender might have been used by participants as a substitute for mean pitch in Condition Two as well (see Section 5.6.1). The only other condition where speaker gender showed significance was the unaltered speech condition, where it interacted with the standard deviation of pitch. Overall, the higher the SD of pitch, the more likely the speaker will be marked as Pākehā. However, this trend was most obvious for female speakers, where monotonous intonation was believed to be very Māori sounding. Participant gender features more in our models of accuracy, which will be discussed in Section 6.2.6.

6.2 Discussion of Models of Accuracy

6.2.1 Māori Integration Index and Participant Ethnicity

Individuals who are well integrated into local Māori networks are socially positioned to have access to both Māori and Pākehā dialects, and as such, they can be expected to be more accurate in a dialect identification task. Experience with linguistic variation has been shown to play an important role in dialect identification performance (e.g., Robertson 1994, Clopper 2004, Szakay 2006*a*).

Exemplar models of speech perception are particularly well suited to account for frequency- and experience-based phenomena (Bybee 2001, Johnson 1997). In an exemplar based model, categories are made up of a large set of remembered exemplars from a wide range of speakers, and the auditory properties that distinguish speakers are retained in these exemplars. Each exemplar has an associated strength or resting activation level. Exemplars encoding frequent and recent experiences have higher resting activation levels than exemplars encoding infrequent and temporally remote experiences. When a new token is encountered, it is classified according to its similarity to the exemplars already stored (Pierrehumbert 2001). Thus, the framework of exemplar theory readily accommodates the finding that the ability to process and identify a given dialect is enhanced by direct experience with that particular variety.

Robertson (1994) and Szakay (2006*a*) demonstrated that New Zealand listeners who are highly integrated into Māori society perform much better at a dialect identification task than those with a low integration index. The results of the present study are consistent with their research. Participants' MII is a significant predictor of accuracy in each of the seven conditions, indicating that highly integrated participants are able to distinguish Māori and Pākehā speakers significantly more accurately than those who are not integrated into Māori networks, even in degraded listening conditions.

In our models of accuracy for the unaltered and low-pass filtered speech conditions, participant MII interacted with participant gender. Both females and males improve at the task as the MII increases, however, the rise is steeper for male participants. In both conditions non-integrated females are more accurate than their male counterparts but the increase in MII has a greater effect on male listeners than on females. Amongst the highly integrated listeners, men even tend to be performing better than women.

In Conditions Four, Three and Two, MII interacts with speaker ethnicity. Again, the model indicates that participants with a high MII are significantly better at distinguishing Māori speakers from Pākehā speakers, and also shows that an increasing value of MII has a particularly great effect on participants' ability to correctly identify Māori speakers. The fact that there is no such effect for Pākehā speakers is not surprising, as a high MII assumes a greater exposure to Māori English, while exposure to the Pākehā variety is presumed to be of similar levels for all New Zealanders, regardless of their MII.

MII proved to be a better indicator of accuracy at the dialect identification task than participant ethnicity alone. Although on average Māori participants performed better in each condition, the most accurate listener was, in fact, a Pākehā female (p55) [1] who is highly integrated into Māori networks (MII = 13.5). She is married to a Māori man and has close ties with his whānau and the wider Māori community.

In addition to the MII, the variable of participant ethnicity also appeared in the models of accuracy in the low-pass filtered and intonation only conditions[2], where it interacted with speaker ethnicity and the percentage of HRTs used. Both Pākehā and Māori participants are most correct at identifying Māori speakers who use a lot of HRTs, and Pākehā speakers who do not use them much. However, they

[1] See Appendix A for all participants' accuracy ratings.

[2] The fact that both MII and participant ethnicity are tolerated in the same statistical model proves that MII is clearly doing more than just separating the two ethnic groups.

rely on HRTs under different conditions. Māori participants are able to rely on HRTs in the intonation only condition, where Pākehā tune into other cues. Pākehā participants, on the other hand, tune into HRTs in the low-pass filtered condition, where Māori accuracy is affected by other variables, such as rhythm.

6.2.2 Rhythm

As discussed in Section 6.1.1, syllabic rhythm is a significant predictor of whether participants identify a speaker as Māori or Pākehā. In three of the seven conditions, it is also a significant predictor of listeners' accuracy at the dialect identification task. PVI features in our models of accuracy in Conditions Seven, Six and Five. In the unaltered speech and low-pass filtered conditions, the model predicts that listeners are more accurate at identifying speakers who are more syllable-timed. In Condition Five, the rhythm plus intonation condition, however, the direction changes, with now stress-timed speakers being identified with greater accuracy.

In Conditions Seven and Six many other cues are available for the listener in addition to rhythm. In the normal speech condition listeners can rely on the pronunciation of vowels and consonants, while Condition Six also preserves *some* segmental information as well as voice quality. As syllable-timing will likely be tied in with many other features of Māori English, it is not surprising that listeners find it easier to correctly identify speakers who are more syllable-timed. In Condition Five, however, there are no segmental or voice quality cues participants can rely on. This condition is much less speech-like than the previous two, with its staccato-like features. It is possible that under such circumstances stress-timed rhythm can be more easily related to human speech than syllable-timing, the latter sounding more like a machine gun. Thus, it is plausible that in Condition Five stress-timed rhythm makes it easier for participants to tune into other diagnostic features of the speech signal, such as pauses and HRTs. This, in turn, would result in better accuracy rates in the case of the stress-timed speakers.

6.2.3 High Rising Terminals

The results of the models of perceived ethnicity revealed that New Zealanders expect a high percentage of HRTs from Māori speakers (see Section 6.1.2). However, the Pākehā male speakers used in this study actually produced more HRTs than the Māori men. This explains the interaction between percentage of HRTs and speaker gender in our model of accuracy for Condition Seven. Male speakers using a high percentage of HRTs tend to be incorrectly identified, while the percentage of HRTs used by females does not affect participants' accuracy. This is because the female speakers in this study perform as listeners expect, with Māori females using HRTs with higher frequencies than Pākehā females. In Conditions Six, Five and Four it is again shown that listeners expect Pākehā speakers to use fewer HRTs than Māori speakers, as those Pākehā with a high percentage of HRTs are incorrectly identified as Māori. However, as mentioned earlier, in the low-pass filtered condition only Pākehā listeners' accuracy is affected by HRTs, while in the intonation alone condition it is only Māori participants that tune into HRTs and as such, only their accuracy is affected.

6.2.4 Pitch

As discussed in Section 6.1.3, listeners heavily rely on the pitch characteristic of the speaker to make judgements about their ethnicity. However, it seems that New Zealanders perceive Māori to generally speak with a lower pitch than Pākehā. The production data in this study revealed that this is, in fact, not the case. There were no significant differences between the two ethnic dialects with regard to minimum pitch and the standard deviation of pitch. Moreover, the results also revealed that Māori mean pitch was actually significantly higher than Pākehā mean pitch. This mismatch between listeners' expectations and speakers' actual production results in high inaccuracy rates in the dialect identification task.

Accuracy was significantly affected by mean pitch interacting with speaker ethnicity in Conditions Seven and Three, where those Māori with a high mean pitch were incorrectly classified as Pākehā. In Condition Six, the model of perceived ethnicity showed that listeners associate a low minimum pitch with Māori speakers. Again, as this is not the case, it results in a greater rate of incorrect answers for speakers with a high minimum pitch. As discussed in Section 6.1.3, Māori speakers are perceived to be speaking in a monotonous way, which leads to listeners' inaccuracy when identifying Māori speakers with a high standard deviation of pitch, as seen in the intonation alone condition. It is somewhat puzzling then that in Condition Seven the model indicates that both Māori and Pākehā are more accurately identified if they have a high SD of pitch. This seems to be at odds with the results from the model of perceived ethnicity, which suggested that speakers with a high SD of pitch will be marked as Pākehā sounding. However, this condition preserves all segmental information, and it is possible that the Māori speakers who happen to have a high SD of pitch in this study also produce vowels and consonants that are more characteristic of Māori English. This would make them easily identifiable as Māori, as when all phonetic information is retained in the signal, segmental cues seem to serve as the primary cue for listeners, with prosodic information being a secondary resource. All in all, the pitch results reveal that Māori are expected to be speaking with a low pitch, hence the incorrect classification of many of them who do not fit this criterion.

6.2.5 Speech Rate and Pauses

Speech rate proved to be an important factor in the dialect identification task and the results showed that, in general, New Zealand listeners associate a slower speech rate with Māori ethnicity. This assumption has an adverse effect on participants' accuracy, as it is not the case that Pākehā speak quickly and Māori speak slowly. In Condition Six our model indicated that slow speakers were most accurately identified. There seem to be two reasons behind this. On the one hand, it is likely that the low accuracy rate for fast speakers is caused by fast Māori speakers being misperceived as Pākehā. On the other hand, a lower speech rate probably makes it easier for participants to tune into other relevant cues in the speech signal that will facilitate dialect identification.

Condition Six also showed that participants perceive few pauses and hesitations as a sign of a Pākehā speaker, which explains why those Pākehā speakers with a high number of pauses are incorrectly identified in this low-pass filtered condition. The results from the model of accuracy in Condition Two can also be taken to support this claim. Here female speakers with many pauses get inaccurately identified. However, as mentioned earlier, it is likely that gender here is used as a substitute for ethnicity, as participants perceive the high pitched tone created for female speakers as Pākehā, and the lower pitch generated for male speakers as Māori. Thus it is possible that what listeners are really doing is mistakenly marking Pākehā speakers with many pauses as Māori.

In Condition Two, in addition to the number of pauses, speech rate also had a significant effect on accuracy, this time interacting with participant gender. Females were more accurate at identifying the ethnicity of fast speakers, while men performed better at identifying slow speakers. Brizendine (2006) claims that females speak twice as fast as men. If women do talk faster in general, it is not implausible that they would find it easier to tune into features of other speakers who are also fast. Men, on the other hand, could be more conducive to picking up relevant cues from slow speakers like themselves. Gender differences like these are not unknown in speech perception experiments, as discussed in the following section.

6.2.6 Gender

Previous research has shown that the gender of participants can affect their responses in various speech perception tasks (e.g., Hay et al. 2006). In the present study participant gender significantly affected accuracy rates in four of the seven conditions. As already discussed in Section 6.2.1, in Conditions Seven and Six participant gender interacted with the MII of the participant. Both females and males improve at the task as the MII increases, however, this rise is steeper for male participants. Amongst those listeners who have no ties with Māori networks, females tend to perform better at the dialect identification task. In fact, the results from Condition Three also reveal that the ethnic judgements of females are more accurate even in a monotonous, rhythm only speech condition. Research has shown that women tend to use language variation to a greater extent than men to express group membership (e.g., Eckert 2000). Just as there is a difference in production, Drager (2005) suggests that there may also be a gender difference in speech perception, such as women being more aware of the relationship between language variation and the social (or in our case ethnic) characteristics of a speaker. Drager uses this argument in support of an exemplar based theory of speech perception, where females might index their exemplars with a greater amount of social detail. This would enable them to perform better at the dialect identification task.

Exemplar theory might also provide an explanation why a high MII has a greater effect on the accuracy of male participants, so that amongst listeners with a high MII, males even tend to perform better than women. King (1999) suggested that Māori English, or 'bro speak', is heavily associated with Māori men. Highly integrated Māori men could be interacting with many more Māori males than females, whose Māori English features are supposedly more salient than those of the Māori women. Highly integrated females, on the other hand, might have closer relationships and more interaction with Māori females, whose ethnic dialect features are not as strong. This in turn could suggest that highly integrated male participants might have more Māori English exemplars stored, and thus be more accurate in a dialect identification task than women.

6.3 General Discussion

6.3.1 Māori Renaissance

As mentioned at the beginning of this chapter, the accuracy rates in the present study for the normal speech condition were much higher than in previous dialect identification experiments in New Zealand. In 1983 Huygens and Vaughan reported that Māori speakers were correctly identified only 25% of the time. In Robertson's 1994 study this rate was 55%, while in 2001 Holmes et al.'s correct identification rates reached a relatively high 74%. In the present study Māori English speakers were identified with an even higher 89% accuracy. While the methodologies used in these studies are not comparable, I feel that there are at least two valid and justifiable reasons behind this steady and considerable increase in correct identification rates of Māori English over the last 25 years.

Firstly, it seems that the Māori renaissance[1] has had a positive effect on people's attitudes towards all things associated with Māori, from the arts to the language variety they use. Māori English is becoming recognised as a valid linguistic variety and acknowledged as an important element of Māori ethnic identity. As such, it has become more accessible to the general New Zealand audience in recent years. The variety is now heard in many forms of the media, from radio to television. There is an increasing number of Māori faces seen on regular television, and the launch of the Māori Television channel itself in 2004 is also likely to have contributed to making Māori English more available to both Māori and Pākehā New Zealanders. In addition, King (1999) suggests that some Māori are able to alternate between Māori English and Pākehā English depending on the situation and the audience, while others have only Māori English at their disposal. Based on my experience, it seems that there is an increasing number of Māori who belong to the latter group, for whom Māori English is in fact the only available variety rather than just a register. This would mean that Pākehā are exposed to more Māori English than they were in the past, as many Māori speakers do not switch to the Pākehā variety when interacting with Pākehā listeners.

Secondly, I would argue that the features of Māori English appear to be becoming more salient than they were a couple of decades ago. For example, the results

[1] The Maori renaissance has its roots in post-war urban alienation of Maori and is a term applied to activism and events during the 1970s and 1980s which focussed on reviving and applying aspects of Maori culture into New Zealand society.

concerning rhythm and age in this study (see Section 3.2.1) indicate an apparent-time change, with younger Māori English speakers producing more syllable-timed speech than older speakers. It is possible that the segmental features of young speakers would also show an increasing divergence from Pākehā English, which in turn would make the dialect easier to identify in a perception experiment.

6.3.2 Suprasegmental Cues and Ethnic Stereotypes

The results from our production experiment are consistent with conclusions drawn in previous studies (e.g., King 1993, Bayard 1995, Holmes and Ainsworth 1997, Bell 2000) showing that the relationship between Māori English and Pākehā English is best viewed as a continuum. The two varieties mainly differ in the relative frequency of certain shared features, that is, the differences are proportional rather than absolute. Although there is no categorical distinction, it seems that listeners are able to classify speakers as Māori or Pākehā with some accuracy.

Despite the relatively poor levels of performance in the degraded listening conditions, the results of the regression analyses suggest that listeners *were* attending to some of the relevant properties in perception that distinguish the two ethnic varieties. As there is no single feature that distinguishes Māori English from Pākehā English, it is perhaps not surprising that in the perception task various listeners make use of different cues in different ways, depending on what's available in the speech signal and also on their own linguistic and social background. This is well illustrated in the case of HRTs, which are available cues in both Conditions Four and Six. However, in the intonation alone condition only Māori participants relied on the percentage of HRTs used by the speaker in their ethnic judgements. Pākehā listeners tended to tune into other cues, such as the SD of pitch. In the low-pass filtered condition, however, an opposite trend emerged with Pākehā listeners now using HRTs as a cue and Māori participants tuning into other cues available in the speech signal. The results of our study also revealed that participants who have had more exposure to Māori English are more capable of relying on syllabic rhythm in a dialect identification task, even in the lower conditions, where those participants with less exposure seem to be tuning into speech rate instead of rhythm itself. These results prove that different listeners utilise different cues for ethnic dialect identification.

Thomas and Reaser (2002) argue that it is also possible that different cues are used to identify different speakers. The results from the present study support this claim in several different ways. For example, the results of the perception experiment show that speakers who are very syllable-timed will be identified as Māori based on their rhythmic characteristics. Those speakers who demonstrate more of a stress-timed rhythmic pattern, however, will be instead judged on their use of HRTs. They might still be identified as Māori regardless of being stress-timed, *if* they use a high percentage of HRTs. Further, the introduction of segmental features in Condition Seven proved to be extremely useful for participants in the perception experiment, and most of the time generated higher accuracy rates. However, two of the Māori speakers were less correctly identified in the normal speech condition than in conditions with prosodic cues only. One female Māori speaker (m18) was correctly identified 74% of the time in the low-pass filtered condition, while in the unaltered speech condition her average accuracy rate dropped to 49%. This suggests that her rhythmic and intonational properties (and voice quality - more on which in Section 6.3.3) are easily identifiable as Māori, however, her vowels and consonants are not typical of the Māori English variety. The accuracy rates for one young male Māori speaker (m20) also show a similar pattern. He is correctly identified 68% of the time in Condition Five, where only rhythm and intonation are available cues for the listener. His mean accuracy rates in Conditions Six and Seven are 61% and 63% respectively. This drop indicates that although he displays Māori English rhythmic and intonational characteristics, he neither possess the voice quality nor the segmental features associated with Māori speakers.[1]

Taken together, these latter results suggest that segmental features override the importance of prosodic cues in listeners' perception as markers of ethnicity in New Zealand. However, the present study also demonstrates that there is no shortage of prosodic variables that are accessible to listeners as potential cues for ethnic identification of speakers' voices. We can group these suprasegmental cues into three main classes based on their usefulness and relevance in ethnic dialect identification in New Zealand. A summary of the main prosodic variables is given below in Figure 6.3.

[1] This speaker's accuracy rate in Condition Three is the same as in Condition Seven, further proving he does have perceivable Māori rhythm.

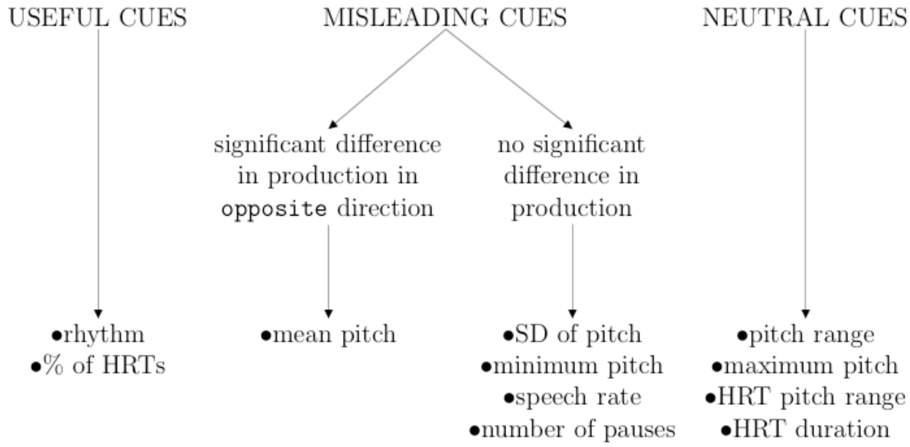


Figure 6.3: Summary of prosodic cues in ethnic dialect identification

Useful Cues

Rhythm and the percentage of HRTs used by the speaker proved to be extremely useful cues for participants in the dialect identification task. In the case of these two prosodic features, not only do we have production evidence of a divergence between Māori English and Pākehā English, this evidence is supported by perception evidence that shows that listeners are aware of these features and associate them with ethnicity. The acoustic analysis demonstrated that Māori English speakers are significantly more syllable-timed than Pākehā English speakers, while the dialect identification experiment showed that listeners are aware of this difference and are able to rely on rhythm in ethnic dialect differentiation. The production data also confirmed the results of previous research on the use of HRTs in New Zealand (e.g., Britain 1992), showing that, overall, Māori use a significantly higher percentage of HRTs than Pākehā. Again, listeners are aware of this tendency and expect a high number of HRTs to be an indicator of a Māori speaker. Moreover, the perception results also revealed the relative importance of rhythm and HRTs. Listeners heavily rely on rhythm in their ethnic judgements when the speaker is very syllable-timed. These speakers will be identified as Māori, regardless of their use of HRTs. However, if the speaker is stress-timed, it is in fact the frequency of HRTs that will be the deciding factor in distinguishing Māori English from Pākehā English. Thus, we can conclude that both of these prosodic features facilitate accurate dialect identification in New Zealand. However, certain other suprasegmental cues achieve the opposite effect.

Misleading Cues

The results of this study have indicated that there is a divergence between the production and the perception of some prosodic cues. What listeners expect to be hearing does not always match what speakers are actually producing. As shown in Section 3.2, there were no significant differences between the two ethnic dialects with regard to speech rate, number of pauses, minimum pitch and the standard deviation of pitch. However, naive listeners expect Māori English speakers to have a lower speech rate, more pauses, lower SD of pitch and a lower minimum pitch than Pākehā English speakers. In addition, the results from the production experiment also demonstrated that Māori English speakers now have a significantly higher mean pitch than Pākehā speakers, yet, listeners expect Māori to be speaking with a lower mean pitch than Pākehā. This mismatch between production and perception has a hindering effect on accuracy in ethnic dialect identification. All in all, Māori English speech seems to be perceived as *slow, hesitant, monotonous and low-pitched*. Those speakers who do not conform with these expectations are often misidentified as Pākehā. That listeners' expectations are not corroborated by evidence from speech production suggests that they are based on ethnic stereotypes.

Research has shown that many ethnic groups hold unfavourable attitudes towards other ethnic groups, and towards ethnic minority groups in particular (e.g., Wilson 1996, Fought 2001). It has also been illustrated that people who speak the English associated with an ethnic minority are more negatively evaluated than are people who speak the majority dialect (e.g., Giles, Williams, Mackie and Rosselli 1995, Nesdale and Rooney 1996). The variation found in speech-evaluation studies reflects social perceptions of the speakers of given varieties and has nothing to do with linguistic qualities of the dialect itself. Thus, listening to a given variety is generally considered to act as a trigger or stimulus that evokes attitudes, prejudices or stereotypes about the relevant speech community (Edwards 1999). For example in our case, a slower speech rate and a more hesitant speech style as features associated with Māori English might reflect listeners' general perception about the Māori community as less self-confident, less intelligent and lazier than Pākehā. These characteristics were in fact associated with Māori speakers in previous attitudinal studies (e.g., Huygens and Vaughan 1983, Robertson 1994). Street, Brady and Putman (1983) demonstrated that listeners find a speaker with a faster speech rate (both actual and perceived) more competent and socially attractive than a speaker with slower rates. In the New Zealand context Bayard (1995) also suggests that *"Paralinguistic features such as slow speech rate, hesitations and low audibility are seemingly far more important in rating the speaker as a 'Māori', and once so rated, evaluations of traits such as income, social class and education decline significantly."* (Bayard 1995:167)

It appears that more than a decade later these stereotypes about slow, hesitant speech as a perceived feature of Māori English are still alive and well in New Zealand. In addition to speech rate and hesitations, the present study also shows that Māori are assumed to be speaking with a low pitch and a flat intonation. These traits have also been shown to be negatively evaluated by listeners. Oksenberg, Coleman and Cannell (1986) demonstrated that it is not only fewer hesitations and faster speech rate that are associated with attractiveness, so is high pitch and great variations in pitch. These are believed to be Pākehā English features. Perception studies conducted in the US yielded similar results with regards to the pitch characteristics of the speakers, where lower F0 levels were associated with African Americans and higher F0 with European Americans (Hawkins 1992, Foreman 2000, Thomas and Lass 2005). Fought (2006) also mentions that there seems to be an association of African American English with low voice pitch and notes that *"...it is not clear to what extent this view corresponds to actual usage by African American English speakers, as opposed to a general stereotype."* (Fought 2006:51). It is possible that the same kind of stereotypes are at work in New Zealand and the US in relation to the perceived pitch levels of ethnic minority dialects.

Neutral Cues

The third group of suprasegmental cues, as shown in Figure 6.3 above, are the ones that listeners were not attending to in their ethnic dialect identification. These include pitch range and maximum pitch as well as some properties of HRTs, such as duration, pitch range and ratio (calculated as pitch increase per duration). These cues neither facilitate nor hinder participants' accuracy in distinguishing Māori English from Pākehā English; listeners simply did not rely on these prosodic cues in this research. Although unexplored in the present study, another possible prosodic cue that listeners are likely to be tuning into is voice quality.

6.3.3 *Future Path: Voice Quality Analysis*

The results of the perception experiment indicated that there is a steep rise in accuracy between Conditions Five and Six. Condition Five retained both rhythm and intonation by creating a hummed version of the original sound file and by replacing all consonants with silence. Condition Six used low-pass filtering at 400Hz, which also retains both rhythm and intonation, but in addition to these two prosodic features, it arguably also preserves the voice quality of the speaker. For example, both shimmer and jitter are preserved in the low-pass filtered signal (Bezooijen and Boves 1986). This information is clearly not available in Condition Five. The fact that the mean accuracy rates are much higher in the low-pass filtered condition than in Condition Five suggests that the voice quality of a speaker is an important and useful cue in ethnic dialect identification in the New Zealand context. An acoustic analysis of voice quality was not carried out as part of the present study, however, impressionistically it appears that many of the Māori speakers used in this research exhibit certain voice quality features that distinguish them from the Pākehā speakers. Without extensive research it is hard to define the exact characteristics of this 'Māori voice quality' but many of our participants commented on features such as husky, hoarse or brittle voice. It seems that listeners are aware of the difference between Māori and Pākehā voice quality and are able to successfully rely on it in ethnic differentiation. Thus, voice quality could be grouped with rhythm and HRTs under the *Useful Cues* in Figure 6.3 above.

Listening to the recordings used in the perception task, it appears that four of our ten Māori speakers in particular display salient features of this 'Māori voice quality' (m07, m18, m19, m21). These speakers are significantly more accurately identified in the low-pass filtered condition than in Condition Five. The case of a young Pākehā male speaker (p09) also confirms our hypothesis that voice quality serves as an important cue in the perception of Māori English and Pākehā English. This speaker is one of the most syllable-timed Pākehā used in this study (PVI = 52.1), which explains why his accuracy rate is only 30% in Condition Five, in other words, he is mistakenly identified as Māori 70% of the time based on his rhythmic and intonational properties. However, in Condition Six, where listeners are able to tune into his lack of 'Māori voice quality', there is over a 50% increase in his correct identification as Pākehā. This seems to suggest that in certain cases voice quality overrides the importance of rhythm as a marker of ethnicity. The accuracy rates for the above mentioned speakers in Conditions Five and Six are presented in

Table 6.4. The table includes speaker gender information, showing that the voice quality features in question is not exclusive to male speakers.

Table 6.4: Accuracy rates in Conditions Five and Six for selected speakers.

Speaker		Accuracy of ethnic identification	
ID	Gender	Condition Five Rhythm + Intonation	Condition Six Rhythm + Intonation + Voice Quality
m07	male	64%	93%
m18	female	53%	74%
m19	female	46%	68%
m21	male	74%	93%
p09	male	30%	65%

Much insight would be gained from an extensive acoustic study on the voice quality differences between Māori and Pākehā in a production experiment, while a perception experiment could investigate how voice quality is exploited for ethnic identification in New Zealand. The coverage of rhythm and pitch dynamics without taking corresponding modulation of the voice source into account is inevitably a partial treatment of prosodic features. This is a difficult area to work in as obtaining reliable measurements is not at all straightforward (Ní Chasaide, Dalton, Ito and Gobl 2004). However, as mentioned in Section 2.3.1, there has been some research carried out on ethnic differences in voice quality. Walton and Orlikoff (1994) show that African American males show more jitter, significantly more shimmer and significantly lower harmonics-to-noise ratio, where a lower ratio indicates greater hoarseness. Jitter is measured by local F0 variation, while shimmer is measured by local amplitude variation. In a perception experiment, Thomas and Lass (2005) found that greater breathiness was associated with African American male speakers. Studies like these are much needed and might yield similar results in New Zealand.

Chapter VII

Conclusion

The aim of this thesis was twofold. Firstly, it set out to explore possible prosodic differences between the two main ethnic varieties of New Zealand English. To investigate this, a production experiment was carried out using 36 speakers. Secondly, it aimed to establish whether listeners are able to identify Māori English and Pākehā English from suprasegmental cues only. For this reason, a perception experiment was conducted using 107 participants.

The production experiment had a special emphasis on the analysis of syllabic rhythm. The results confirm the findings of previous research and show that Māori English is significantly more syllable-timed than Pākehā English. The results also provide apparent-time evidence that New Zealand English is currently undergoing a rhythmic shift, as younger speakers of both Māori and Pākehā English produced significantly more syllable-timed speech than older speakers. Only time will tell whether Pākehā English rhythm will catch up with Māori English, or whether the two dialects will remain distinct with regards to their rhythmic properties.

Another - and somewhat unexpected - prosodic difference emerged from the production data. Māori participants produced significantly higher mean pitch values than Pākehā speakers. This also showed an age effect, such that Māori mean pitch seems to be getting higher over time, while Pākehā mean pitch appears to be lowering. Other features relating to pitch, such as SD of pitch or pitch range, did not show significant differences between the two ethnic dialects. The results regarding High Rising Terminals are consistent with previous research and indicate that, on average, Māori speakers use higher frequencies of HRTs than Pākehā speakers.

The perception experiment was designed to examine whether New Zealanders can correctly identify a speaker's ethnicity based on suprasegmental information only. It also aimed to investigate whether listeners are aware of the dialectal differences shown by the results of the production experiment and whether they can accurately use them to facilitate the ethnic identification of speakers.

Seven different speech conditions were created, each retaining different prosodic information. These included resynthesised flat rhythm only conditions, an intonation only, a rhythm plus intonation, as well as a low-pass filtered condition. Accuracy rates were predictably highest in the normal speech condition, where segmental information was also available. However, the results from the manipulated speech conditions revealed that listeners are able to tune into the rhythmic characteristics of a speaker and can use rhythm with some level of accuracy as a cue to identify whether the speaker is Māori or Pākehā. The results have also shown that listeners who have had greater exposure to Māori English are able to rely on rhythm more accurately than those participants with low Māori integration. In fact, highly integrated participants were shown to perform much better at a dialect identification task, even in degraded listening conditions.

The percentage of High Rising Terminals used by the speaker also proved to be a crucial cue in identifying Māori English from Pākehā English. The results revealed that listeners are aware of the fact that Māori speakers produce a higher percentage of HRTs overall and they make their ethnic judgements accordingly. An exciting interaction between rhythm and HRTs was also discovered in the perception experiment. Listeners are able to utilise different cues for dialect identification based on what is available in the particular speaker's speech. Those speakers who are very syllable-timed will be identified as Māori based on their rhythmic patterns. However, if the speaker is more stress-timed, then listeners will rely on the percentage of HRTs used by the speaker to facilitate dialect identification.

Not all suprasegmental cues were equally useful for the listener in the perception task. Some cues, such as mean pitch, were clearly misleading due to a clash between production and perception. Taken together, the results from the production and the perception experiment suggest that listeners are not consciously aware of the significantly higher mean pitch values of Māori English. In fact, New Zealanders expect Māori speakers to be speaking with a low mean pitch. As there seems to be a change over time in production with regards to mean pitch, it is possible that with time listeners' perception will adjust and they will become consciously aware of the difference in mean pitch between the two ethnic varieties of New Zealand English.

The present research did not carry out an acoustic analysis of voice quality, nor did it investigate its possible relevance in a dialect identification task. However, comparing the accuracy levels across different speech conditions and different speakers suggests that it is likely to be of great importance. Therefore, further research on the differing properties of Māori and Pākehā English would greatly benefit from a detailed analysis of voice quality, both in production and perception.

Overall, this thesis has contributed to the literature on Māori English by investigating its prosodic qualities and has been able to reveal the degree to which New Zealanders are able to use particular linguistic features to distinguish between Māori and Pākehā speakers. Methodologically, it has pioneered innovative techniques to isolate the precise features that listeners use to identify the ethnicity of a speaker in the New Zealand context.

References

Abercrombie, D. (1967), *Elements of General Phonetics*, Edinburgh University Press, Edinburgh.

Ainsworth, H. (1993), 'Rhythm in New Zealand English', Manuscript, Victoria University of Wellington.

Bauer, L. (1994), English in New Zealand, *in* R. Burchfield, ed., 'The Cambridge History of the English Language', Vol. 5, Cambridge University Press, Cambridge, pp. 382–429.

Bauer, W. (1997), *The Reed Reference Grammar of Māori*, Reed Publishing, Auckland.

Bayard, D. (1995), *Ethnicity, accent, and 'One New Zealand': He Iwi Tahi?*, Dunmore Press, chapter 7 in Kiwi Talk.

Beckman, M. E. (1992), Evidence for speech rhythms across languages, *in* Y. Tohura, E. Vatikiotis-Bateson and Y. Sagisaka, eds, 'Speech Perception, Production and Linguistic Structure', Omsha, Tokyo, pp. 457–463.

Bell, A. (2000), Maori and Pakeha English: A case study, *in* A. Bell and K. Kuiper, eds, 'New Zealand English', Victoria University Press, Wellington, pp. 221–248.

Benton, R. (1966), *Research into the English Language Difficulties of Maori Schoolchildren 1963–1964*, Maori Education Foundation, Wellington.

Bezooijen, R. and Boves, L. (1986), 'The Effects of Low-Pass Filtering and Random Splicing on the Perception of Speech', *Journal of Psycholinguistic Research* **15**, 403–417.

Bezooijen, R. and Gooskens, C. (1999), 'Identification of Language Varieties: The Contribution of Different Linguistic Levels', *Journal of Language and Social Psychology* **18**, 31–48.

Biemans, M. (2000), *Gender variation in voice quality*, Vol. 38 of *Dissertation Series*, LOT (Landelijke Onderzoekschool Taalwetenschap).

Bloch, B. (1950), 'Studies in colloquial Japanese IV: Phonemeics', *Language* **26**, 86–125.

Boersma, P. and Weenink, D. (2006), 'Praat: doing phonetics by computer (Version 4.4.12)'. [Computer program]. Retrieved 1 March 2006, from http://www.praat.org/.

Britain, D. (1992), 'The Use of High Rising Terminals in New Zealand English', *Language Variation and Change* .

Brizendine, L. (2006), *The Female Brain*, Broadway.

Bybee, J. (2001), *Phonology and Language Use*, Cambridge: Cambridge University Press.

Clopper, C. G. (2004), Linguistic Experience and the Perceptual Classification of Dialect Variation, PhD thesis, Indiana University.

Crystal, D. (1996), 'The past, present and future of English rhythm', *Speak Out!* **18**, 8–13.

Daly, N. and Warren, P. (2001), 'Pitching it differently in New Zealand English: speaker sex and intonation patterns', *Journal of Sociolinguistics* **5**, 85–96.

Dauer, R. M. (1987), 'Phonetic and phonological components of language rhythm', Paper presented at the 11th International Congress of Phonetic Sciences.

de Mareüil, P. B., Brahimi, B. and Gendrot, C. (2004), Role of segmental and suprasegmental cues in the perception of Maghrebian-accented French, *in* 'Interspeech 2004-ICSLP. 8th International Conference on Spoken Language Processing', Jeju, Korea, pp. 341–344.

de Saint-Exupery, A. (1943), *The Little Prince*.

Dellwo, V. (2004), Rhythm & Speech Rate: A varation coefficient for deltaC, *in* 'Proceedings of the 38th linguistic Colloquium', Budapest.

Dickens, M. and Sawyer, G. (1952), 'An Experimental Comparison of Vocal Quality among Mixed Groups of Whites and Negroes', *Southern Speech Journal* **17**, 178–185.

Drager, K. (2005), The influence of social characteristics on speech perception, Master's thesis, University of Canterbury.

Eckert, P. (2000), *Linguistic Variation and Social Practice*, Blackwell Publishers Ltd., Massachusetts.

Edwards, J. (1999), 'Refining Our Understanding of Language Attitudes', *Journal of Language and Social Psychology* **18**, 101–110.

Eek, A. and Help, T. (1987), The interrelationship between phonological and phonetic sound changes: a great rhythm shift of Old Estonian, *in* 'Proceedings of the 11th International Congress of Phonetic Sciences', Tallin, Estonia, pp. 218–233.

Foreman, C. (2000), 'Identification of African-American english from Prosodic Cues', *Texas Linguistic Forum* **43**, 57–66.

Fought, C. (2001), Ethnicity, *in* J. Chambers, P. Trudgill and N. Schilling-Estes, eds, 'Handbook of Language Variation and Change', Blackwell Publishers Ltd., chapter 18, pp. 445–472.

Fought, C. (2006), *Language and Ethnicity*, Cambridge University Press.

Frota, S. and Vigário, M. (2001), 'On the correlates of rhythmic distinctions: The European / Brazilian Portuguese case', *Probe* **13**, 247–275.

Frota, S., Vigário, M. and Martins, F. (2002), Language Discrimination and Rhythm Classes: Evidence from Portuguese, *in* 'Proceedings of the 1st International Conference on Speech Prosody', Aix-en-Provence, France, pp. 315–318.

Giles, H., Williams, A., Mackie, D. M. and Rosselli, F. (1995), 'Reactions to Anglo- and Hispanic-American accented speaker: Affect, identity, persuasion, and the English-only controversy', *Language and Communication* **15**, 107–120.

Gooskens, C. and van Bezooijen, R. (2002), The role of prosodic and verbal aspects of speech in the perceived divergence of Dutch and English language varieties, *in* J. Berns and J. van Marle, eds, 'Present-day Dialectology. Problems and Findings', Mouton de Gruyter, Berlin, pp. 173–192.

Gordon, E., Hay, J. and Maclagan, M. (forthcoming), The ONZE corpus, *in* J. C. Beal, K. P. Corrigan and H. Moisl, eds, 'Models and methods in the handling of unconventional digital corpora', Palgrave.

Grabe, E. and Low, E. (2002), Durational Variability in Speech and the Rhythm Class Hypothesis, *in* C. Gussenhoven and N. Warner, eds, ‘Papers in Laboratory Phology’, Vol. 7, Mouton de Gruyter, pp. 515–546.

Gussenhoven, C. (2002), Intonation and Interpretation: Phonetics and Phonology, *in* B. Bel and I. Marlien, eds, ‘Proceedings of Speech Prosody 2002’, Aix-en-Provence, France, pp. 47–57.

Han, M. S. (1962), ‘The feature of duration in Japanese’, *Onsei no kenkyuu* **10**, 65–80.

Hawkins, F. D. (1992), Speaker Ethnic Identification: The Roles of Speech Sample, Fundamental Frequency, Speaker and Listener Variation s, PhD thesis, University of Maryland at College Park.

Hay, J., Nolan, A. and Drager, K. (2006), ‘From Fush to Feesh: Exemplar Priming in Speech Perception’, *The Linguistic Review* **23**(3), 351–379.

Holmes, J. (1996), ‘Losing voice: is final /z/ devoicing a feature of Maori English?’, *World Englishes* **15**(2), 193–205.

Holmes, J. (2005), ‘Using Māori English in New Zealand’, *International Journal of the Sociology of Language* **172**, 91–115.

Holmes, J. and Ainsworth, H. (1996), ‘Syllable-timing and Maori English’, *Te Reo: Journal of the Linguistic Society of New Zealand* **39**, 75–84.

Holmes, J. and Ainsworth, H. (1997), ‘Unpacking the research process: investigating syllable-timing in New Zealand English’, *Language Awareness* **6**(1), 32–47.

Holmes, K., Murachver, T. and Bayard, D. (2001), ‘Accent, Appearance and Ethnic Stereotypes in New Zealand’, *New Zealand Journal of Psychology* **30**(2), 79–86.

Huygens, I. and Vaughan, G. (1983), ‘Language attitudes, ethnicity and social class in New Zealand’, *Journal of Multilingual and Multicultural Development* **4**, 207–223.

Johnson, K. (1997), Speech perception without speaker normalization: An exemplar model, *in* K. Johnson and J. W. Mullennix, eds, ‘Talker Variability in Speech Processing’, San Diego: Academic Press, pp. 145–165.

King, J. (1993), 'Maori English: a phonological study', *New Zealand English Newsletter* **7**, 33–47.

King, J. (1999), 'Talking Bro: Māori English in the University Setting', *Te Reo: Journal of the Linguistic Society of New Zealand* **204**, 20–38.

Labov, W. (1972), *Sociolinguistic patterns*, PA: University of Pennsylvania Press.

Labov, W. (1990), 'The intersection of sex and social class in the course of linguistic change', *Language Variation and Change* **2**, 205–254.

Ladefoged, P. (1975), *A course in phonetics*, Harcourt Brace Jovanovich, New York.

Lass, N. J., Almerino, C. A., Jordan, L. F. and Walsh, J. M. (1980), 'The Effect of Filtered Speech Alterations on Speaker Race and Sex Identifications', *Journal of Phonetics* **8**, 101–112.

Lass, N. J., Mertz, P. J. and Kimmel, K. L. (1978), 'The Effect of Temporal Speech Alterations on Speaker Race and Sex Identifications', *Language and Speech* **21**, 279–290.

Laver, J. (1994), *Principles of Phonetics*, Cambridge University Press, Cambridge.

Leyden, K. v. (2004), Prosodic Characteristics of Orkney and Shetland Dialects. An Experimental Approach., PhD thesis, Leiden University.

Lloyd James, A. (1940), *Speech signals in telephony*, London.

Loveday, L. (1981), 'Pitch, politeness, and sexual role: an explanatory investigation', *Language and Speech* **24**, 71–88.

Low, E. L. and Grabe, E. (1995), Prosodic patterns in Singapore English, *in* 'Proceedings of the Intonational Congress of Phonetic Sciences', Vol. 3, Stockholm, pp. 636–639.

Low, E. L., Grabe, E. and Nolan, F. (2000), 'Quantitative Characterizations of Speech Rhythm: Syllable-Timing in Singapore English', *Language and Speech* **43**(4), 377–401.

Majewski, W., Hollien, H. and Zalewski, J. (1972), 'Speaking fundamental frequencies of Polish adult males', *Phonetica* **25**, 119–125.

Miller, M. (1984), 'On the perception of rhythm', *Journal of Phonetics* **12**, 75–83.

Milroy, L. (1987), *Language and Social Networks*, Blackwell Publishers Ltd., Oxford.

Nesdale, D. and Rooney, R. (1996), 'Evaluations and stereotyping of accented speakers by pre-adolescent children', *Journal of Language and Social Psychology* **15**, 133–154.

Ní Chasaide, A., Dalton, M., Ito, M. and Gobl, C. (2004), Analysing Irish prosody: A dual linguistic/quantitative approach, *in* 'Proceedings of the SALTMIL Workshop: First Steps in Language Documentation for Minority Languages', Lisbon, pp. 60–63.

Oksenberg, L., Coleman, L. and Cannell, C. F. (1986), 'Interviewers' Voices and Refusal Rates in Telephone Surveys', *The Public Opinion Quarterly* **50**(1), 97–111.

Pierrehumbert, J. (2001), Exemplar dynamics: Word frequency, lenition and contrast, *in* J. Bybee and P. Hopper, eds, 'Frequency effects and the emergence of lexical structure', John Benjamins, Amsterdam, pp. 137–157.

Pike, K. (1946), *The Intonation of American English*, 2nd edn, University of Michigan Press, Ann Arbour.

Preston, D. R. (1986), 'Five visions of America', *Language in Society* **15**, 221–240.

R-Development-Core-Team (2004), 'R: A language and environment for statistical computing', R Foundation for Statistical Computing. URL http://www.R-project.org.

Ramus, F. and Mehler, J. (1999), 'Language identification with suprasegemntal cues: A study based on speech resynthesis', *Journal of the Acoustical Society of America* **105**(1), 512–521.

Ramus, F., Nespor, M. and Mehler, J. (1999), 'Correlates of linguistic rhythm in the speech signal', *Cognition* **73**, 265–292.

Roach, P. (1982), On the distinction between stress-timed and syllable-timed languages, *in* D. Crystal, ed., 'Linguistic Controversies', Edward Arnold, London, pp. 78–98.

Robertson, S. (1994), Identifying Maori English: A study of ethnic identification, attitudes, and phonetic features, Master's thesis, Victoria University of Wellington, New Zealand.

Stanton, K. (2006), A Study of High Rising Terminals in Various Dialects of English. Unpublished LING303 Class Project, University of Canterbury.

Street, R. L., Brady, R. M. and Putman, W. B. (1983), 'The Influence of Speech Rate Stereotypes and Rate Similarity or Listeners' Evaluations of Speakers', *Journal of Language and Social Psychology* **2**(1), 37–56.

Stroud, R. V. (1956), A Study of the Relations between Social Distance and Speech Differences of White and Negro High School Students of Dayton, Ohio, Master's thesis, Bowling Green State University.

Szakay, A. (2006*a*), 'Identifying Māori English from suprasegmental cues', Paper presented at the 10th New Zealand Language and Society Conference.

Szakay, A. (2006*b*), Rhythm and pitch as markers of ethnicity in New Zealand English, *in* P. Warren and C. Watson, eds, 'Proceedings of the 11th Australasian International Conference on Speech Science & Technology', University of Auckland, pp. 421–426.

Tamasi, S. L. (2003), Cognitive Patterns of Linguistic Perceptions, PhD thesis, University of Georgia.

Thomas, E. R. and Lass, N. J. (2005), 'Cues for Identification of African American and European American Speakers', Poster presented at the Annual Convention of the American Speech-Language-Hearing Association.

Thomas, E. R. and Reaser, J. (2002), 'Delimiting perceptual cues used for the ethnic labeling of African American and European American voices', *Journal of Sociolinguistics* **8**(1), 54–87.

Walker, A. (2006), A pitch of class: Pitch and social factors in New Zealand English. Paper presented at the 3rd International Postgraduate Linguistics Conference, Victoria University of Wellington.

Walton, J. and Orlikoff, R. (1994), 'Speaker race identification from acoustic cues in the vocal signal', *Journal of American Speech and Hearing Research* **37**, 738–745.

Warren, P. (1998), 'Timing Patterns in New Zealand English Rhythm', *Te Reo: Journal of the Linguistic Society of New Zealand* **41**, 80–93.

Warren, P. and Britain, D. (2000), Intonation and prosody in New Zealand English, *in* A. Bell and K. Kuiper, eds, 'New Zealand English', John Benjamins, Amsterdam, pp. 146–172.

Wilson, T. C. (1996), 'Cohort and prejudice: Whites' attitudes towards Blacks, Hispanics, Jews, and Asians', *The Public Opinion Quarterly* **60**, 253–274.

Appendix A

Average accuracy ratings for all participants

Table A.1: Summary of accuracy ratings for each participant in all conditions.

Participant				Conditions							
ID	ethn	gender	MII	One	Two	Three	Four	Five	Six	Seven	All
1	p	m	0	0.45	0.5	0.4	0.35	0.55	0.55	0.9	0.53
2	p	f	0	0.5	0.4	0.5	0.3	0.55	0.7	0.95	0.56
3	p	f	2.5	0.55	0.35	0.5	0.4	0.6	0.6	0.95	0.56
4	p	f	1	0.4	0.45	0.45	0.3	0.4	0.5	0.95	0.49
5	p	f	1	0.35	0.55	0.45	0.8	0.4	0.7	0.95	0.60
6	m	m	13.5	0.65	0.35	0.65	0.55	0.65	0.55	0.95	0.62
7	m	f	13	0.2	0.5	0.6	0.55	0.75	0.65	0.95	0.60
8	p	m	10.5	0.6	0.45	0.5	0.5	0.6	0.75	1	0.63
9	p	m	1.5	0.4	0.45	0.45	0.3	0.6	0.6	0.95	0.54
10	p	f	3	0.4	0.6	0.45	0.4	0.5	0.5	0.9	0.54
11	p	f	1	0.45	0.45	0.4	0.4	0.65	0.65	0.95	0.56
12	p	f	4.5	0.45	0.55	0.55	0.55	0.7	0.65	0.9	0.62
13	p	f	3	0.55	0.5	0.65	0.45	0.5	0.8	1	0.64
14	p	f	1.5	0.5	0.65	0.75	0.6	0.55	0.55	0.9	0.64
15	m	m	3	0.4	0.35	0.45	0.35	0.6	0.45	0.85	0.49
16	m	f	9	0.4	0.45	0.55	0.35	0.6	0.55	0.85	0.54
17	p	m	3	0.5	0.25	0.55	0.5	0.7	0.5	0.8	0.54
18	p	m	1.5	0.55	0.55	0.4	0.65	0.45	0.45	0.9	0.56
19	p	m	2.5	0.55	0.45	0.6	0.7	0.5	0.7	0.85	0.62
20	p	f	0	0.7	0.55	0.35	0.8	0.55	0.75	0.75	0.64
21	p	f	0	0.55	0.55	0.5	0.45	0.55	0.6	0.9	0.59
22	p	m	3	0.55	0.45	0.55	0.15	0.45	0.7	0.75	0.51
23	p	m	2.5	0.5	0.4	0.45	0.5	0.4	0.45	0.8	0.50
24	p	f	4	0.4	0.6	0.6	0.6	0.45	0.75	0.85	0.61
25	m	m	9	0.35	0.5	0.2	0.6	0.55	0.55	0.95	0.53
26	m	m	13.5	0.3	0.65	0.4	0.35	0.35	0.75	1	0.54
27	m	m	12	0.25	0.8	0.4	0.6	0.6	0.55	0.95	0.59
28	p	f	7.5	0.75	0.65	0.45	0.45	0.55	0.5	0.65	0.57
29	p	f	0	0.6	0.6	0.6	0.35	0.4	0.6	0.95	0.59
30	p	m	2.5	0.2	0.55	0.7	0.65	0.5	0.75	1	0.62

Continued on next page

Participant				Conditions							
ID	ethn	gender	MII	One	Two	Three	Four	Five	Six	Seven	All
31	p	f	1.5	0.5	0.55	0.4	0.55	0.6	0.8	0.9	0.61
32	p	f	1.5	0.5	0.6	0.6	0.45	0.5	0.6	0.9	0.59
33	p	f	2.5	0.55	0.6	0.4	0.6	0.5	0.9	0.95	0.64
34	p	f	5.5	0.3	0.35	0.55	0.6	0.7	0.65	0.95	0.59
35	m	m	14.5	0.55	0.6	0.75	0.7	0.5	0.55	0.95	0.66
36	m	m	9	0.5	0.5	0.3	0.4	0.55	0.45	0.85	0.51
37	m	f	11.5	0.5	0.6	0.5	0.55	0.45	0.5	0.95	0.58
38	m	m	14	0.3	0.4	0.35	0.65	0.55	0.7	0.9	0.55
39	m	f	14	0.5	0.5	0.65	0.55	0.5	0.4	0.95	0.58
40	m	f	16	0.75	0.65	0.55	0.45	0.45	0.55	0.95	0.62
41	m	f	11.5	0.4	0.45	0.55	0.7	0.55	0.7	0.85	0.60
42	m	f	16	0.4	0.4	0.6	0.7	0.3	0.5	0.8	0.53
43	p	f	1	0.2	0.45	0.55	0.5	0.5	0.75	1	0.56
44	p	f	0	0.6	0.4	0.65	0.5	0.5	0.7	0.85	0.60
45	p	m	1.5	0.55	0.45	0.4	0.25	0.25	0.6	0.9	0.49
46	p	f	6.5	0.55	0.5	0.4	0.45	0.35	0.6	0.85	0.53
47	m	m	4	0.6	0.55	0.55	0.55	0.4	0.75	1	0.63
48	p	f	5	0.45	0.5	0.5	0.5	0.65	0.75	0.85	0.60
49	m	m	4.5	0.4	0.35	0.45	0.6	0.6	0.4	0.9	0.53
50	p	f	1.5	0.45	0.4	0.3	0.45	0.5	0.4	0.85	0.48
51	p	m	1.5	0.45	0.45	0.65	0.25	0.4	0.35	0.85	0.49
52	p	f	1.5	0.4	0.4	0.5	0.4	0.6	0.6	0.85	0.54
53	m	m	5.5	0.75	0.5	0.55	0.35	0.7	0.65	0.7	0.60
54	p	m	5.5	0.5	0.3	0.5	0.3	0.5	0.65	0.85	0.51
55	p	f	13.5	0.45	0.8	0.7	0.7	0.8	0.9	1	0.76
56	m	f	8	0.6	0.4	0.55	0.6	0.5	0.75	0.95	0.62
57	p	m	7.5	0.4	0.5	0.65	0.45	0.45	0.4	1	0.55
58	p	m	11.5	0.3	0.45	0.7	0.65	0.55	0.55	0.95	0.59
59	m	m	11.5	0.55	0.55	0.45	0.5	0.55	0.85	0.95	0.63
60	m	m	13	0.55	0.65	0.55	0.55	0.7	0.7	0.95	0.66
61	m	m	14.5	0.55	0.35	0.45	0.55	0.35	0.7	0.95	0.56
62	m	f	11.5	0.5	0.5	0.45	0.45	0.5	0.85	0.95	0.60
63	m	f	14	0.6	0.55	0.55	0.6	0.6	0.75	1	0.66

Continued on next page

Participant				Conditions							
ID	ethn	gender	MII	One	Two	Three	Four	Five	Six	Seven	All
64	m	f	14.5	0.6	0.55	0.55	0.6	0.65	0.75	1	0.67
65	m	f	12	0.55	0.5	0.45	0.6	0.6	0.75	0.95	0.63
66	m	m	10	0.3	0.4	0.35	0.65	0.6	0.7	0.9	0.56
67	m	f	13	0.5	0.5	0.6	0.6	0.6	0.65	0.95	0.63
68	m	f	13	0.3	0.6	0.55	0.55	0.6	0.75	1	0.62
69	m	f	6	0.45	0.5	0.4	0.4	0.45	0.6	0.95	0.54
70	m	m	16	0.55	0.5	0.55	0.6	0.65	0.75	1	0.66
71	p	m	10	0.6	0.45	0.5	0.55	0.6	0.75	1	0.64
72	p	m	3	0.4	0.55	0.45	0.4	0.5	0.5	0.9	0.53
73	p	m	2.5	0.5	0.4	0.45	0.5	0.4	0.45	0.8	0.50
74	m	m	4	0.5	0.45	0.5	0.5	0.4	0.45	0.85	0.52
75	p	m	1	0.45	0.5	0.4	0.4	0.55	0.55	0.9	0.54
76	p	f	6	0.45	0.5	0.55	0.5	0.65	0.75	0.85	0.61
77	m	m	16	0.45	0.8	0.65	0.7	0.8	0.9	1	0.76
78	m	f	16	0.5	0.75	0.65	0.7	0.8	0.9	1	0.76
79	p	m	9	0.6	0.45	0.5	0.5	0.6	0.75	1	0.63
80	p	f	3	0.5	0.45	0.5	0.3	0.55	0.75	0.95	0.57
81	p	m	1	0.45	0.5	0.4	0.4	0.55	0.55	0.9	0.54
82	p	f	4	0.45	0.5	0.55	0.5	0.65	0.75	0.85	0.61
83	m	m	14.5	0.45	0.8	0.65	0.7	0.8	0.9	1	0.76
84	m	f	16	0.5	0.75	0.65	0.7	0.8	0.9	1	0.76
85	p	m	9	0.6	0.45	0.5	0.5	0.6	0.75	1	0.63
86	p	f	3	0.5	0.45	0.5	0.3	0.55	0.75	0.95	0.57
87	p	m	2	0.45	0.5	0.4	0.35	0.55	0.55	0.9	0.53
88	p	f	0	0.5	0.4	0.5	0.3	0.55	0.7	0.95	0.56
89	p	f	2.5	0.55	0.35	0.5	0.4	0.6	0.6	0.95	0.56
90	p	f	1	0.4	0.45	0.45	0.3	0.4	0.5	0.95	0.49
91	p	f	1	0.35	0.55	0.45	0.8	0.4	0.7	0.95	0.60
92	m	m	11	0.55	0.55	0.45	0.5	0.55	0.85	0.95	0.63
93	m	m	13.5	0.55	0.65	0.55	0.55	0.7	0.75	1	0.68
94	m	m	14.5	0.55	0.35	0.45	0.6	0.6	0.75	1	0.61
95	m	f	12	0.5	0.5	0.45	0.45	0.6	0.9	0.95	0.62
96	m	m	10	0.5	0.5	0.35	0.55	0.55	0.6	0.9	0.56

Continued on next page

Participant				Conditions							
ID	ethn	gender	MII	One	Two	Three	Four	Five	Six	Seven	All
97	m	f	11.5	0.5	0.6	0.5	0.55	0.45	0.5	0.95	0.58
98	m	m	14	0.3	0.4	0.35	0.65	0.65	0.7	0.95	0.57
99	m	f	15	0.55	0.55	0.65	0.6	0.6	0.75	1	0.67
100	m	f	16	0.75	0.65	0.55	0.6	0.6	0.75	0.95	0.69
101	m	f	11.5	0.4	0.45	0.55	0.7	0.55	0.7	0.85	0.60
102	m	f	9	0.5	0.4	0.5	0.6	0.55	0.65	0.85	0.58
103	m	m	11.5	0.55	0.55	0.45	0.5	0.55	0.85	0.95	0.63
104	m	m	13	0.55	0.65	0.55	0.6	0.7	0.75	0.95	0.68
105	m	m	14.5	0.55	0.35	0.45	0.55	0.35	0.7	0.95	0.56
106	m	m	12	0.5	0.5	0.45	0.6	0.65	0.85	0.95	0.64
107	m	f	15.5	0.6	0.55	0.55	0.6	0.7	0.8	1	0.69

Appendix B

Average accuracy ratings for all speakers

Table B.1: Summary of accuracy ratings for each speaker in all conditions by Maori and Pakeha participants.

speaker	Condition 1		Condition 2		Condition 3		Condition 4		Condition 5		Condition 6		Condition 7		Total
	m	p	m	p	m	p	m	p	m	p	m	p	m	p	
m02	0.52	0.55	0.73	0.65	0.60	0.36	0.67	0.38	0.63	0.58	0.92	0.67	1.00	1.00	0.66
m03	0.48	0.44	0.40	0.35	0.31	0.38	0.73	0.49	0.38	0.62	0.50	0.53	1.00	0.89	0.54
m07	0.46	0.55	0.48	0.42	0.42	0.60	0.67	0.65	0.77	0.51	0.96	0.91	1.00	1.00	0.67
p09	0.38	0.35	0.37	0.40	0.65	0.31	0.29	0.33	0.29	0.31	0.69	0.62	0.87	0.98	0.49
m10	0.50	0.36	0.69	0.53	0.50	0.53	0.50	0.47	0.60	0.65	0.65	0.89	0.96	0.93	0.63
m18	0.56	0.53	0.42	0.62	0.56	0.49	0.69	0.42	0.58	0.49	0.83	0.65	0.62	0.36	0.56
m19	0.56	0.58	0.38	0.44	0.44	0.53	0.52	0.55	0.42	0.49	0.73	0.64	0.96	1.00	0.59
m20	0.48	0.56	0.67	0.55	0.62	0.64	0.67	0.51	0.69	0.67	0.81	0.42	0.79	0.47	0.61
m21	0.58	0.55	0.79	0.56	0.63	0.56	0.73	0.24	0.90	0.58	0.94	0.91	1.00	1.00	0.71
p23	0.44	0.60	0.65	0.47	0.60	0.49	0.38	0.29	0.42	0.62	0.54	0.73	0.98	0.95	0.58
p24	0.62	0.47	0.50	0.62	0.46	0.55	0.58	0.45	0.63	0.56	0.67	0.69	1.00	0.98	0.63
p25	0.52	0.47	0.69	0.64	0.48	0.71	0.71	0.60	0.52	0.69	0.75	0.78	1.00	0.96	0.68
p26	0.23	0.49	0.54	0.47	0.54	0.65	0.42	0.73	0.81	0.49	0.67	0.53	0.92	0.96	0.60
m27	0.67	0.36	0.42	0.42	0.54	0.38	0.77	0.49	0.56	0.31	0.58	0.58	0.94	0.91	0.56
m28	0.63	0.45	0.62	0.35	0.65	0.47	0.46	0.31	0.38	0.38	0.60	0.44	1.00	1.00	0.55
p31	0.42	0.53	0.52	0.55	0.38	0.56	0.52	0.35	0.65	0.58	0.85	0.84	0.98	1.00	0.62
p33	0.44	0.31	0.27	0.40	0.38	0.69	0.29	0.53	0.29	0.45	0.52	0.47	0.83	0.82	0.48
p34	0.60	0.49	0.38	0.53	0.58	0.40	0.65	0.42	0.81	0.55	0.42	0.42	1.00	0.98	0.59
p35	0.54	0.62	0.71	0.45	0.37	0.44	0.63	0.58	0.56	0.64	0.42	0.51	0.98	0.98	0.60
p36	0.21	0.33	0.29	0.29	0.44	0.38	0.37	0.58	0.62	0.42	0.65	0.45	0.94	0.93	0.49
Total	0.49	0.48	0.53	0.48	0.51	0.51	0.56	0.47	0.58	0.53	0.69	0.63	0.94	0.91	0.59

Appendix C

Reading Passage

"And now of course six years have already passed. I have never told this story before. The friends who saw me again on my return were very happy to see me alive. I seemed sad but I said to them: 'It's exhaustion'. Now I have got over my loss a little, which is to say not entirely. But at least I know that he returned safely to his planet because I couldn't find his body in the morning."

Appendix D

Consent Form

UC
UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

CONSENT FORM

Dialect Identification Experiment

I have read and understood the description of the above named project. On this basis I agree to participate as a subject in the project and I consent to publication of the result of the project with the understanding that anonymity will be preserved.

I agree that the results of this experiment be:

1. held at the University of Canterbury linguistics archives
2. made available to bona fide researchers
3. quoted anonymously in published work
4. used for teaching purposes

I understand also that I may at any time withdraw from the project, including withdrawal of any information I have provided.

Signature:

Date:

Participant Number:

Appendix E

Information Sheet

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

INFORMATION SHEET

Dialect Identification Experiment

You are invited to participate in a research project investigating New Zealand English.

Your participation will involve listening to passages spoken by New Zealanders and indicating on the answer sheet whether you think the speaker sounded Pakeha or Māori.

You have the right to withdraw from the project at any time, including withdrawal of any information provided.

The results of the project may be published but you may be assured of the complete confidentiality of data gathered in this investigation: the identity of participants will not be made public without their consent. To ensure anonymity and confidentiality, you will be identified by number and not by name.

The project is carried out by Anita Szakay, who is a Master's student at the Department of Linguistics. She can be contacted at asz13@student.canterbury.ac.nz or at (03) 364-2987 ext 8036. She will be pleased to discuss any concerns you may have about participation in the project.

This project has been reviewed and approved by the Human Ethics Committee of the University of Canterbury.

Appendix F

Answer Sheet

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

INSTRUCTION SHEET

Dialect Identification Experiment

You are asked to listen to 140 short passages spoken by New Zealanders in English. Each speaker is talking on the topic of sport and rugby. After listening to each passage you will have to indicate on a scale of 1 to 4 how Pakeha or Māori sounding you judge the speaker to be. Most of the time this will seem impossible as the passages have been manipulated and you won't be able to understand what's being said. Just go with your first intuition and try your best. There are no 'right' or 'wrong' answers, we are simply interested in what you **think** about the speaker.

SECTION A

This section will take 40 minutes to complete. The passages will be presented in 7 blocks of 20. The first 4 blocks make up Part One with 80 passages. Part Two has the final 60 passages in 3 blocks. At the beginning of each new block you will be presented with an example passage taken from Romanian to illustrate how the manipulated passage relates to normal human speech. Each new condition sounds more speech-like than the previous ones with the final condition presenting the unmodified passages. So your task will get easier as you work through it.
Part One will take 22 minutes to complete while Part Two is 16 minutes long. There will be a 2 minute break between the two parts.
You will indicate your answer by circling a number. Example:

			very Pakeha sounding	somewhat Pakeha sounding	somewhat Māori sounding	very Māori sounding
#1	young	female	1	2	3	4

SECTION B

This section will take 2 minutes to complete. You will be asked to provide information about yourself. You will **not** be asked to provide your name or contact details.

Participant Number:

CONDITION 1

Example... (Romanian)

			very Pakeha sounding	somewhat Pakeha sounding	somewhat Māori sounding	very Māori sounding
# 1	older	female	1	2	3	4
# 2	young	male	1	2	3	4
# 3	older	female	1	2	3	4
# 4	older	male	1	2	3	4
# 5	young	male	1	2	3	4
# 6	older	female	1	2	3	4
# 7	young	male	1	2	3	4
# 8	young	female	1	2	3	4
# 9	older	male	1	2	3	4
# 10	young	male	1	2	3	4
# 11	older	male	1	2	3	4
# 12	young	female	1	2	3	4
# 13	young	female	1	2	3	4
# 14	young	male	1	2	3	4
# 15	young	female	1	2	3	4
# 16	young	male	1	2	3	4
# 17	older	male	1	2	3	4
# 18	older	male	1	2	3	4
# 19	older	female	1	2	3	4
# 20	older	male	1	2	3	4

CONDITION 2

			very Pakeha sounding	somewhat Pakeha sounding	somewhat Māori sounding	very Māori sounding
# 21	young	male	1	2	3	4
# 22	older	male	1	2	3	4
# 23	older	female	1	2	3	4
# 24	older	male	1	2	3	4
# 25	older	male	1	2	3	4
# 26	young	male	1	2	3	4
# 27	older	female	1	2	3	4
# 28	young	female	1	2	3	4
# 29	young	female	1	2	3	4
# 30	young	female	1	2	3	4
# 31	older	female	1	2	3	4
# 32	young	male	1	2	3	4
# 33	young	female	1	2	3	4
# 34	young	male	1	2	3	4
# 35	older	female	1	2	3	4
# 36	older	male	1	2	3	4
# 37	older	male	1	2	3	4
# 38	young	male	1	2	3	4
# 39	older	male	1	2	3	4
# 40	young	male	1	2	3	4

CONDITION 3

			very Pakeha sounding	somewhat Pakeha sounding	somewhat Māori sounding	very Māori sounding
# 41	older	male	1	2	3	4
# 42	young	male	1	2	3	4
# 43	older	female	1	2	3	4
# 44	young	male	1	2	3	4
# 45	young	male	1	2	3	4
# 46	older	female	1	2	3	4
# 47	older	male	1	2	3	4
# 48	older	female	1	2	3	4
# 49	older	male	1	2	3	4
# 50	older	male	1	2	3	4
# 51	young	male	1	2	3	4
# 52	young	female	1	2	3	4
# 53	young	female	1	2	3	4
# 54	young	female	1	2	3	4
# 55	older	male	1	2	3	4
# 56	young	female	1	2	3	4
# 57	young	male	1	2	3	4
# 58	older	male	1	2	3	4
# 59	young	male	1	2	3	4
# 60	older	female	1	2	3	4

CONDITION 4

Example... (Romanian)

			very Pakeha sounding	somewhat Pakeha sounding	somewhat Māori sounding	very Māori sounding
# 61	young	female	1	2	3	4
# 62	older	male	1	2	3	4
# 63	older	female	1	2	3	4
# 64	older	male	1	2	3	4
# 65	young	male	1	2	3	4
# 66	young	female	1	2	3	4
# 67	young	male	1	2	3	4
# 68	young	male	1	2	3	4
# 69	older	male	1	2	3	4
# 70	older	male	1	2	3	4
# 71	young	female	1	2	3	4
# 72	young	male	1	2	3	4
# 73	older	male	1	2	3	4
# 74	young	male	1	2	3	4
# 75	older	female	1	2	3	4
# 76	older	female	1	2	3	4
# 77	older	female	1	2	3	4
# 78	young	female	1	2	3	4
# 79	young	male	1	2	3	4
# 80	older	male	1	2	3	4

TWO MINUTE BREAK

CONDITION 5

Example... (Romanian)

			very Pakeha sounding	somewhat Pakeha sounding	somewhat Māori sounding	very Māori sounding
# 1	older	male	1	2	3	4
# 2	older	male	1	2	3	4
# 3	young	female	1	2	3	4
# 4	young	female	1	2	3	4
# 5	young	male	1	2	3	4
# 6	older	female	1	2	3	4
# 7	older	female	1	2	3	4
# 8	young	male	1	2	3	4
# 9	older	male	1	2	3	4
# 10	older	male	1	2	3	4
# 11	older	female	1	2	3	4
# 12	young	male	1	2	3	4
# 13	young	male	1	2	3	4
# 14	young	female	1	2	3	4
# 15	older	male	1	2	3	4
# 16	young	male	1	2	3	4
# 17	older	male	1	2	3	4
# 18	older	female	1	2	3	4
# 19	young	male	1	2	3	4
# 20	young	female	1	2	3	4

CONDITION 6

Example... (Romanian)

			very Pakeha sounding	somewhat Pakeha sounding	somewhat Māori sounding	very Māori sounding
# 21	young	male	1	2	3	4
# 22	older	male	1	2	3	4
# 23	older	female	1	2	3	4
# 24	older	male	1	2	3	4
# 25	older	male	1	2	3	4
# 26	young	male	1	2	3	4
# 27	older	female	1	2	3	4
# 28	young	female	1	2	3	4
# 29	young	female	1	2	3	4
# 30	young	female	1	2	3	4
# 31	older	female	1	2	3	4
# 32	young	male	1	2	3	4
# 33	young	female	1	2	3	4
# 34	young	male	1	2	3	4
# 35	older	female	1	2	3	4
# 36	older	male	1	2	3	4
# 37	older	male	1	2	3	4
# 38	young	male	1	2	3	4
# 39	older	male	1	2	3	4
# 40	young	male	1	2	3	4

CONDITION 7

			very Pakeha sounding	somewhat Pakeha sounding	somewhat Māori sounding	very Māori sounding
# 41	older	male	1	2	3	4
# 42	young	male	1	2	3	4
# 43	older	female	1	2	3	4
# 44	young	male	1	2	3	4
# 45	young	male	1	2	3	4
# 46	older	female	1	2	3	4
# 47	older	male	1	2	3	4
# 48	older	female	1	2	3	4
# 49	older	male	1	2	3	4
# 50	older	male	1	2	3	4
# 51	young	male	1	2	3	4
# 52	young	female	1	2	3	4
# 53	young	female	1	2	3	4
# 54	young	female	1	2	3	4
# 55	older	male	1	2	3	4
# 56	young	female	1	2	3	4
# 57	young	male	1	2	3	4
# 58	older	male	1	2	3	4
# 59	young	male	1	2	3	4
# 60	older	female	1	2	3	4

END OF SECTION A

Appendix G

Background Information Sheet

1. Which age group do you belong to?

   10–19 20–29 30–39 40–49 50–59 60+

2. You are:

   female male

3. Your ethnicity is:

   Māori Pakeha Other

4. If you have a partner, their ethnicity is:

   Māori Pakeha Other N/A

5. Your highest education is:

   primary school high school diploma degree

6. How well do you speak Te Reo Māori?

   0 (none) 1 (basic) 2 3 4 5 (fluent)

7. How often do you listen to Māori radio stations? (eg. Tahu FM)?

   never sometimes often

8. How often do you watch The Māori Television or other Māori TV programmes?

   never sometimes often

9. Do you ever visit a marae?

   never sometimes often

10. People you spend most of your time with (friends, colleagues etc...) are:

    Māori Pakeha Pasifika other

11. In general, to what extent do you perceive yourself to have been exposed to Māori English?

    never seldom sometimes often

END OF EXPERIMENT

Participant Number: