Copyright Statement

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

This thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognise the author's right to be identified as the author of this thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author's permission before publishing any material from their thesis.

To request permissions please use the Feedback form on our webpage. [http://researchspace.auckland.ac.nz/feedback](http://researchspace.auckland.ac.nz/feedback)

General copyright and disclaimer

In addition to the above conditions, authors give their consent for the digital copy of their work to be used subject to the conditions specified on the [Library Thesis Consent Form](http://researchspace.auckland.ac.nz/librarythesisconsentform) and [Deposit Licence](http://researchspace.auckland.ac.nz/depositlicence).
ELICITING AND ANALYSING PERCEPTIONS OF PROSODIC PROMINENCE

A MĀORI CASE STUDY

Laura Frances Thompson

Abstract

This project outlines a methodology developed for analysing perception of prosodic prominence based on Māori, the indigenous language of New Zealand. Development of the study began with a pilot, which employed a perception test involving a recorded tapping exercise. The second phase implemented a participant-friendly, purpose-built web survey, increasing efficiency and accessibility. The third phase increased the scale and developed a statistical processing method for determining significantly perceptually prominent syllables (PPMs) based on the perception test responses. Ninety-two New Zealand and overseas listeners responded to 30 short recordings of continuous Māori speech from the MAONZE database (King et al., 2011a). Since native language cues affect prominence perception in any language, participants were further divided into three self-rated Māori proficiency groups. Results indicated a pattern of difference between high proficiency and lower or zero proficiency participants, both in response to the perception task, the effects of which were handled by the result-processing, and in PPM identification, though the latter difference was not always significant. Analysis of PPMs from all groups indicated strong alignment with syllables where existing expected phrase and word stress overlap, and with word stress (but not phrase stress) alone; a sensitivity to heavy (2+ morae) syllable weight and whole syllable duration; and influence from F0 movement around the phrase F0 peak or from the F0 peak itself. There were indications that the strongest prominence effect is cumulative, but that word stress cues are, or have become, stronger than phrase stress cues, and that this may be increasing with time.
for my family

who have always listened
and told me
what stood out
Acknowledgements

A great many people have contributed to the completion of this project, and it is here that I attempt to express my sincere appreciation to them all. Without their generosity of time and spirit, this project could not have progressed. I have been very fortunate. It is the greatest wish (and most fervent hope) of any researcher doing this type of work that others will indulge them and jump on board, and I am indebted to all of the participants, both in New Zealand and overseas, who took the time to help with the studies by doing the perception tests and giving me feedback. My academic supervisors, Dr Catherine Watson and Dr Helen Charters, have shared their wisdom, experience, insight and expertise with me throughout the project. They have read, commented, discussed, and questioned; given me so much encouragement and motivation. They cannot know how many valuable things I have learned from our collaboration and their guidance. The MAONZE research team: Professor Ray Harlow (University of Waikato), Associate Professor Jeanette King (University of Canterbury), Professor Margaret Maclagan (University of Canterbury), Dr Peter Keegan (University of Auckland), and Dr Catherine Watson (University of Auckland) have given me not only data, but new opportunities, support, feedback and Māori language advice. On top of this, they have helped me with networking and been willing experimental subjects! They, and my fellow MAONZE students, Stephen Bier and Stephanie Kaefer, are a great bunch of people, and it has been a pleasure to work with them. Dr Winifred Bauer of Victoria University, Wellington, also gave me valuable help in networking to find more proficient Māori-speaking participants. I am in the debt of both Professor Amalia Arvaniti of the University of Kent and Dr Sasha Calhoun of Victoria University, Wellington for their insightful comments and thought-provoking discussion on the work. The anonymous reviewers and the less-anonymous audience members at various conferences and workshops; some FAB members of the Department of Electrical and Computer Engineering and the Department of Applied Language Studies and Linguistics at the University of Auckland; and the people at IPS München on that very hot day in 2011, have provided some of the most interesting and entertaining comments I have ever received on my research. This project was completed with the help of a University of Auckland Doctoral Scholarship, and I am grateful to the Statistical Consulting Centre in the Department of Statistics for advice on a subject to which I was very new. To Dr Donna Starks, thank you for starting me along a path and continuing to encourage me to pursue it. And last, but never, never, least, there are my family & friends, both near and so very far away. They have listened, smiled, and endlessly humoured an occasionally
irascible PhD student. Thank you all for your patience, love, support, and cheerleading; for your tolerance of the towers of books and papers; and for not being afraid to put a bomb under me when I needed it.
# Table of Contents

List of Figures ........................................................................................................... xiv

List of Tables ............................................................................................................. xvi

List of Glosses ........................................................................................................... xix

Glossary ..................................................................................................................... xx

1. Introduction ............................................................................................................ 1

2. Background: Project Concepts and the Māori Language ................................. 9
   2.1. Introduction ....................................................................................................... 9
   2.2. Māori Language: Overview ........................................................................ 10
      2.2.1. Phonemic Inventory and Syllable Structure .......................................... 12
      2.2.2. Syntactic Organization: The Māori Phrase ............................................ 13
   2.3. Project-Related Prosodic Concepts ............................................................... 15
      2.3.1. Core Concepts ....................................................................................... 15
      2.3.2. Rhythm, Groupings, and Timing ......................................................... 18
      2.3.3. Description and Measurement of Prosodic Features .......................... 22
   2.4. Māori Language: Prosodic Description ..................................................... 30
      2.4.1. Stress Assignment ............................................................................... 30
      2.4.2. Intonation ............................................................................................. 34
      2.4.3. Timing, Rhythm Class, and a Note on Mita ......................................... 36
   2.5. The MAONZE Database ............................................................................ 39
      2.5.1. Speakers ............................................................................................... 39
      2.5.2. Recordings ............................................................................................ 40
      2.5.3. Use of the Database in the Present Study ........................................... 40

3. Preliminary Studies: Developing the Experimental Protocol ....................... 41
   3.1. Introduction .................................................................................................... 41
   3.2. Pilot Study: Tapping ..................................................................................... 42
      3.2.1. Stimuli .................................................................................................... 42
3.2.2. Participants ................................................................. 43
3.2.3. Experiment Setup and Delivery Method ............................. 44
3.2.4. Result Processing ........................................................... 45
3.2.5. Results ........................................................................... 47
3.2.6. Evaluation of the Methodology and Proposed Modifications .......... 54
3.3. The Second Experiment: Prominence on the Web.................... 58
3.3.1. Stimuli ........................................................................... 58
3.3.2. Development of the Survey for the Web ................................. 60
3.3.3. Survey Participants ............................................................ 66
3.3.4. Results from the First Web Experiment ............................... 67
3.3.5. Evaluation of the Methodology and Proposed Modifications .......... 77

4. The Prominence Elicitation Tool: Main Study Methodology and Dataset .... 83
4.1. Introduction ........................................................................ 83
4.2. Stimuli ............................................................................... 84
4.3. Participants .......................................................................... 87
4.3.1. Background to the Proficiency Group Divisions .................. 89
4.4. Further Development of the Web Survey ................................ 92
4.4.1. Structure and Participant Access ........................................ 92
4.4.2. Managing Submissions .................................................... 94
4.4.3. Cleaning of the Raw Data ................................................ 95
4.5. Preparing the PROM Dataset for Analysis ............................. 95
4.5.1. Refining and Defining the Criteria for Perceived Prominences (PPMs) .... 96
4.5.2. Triplets, Peaks and the Peak Criterion ................................. 98
4.5.3. Triplet Mode: A Triplet-Based Statistical Approach to Refinement of the Peak List............................... 101
4.5.2. Response of the Proficiency Groups to the Perception Task ....... 104
4.5.2. Proficiency Group Effects on the Triplet Mode Statistics ......... 106
4.5.2. Quinlet Mode .................................................................. 108
4.6. The PROM Database: Construction and Analysis Methods ........... 110
4.6.1. Creating the PROM Database in Emu ................................ 110
4.6.2. Use of R to Query the Emu Database .................................... 114
4.6.3. Statistical Assessment of Results ........................................ 116
4.7. Results & Analysis: PROM Dataset Overview ....................... 117
6.8. Vowel Quality .............................................................................................................. 166
6.9. F0 Behaviour ............................................................................................................... 169
6.10. Summary .................................................................................................................... 172

7. Conclusion: Discussion, Evaluation, and Future Directions ......................... 177
  7.1. Introduction .................................................................................................................. 177
  7.2. Issues in Participant Prominence Identification ........................................................... 179
  7.3. Perceived Prominences: Distribution ........................................................................ 181
  7.4. Perceived Prominences: Stress Category Analysis ...................................................... 182
  7.5. Perceived Prominences: Acoustic Analysis ................................................................. 184
  7.6. Speaker Vintage Comparison ....................................................................................... 186
  7.7. Proficiency Group Comparison ................................................................................... 188
  7.8. Evaluation and Development ....................................................................................... 189
  7.9. Applications .................................................................................................................. 193

Appendix A: Ethics and Survey Material ................................................................. 196
  A1. Online Participant Information Sheet with Ethics Approval and
      Consent Checkbox ......................................................................................................... 196
  A2. Participant Questionnaire Used in the Web Surveys .................................................... 198

Appendix B: Stimuli Used in the Pilot Study and First Web Survey .......... 201
  B1. Full Text of Letter and Story and Full Translation of Story ......................................... 201
  B2. Glossing of Māori Stimuli ............................................................................................ 203

Appendix C: Continuous Speech Stimuli from the Main Experiment .......... 204

Appendix D: Syllable Nuclear Duration Tables for Group Z and A, B, C ...... 213

List of References .............................................................................................................. 215
List of Figures

Figure 3.1. Summary of the stimulus set for the tapping experiment ........................................ 43
Figure 3.2. Partial Emu signal view screencapture showing different levels and taps in database. .................................................................................................................. 46
Figure 3.3. Partial Emu hierarchy view screencapture showing different levels and taps in database. ........................................................................................................... 46
Figure 3.4. Tap distribution on Māori stimuli: filtered and unfiltered conditions. ................. 53
Figure 3.5. Tap distribution on English stimuli: filtered and unfiltered conditions. ............... 53
Figure 3.6. Stimuli used in the second experiment ................................................................. 59
Figure 3.7. The first web survey interface, with controls, syllabification, word-grouping, and checkboxes. .......................................................................................... 63
Figure 3.8. Example of the raw result text file generated by the web form. ......................... 64
Figure 3.9. Example of a section of the result spreadsheet. .................................................. 65
Figure 3.10. Utterance result chart showing more and less prominent syllables and sequences. .................................................................................................................. 68
Figure 3.11. Distribution of prominences identified in Māori stimuli across Biggs stress categories. ........................................................................................................... 71
Figure 3.12. Differences between proficiency groups in syllable box checking. .............. 74

Figure 4.1. Summary of sentence stimuli used in the second web survey. ......................... 85
Figure 4.2. Survey advertisement card prepared for distribution at a conference. .............. 87
Figure 4.3. Partial screencapture showing two stimuli from the second web survey. ........ 92
Figure 4.4. Result chart from the first web survey: non-adjacent distribution of >60% syllables; excluded <60% peaks. .................................................................................. 97
Figure 4.5. Triplet shapes. .................................................................................................... 98
Figure 4.6. Phrase with partial context, showing overlapping syllable triplets. ................. 98
Figure 4.7. Phrase e toru with partial context, showing overlapping syllable triplets. ...... 99
Figure 4.8. Example of a pair-top peak (wha-ka) from the Group Z results. .................... 100
Figure 4.9. Peak shapes. ..................................................................................................... 101
Figure 4.10. Rate of syllable selection by proficiency group. ........................................... 104
Figure 4.11. Examples of non-significant (Triplet Mode) peaks with very high selection rates. ................................................................................................................ 107
Figure 4.12. Triplet (grey brackets) and quinlet (black brackets) for the same
peak syllable. ...........................................................................................................108

Figure 4.13. Emu signal view: selected levels, waveform, spectrogram, and formant tracks..........................................................113

Figure 4.14. Emu hierarchy view, showing linked levels. .............................................113

Figure 4.15. Syllable distribution in the PROM Dataset by Biggs stress category. ..........119

Figure 4.16. Examples of Emu-generated F0 contour for flat-, fall- and rise-type phrases. ................................................................................................................120

Figure 4.17. Distribution of H* across stress categories in the PROM Dataset. ............121

Figure 5.1. PPM distribution across Biggs stress categories. ........................................133

Figure 5.2. PS-Group+WS monophthongs: PROM vs. MAONZE. .................................139

Figure 5.3. PROM Group Z PPM monophthongs vs. MAONZE PS-Group+WS
monophthongs. ...........................................................................................................140

Figure 5.4. Vowel space comparison of Group Z PPMs and non-PPMs. .........................141

Figure 5.5. Distribution of H*-PPMs by Biggs stress category. .......................................145

Figure 6.1. Alignment of Group A, B and C PPMs with Biggs stress category,
as % of Group PPM sets. .........................................................................................159

Figure 6.2. Short (left) and long (right) monophthong PPM vowel spaces
by proficiency group. .................................................................................................167

Figure 6.3. H* PPMs for the stress categories as % of the available H* syllables
in that category. ...........................................................................................................172
List of Tables

Table 3.1. Tap Distribution in Responses to the Māori Stimuli ........................................... 48
Table 3.2. Tap Distribution in Responses to the English Stimuli ........................................ 50
Table 3.3. Total Syllables and Taps for Different Filtering Conditions, Showing Proportions ........................................................................................................ 52
Table 3.4. Prominence Agreement in Response to Older and Younger Speakers: All Stimuli (Māori & English) ......................................................................................... 69
Table 3.5. Prominence Agreement in Response to Older and Younger Speakers: Māori vs. English Stimuli ......................................................................................... 70
Table 3.6. Māori Language Proficiency Groups Based on Self-Rating ........................................ 73
Table 3.7. Prominence Identification by Proficiency Group at 60% and 50% Thresholds ........................................................................................................ 75
Table 3.8. Prominence Identification by Proficiency Group on Expected Stress Syllables at 60% and 50% Thresholds ........................................................................ 76

Table 4.1. English Proficiency Self-Ratings from the Second Web Survey Questionnaire 89
Table 4.2. Māori Proficiency Self-Ratings from the Second Web Survey Questionnaire ... 89
Table 4.3. Breakdown of Participant Numbers in Māori Proficiency Groups ..................... 90
Table 4.4. Slice of Dataset Showing Part of Two Utterances, With Syllables, Triplet Context, and Added Zero-Centres ........................................................................ 101
Table 4.5. Slice of Dataset Showing Adjacent Syllables with Identical Vote Counts ......... 103
Table 4.6. Concatenated Pair-Top in New Triplet, with Recalculated X², p-Value, and Adjusted p-Value ...................................................................................................... 103
Table 4.7. Summary of Triplet Mode Proficiency Group PPM Results .............................. 106
Table 4.8. Summary of Proficiency Group PPMs: from Triplet Mode to Quinlet Mode . 109
Table 4.9. Summary of PROM Dataset Syllable Characteristics ...................................... 117
Table 4.10. Average Mean F0 and Mean Range of F0 for Phrase Contour Categories in PROM Dataset .................................................................................................... 120
Table 4.11. Average Mean F0 of Contour Category vs. Mean F0 of H* Events in PROM Dataset ............................................................................................................. 121
Table 4.12. PPM Selection by All Participant Groups for Three Speaker Vintages ........ 124
Table 4.13. Distribution of Word Stress (WS) PPMs by Speaker Vintage and Participant Group ............................................................................................................... 124
Table 4.14. Distribution of Non-Stress (OS) PPMs by Speaker Vintage and Participant Group ................................................................. 124
Table 4.15. Distribution of Phrase+Word Stress (PWS) PPMs by Speaker Vintage and Participant Group ......................................................... 125
Table 4.16. Distribution of Phrase Stress (PS) PPMs by Speaker Vintage and Participant Group ................................................................. 125

Table 5.1. Summary of Group Z PPM and Non-PPM Characteristics ................................................................. 129
Table 5.2. Group Z PPM Distribution Across Utterances ................................................................................................. 129
Table 5.3. Group Z PPM Distribution Across Phrases ................................................................................................. 130
Table 5.4. Selection of Biggs Stress Category Syllables as PPMs Out of All Available in Category in Dataset ................................................................................................. 134
Table 5.5. Mean Whole Syllable Durations of Heavy and Light Syllables ................................................................................................. 136
Table 5.6. Mean Whole Syllable Durations of Heavy and Light Syllables by Presence or Absence of Onset ................................................................................................. 136
Table 5.7. Monophthong Token Counts for Datasets in Vowel Space Plots ................................................................................................. 139
Table 5.8. Distribution of PPMs According to Phrase Contour Types ................................................................................................. 143
Table 5.9. Mean F0 for H* Events in PPM Syllables and Non-PPM Syllables ................................................................................................. 144

Table 6.1. Breakdown of Participant Numbers in Māori Proficiency Groups ................................................................................................. 150
Table 6.2. Summary of Group A, B and C PPM Characteristics ................................................................................................. 152
Table 6.3. Overlapping PPMs by Participant Group ................................................................................................. 153
Table 6.4. Exclusive PPMs by Participant Group ................................................................................................. 153
Table 6.5. Group A, B and C Rate of Selection: PPMs per Utterance ................................................................................................. 155
Table 6.6. Group A, B and C Rate of Selection: PPMs per Phrase ................................................................................................. 155
Table 6.7. Silence-Adjacent PPM Selection (Edges and Pauses) by Proficiency Group ................................................................................................. 156
Table 6.8. Alignment of Group PPMs with Biggs Stress Categories and as % of Available Syllables of Category Type ................................................................................................. 159
Table 6.9. Group A, B, C PPMs by Syllable Weight (of Total Available PPMs in Weight Category in Dataset) ................................................................................................. 161
Table 6.10. PPMs by Presence of Onset per Weight Category (of Groups’ Total PPMs) ................................................................................................. 162
Table 6.11. Possible False Diphthong Sequence PPMs by Proficiency Group ................................................................................................. 164
Table 6.12. PPM and Non-PPM Vowel Token Sets for All Proficiency Groups ................................................................................................. 167
Table 6.13. Distribution of PPMs According to Phrase Contour Types for the Three Proficiency Groups ................................................................................................. 169
Table 6.14. Location of PPMs Relative to H* Events for the Three Proficiency Groups . 170
Table 6.15. Group A, B and C Mean F0 for PPM and Non-PPM H* Events ................. 171

Table D1. Group Z Mean Nuclear Durations by Nuclear Vowel Category and Syllable PPM Status ................................................................. 213
Table D2. Group A Mean Nuclear Durations by Nuclear Vowel Category and Syllable PPM Status ................................................................. 213
Table D3. Group B Mean Nuclear Durations by Nuclear Vowel Category and Syllable PPM Status ................................................................. 214
Table D4. Group C Mean Nuclear Durations by Nuclear Vowel Category and Syllable PPM Status ................................................................. 214
List of Glosses

<table>
<thead>
<tr>
<th>Gloss</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,2,3</td>
<td>person</td>
</tr>
<tr>
<td>Ag.Emph.</td>
<td>agent emphatic</td>
</tr>
<tr>
<td>Ana.</td>
<td>anaphoric</td>
</tr>
<tr>
<td>Caus.</td>
<td>causative</td>
</tr>
<tr>
<td>Det.</td>
<td>determiner</td>
</tr>
<tr>
<td>Dir.</td>
<td>directional</td>
</tr>
<tr>
<td>Foc.</td>
<td>focus</td>
</tr>
<tr>
<td>Excl.</td>
<td>exclusive</td>
</tr>
<tr>
<td>Indef.</td>
<td>indefinite</td>
</tr>
<tr>
<td>Inf.</td>
<td>infinitive</td>
</tr>
<tr>
<td>Interj.</td>
<td>interjection</td>
</tr>
<tr>
<td>Mann.</td>
<td>manner</td>
</tr>
<tr>
<td>Neg.</td>
<td>negative</td>
</tr>
<tr>
<td>Obj.</td>
<td>object</td>
</tr>
<tr>
<td>P</td>
<td>preposition</td>
</tr>
<tr>
<td>Pass.</td>
<td>passive</td>
</tr>
<tr>
<td>Pl.</td>
<td>plural</td>
</tr>
<tr>
<td>Poss.</td>
<td>possessive</td>
</tr>
<tr>
<td>Prox.</td>
<td>proximal</td>
</tr>
<tr>
<td>TA</td>
<td>tense/aspect</td>
</tr>
<tr>
<td>Top.</td>
<td>topic</td>
</tr>
<tr>
<td>Rel.</td>
<td>relativiser</td>
</tr>
<tr>
<td>Sg.</td>
<td>singular</td>
</tr>
</tbody>
</table>

Note: Glossing follows Harlow (2007), with additions.
# Glossary

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>adj.</td>
<td>adjective</td>
</tr>
<tr>
<td>(Biggs) stress category</td>
<td>Categories of word and phrase stress designated by Bruce Biggs (1969) and additional categories relating to these that are used in this thesis (see PS, PS-Group, PWS, WS, WS-Group, and OS)</td>
</tr>
<tr>
<td>edge-silence</td>
<td>(in this study) silence at the beginning or end of a sentence</td>
</tr>
<tr>
<td>Group A</td>
<td>participants in the main experiment who self-rated as having high Māori language proficiency</td>
</tr>
<tr>
<td>Group B</td>
<td>participants in the main experiment who self-rated as having minimal Māori language proficiency</td>
</tr>
<tr>
<td>Group C</td>
<td>participants in the main experiment who self-rated as having no Māori language proficiency</td>
</tr>
<tr>
<td>Group Z</td>
<td>all participants in the main experiment</td>
</tr>
<tr>
<td>HE</td>
<td>Historical Elders (speaker vintage)</td>
</tr>
<tr>
<td>heavy syllable</td>
<td>syllable containing two or more morae (up to three in Māori)</td>
</tr>
<tr>
<td>H*</td>
<td>high pitch point; pitch peak</td>
</tr>
<tr>
<td>kaumātua</td>
<td>a Māori elder</td>
</tr>
<tr>
<td>L1</td>
<td>first or native language</td>
</tr>
<tr>
<td>L2</td>
<td>second language</td>
</tr>
<tr>
<td>light syllable</td>
<td>syllable containing one mora</td>
</tr>
<tr>
<td>MAONZE</td>
<td>Māori and New Zealand English</td>
</tr>
<tr>
<td>mora (μ)</td>
<td>(in Māori) a unit consisting of a single vowel and optional onset: (C)V</td>
</tr>
<tr>
<td>n.</td>
<td>noun</td>
</tr>
<tr>
<td>OS</td>
<td>syllables not expected by rule to carry stress of any kind</td>
</tr>
<tr>
<td>pause-silence</td>
<td>(in this study) silence occurring within a sentence or phrase</td>
</tr>
<tr>
<td>PE</td>
<td>Present-Day Elders (speaker vintage)</td>
</tr>
<tr>
<td>PPh</td>
<td>Phonological Phrase (after de Lacy (2003) and others)</td>
</tr>
<tr>
<td>PPM</td>
<td>perceived prominence</td>
</tr>
<tr>
<td>PROM</td>
<td>the database used in the main experiment in this project</td>
</tr>
<tr>
<td>PS</td>
<td>phrase stress</td>
</tr>
<tr>
<td>PS-Group</td>
<td>phrase stress associated (PS and PWS) syllables</td>
</tr>
<tr>
<td>PWS</td>
<td>coincidence of phrase stress and word stress</td>
</tr>
<tr>
<td>Term</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>PY</td>
<td>Present-Day Young (speaker vintage)</td>
</tr>
<tr>
<td>v.</td>
<td>verb</td>
</tr>
<tr>
<td>WS</td>
<td>word stress</td>
</tr>
<tr>
<td>WS-Group</td>
<td>word stress associated (WS and PWS) syllables</td>
</tr>
<tr>
<td>speaker vintage</td>
<td>age of a set of speakers used in the study, and age at recording (see HE, PE, PY)</td>
</tr>
<tr>
<td>syllable (σ)</td>
<td>(in Māori) a unit consisting of a vocalic nucleus and optional onset, taking any form permitted by the structure (C)V(V(V))</td>
</tr>
<tr>
<td>triplet</td>
<td>an overlapping set of three consecutive syllables in the dataset</td>
</tr>
<tr>
<td>quinlet</td>
<td>an overlapping set of five consecutive syllables in the dataset</td>
</tr>
<tr>
<td>XP</td>
<td>a syntactic phrase</td>
</tr>
</tbody>
</table>
INTRODUCTION

All God’s chillun got rhythm, all God’s chillun got swing,
Maybe haven’t got money, maybe haven’t got shoes,
All God’s chillun got rhythm for to push away their blues

(Jurmann, Kahn & Kaper, 1937)

Rhythm is everywhere. It is in the dripping of water, the thump of a bassline, the tapping of a foot or the cycle of day and night. It is in music and science and all in-between: sometimes natural, sometimes artificial; so often human. This latter is perhaps best shown by its presence in that most personal and human of things: language.

Sensitivity to rhythm, and particularly rhythm in language, is thought to be a universal, and to start very early in life (Nazzi, Bertoncini, & Mehler, 1998). Even where they cannot describe it, native speakers know their language’s rhythm; its nuances, and they know when there is deviation, as might be produced by non-native speakers or in error. They know that this rhythm, its patterns, and its characteristics can convey meaning, emotion, and all manner of other messages.

But what are the characteristics of rhythm in language? What makes it happen and why do we hear it? Over many years, a vast amount of scholarship in the fields of prosody and phonetics has been devoted to this question, and by extension to the definition of the term itself, which is no less complex than the human brain that devised it.

If rhythm is most succinctly described as a regular recurrence of beats, as by Kohler (2009a, p. 31), then perhaps rhythm in language may be described in terms of patterns of prominence (and non-prominence), suggesting that the two are very closely linked. Prominence, here, is simply relative salience, the key word being “relative”. Prominence cannot exist without the contrast of
lesser salience alongside. The way in which prominent units in speech (such as syllables) are organised in relation to larger grammatical or phonological units (such as phrases) may vary widely according to language. Equally, the features at the acoustic level, including duration, loudness and pitch, that cause listeners to hear these units as prominent (also known as “parameters”, “correlates”, or from the perceptual point of view, “cues”) may vary cross-linguistically in the same way. Most often, in any language, there is found to be a confluence of features: rarely is a single one responsible for the effect of prominence, and the features that operate together may also differ on levels within the prosodic structure, from the segment all the way up to the level of the entire utterance. Thus, there cannot be one universal description of the correlates of prominence and the organization of prominences: each language must have its own. The methods of measuring any of these aspects, from the correlates of prosodic prominences to their location, grouping, and the meaning of that organization, are numerous in themselves, encompassing natural and artificial speech, perception tasks, production tasks, acoustic and mathematical studies, and entire theories of, for example, stress and intonation. A fuller discussion of prominence, rhythm, related concepts and measurement methods may be found in Chapter Two.

The culture of the Māori people of New Zealand is an oral culture, filled with rhetoric and song. Before the arrival of the English language with missionaries in the 19th century, Māori language (te reo, “the language”) was not written down. The stories, songs and folklore were passed from one generation to the next in speech, and there was, and still is, a strong focus on rhetorical skills and the ability to recite information such as whakapapa (genealogy) from memory. Speaking on a marae (a meeting place) is an honour and in many cases there is a required structure to the oratory. Prayers and speeches of welcome are made by elders, called kaumātua, who are well-versed in the art of speaking; this skill is part of what makes them respected as elders in the community. It is also they who have the principal responsibility for passing on the taonga, or treasure, of the language to the younger generation. Language is, for many Māori, spiritual: a connection with the ancestors and the past; with their heritage (King, 2009). This is beautifully exemplified by the following proverb:
As an indigenous language in contact with a majority language (that is, English) for nearly 200 years, Māori has been subject to conditions that would precipitate language change. Indeed, there is some evidence that such change has occurred, whether or not this is internal (Labov, 1994), or due to the influence of English, or both (Watson, King, Bier, Maclagan, Harlow, Thompson, & Keegan, 2011). This will be explained in more detail in the next chapter.

When the Māori elders, with their strong connection to the oral tradition, along with other speakers, both native and non-native, notice a change in the language to the point of comment, questions naturally arise about the nature and extent of that change. As King, Harlow, Watson, Keegan, & Maclagan put it, “older generation speakers of all languages typically complain about innovations in pronunciation made by the young” (2009, p. 85). This state of affairs is by no means exclusive to Māori; in fact, it could probably be said to be a genuine “language universal” that has been happening since time immemorial. Records regarding Māori specifically show a long history of it (King et al., 2009), and this is especially unsurprising in a strongly oral culture.

What the elders say has changed is what they call the mita of the language (te mita o te reo). Despite the sound of this word, which bears a resemblance to the English “metre”, as in poetry, it does not refer exclusively, or even mostly, to rhythm. It is discussed further in Chapter Two. Suffice it to say here that it may encompass all aspects of language, from the segment to the lexicon and grammar, but may certainly also include suprasegmentals. What the non-native, non-proficient, or naïve observers will say has changed is sometimes “rhythm”, in as many words, but this ought to be taken with a caveat. These commentators are not as attuned as Māori speakers to other aspects of the language: they do not necessarily know meaning or grammar or vocabulary, so their reaction is based on the speech signal alone. On the other hand, they will say that older and more recent recordings or speakers “sound different” or have “more” or “less” rhythm. Just as speakers of a given variety of English will insist that they “don’t have an accent” – it’s all the other varieties that have the accents – what one hears often, one notices less. As most of these non-proficient commentators are English speakers, the descriptions of more or less rhythm may be an
indication that what they hear from the recent Māori speakers is, or has become, more like their primary language.

Based on these comments, then, whatever other change is evident in, for example, vowel quality or duration or in the lexicon, something in the prosody of Māori is also perceived to have changed, and there is a suggestion that contact with English may have had something to do with this (King et al., 2009, p. 92; Watson et al., 2011, p. 88). Māori does not lack documentation of rhythm and “stress”, nor potential correlates for that stress based on impressionistic analysis. See Chapter Two for a concise introduction to this aspect of the language. There is, however, no in-depth acoustic analysis, and the present descriptive structure of the stress system has been in place since the mid 20th century (Biggs, 1969). This does not imply that the description is necessarily inaccurate, but if change is occurring (or has occurred), it may be measurable, particularly by acoustic means. If measurable change is not found, acoustic analysis still provides an addition to the overall illustration of Māori prosody.

Since the question of change arose with listeners, it is to listeners that we turn to help attempt to answer that question. The studies in this thesis, and the approach used in them, were informed by the informal description of the change as a difference in rhythm, and the description of rhythm as patterns of prominence and therefore the collection of possible perceptual cues or acoustic features which create that prominence. The approach designed here was intended to begin to address the broader question of rhythm in Māori by investigating prominence specifically. A central goal was to find an efficient way of gaining participant perceptions of prominence in Māori, and analysing the qualities of that prominence, both independently and in comparison or contrast to existing description. The aims of the project are set out below.

(1) To establish the location of perceived prosodic prominences in Māori speech.
(2) To determine whether or not these perceived prominences were consistent with existing descriptions of stress locations and if not, what had changed.
(3) To determine what acoustic features were involved in the production of these perceived prominences, and whether or not this was consistent with the existing impressionistic description.
(4) To determine the effect of listener proficiency on responses to Māori speech. This avenue of investigation was included because the same basic description of the difference between older and younger speech came from both more and less proficient listeners.

(5) To ascertain whether or not there was a difference in response to older and younger speakers in terms of prominence identification.

(6) To create a methodological approach to enable the investigation of perceived prominences on a large scale, including a tool for elicitation of those prominences that was simple, user-friendly and effective.

This introduction has provided some context, background, and motivations, and introduced the aims of the project. More detail on the content of each chapter can be found below, but the following is a quick guide to the content of the thesis as relates specifically to the aims that have just been laid out. For or background to the Māori language, definitions and concepts used in this study, see Chapter Two. For the methodological development and evaluation as they relate to aim (6) above, see Chapters Three, Four and Seven. For results and analysis to answer aims (1) to (5), see Chapters Three, Four, Five, Six; particularly the latter three. For aim (4), listener proficiency, specifically, see Chapter Six. For aim (5), the response of all proficiency groups to older and more recent speech, see Chapter Four. Chapter Seven contains the evaluation of the main experiment.

In Chapter Two, the reader will find more detailed information on the Māori language and its structure. This is not intended as a reference grammar, as there are plenty of excellent resources available containing descriptions and grammatical analysis of the language in its entirety. Some recommendations appear in the relevant section. What is provided in Chapter Two simply aims to give the reader knowledge of the elements that will be relevant later on: first, the basic phonology and syntactic organization of Māori. Following this, there is an overview of the field of prosodic research and concepts within it. This is, of necessity given the vastness of the field, limited to what is relevant for this study. Some time is spent distinguishing and defining such terms as stress, prominence, accent, rhythm and intonation, which may have as many different nuances to their definition as there are researchers to study them. It is made clear how they will be defined for present purposes. Further to this, there is a discussion of the acoustic correlates that one may measure in studies of this type, the possible groupings for prominences created by these features, and an outline of different means by which prominences or correlates in a language may
be investigated, including acoustic analysis, perception and production tests of various types, and metrics of recent years. There is also some discussion of native and non-native perception. All of this is related to the approach taken in the present project. Finally, there is a summary of the prosodic analysis that has been carried out on Māori, including Biggs’ (1969) stress rules, followed by an introduction to the MAONZE project and the database connected with it, from which all of the stimuli used in the experiments in the present project were sourced.

The next two chapters deal primarily with the methodology that was used. Chapter Three contains the pilot experiments that were influential and instrumental in the development of the approach and of the perceived prominence elicitation tool itself. These two experiments, a small-scale pilot study involving a tapping-based methodology, and the first incarnation of a web-based perception test, are described in full, including their participants, the delivery methods and stimuli used, the analysis, and the results, which show some similar tendencies to those observed in the main study. The initial experiments are evaluated and the methodological developments to the overall project that were motivated by the process and the results are also described. Detail of the final methodology and web survey tool, as used in the main investigation, may be found in Chapter Four, along with an overview of the dataset (the characteristics of syllables in the utterances available to the participants) and the results concerning the response to older and more recent speech.

All of the studies had some elements in common. Since there was a possibility that perceived prosodic prominences in Māori might no longer match the expected stress locations from existing description, a set of listeners were asked to identify prominences in Māori as they heard them. A unit was selected around which to base the central question concerning prominence: the syllable, for reasons that will be outlined in later chapters. Stimuli of an appropriate nature were required, and these were taken from an existing database of historical and modern Māori recordings (the MAONZE database: see Chapter Two). The stimuli chosen reflected a range of speakers, speaker ages and recording dates while restricting variables such as gender, recording quality and utterance type. In the final experiment, participants also needed to be subject to some targeting, in order to obtain data from different proficiency levels, including, most especially, native or
highly proficient Māori speakers. See Chapter Two for further discussion of native speaker perception, and Chapter Four for participant targeting and proficiency levels.

The perception information itself needed to be acquired in as efficient a manner as possible. The test needed to be distributable, practical and participant-friendly, and to encourage a useful number of responses (that is, as large a number as possible). After the pilot-testing, the answer was found in a web-based survey. The result processing methods needed to be equally efficient and suitable for handling a large amount of data, as well as different behaviour in the task from participants of varying proficiency levels. A firm set of criteria for what constituted a prominence also needed to be developed. This was partly based on the participants' raw responses, and ensured that the prominences were reliable and robust candidates for empirical and acoustic analysis. Once the number and location of the perceived prominences were known, they were subjected to analysis intended to answer the questions outlined in the aims above. Their overall rate of selection and alignment to the existing expected stress positions was compared. Acoustically, they were subjected to analysis of duration, F0 behaviour and vowel quality. The responses to older and more recent speech were compared and contrasted in search of evidence of change (see Chapter Four, Section 4.10).

Chapters Five and Six present the results and analyses for the main experiment in the project, in which the web survey was distributed to the widest audience, who were both local to New Zealand and international. In Chapter Five, the reader will find the results of those analyses, based on the responses of the entire participant set. Chapter Five contains the same pattern of analysis on the responses from the participants as split into three groups according to their self-rated level of Māori language experience, in order to examine the possible effects of Māori language proficiency on prominence identification.

In the final chapter, the reader will find a return to the questions and aims of the study that have been posed here at the outset. A summary and evaluation is provided, of the experiments in the project and the results gained from them. This chapter contains some reflection of the results of the study and what they tell us about prosody and prominence in Māori, whether or not the stress rules may need revising, whether or not the cues that appear to be in use are consistent with
existing description, and whether or not this allows us to draw any conclusions about change in this area of the language. Finally, there is some discussion of the potential for further refinement and expansion of the methodology; the applications for studies such as this, both for Māori and other languages; and some possible future directions for the project.
2

BACKGROUND

PROJECT CONCEPTS AND THE MĀORI LANGUAGE

2.1. Introduction

Māori, or te reo Māori, is the indigenous language of New Zealand. It belongs to the Polynesian subgroup of the Austronesian language family (Harlow, 2007, p. 1), and has been in contact with and influenced by English for about 200 years. This contact has resulted in both qualitative and quantitative change (see for example King et al., 2009; Maclagan et al., 2009a; Watson, Maclagan, Harlow, Bauer, King, & Keegan, 2008). Describing and investigating this change is a complex task, both for the linguistic researcher and the general observer. The studies in this thesis were partly motivated by the question of whether or not some of the perceived change is prosody-related, and are therefore concerned with a method of investigation that will enable the description required to answer this question. The core questions in the project have already been introduced; this chapter provides practical background and definitions for navigating the experiments and results presented later on.

After a brief introduction to the Māori language and relevant aspects of its phonology and syntax (§2.2), we move on to a discussion of some of the concepts and terms used in the field of prosodic research that will be relevant for the present study (§2.3). This section also includes an account of a selection of the many and varied methods of measurement available in investigation of perceived prominence and the correlates that may cause it, along with aspects of native speaker perception of prosodic prominence. Following this, some aspects of prosody specific to Māori are described, including existing stress rules and prosodic analysis (§2.4). Finally, there is an
introduction to the nature of the MAONZE database, from which the stimulus recordings used in the experiments in the study were taken (§ 2.5).

2.2. Māori Language: Overview

At the time of the 2013 New Zealand Census, Māori was spoken by approximately 3.3% of the whole New Zealand population. This equated to around 148,000 Māori speakers, but all of these would also speak English.¹ In the same Census, of the 598,605 people identifying as Māori, 21.3% claimed to be able to converse in Māori: around 125,000 speakers. Most of these were older people (Statistics New Zealand, 2013a; 2013b; 2014).

Māori is not endangered, but it is currently listed by the UNESCO Atlas of the World’s Languages in Danger (formerly the Red Book) as vulnerable² (Moseley, 2010), and its status did become highly precarious for a time. There has been language contact with English since the late 18th century, and between a declining Māori population in the second half of the 19th century, the increasing English-speaking population, and significant changes in the domains where Māori language was used, including schools (Harlow, 2007), the language suffered. See Harlow (2007, p. 192ff) for a much fuller description of the language situation across the years. Revitalisation efforts in recent years, since the 1980s, have been strong. The Māori Language Act (1987) established Māori as an official language of New Zealand, along with an organization intended to promote the use of the language: Te Taura Whiri i te Reo Māori, or the Māori Language Commission. There was an increase in the number of kōhanga reo and kura kaupapa (Māori language immersion schools), along with the promotion of kapahaka (traditional performance involving dance and song) among young people, and the ability to study the language and culture to doctoral level. Teaching, learning, and use of the language, both in the home and outside, is supported by an enthusiastic group of both older and younger Māori speakers: the “language fanatics” mentioned by King (2009, p. 97).

¹ Following Census figures from 2013, the population of New Zealand was estimated, at 31 December 2014, to be around 4,553,700 (Statistics New Zealand, 2014).
² “Vulnerable” is defined as “most children speak the language, but it may be restricted to certain domains (e.g., home)” (Moseley, 2010).
Variation in the language is regional: broadly, dialects spoken in the western part of the country are different from those in the east (Biggs, 1961, p. 2). The differences, often associated with tribal identity-marking, are minimal, and the dialects are mutually intelligible (Harlow, 2007, p. 44). The variation appears throughout the language, from vocabulary to grammar and phonology, though prosodic variation is not noted.

Within the 200 years of contact between Māori and English, there has been influence of various kinds in both directions. While the Māori influence on New Zealand English is lexical only, in the form of Māori loanwords, there is also a variety of English known as Māori English, which is sometimes regarded as a subvariety of New Zealand English, and sometimes regarded as a variety in its own right (Gordon, Campbell, Hay, Maclagan, Sudbury, & Trudgill, 2004). Its chief function is as an identity marker. Its features include various phonetic differences; a rhythm that is more “syllable-timed” than “stress-timed”; a high-rising terminal pattern in non-interrogatives; loanwords, and high use of the tag *eh?* (Gordon et al., 2004, p. 70). Some of these are influenced by the Māori language.

The influence of English on Māori, however, is greater, and relatively rapid. Changes have been both qualitative (including vowel quality, with the overall short vowel space becoming more peripheral, raising of /e/, aspiration of /t/, and related fronting of /u:/) and quantitative, including some loss of the phonemic vowel length distinction (Harlow et al., 2011; Maclagan, Harlow, King, Keegan, & Watson, 2004; Maclagan, Watson, Harlow, King, & Keegan, 2009a; Watson, Maclagan, King, & Harlow, 2006; Watson, Maclagan, King, Harlow, & Keegan (in press); and see below for further explanation). The changes may also have prosodic effects, including the location of word stress (King et al., 2009; for more on Māori stress, see Section 2.4.1 below).

There are many handbooks, textbooks and grammars of the language available, both older and more recent, and the reader is referred to those for greater detail under varying approaches. This section introduces elements that will be relevant to the analyses used in this project. Here, there is a basic description of Māori phonology and syntactic structure; later in the chapter, following the

---

3 See, for example, Bauer, 1993; Biggs, 1969 & 1998; Harlow, 2007; the textbook series by Moorfield (2001). For older references, see Biggs (1961) and Schütz (1985) for a list of very early work, and also those mentioned in Harlow (2007).
discussion of prosodic concepts in general, relevant aspects of Māori prosody will be introduced (§ 2.4).

### 2.2.1. Phonemic Inventory and Syllable Structure

Māori has 10 consonants: /p t k m n ŋ r w f h/, and five vowels: /a e i o u/. Each of the vowels has a two-way phonemic length distinction, though the differences are becoming reduced in all of the pairs except /a - a:/, which is the pair with the highest functional load (Harlow, Bauer, Maclagan, Watson, Keegan, & King, 2011; Maclagan et al., 2004; Watson et al., 2006).

The long monophthongs are considered to be, underlingly, two identical short monophthongs. There are morpheme-internal long monophthongs, but these can also be created when two identical short monophthongs occur either side of a morpheme boundary. The two types are the same where vowel quality and duration are concerned, and stress assignment rules (§ 2.4.1) apply to them in the same ways (Harlow, 2007, p. 66-67).

This applies equally to diphthongs, which are underlingly two unlike short monophthongs. However, not all sequences of unlike vowels form diphthongs. A vowel plus a higher vowel, as well as /oe, io, iu/, can form one, but do not always do so. This is dependent on context; for example, reduplication, affixation, or the position of particles. For detailed discussion, see Bauer (1993, pp. 534ff) or Harlow (2007, pp. 71ff). Other vowel sequences besides those mentioned above will remain separate. There are also long diphthongs, which do not occur often. These include any sequence of /aa/ plus a higher vowel, as well as /eei/ and /oou/ (Harlow, 2007, p. 69).

In terms of vowel quality, long vowels are more peripheral in the vowel space (Harlow, 2007, p. 77), though there has been a shift towards the periphery in the short vowel space (Maclagan et al., 2009b). Māori is not generally considered to have vowel reduction (Bauer, 1993); however, some evidence of centralization in vowels classified as “unstressed” has been uncovered (Kaefer, King, Watson, Maclagan, Harlow, & Keegan, 2010). There is also devoicing, most commonly heard phrase-finally in short vowels, but in rapid speech this may extend to other vowels and to “whole post-stress syllable sequences” (Harlow, 2007, p. 76).
Māori syllables can take any of the forms permitted by the structure (C)V(V(V)). The smallest instantiation of this is the mora, (C)V: a short monophthong (nucleus) plus an optional onset consonant. Codas and consonant clusters do not occur. A syllable may, therefore, be the same as the mora, and these monomoraic syllables are considered “light”. Through the syllabification process, which also forms long monophthongs and diphthongs (Harlow, 2007, p. 72), a larger syllable is formed from two or three underlying morae. These bimoraic or trimoraic syllables are considered “heavy”. For more on syllable weight, see Section 2.3.2.1. Some examples of words split into syllables (σ) and morae (μ) are in (1) below. From right to left: the Māori word; the English translation; syllabification, with syllables separated by a plus sign (+) and morae by a period (.); syllable count; and mora count.

(1)  
tē        “the”          te  1σ  1μ  
tū        “stand”        tu.u  1σ  2μ  
kāinga    “home”         ka.a.i+nga  2σ  3μ+1μ  
kāwaha    “door”         ku.u+wa+ha  3σ  2μ+1μ+1μ  

The reason for having two separate units, mora and syllable, is that neither alone can account for all of the rules in the language (see, for example, Bauer, 1981). Rules of reduplication and the forms taken by some constructions, such as the imperative, rely on the number of morae in the words concerned. Equally, the mora is used as a unit of metre in Māori poetry: “a high proportion of traditional waiata contain just eight vowels to each half-line of the text... Each long vowel counts as two...” (Biggs, 1980, p. 48). This is known as the “Rule of Eight”, and McLean (1982) also identified some instances of a rule of twelve. In both cases, the unit is the mora. The syllable, on the other hand, is the central unit for the rules of stress assignment, even though the mora is involved. Also relevant for stress assignment in Māori is the concept of the phrase.

2.2.2. Syntactic Organization: The Māori Phrase

Māori syntactic phrases and phonological phrases are recognized as being very closely linked (Bauer, 1993; Biggs, 1961). Māori syntactic phrase structure may be described in a number of

---

4 A kind of song. For printed examples, see McLean & Orbell (1975).
ways, but the most useful and relevant method here is the form used by Biggs (1969; 1998), since it relates directly to his stress assignment rules. Under Biggs’ description (1969, pp. 18-19), the Māori phrase is a unit with up to three parts, and all phrases are either verbal or nominal. The nucleus is obligatory and usually contains one or more content words, called “bases”; for example, nouns. There are select circumstances where the nucleus may be empty (as for certain determiners: see example (5) below). Either side of the nucleus are optional peripheries, which hold function words or “particles”. Prefixes, such as the causative whaka-, and suffixes, such as the passive -ia / -bia / -tia, may also be present. The allocation of the lexical and functional items to their respective phrasal parts is strict. Some simple examples are found in (2) – (5) below, where square brackets contain the parts of the phrase and the phrase as a whole, which are labelled as follows: PHR is the whole phrase, NUC is the nucleus, PRE is the preposed periphery (before the nucleus) and POST is the postposed periphery (after the nucleus). Examples (2) and (3) are phrases with both peripheries; example (4) has one periphery, which in this case happens to be preposed; and example (5) is a phrase with an empty nucleus, with the context of the preceding phrase. Empty parts of the phrases are marked with Ø.

(2)  [ [ ki ] PRE [ re ] PRE ] PHR
     [ [ whare ] NUC ] house
     “to this house”

(3)  [ [ kua ] PRE ] TA
     [ [ piri ] NUC ] cling
     “[they] still clinging”

(4)  [ [ ka ] PRE ] TA
     [ [ mutu ] NUC ] end
     [ [ Ø ] POST ] PHR
     “[it] ended”

     [ [ Ø ] POST ] PHR
     [ [ tērā ] PRE ] [ [ Ø ] NUC ] [ [ Ø ] POST ] PHR
     Det.
     “it was we [who...]”
Further Māori language related information, including Biggs’ stress assignment rules and intonation, will be presented in Section 2.4 following a more general discussion of some prosodic concepts relevant to the study.

2.3. Project-Related Prosodic Concepts

The concept at the core of the project in this thesis is prominence: that relative salience and collection of acoustic features that is connected, by the intricate web of prosody, to what we know of as stress, accent, intonation, rhythm, grammar and pragmatics. This very introduction hints, in a rather obvious way, at the path this project intends to take through the terminological minefield of prominence, stress, and accent in particular. The very wide variation in usage and definition of these terms, as well as those of prosody and rhythm, can make sorting one approach from another incredibly tricky. For example, some approaches treat “accent” and “stress” as being essentially the same, or interchangeable. Different approaches, using both, make very strict distinctions. Prominence itself is not always defined; rather, its existence is simply assumed and often equated with stress or accent without explanation. In other cases, it is used instead of either term: as Duanmu has it, “many linguists believe that accent and stress are different realizations of the same thing, namely, phonological prominence” (2007, p. 4). To avoid confusion, no attempt is made to summarise all of the differences here; instead, a small overview is provided, with a focus on defining the terms as they will be used in the present study.

2.3.1. Core Concepts

2.3.1.1. Prominence

The concept of prominence refers to the relative salience of some unit, often a syllable, compared to those around it. Virtually everywhere in which the term is used, the importance of this relativity and contrast is emphasised. As Jespersen put it in 1933 in his Notes on metre, the relations between syllables of different prominence levels are not perceived “with any degree of certainty except when the syllables concerned are contiguous” (as cited in Gross, 1979, p. 111). Salience may imply that the syllable is somehow more than its surroundings, and Jespersen also made the comment that these syllables tended to be pronounced “with higher pitch than weak syllables” (Gross, 1979, p. 124). As Kohler wrote much more recently, prominence and its
patterns are “triggered by physical properties among which vowel and consonant durations are but some besides several others” (2009b, p. 6). What this illustrates is that the prominence of syllables, though in itself effectively abstract, has a cause, and that cause is the relative behaviour of certain articulatory, acoustic and perceptual attributes. These attributes, also known variously as “correlates”, “features” and “cues”, are produced either consciously or subconsciously by the speaker, and perceived, consciously or subconsciously, by the listener. As the observations by Jespersen, Kohler, and many others both before and after them indicate, the correlates include pitch (vocal fold vibration or fundamental frequency: F0), and length (timing of articulation and duration or weight of syllables or segments), but also loudness (effort or intensity), among several other possibilities (§ 2.3.2.1). The causes and effects and patterns of prominence vary cross-linguistically. While prominence is generally described as being localised on certain small phonological units, such as the segment, mora or syllable, since it is inherently relative, it is also possible for larger units within an utterance to be relatively stronger or weaker than each other.

In the present study, the concepts of prominence and stress will be separated from each other; prominence, here, is only relative salience, potentially created by the features mentioned above. When the phrase “a prominence” is used, it simply means a syllable that was identified by the participants in the tests as more salient than its neighbours. These syllables will also collectively be called “prominent syllables”, “perceived prominences” or, for short, “PPMs”. Chapters Three and Four provide detail of the development of a methodological project definition for prominence.

2.3.1.2. Stress

Stress is the most commonly used term among “prominence”, “stress” and “accent”, and is, broadly, defined in one of two ways. Either it is a concrete property of a unit (an acoustic feature in itself, realized by the correlates) or it is an abstract property of that unit (a feature that specifies relative prominence in a prosodic structure, which is then realized by the correlates) (Cutler & Ladd, 1983, p. 4). The latter, abstract, structural view is somewhat more common (for example, Beckman & Edwards, 1994; Sluijter & van Heuven, 1996). As with prominence, the associated unit is, most often, the syllable. Most definitions of stress come with at least a generalised mention of correlate behaviour, implying, again, that contrast and relativity in that behaviour is
part of the concept: longer duration, higher pitch, greater intensity. Stress may have a number of functions, both lexical and grammatical, depending on the language, and it may also be either variable, which is to say, its position may alter depending on grammatical, morphological or phonological circumstances, or it may be fixed: always occurring on a given syllable within a unit.

What is called “stress” in this project belongs firmly in the abstract. As mentioned above, it will also be separated from prominence. Here, “stress” refers only to the predicted word and phrase stress locations outlined by Biggs’ (1969) stress rules (§ 2.4.1). While these were based on prominence effects as he heard them, including criteria based on a number of correlates, one of the questions central to the project concerns whether or not perceived prominences match the predicted stress locations. It was not possible to say at the outset of the experiments that prominence was entailed by the existing rule-based designation of a syllable as stressed, or would be perceived on those syllables by the participants. The stresses were not indicated to the listeners involved in the perception task; they were instead free to identify prominence on any syllable on which they heard it. The rule-based stressed syllables were analysed for their qualities, and the identified prominences (whether or not they fell on syllables with expected stress) were analysed separately. As a natural consequence of the core project question, part of the analysis for both was their alignment with each other, but this meant it was necessary to keep the two concepts apart.

2.3.1.3. Accent and Intonation

Just as stress and prominence are sometimes treated together, so too are stress and accent. However, where it is not used in place of the term “stress”, accent is often defined more specifically. While the prominence that is used as an indication of (abstract) stress is said to be caused by a number of different cues, accent is frequently restricted in its definition to units around which there is some form of pitch movement (Hermes, 2006; Sluijter & van Heuven, 1996). Accent can apply at a lexical level, or in larger, phrasal units, and has a number of possible tonal configurations (using high (H) and low (L) tones: see next paragraph for more on this method of description). In languages like Japanese, the location of the pitch-accent by mora in a word can be lexically contrastive: for example, the difference between あめ (HL) “rain” and あめ (LH) “candy”, in which, when standing alone, the former has an accent (bold) where the pitch falls, and the latter is “unaccented” (Tsujimura, 1996, pp. 76-78). In many languages, accented
syllables are also sometimes already classified as stressed due to other prominence-lending criteria, so the presence of accent can be described as making them more prominent, and the strongest element at a higher level will have greater prominence than the strongest at any lower level (Beckman & Pierrehumbert, 1986). Some theories disallow the presence of stress or accent at higher levels of a prosodic hierarchy (§ 2.3.2) on units that do not have stress or accent at lower levels, as in the Continuous Column Constraint (Hayes, 1995).

Further use of pitch movement to differentiate meaning is found in intonation, a means by which pragmatic meanings that apply to whole phrases or utterances are conveyed by suprasegmental phonetic features in a structured way (Ladd, 2008; Ladefoged & Disner, 2012). Intonation takes the form of different pitch contour shapes, or tunes. The tunes include pitch accents and boundary tones, which, in some methods of description (for example, Pierrehumbert, 1987, and Ladd, 1992, 2008), are marked with H (high) or L (low), and various diacritics, such as an asterisk (*) for the accented syllable, a raised hyphen (ˉ) for a following tone in the tune, and a percent sign (%) for the boundary tone of an intonational phrase. The tunes serve to indicate, for example, sentence or phrase type: declarative, interrogative, imperative, etc, as well as focus, emotion, and other communicative and discourse functions.

Accent and intonation, as described this way, are not specifically considered in the present study. The goal is a description of perceived prominences and some of their correlates, of which pitch behaviour is considered to be one, rather than an intonational description of Māori. For such a description, the reader is encouraged to consult Bauer (1993) or de Lacy (2003). Here, the behaviour of pitch and the identification of a high pitch point within a phrase in Māori, as well as the general phrase contour shape, are used in the analysis in the comparisons with existing prosodic description (§ 2.4.2).

2.3.2. Rhythm, Groupings, and Timing

Kohler calls prominence, and its recurring patterns, “the essential ingredient of rhythm” (2009b, p. 6). As the above definitions suggest, the smaller prominent and non-prominent units are organized into other, larger units of varying sizes and types. The larger units are often, but not always, equal to grammatical units such as words or sentences, but not all morphological or
syntactic units need to be marked by features of the spoken utterance (Shattuck-Hufnagel and Turk, 1996, p. 198). Prosody is described as

... both (1) acoustic patterns of F0, duration, amplitude, spectral tilt, and segmental reduction, and their articulatory correlates, that can be best accounted for by reference to higher-level structures, and (2) the higher-level structures that best account for these patterns.

(Shattuck-Hufnagel & Turk, 1996, p. 196)

In English and some other languages, the concept of higher-level structures (i.e. larger groupings), can be expressed using the unit called the foot or stress-foot. The Abercrombian Foot (Abercrombie, 1967) contains one stressed syllable followed by any number of unstressed syllables, and there are other, more defined variations well known to the student of poetry: iamb, trochee, spondee, dactyl. The way in which these patterns recur is known as metre (meter): for example, the famous iambic pentameter. Importantly, feet of this type are not necessarily word- or phrase-sized: they may include several such units (words or phrases) or be smaller even than a single word.

An attempt at typology based on these groupings came with the Rhythm Class Hypothesis, which traditionally put languages into different rhythmic categories based on the timing pattern of the groupings; principally, the categories of “stress-timing” (English, German, etc.) and “syllable timing” (Spanish, French, etc.). A third category of “mora-timing” was added to account for languages such as Japanese, which make use of that unit (Bloch, 1950; a review may be found in Warner & Arai, 2001). Maori has also been described traditionally as mora-timed, on this scale, because, like Japanese, it makes use of the mora in phonological rules (§ 2.2.1).

Though in fact it refers only to timing patterns, this popular and persistent classification of “rhythm” originated in work by Lloyd James (1940), following through Pike (1945) and Abercrombie (1967), to name a few. An important point here is that, as Kohler (2009a, p. 30) points out, Pike did observe that all languages have more than one type of timing, while favouring one or another, though later commentators treated the categories as mutually exclusive. Abercrombie: “As far as is known, every language in the world is spoken with one kind of rhythm or with the other…” (1967, p. 97). The classification is based on the idea of isochrony: that is,
each unit (foot, syllable, mora) ought to take the same amount of time. So, in stress-timed languages, the duration of the metrical feet should be equal; in syllable-timed languages, each syllable should be equal, and so on. However, despite experimental evidence both for and against the timing units, isochrony itself has proved highly elusive on closer examination of all of the categories proposed. It is described as a tendency, rather than being absolute (Beckman, 1992; Laver, 1994), and is almost certainly more perceptual than directly measurable (Dauer, 1983; Lehiste, 1977, 1979; Roach, 1982). Later, it was suggested by Dauer (1987) that languages fell on a continuum between rhythm types, and, similarly, the possibility of a larger set of rhythm classes than the three originally proposed has also been put forward (as discussed in Ramus, Nespor, & Mehler, 1999).

The concept of rhythm class and timing categories reduces rhythm in language to duration-based evaluation: they do not take into account other correlates that may be involved, nor do they describe the prosodic structure further or at higher levels. In recent work (Arvaniti, 2009, p. 61), it has been suggested that the idea of this terminology should be abandoned as a necessary result of separating rhythm from timing, precisely because timing is only one part of a much larger rhythmical picture. Dauer (1983, 1987) proposed eight parametric criteria for determining prominences: these included pitch, duration, vowel quality, accent and syllable structure. Significantly, these criteria do not focus only on timing and duration, even if, in themselves, they still do not account for the entire rhythmic picture. Rhythm, then, would appear to be composed of a number of elements: not timing, not intonation, not individual correlates or prominences by themselves, but the way they all come together based on a pattern. In different languages, one element may have more influence than another, or the layering may be different at the different levels on the hierarchy, but it is the layering of all these elements that creates the pattern we perceive as rhythm.

Though not intended to describe rhythm per se, a different and useful way of representing the organization of all of the relevant prosodic constituents or domains (units) for a language is the prosodic hierarchy (Selkirk, 1980), intended to account for the organization of speech elements such as intonation patterns, timing, and prominences (Shattuck-Hufnagel & Turk, 1996). As mentioned earlier, these prosodic constituents or domains may be more or less related to
morphological or syntactic structure, and logically, not all of the units that are available to a prosodic hierarchy are required by all languages. Several forms of hierarchy have been proposed. Shattuck-Hufnagel and Turk (1996, pp. 205-224) give an excellent review of the concept and constituents of these hierarchies, and some of the constraints that have been proposed for their use. The diagram in (6) below, and its accompanying text, are merely a summary of what such a hierarchy may look like.

(6) Utterance
   | Intonational Phrase
   | Phonological Phrase
   | { Clitic Group, Prosodic Word, Accentual Phrase… }
   | Foot
   | Syllable
   | Mora
   | Segment

At the top of a hierarchy are large units such as the Utterance or the Intonational Phrase, which can often correspond to a single sentence. Below this are various forms of Phonological Phrase, which can be either syntactically or accentually defined. The number and type of these middle units vary rather widely according to theory and language. Some theories refer only to a Phonological Phrase; others find Clitic Groups, Prosodic Words (Nespor & Vogel, 2007), and further Intonational and Accentual Phrases (Beckman & Pierrehumbert, 1986), according to the analysis concerned. At the lowest levels of the hierarchy are found the smallest units: generally these are the foot (if relevant), the syllable, and sometimes the mora or segment (see McCarthy & Prince (1999) for a list of references to theories allowing these). For a long time, it was considered that these structures had finite depth and did not allow recursivity (that is, for a node of one type to dominate nodes of the same type; for a Phonological Phrase, for example, to contain other Phonological Phrases) or layer-skipping (the ability to bypass, in some instances, a level of the hierarchy); this is the Strict Layer Hypothesis that originated with Selkirk (1984). However, more recently, convincing arguments have been made for the opposite case. One such is that of Ladd
(2008, pp.297-298), who proposes what he calls the Compound Prosodic Domain, in which recursivity is permitted, and by extension, prosodic structures do not have fixed depth. If circumstances are allowed where categories can fit inside others of the same type, this also allows for fewer types of prosodic domain. The many layers represented in Figure 2.1, themselves only a representative summary of the myriad different types of prosodic domain, may therefore quite easily be collapsed to some degree in description of the prosody of a language. Selkirk’s (2011) Match theory, with three levels (word, phrase, and clause) and Ito and Mester (2012) suggest that prosodic structure ought to behave, formally, in a similar manner to syntactic structure: again, allowing recursivity and a limited number of categories. Ito and Mester’s simplified prosodic hierarchy is shown in the diagram in (7) below.

(7)  υ  Utterance  
|  ι  Intonational Phrase  
|  φ  Phonological Phrase  
|  ω  Prosodic Word  
|  f  Foot  
|  σ  Syllable  
|  μ  Mora

In this study it is not proposed to document the prosodic hierarchy for Māori: comments on such structuring can be found in Bauer (1993) and de Lacy (2003), among other work. However, concepts from this section and levels from within the hierarchy will be referenced in the analysis and later in this chapter with regard to Māori (specifically, in § 2.4 below), and it is useful to outline them in general terms here.

2.3.3. Description and Measurement of Prosodic Features

There are many different methods of measuring the elements discussed above, from prominences and patterns to the features that produce them. For example, perception and production studies, and acoustic analysis. Prominences can be measured on different levels. On one level, they are located in larger phonological, morphological or syntactic units and structures, which parallel
levels on the prosodic hierarchy. For example, as in languages where prominent syllables or
patterns of prominence may have a predictable location within a word or other unit. On another
level, there are the acoustic features that make the prominences salient: relative duration and/or
syllable weight, pitch, intensity, vowel quality, and so on.

2.3.3.1. Correlates, Cues, Features, and Parameters

Depending on the approach and the study, all four of the terms in this section title can refer to
the same group of elements. The nature and behaviour of these elements, however they are called,
result in prominence effects. Which term is used for which element ought to depend on the
angle from which it is being described (acoustic, perceptual, etc.), but this convention is not
always strictly followed. Examples are F0 (fundamental frequency) and pitch; duration and
length; and loudness and intensity.

In both production and perception, the relevance of any correlate may vary both within and
across languages. As Arvaniti puts it, “languages use different parameters to make some syllables
more prominent than others, and the same parameter may be allocated different degrees of
importance in different languages” (2009, p. 60). Besides duration, pitch and loudness, there are
other correlates whose relative reliability and importance has been investigated. These may
include spectral tilt or emphasis (Heldner, 2003; Sluijter & van Heuven, 1996); vowel quality;
and various articulation or co-articulation effects; for example, the domain-initial strengthening
of consonants found by Cho & Keating (2001).

Earlier, in the discussion of accent, it was mentioned that in many languages, syllables which are
accented (an F0 feature) may already be classified as prominent (or stressed) due to other acoustic
features they may have, causing the overall prominence effect to be stronger. Put simply, layers of
prominence-causing features create greater prominence, and some theories (Hayes, 1995) require
these overlap circumstances where there is the possibility of prominence on more than one level
of the hierarchy. Some correlates, more than others, are subject to intrinsic, environmental, and
physiological effects, and to interaction with other features: for example, the additional loudness
that may be perceived in association with higher pitch (Ladefoged & Disner, 2012), or the
measurement of spectral tilt rather than overall intensity in order to combat the susceptibility of
the latter to readings skewed by background noise and the distance of the speaker from the microphone (Sluijter & van Heuven, 1996). This has implications for the study of prominence in any language: it can affect the results of language identification and discrimination exercises, particularly where native and non-native speakers, or speakers of different proficiencies and exposure levels, are involved. The following is a brief summary of the correlates that will be relevant for the present project: duration, pitch and vowel quality.

Duration is also known as timing or length. Timing classification and the question of isochrony were discussed above, but measurement of duration may include syllable weight as well as the raw duration of units or segments. Increased duration is often noted in conjunction with stress, as are heavy syllables. What constitutes a heavy or light syllable is defined differently across languages, according to syllable structure. However, broadly, as touched on briefly in Section 2.2.1 above, there is a moraic weight distinction, in which one mora constitutes light and two or more means heavy. A simple nucleus-based weight distinction divides short vowel and long vowel nuclei and ignores any coda. If codas are permitted, either they are counted towards the mora total or they are not, depending on the language. When there are more than two morae, such syllables are sometimes classified as superheavy (see, for example, Hyman, 1977, p. 10). The prevailing view is that the onset is weightless (Hyman, 1977; Hayes, 1995), but in some languages, onsets are considered to contribute to syllable weight and by extension to the location of stress (Gordon, 2005; Topintzi, 2010). Heavy syllables are very often associated with stress. Hyman remarks that they may even carry what he calls “redundant” (meaning, non-phonemic) stress, which attests to their perceptual prominence (1977, p. 5). The connection made between syllable weight and prominence may also be more broadly applied. For example, in Japanese, when weight adjustments are made in specialized language situations, such as language games or baby talk, in order to produce a specific weight pattern, there is a matching shift in the accent location (Duanmu, 2007, pp. 15-16).

Pitch, also referred to as F0 (fundamental frequency), is produced by vibration of the vocal folds. F0 varies inherently between men and women, and between different sonorant segments. It is usually measured in Hertz (Hz), but non-linear measures such as Mel or Bark can also be used (Clark, Yallop, & Fletcher, 2007). It can be described in terms of high and low tones, and
contour combinations thereof, as described earlier (§ 2.3.1.3). Higher pitch is often connected with prominence, but since contrast is central to any perception of salience, it is not necessarily the absolute F0 value associated with a unit that is relevant, but rather a change or movement of some sort on or around the unit. Prosodically, F0 and its movement are often said not to be a (direct) correlate of lexical stress, but instead of accent as it is defined above: a higher level manifestation of prominence (Beckman & Edwards, 1994; Sluijter & van Heuven, 1996). F0 values or movement, however, are neither exclusive to accent, tone and intonation, nor excluded from the features that cause the prominence generally said to be associated with stress: they may layer on each other depending on the function of the units involved. Pitch may be a correlate of prominence on small units as well as building contours to serve the wider intonational and communicative functions that will not be discussed in detail here, owing to their lack of inclusion in the analysis to be presented. This study focused on phrase contour shapes, pitch range, and the qualities of pitch peak syllables in the data relative to the perceived prominences.

Vowel quality concerns formant values and the shape of the vowel space, for which an F1-F2 plot can be produced. With F2 on the $x$-axis and F1 on the $y$-axis, this vowel space closely resembles the well-known vowel “triangle” or “quadrilateral”, based on vowel height and backness. Vowel quality may be linked to duration, in that fuller vowels can take longer to articulate, and there is a connection between stress and vowel reduction in many languages, including English. Stressed vowels are generally full vowels, while unstressed vowels are reduced. The fuller vowels are expected to appear more peripherally in the vowel space, as are longer vowels. Reduced or short vowels may appear more centralised in the vowel space (Harrington, 2010a).

Ladefoged and Disner make a generalisation that “in nearly every language … what we hear as stress is more a matter of increasing the pitch and length of the syllables concerned rather than of increasing their loudness” (2012, p. 24). They are, of course, referring to the interaction of cues. Whether this assessment is in fact true cross-linguistically is not for comment here, and in Chapter Seven there are some reflections on the inclusion of loudness in the future of the project presented in the thesis. However, considering the above, along with (a) a known vowel space shift and observation of vowel reduction in Māori, where it is not expected to occur (§ 2.2.1); (b) the known association of heavy syllables and greater duration with stress or prominence; and (c) the
association in description of Māori of a peak in F0 with stress and phrase contour (§ 2.4.1-2), the 
three correlates described here were selected for use in the analysis of the perceived prominences.

2.3.3.2. Methods of Measurement for Correlates and Prominence

The appropriate measurements for correlates and prominence depend on the information that is 
already available. If the locations of prominence are already known, as in a prosodically well-
documented language, their qualities can be measured, or tested directly. If the locations of 
prominence are not known for certain, or their position is thought to have changed, then they 
must be determined by means of listener perception information and/or production exercises, and 
the acoustic examination of the various correlates. This can be done by a trained phonetician with 
input from native speakers, but particularly in cases where change is suspected, a large scale 
response can be very informative and easily provide a form of consensus. Use of native and non-
native (or proficient and non-proficient) listeners may or may not elicit different responses; where 
there is a difference, the comparison between these responses may be helpful in itself, especially 
where a shift in cues or outside influence is suspected. The latter method, a large-scale response, 
was used in the present study, where the analysis relies on the listeners’ consensus on prominence 
position.

Acoustic analysis of the correlates under discussion can be carried out using speech analysis 
software such as Praat (Boersma & Weenink, 2013) or Emu (http://emu.sourceforge.net), as in 
the present study. This software allows annotation of the stimuli involved, management of a 
database in which they are stored, and will provide measurements of acoustic information such as 
F0, duration, formants, intensity and spectral slope when required. Stimuli may be from existing 
corpora, or produced by human speakers or artificial means. It is possible to include or exclude 
information that may influence the perception of the signal, in order to determine the relative 
contribution of different correlates and other factors to perception for a given group of listeners. 
For example, a speech signal may be low-pass filtered to a certain level to exclude segmental 
information, leaving only pitch. Examples relevant to Māori include Maclagan et al. (2009b) and 
Watson et al. (2011). Within the present study, low-pass filtering of the type described above was 
also trialled during the methodological development process (see Chapter Three, Section 3.2.1).
Rhythm metrics (see, among others, Ramus, Nespor, & Mehler, 1999; Grabe & Low, 2002; Dellwo, 2006; and White & Mattys, 2007) compare consonantal and vocalic proportions and variation by means of formulae, with varying degrees of correction for the effects of speech rate, returning a typological result relating to timing pattern (stress- and syllable-timing). However, they have met with criticism for being only partly related to the rhythm they purport to measure (Arvaniti, 2009; Cummins, 2002). Grabe and Low’s (2002) metric, the Pairwise Variability Index, or PVI, has been tested on, and shown to be unsuitable for, Māori, which it places with stress-timed languages, despite evidence that listeners can distinguish between Māori and English when segmental information is removed (Maclagan et al., 2009b). This PVI result is due to the large number of vowel clusters in Māori; these can be quite substantial (e.g. kāaeaea, (n.) “bush hawk”, (v.) “to look greedily; act hawk-like”), and often cross word or morpheme boundaries (e.g. whakaaro ai au, “why I thought [to do sth.]”). Since metrics are intended to produce a general classification and do not aid in the quantitative and qualitative measurement of prominences, they were not used in this study.

In order to obtain prominence information, speakers can be asked to produce utterances under circumstances designed to elicit specific potential prominence, stress or intonation patterns, such as questions and answers, fill-in-the-blanks, and carrier sentences. There are also repetitive production tasks such as speech cycling, in which phrases are produced in time with a periodic stimulus (Cummins, 1997; Tajima, 1998; Zawaydeh, Tajima, & Kitahara, 2002), and reiterant speech (for example, sasasa), though as Fougeron and Keating point out, the latter causes some speakers difficulty and may result in “exaggerated rhythmic alternations” (1997, p. 3738). The measurement of features such as articulation effects or voice quality, can include articulatory measurement tools like the electropalatograph (Fougeron & Keating, 1997). There are also automated prominence detection methods under development, whose algorithms measure and compare relative acoustic values such as F0 and duration across a group of syllables, and estimate prominence locations based on this. An example of such a tool is the ANALOR software based on French (Avanzi, Lacheret-Dujour, & Victorri, 2010).

Perceived prominence elicitation may also involve perception of stimuli, either natural or artificial, by trained or untrained listeners. The listeners can be asked to identify prominences they hear by
mimicry or choosing prominent units from a selection of visually or auditorily displayed options: these might be the syllables in a word, phrase or utterance, or the word, phrase or utterance itself, as in, for example, Mo (2010). Listeners can also be asked to tap or clap along with a signal: they must synchronise to an auditory beat (as opposed to, for example, a visual one), and continue to track that beat in order to process it and respond (Tierney & Kraus, 2013). This method has been used in both elicitation of rhythm and prominence in general, with results that suggest the task is sensitive to listeners’ native language and to the language of the stimuli (e.g., Lidji, Palmer, Peretz, & Morningstar, 2011; Scott, Isard, & de Boysson-Bardies, 1985; see also next section, and the pilot study of the present project in Chapter Three, Section 3.2). As this project was concerned with listener perception, both tapping and a variation of the task involving selection of options for prominence were used in prominence elicitation here.

2.3.3.3. Native Speakers and Prosodic Assessment

Language background and exposure also play a role in this type of experiment. Kohler (2009a) and Arvaniti (2009) (though this view is held by many others) underline the importance of the intuition of native speakers in prominence identification or prosodic assessment of any sort. They should be displaced by neither non-natives nor machines nor mathematical formulae; however, the nature and purpose of any experiment will have bearing on this. Non-native, L2 or zero-exposure participants are often included to provide comparison, even where they are not the primary source of results. It must also be acknowledged that not only proficiency but level of exposure to the language under consideration or to other languages do make a difference to perception.

If an ability to perceive contrast is available to all humans, and prominence is inherently relative, then anyone ought to be able to perceive prominence in a language that is not their own. Whether or not this is the same contrast or prominence that is identified by native speakers on listening to the same language is, as explained below, dependent on the resources available to the listener. There is ample evidence that humans possess an ability to distinguish languages and varieties of languages from each other, with and without segmental information, and with and without knowledge of the language(s) concerned. There is a limited set of available acoustic cues, and certain of these, namely, F0, duration and intensity, are employed, and salient, cross-
linguistically, even to infants (Nazzi et al., 1998). However, the use of the cues is, of course, controlled by the prosodic phonology of a specific language (Kim, Broersma, & Cho, 2012). While non-natives can and do identify prominence and contrast in general, in order to do so, all listeners rely on perceptual cues with which they are familiar, regardless of the language to which they are listening. This has been shown, for example, in studies on perception of prominence in Korean by English, Japanese and Korean listeners (de Jong, 1994) and French and English listeners (Frost, 2011); use of F0 by Dutch, English and French listeners (Tyler & Cutler, 2009); and speech segmentation units used by Japanese, French and English listeners (Otaka, Hatano, Cutler, & Mehler, 1993). The listeners’ perception was, in each case, based on their own native criteria or cues, producing different results. Specifically to Māori, Bauer (1993, p. 557) makes the observation that “English speakers and Māori semi-speakers do not always pay attention to the features salient from the Māori point of view [§ 2.4.1.3], but are influenced by loudness and relative pitch in their assessment of stress.” However, the results of other studies have been mixed. For example, Scott et al. (1985) found that the rhythm detection responses of French and English listeners in a tapping exercise were very similar when they were exposed to stimuli from both languages, as well as to stimuli with no segmental information. It appears that it is not only native versus non-native cues, but also the languages under investigation and the experimental process that make a difference.

Preliminary work on the present study showed agreement of an acceptable level among an assorted group of listeners, including those of both native and non-native linguistic backgrounds, on what was prominent in the stimuli provided (Thompson, Watson, Charters, Harlow, Keegan, King, & Maclagan, 2010a). Later work made a specific comparison of the responses according to proficiency level, and found both similarities and differences (Thompson et al., 2010b, 2010c). Moving forward, there was motivation to test whether or not, and if so, how, proficiency and exposure to the Māori language affected participant responses to the prominence perception task. Methodologically speaking, this was an important question, if the process developed here were to have any application beyond, specifically, the language for which it was designed. See Chapters Five and Six for results and further discussion on this point. For now, we return to the description of Māori: this time, its prosody.
2.4. Māori Language: Prosodic Description

The following sections will cover the rules of stress assignment, as well as description of phrase and intonation contours. While there is existing description of Māori prosody (for example, Bauer, 1993; Biggs, 1961, 1969, 1998; de Lacy, 2003, and Schütz, 1985 on loanwords), in-depth acoustic prosodic analysis is limited. As Schütz (1985) describes, stress, or what he calls “accent”, has been observed in Māori since the early 19th century, with the work of the missionary Thomas Kendall (1815), who attempted to document the accent position on some words in his language primer. Subsequent description was variable (largely, either present or absent), but Ngata (1926, p. 8) gave a few guidelines on the position (specified by syllable) of primary and secondary “accent” in a word. He also emphasized the “purity” or “fullness” of vowel production in connection with this accent. The early works were concerned with education of the non-Māori speaker and facilitating communication; thus, there is often a focus on correctness in pronunciation that leans more towards vowel quality than rule development. Linguistic discussion of stress position in Māori references the core work of Biggs, (1961, 1969 (revised 1998), 1973 and Hohepa (1967). The essence of these will be presented here, along with some previous investigations into the nature of rhythm in Māori, and a brief discussion of where Māori is situated in the traditional rhythm class typology. Finally, we will touch on the concept of mita, sometimes related to rhythm but in fact much broader, and its connection to the present study.

2.4.1. Stress Assignment

Based largely on the work of Biggs and Hohepa mentioned above, there are two accepted types of stress in Māori: word stress and phrase stress. These are not contrastive, and their locations are, it is argued, predictable from grammatical, morphological and phonological structure. Bauer (1993, p. 560) comments that no intonation peak is clearly observable in Māori sentences (where these are larger than phrases), suggesting no category of sentence stress. Working from the largest to the smallest units relevant in describing the stress categories in Māori, first we have the utterance or sentence. While there is no sentence stress, the position within the sentence of the phrase, which is the next unit, affects the location of phrase stress. Below the phrase is the word, within which word stress is assigned according to a further set of rules. The units within words which bear stress are syllables. The mora is required for the description of stress in so far as it qualifies syllable
weight, which is relevant to the word stress rules. As mentioned earlier (§ 2.2.1), monomoraic syllables are light, and bi- or trimoraic syllables are heavy.

The first detailed analysis of the connection between syntactic and prosodic organization in Māori was that of Bruce Biggs (1921-2000), who discussed the issue in his 1961 work *The Structure of New Zealand Māori*. In this work, he described a system of *pause-bounded* contours. At the top level were “isolable contours” (roughly, whole utterances, sentences or intonational phrases). Within these were “contour spans”: also pause-bounded, and usually corresponding to the grammatical phrase (§ 2.2.2), and “contour words”. The latter appeared to align with the left edge of a grammatical phrase or “contour span”; that is, they were concerned only with the preposed periphery and the nucleus of the phrase, but the distinction between contour words and contour spans was not especially clear. In some instances, as in an example given by Biggs of the imperative *haere!* (“go!”), it is obvious that all of the three units just described could be exactly the same (1961, p. 15). This is, in fact, still the case in later analyses, of which more below.

In the 1961 work, Biggs identified “contour stress”, “primary stress”, and “secondary stress”, in decreasing order of what he refers to as strength, which is connected to pitch height (p. 11). Broadly speaking, the first two correspond to the current categories of phrase and word stress. It is not clear that secondary stress is associated with anything other than heavy syllables. Hohepa (1967, p. 10) also identified more than two levels of stress (in fact, four), but Bauer (1993, p. 556) commented that she was unable to identify more than two levels (or three categories: specifically, “primary”, “secondary” and “unstressed”). This was also the case in the extensive revision of stress categories that Biggs published in his 1969 textbook, *Let’s Learn Māori*. This was, in part, a response to Hohepa (1967), the description in whose own work was partly a response to Biggs. Hohepa pointed out that Biggs’ “phonemic” stress distinctions compared words and phrases, not words and words, and argued that stress was, in fact, predictable on both the word and phrase level (p. 26). It is the revised set of rules outlined in Biggs (1969) that are now generally accepted. It should also be noted that Biggs’ rules for stress assignment, from his first analysis forward, are impressionistically based. The stresses were not originally measured acoustically; rather, they are based on where (by syllable) and how (by prosodic cue) he heard salience or prominence in words and phrases. The corpus on which Biggs’ analysis was based was
largely comprised of material from speakers older than he, or somewhat contemporary (1961, pp. 5-6): this makes them equivalent to the historical elder group (and perhaps in a few cases the present-day elders) in the MAONZE database used in this study (see Section 2.5). Biggs himself would be roughly contemporary with the present-day elder group, being approximately 10-15 years older than them. The word and phrase stress rules are described below: most of the description is based on Biggs (1969, p. 132-133 and 1998, p. 172-173).

2.4.1.1. Word Stress

The location of word stress (hereafter, WS) is predictable from rules based on syllable weight. In monomorphemic words, WS is assigned according to a syllable weight hierarchy (Bauer, 1993, p. 557), where heavy > light and, in more detail, long vowel > diphthong > short vowel. The highest available syllable type that is found within four morae from the end of a word will have the WS. Long diphthongs are not discussed; however, given their size (3μ), it is not possible for there to be a conflict between these and another heavy syllable (minimum 2μ) within the four morae concerned. Where all syllables in a word are light (as in, for example, whenua ‘land’), the WS will be found on the first available syllable/mora, up to the fourth from the end. In (8) below, the WS syllables are shown in boldface. Syllables are separated by plus signs (+), and morae by periods (.)

(8) ko.i+ra.a  koirā  “and so”  
    ko.o+re.ro  kōrero  “story”  
    whe+nu.a  whenua  “land”  
    mi+hi  mīhi  “greeting”

Where there are more than four morae in the word, the process is repeated on the next four morae; generally such words are proper nouns (9) or loanwords.

(9) Nga.a+ru+a+wa.a+hi+a  “Ngāruawāhia” (a town in the North Island)

These weight-based stress assignment rules also apply to some long vowels or diphthongs created across morpheme boundaries. Prefixes and reduplications have different rules. Example (10) contains illustrations of these.
In the analysis of Māori prosodic phrasing by de Lacy (2003), which uses terms referring to levels on a prosodic hierarchy, WS is assigned within a prosodic word, which in this case effectively aligns with a content word, or the phrasal nucleus (p. 63).

2.4.1.2. Phrase Stress

The next domain of stress assignment is the phrase, as defined in Section 2.2.2. All phrases are expected to have one phrase stress (hereafter, PS). The location of this PS is also predictable from a combination of word stress location and the position of the phrase in the sentence. In sentence-final phrases, the PS is expected on the WS of the last content word, not including enclitic pronouns (Harlow, 2001, p. 16). In non-final phrases, PS is expected on the penultimate mora of the phrase, which may depart from the content word WS position. A comparative example is in (11), using the phrase from the syntactic description in Section 2.2.2). Phrase divisions are vertical lines; word stress is in boldface, as before; and phrase stress is underlined.

(11) Non-final: | ki te whare nei | “to this house” (P Det. house Prox.)
Final: | ki te whare nei |

The phrase is, under all descriptions, bounded by pauses. As Biggs (1961, 1969, 1998), Bauer (1993), and de Lacy (2003) describe it, the pause-defined phonological phrase (here, PPh, after de Lacy (2003) and others) has exactly the same shape as the syntactic phrase (XP) described above: nucleus and optional peripheries.

It is also allowed that in rapid speech, more than one XP may be included in a PPh. Bauer (1993, p. 559) limits this to two such XPs, and notes that the conflation is not always reflected in the intonation pattern (§ 2.4.2). Biggs, on the other hand, permits “several” (1998, p. 172) or, originally, up to six XP in one PPh (1961, p. 16). However, he is not specific about precisely how such groupings are identified; presumably it is based on the length of the pauses at either edge, relative to that of the internal pauses. De Lacy observed a general post-PPh pause of about 20-
30ms in the “moderately paced natural speech” of one informant (2003, p. 64). It is not clear from any analysis if there is a limit to the length of the syntactic phrases involved, whether in milliseconds, words, syllables or morae, and if this affects how many XP may be included in a single PPh. The criteria being ill-defined, these circumstances were not applied in the analysis used in the present project.

2.4.1.3. Correlates of Prominence and Stress

Where description of the correlates of stress and prominence in Māori is concerned, there is, as mentioned above, a connection between word stress and duration, via syllable weight. According to Bauer (1993, p. 555-556), the most constant correlates are “pitch fall and length of vocalic element”. She also comments that these may be accompanied by various articulation effects: generally forms of strengthening and lengthening in consonants. Loudness is observed, but not noted as a particular correlate. Phrase stress is often connected to word stress, owing to its assignment to a word stress position in final phrases, but is in fact most often associated with pitch as a correlate: it is described either in relation to a fall, or as a pitch peak. The next section contains a detailed account of this association.

2.4.2. Intonation

The amount of variation that exists in Māori intonation is a point of comment (Bauer, 1993; Harlow, 2007). However, there is some consensus on general tunes. Biggs’ original description (1961, p. 11) noted “decreasing loudness and falling pitch on the final syllable or two of the [isolable] contour”, which we may equate to a pitch peak followed by a fall near the end of an utterance.

In more recent and detailed intonational analyses, there are considered to be four main tune types that a PPh, rather than an entire utterance, may have (though Biggs’ description essentially still stands). The first is the default: each PPh in a declarative is expected to have a rise-fall contour, or, in de Lacy’s (2003) analysis, a H*L⁻ tune. This will be associated with a decrease in loudness in final phrases (Bauer, 1993, p. 559). De Lacy describes the pitch peak (“high target”, or H*) as occurring on the “most prominent” syllable in the phrase, with a “gradual” descent towards the low target found at the phrase end (2003, p. 64). This is entirely consistent with Biggs’
description: “each contour [=phrase] has a peak of intonation which is heard as the most prominent point in the phrase… In a [sentence-final] contour the pitch of the voice falls (except in questions and exclamations)” (1998, pp. 172-173). This brings the discussion to the other tune types: the continuation rise, in which, as Biggs describes it, the pitch is “held up or rises to the end of the contour” (1998, p. 173). The imperative starts on a high pitch, and drops across the phrase. If there is a second phrase, it will have the declarative tune (Bauer, 1993, p. 559). A further tune is the interrogative, usually involving a rise, but which, as Bauer (1993, pp. 2ff) explains, has much variation according to question type and speaker. There may be a rising terminal, a fall, a rise on the (non-final) question word, higher pitch throughout, or the speaker may simply retain declarative intonation and rely on lexical means to mark the question. Increasing presence of the high rising terminal in non-interrogatives, a feature of New Zealand English and Māori English that was mentioned earlier in the section, has also been noted in Māori speech (Bauer, 1993, p. 560).

Declarative (falling) tunes can be seen in the phrases in both (12) and (13) below, and an interrogative (rising) tune is visible in the final phrase (konei) of the question shown in (12). Phrase boundaries are marked here with ( | ).

(12)

![Diagram](image1)

He tuhanga rei na | e hau | kei konei?

“Do you have siblings here?”

(13)

![Diagram](image2)

Te wāhanga tuatahi | e pā ana | ki ngā mabi whai.

“The first part relates to string games.”

Also marked on the above diagrams are six H*: these represent the theoretically prominent pitch peaks. Phrase stress is marked on the example sentences with an underline, and word stress in boldface. There is general consensus from both impressionistic (Bauer 1993; Biggs 1961, 1969,
1998) and acoustic (de Lacy, 2003) points of view, that such peaks (and the following falls, if present) are associated with the phrase stress. This can be seen in the first and final phrases of (12) and all three phrases in (13) above. Bauer also points out that the H* may sometimes appear on the syllable preceding the phrase stress (1993, p. 559), but this is not indicated as a particular rule. An example of an exception (or possible change) is seen in the middle phrase above (ōhau): phrase stress and H* do not coincide; nor does H* precede the phrase stress. One of the goals of the main study presented in this thesis is to investigate whether or not F0 behaviour and listeners’ perceived prominence results are consistent with these observations.

As noted in the section on stress above, according to Bauer (1993, p. 560), there is no apparent intonation peak at the sentence (or, intonational phrase) level in Māori. There is sometimes a general pattern of descending pitch across the phrases in a declarative utterance, but this, like other elements of Māori intonation, is not consistent enough to be called a rule. The right boundary of an intonational phrase is, as discussed in all the work mentioned here, often marked by devoicing (Biggs, 1961; de Lacy, 2003), though such devoicing may also occur in other environments.

2.4.3. Timing, Rhythm Class, and a Note on Mīta

While the rhythm and prosodic features of Māori English have been investigated (Szakay, 2007), the study of rhythm in the Māori language itself is relatively new (King et al., 2009; Maclagan et al., 2009b; Watson et al., 2008, Watson et al., 2011).

To situate Māori in the traditional rhythm class typology, it is usually called a mora-timed language (Bauer, 1981; Biggs, 1980). As was partially discussed earlier in the section on syllable structure, the description of Māori as mora-timed is based on morpho-phonological rules governing reduplication and the formation of imperatives, vocatives and other constructions (Bauer, 1981; Harlow, 2007), as well as on the “Rule of Eight” that accounts for certain lines of poetry (Biggs, 1980). As in Japanese, another mora-timed language (Warner & Arai, 2001), such rules must refer both to the syllable and the mora. As Bauer (1981) puts it, in Māori, the syllable alone is not sufficient to account for all phenomena, but must be used for stress rules as the mora version is uneconomical.
If timing classifications are considered to sit on a continuum between stress-timing and mora-timing, with syllable-timing in the middle, it is thought that the influence of English on Māori may be causing a shift towards syllable-timing in the latter. This is based on evidence from the loss of the “ka-rule”, which is one of the “other grammatical constructions” mentioned above. Ka is an aspectual particle which will appear with a long vowel when what follows has only two morae, and a short vowel when the following material is longer than two morae. Because words of two and three morae can both have two syllables, a movement towards syllable-timing could be contributing to documented reduction in use of this rule (Harlow et al., 2011; Watson et al., 2008). While timing cannot be directly equated with rhythm (§ 2.3.2.2), it is possible that under the English influence, rhythm as a whole, or elements that produce rhythm, may be included among the other changes. It is here that a discussion of the concept of mita may be helpful.

Māori elders\(^5\) are aware of a distinctive rhythm to their language. They are also, anecdotally, sensitive to a change in what they refer to as its mita (King et al., 2009; Maclagan et al., 2009b). It is thought that the origin of the word may be related to the English metre, but this is not certain. It is sometimes described as the “sound” or “accent” of a language, and may apply to any language: it is not exclusive to Māori. Some commentary on the prosodic aspects of Māori has linked the concept to rhythm. The definition of the latter term can be rather varied, but even at its broadest, it does not encompass mita. When it comes to a precise definition of either the change or the term, difficulty arises. Mita, it would seem, spans a range from regional lexical or phonetic variation to prosodic elements such as stress, rhythm and pitch (Maclagan et al., 2009b). Because of the difficulty in defining the term, it is not particularly useful in the present discussion of Māori, and the account of it is simply to provide an explanation in the event that the reader wishes to explore the documentation of Māori further.

While the concept of mita itself is not central to this project, it is the concept of some change connected with it, whatever that change might be, that provided the impetus for investigating the nature of prosodic prominence in Māori. With Biggs’ stress rules already established (Biggs, 1969, 1998), and the related description of the characteristics of the stressed syllables as described earlier,

---

\(^5\) Elders (Māori: kaumātua) have a respected place in Māori society. They are considered to be leaders and guardians of tradition for tribes and families. They have strong cultural and historical knowledge and speaking skills.
one could simply take the existing illustration of stress and prominence in Māori at face value. However, the problems encountered in describing the change are an indication that it could be or involve a prosodic element. Some listeners may only comment that, for example, “it just sounds different”, but others go as far as to say, specifically, that older and younger speakers have different “rhythm”. This was the case in informal feedback from a small number of participants in the study presented in this thesis, where they (accurately) based their evaluation of age on voice quality in the recordings they heard. However, when questioned on what made them say that, they could not describe it further, falling back on “it’s just different”. These comments are interesting in their own right. However, they can also be combined with vowel quality studies (Harlow et al., 2009) and language identification task results (Maclagan et al., 2009b; Watson et al., 2011). When played low-pass filtered utterances in English, Māori and Japanese, listeners can tell, above chance, that the first two are different; more confusion occurs between Māori and Japanese. On top of this, when played either Māori or English utterances spoken by older Māori, listeners were more likely to identify the language as Māori. When played Māori or English utterances from younger speakers, they were more likely to identify the language as English. There is, then, definite evidence of change and influence from English, but still a clear indication of difference and the ability to hear it. With this support for the assertions about change in Māori, there is a clear suggestion that it is worthwhile to pursue the issue further. If the stresses described above are affected by this change, what can listeners’ perceptions tell us about it? Are the stresses still perceived in the same places as they were when the rules were formulated? Are their characteristics, the acoustic features to which they are linked, still as described? A number of studies addressing these questions have been launched using material in the form of recorded utterances from the MAONZE database.

2.5. The MAONZE Database

The MAONZE (or, Māori and New Zealand English) database, begun in 2001 by the members of the MAONZE project team (King, Maclagan, Harlow, Keegan, & Watson, 2011a, 2011b; Keegan, Watson, Maclagan, & King, 2014), is a database of audio recordings containing Māori and English speech, along with transcriptions of those recordings. The database project was begun with the purpose of analyzing changes in the Māori language over time. For a full
description of the database and its construction, see King et al. (2011a). All of the recorded stimuli used in the experiments in this thesis come from speakers in the MAONZE database, introduced in this section.

2.5.1. Speakers

The 62 speakers in the database are of both genders, come from different age groups, and were recorded in different time periods. Their birthdates span a timeframe of more than 100 years, from the late 19th century (1871) through to the late 20th century (1992). The recorded speakers include both first language (L1, native) Māori speakers, as well as some L2 speakers. There are both Māori and English recordings for almost all of the speakers.

The oldest speakers, most of whom were born in the 1880s, were recorded by the Mobile Disc Recording Unit, which toured New Zealand in the 1940s conducting interviews for broadcast (see Gordon et al. (2004, p. 38ff), for a description of the collection of these recordings); the discussion focuses largely on the English speakers, but the recording circumstances were the same for all). These Mobile Unit speakers are referred to in the text of this thesis and of studies under the umbrella of the MAONZE Project as the historical elders. All of them are L1 Māori speakers who did not learn English until they went to school or were adults (King et al., 2011a).

The second group of speakers are the present-day elders: these men and women were born between 1920 and 1940, and recorded between 1954 and 2010 (most between 2000-2010), making them older adults between 49 and 101 at the time of recording. Like the historical elders, these speakers also have L1 Māori and would not have learned English until at school. Most did not speak English regularly until they were adults; however, compared to the historical elders, they would have had less exposure to Māori across their whole lives (King et al., 2011b).

The third group are young males and females born between 1969 and 1992. Some of these are L1 speakers, having been raised in a Māori-speaking environment and attended Māori-speaking schools. Some are L2. All of them have high exposure to English and speak it with native fluency (King et al., 2011b), where this was not always the case for the elders.
2.5.2. Recordings

The recordings in the database take slightly different forms. Interview-type recordings are available for all of the groups. The older recordings are more formal than the recent ones, and it is possible that two of the historical speakers had notes (King et al., 2011b). However, all recordings contain continuous speech. The recent interviews conducted by the MAONZE team include a read passage (one Māori, one English) and some word lists in addition to the interview portion. The recordings are all transcribed and searchable using a web-based interface. Where the recording signal is concerned, there are further differences arising from the age of the recordings. The historical recordings are analogue, made on acetate disks; the modern ones are on digital audio tape. All are converted to .wav format for the database. The mobile unit recordings are stored at 16kHz owing to the nature of the originals, with the modern ones at 20 or 22.05kHz (King et al., 2011b). Some of the recordings (the modern ones more than the historical ones) contain background noise, since they were not made in a specialist recording studio: the participants chose the location.

2.5.3. Use of the Database in the Present Study

In the experiments that make up the project presented in this study, different speakers from the database were used, including male (L1) speakers from all three groups. Female speakers from the two present-day speaker categories were also used in one of the pilot experiments, but the lack of availability of recordings in the oldest (historical) category at the time of the study precluded the use of female speakers in the main experiment. More detail on the speaker and recording selection from these groups to create stimuli for the study is provided in the following chapters, where the three phases of the project are described in full.
PRELIMINARY STUDIES

DEVELOPING THE EXPERIMENTAL PROTOCOL

3.1. Introduction

The motivations for the project, as outlined in Chapter One, were as follows: to establish the current location of perceived prosodic prominences in Māori speech and their consistency with existing description of stress locations; to determine the acoustic features associated with these perceived prominences and their consistency with existing impressionistic description; to investigate whether or not either listener proficiency in the Māori language or age of speaker and recording made a difference to the prominences that were identified; and to develop a methodological approach and tool by which these prominences could be elicited and analysed effectively.

Initial experiments in the project were centred on determining whether or not participants would identify prominences consistently, and creating a tool, in the form of a perception test, to measure this as accurately and efficiently as possible. This was then developed, through different phases of testing, evaluation, and use, into a methodology capable of addressing the questions above.

This chapter is about the first two phases: the preliminary studies. The original pilot study used a tapping-based method. The second phase of development used a web-based method. The stimuli and participants in each study will be introduced, and the delivery methods will be described, including the development of the second test for the web. The results of both studies will be
presented in brief, along with evaluations of the methods used and discussion of the modifications that were considered necessary.

Chapter Four will describe the final methodology for the main experiment in the project, including the methodological adjustments that arose from the effects of participant proficiency differences. Chapter Five presents the results of the main experiment, based on the responses of all participants. Chapter Six presents the results of the investigation into differences in response based on participant language proficiency.

3.2. Pilot Study: Tapping

The first experiment in the project was a small-scale pilot study, designed to establish whether or not tapping would be a suitable method of getting listeners to identify elements that stood out to them in spoken phrases. If so, would they do so consistently and in agreement with each other, and would the standout elements cluster in certain locations, such as syllables or morae? Despite informal pre-experimental predictions about these locations on the part of the researcher, participants were not led to focus on any particular element or prosodic unit.

3.2.1. Stimuli

The stimuli used in this study were short portions of read speech from three present-day elder (PE) and three present-day young L1 (PY) Māori speakers, taken from the MAONZE database (see Chapter Two for MAONZE speaker groups). To maintain consistency, and because the number of stimuli to be used in the study was relatively small, all of the speakers were male. Their recordings were chosen based on clarity and lack of interruption. Two sentences were taken from each of the six speakers: one English sentence, from a letter, and one Māori sentence from a story passage. Using read speech made it possible to obtain multiple versions of the same sentence from all of the different speakers, in order to limit the effects of idiosyncrasy. The full texts of the story and the letter, with a translation for the Māori story, are in Appendix B1. The two sentences that were used in the study are shown in (1a) and (1b) below. Glossing for the Māori sentence can be found in Section 3.2.5.1.
(1a) *Ahakoa ko te hōtoke, he tino mao te rangi.*

“Although it was winter, the sky was very clear.”

(1b) Although the weather is nice at the moment, the forecast is for hail.

Using audio editing software (see next section), two versions were made of each recording: one unfiltered and one low-pass filtered to 400Hz (with 50Hz smoothing). The purpose of having the filtered versions was to see if the filtered condition would alter the perceived location of standout points; that is, whether or not segmental or semantic information had any effect.

![Figure 3.1. Summary of the stimulus set for the tapping experiment.](image)

In total, there were 24 stimuli, taken from six different speakers: 12 Māori and 12 English. The length of the stimuli ranged from 3 to 6 seconds, with an average of 4.3 seconds. The total playing time of the 24 stimuli was 104 seconds. The stimulus set is summarized in Figure 3.1 above.

### 3.2.2. Participants

The group of participants in the pilot was very small: only five, who came from within the research group with which the investigator is associated. All were linguists with varying degrees of Māori proficiency, ranging from limited to very high. A benchmark result was recorded at the outset while pre-testing the exercise. To create this, the investigator recorded her own responses three separate times. This brought the total of submitted sets of participant responses in the study to seven, where three were from the same person. The actual number of participants was intended to be larger; the main reasons this did not happen were the impracticalities discovered in the

---

1 Because of the in-group participants, no ethics approval for the study was required at this early stage. For subsequent experiments, ethics approval was sought and granted by the University of Auckland’s Human Participants Ethics Committee (see Section 3.3.2.2 and Appendix A).
experiment and the time involved. These problems are discussed further in the evaluation below (§ 3.2.6).

3.2.3. Experiment Setup and Delivery Method

Physically, the experiment required a soundproof booth, a microphone on a stand, headphones, sticks for tapping, and a computer with sound software to provide the stimuli and record the responses. For the latter, the free, open-source audio editing and recording software Audacity (http://audacity.sourceforge.net/) was used.

The four listeners (besides the investigator) who participated in the experiment were played the stimuli a number of times for practice, and asked to familiarize themselves with the parts of the utterances that “stood out” to them. This instruction was intentionally minimal and intended to avoid leading the participants toward any particular cue. The participants were encouraged to practice tapping along with what they heard as the stand-out parts of the utterance using the sticks provided, and were given as much practice as they needed to feel comfortable with the exercise. Their practice taps were not recorded.

After the practice, the participants were recorded tapping along with each of the stimuli, in front of the microphone. They were given a second opportunity to record if they wished to do so. For each participant, the taps for each of the 24 stimuli were recorded on separate tracks using Audacity’s overdub function. This allows one track (here, the stimulus) to play while another (the tap) is recorded, with a simultaneous start. The resulting tap tracks were saved as separate files, whose filenames included identical codes to those of the relevant stimulus files.

All participants heard the stimuli in the same order, which was “artificially random”. That is, while an effort was made to alternate older and younger speakers, English and Māori stimuli, and filtered and unfiltered utterances, the primary concern was ensuring that participants never heard the filtered and unfiltered versions of the same recording in sequence.

The experiment was not automated, so the participant and the investigator both needed to be present. Since the participant’s hands were occupied with the sticks, they could not also use a
mouse with the computer to record the tracks. The investigator’s own experience recording the benchmark responses, which required careful balancing of sticks and a high degree of familiarity with the software involved, suggested that untrained participants would not cope easily with the exercise if required to do it alone.

Owing to different participant requirements for practice, tap recording, and re-recording when requested, the exercise took vastly varying amounts of time for each participant. This resulted in one participant not finishing the exercise for all 24 recordings.

3.2.4. Result Processing

After the data collection came a multi-stage process involving database creation and markup of the utterances for relevant units such as words, syllables, and phonemes, as well as tap locations, and the assessment of any alignments between these units and the taps.

The speech analysis software Emu (http://emu.sourceforge.net/) was used to create a database in which the recording files were manually segmented to show the location of the words, syllables and phonemes within the sound files. In brief, this means that it was possible to view the recorded speech as both waveform and spectrogram, with the start and end points and the content of the relevant units (words, syllables and phonemes) labelled according to standard criteria. For a more detailed explanation of this process as used in the present study, see Chapter Four. For even greater technical detail, see Harrington (2010b). Such markup is impossible on low-pass filtered recordings because of the (intentionally) missing visual and audio information. To solve this, first the markup for each unfiltered recording was completed, and then the time points for the various units were transferred to the equivalent low-pass filtered recordings, which were identical in length.

The recorded tapping files, which contained only the locations of the taps from each participant, were also marked in this way. Emu creates a set of files to store the labels for each level (or unit) involved in the markup (words, syllables, taps, and so on). There was, therefore, a tap file for each of the stimuli, for each participant (7 tap files per stimulus recording).
Since one of the main goals of the exercise was to observe clustering in the tap responses, all of the tap files, containing the time points for the taps from each participant, were concatenated into a single tap file and linked with the master database, where they appeared as an additional “tap level” of markup on the relevant stimulus file. The alignment of the different participants’ taps with the parts of each utterance could then be easily observed. Figure 3.2 below is a partial screen capture of the Emu software signal view of a fragment of one of the Māori stimuli, with completed markup. Normally, the waveform and spectrogram would be visible below the different levels, which are magnified here. Some tap clustering is visible on the tap level.

![Figure 3.2. Partial Emu signal view screen capture showing different levels and taps in database.](image)

Emu allows hierarchical association between the levels. This association was done automatically using a script that links the elements according to their times, and the result was checked by hand for errors. The labelled taps were associated with the units on the levels above them in this way. The Emu hierarchy view of the same section of the Māori utterance above is in Figure 3.3 below. In this screen capture, the sentence level is also visible; in this database, it was used for the stimulus codes.

![Figure 3.3. Partial Emu hierarchy view screen capture showing different levels and taps in database.](image)
For convenience and consistency, taps were assigned to the syllable in which they fell. The participants were not specifically instructed to aim for syllables when they tapped, but since all of the participants were trained linguists, intentional syllable-orientation may have occurred in responses to the unfiltered stimuli. For the filtered stimuli, it was not possible to aim for specific syllables, except perhaps by learning the utterance pattern. In the database, however, the alignment of syllables and taps was visible for both unfiltered and filtered stimuli, and based on this, the distribution of the taps on the syllables was recorded in a spreadsheet. The results are presented in the following section.

### 3.2.5. Results

All of the results shown here include the responses from the six completed tapping tests. The incomplete test result was omitted.

#### 3.2.5.1. Results for the Māori Stimuli

Below are results from the responses to the Māori stimuli, all of which were versions of the sentence shown in (2). This is reproduced from (1a) above and glossed here for convenience. Vertical lines ( | ) indicate grammatical phrasing, boldface indicates expected word stress, and underline indicates expected phrase stress, as per Biggs (1969).

(2) **Ahakoa ko te hōtoke, he tino mao te rangi.**

    although Foc. Det. winter Det. very clear Det. sky

    “Although it was winter, the sky was very clear.”

The Māori stimulus set allowed for 12 instances of each syllable in the sentence. Across the six recorded responses, including all of the investigator’s benchmark results, this meant that any of the syllables could have been tapped on a possible 72 times in total. For each of the filtered and unfiltered conditions, any syllable could have been tapped on up to 36 times in total (6 stimuli x 6 responses). In Table 3.1 below, the numbers represent the total taps associated with each syllable, out of the possible 36 or 72. The totals are presented for the 6 unfiltered (UF) Māori stimuli together, the 6 filtered (LP) Māori stimuli together, and both conditions (All). The final
The benchmark result is indicated by **bold italics**, and represents 2 or 3 out of 3 taps on that syllable across the investigator’s three recorded results. The hash mark (#) indicates a narrative pause which was produced by all of the speakers. It was included in the database markup (visible at the right in Figure 3.3 above), and some participant taps did fall in this zone.

Table 3.1.

*Tap Distribution in Responses to the Māori Stimuli*

<table>
<thead>
<tr>
<th>Condition</th>
<th>a ha ko a</th>
<th>ko</th>
<th>te hō to ke</th>
<th>#</th>
<th>he ti no mao</th>
<th>te ra ngi</th>
<th>Total Taps</th>
</tr>
</thead>
<tbody>
<tr>
<td>UF</td>
<td>0 2 22 8 3 2</td>
<td>16 15 8 1 1 1 8</td>
<td>29 1 4 32</td>
<td>153</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LP</td>
<td>1 0 11 12 7 3</td>
<td>14 13 12 5 0 0 9</td>
<td>15 1 10 24</td>
<td>137</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>1 2 33 20 10 5</td>
<td>30 28 20 6 1 1 17</td>
<td>44 2 14 56</td>
<td>290</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The participants are not distinguished individually on this chart. The purpose was not to build participant profiles, but to show general consensus or lack thereof.

Across all six readings, all four included participants (six responses, including the benchmark), and both conditions, the syllables considered to stand out most in the Māori utterance were [aha]-ko-[a], mao and -ngi. Clustering of various sizes is evident in several zones, and some of the syllables on and around which the tap clusters fall do have characteristics in common.

In terms of alignment with expected phrase and word stress as described by Biggs (1969; see also Chapter Two), higher numbers of taps do cluster around these locations, particularly in proximity to the phrase stress (PS, underlined). For example, there are clusters around -ko- (PS) and -a in the first phrase; hō- (WS, boldface) and -to- (PS) in the second; mao (PWS) in the third; and ra- (PWS) and -ngi in the final phrase. All of the available phrase stresses show some degree of clustering, though in the case of ra it is small relative to that found on -ngi. Since phrase stress is predicted to be associated with a high pitch point, these zones may have stood out if that high pitch point were present in the phrases and if the listeners were sensitive to it.
In further connection with expected stress location rules, and this time suggestive of a possible participant sensitivity to syllable weight or duration, hō- and mao are both heavy syllables, while ko- forms part of the non-diphthong vowel sequence /oa/. Though ko- is not heavy, and all of the participants were linguists with enough knowledge of Māori language to know that /koa/ was not a diphthong, it is possible that the second, onsetless syllable created a heavy-syllable-like effect, which could have affected the response. The distribution of taps across the sequence /koa/ can be compared between the filtered and unfiltered conditions. When it was unfiltered, with full access to the segmental information available, participants “hit” the first syllable more often than the second. In the filtered condition, without segmental information, the hits on each syllable are almost the same. Both results could suggest response to the two syllables as a single unit.

Acoustic analysis was not carried out on these stimuli and the connected results, chiefly because the numbers were not large enough to allow robust analysis. However, there were indications, in the tap clustering and its association with expected stress locations, of possible participant sensitivity to syllable weight, duration and pitch in choosing standout portions of the stimuli. There may also be indications of delay effects, which can be a product of reaction time in the tapping process (see the evaluation in § 3.2.6 for more on this).

In the final phrase, the high level of tap clustering on -ngi is not immediately explainable. It is a light syllable, and is neither expected word stress nor phrase stress: both of these are expected on the syllable preceding it. In two of the recordings, is it associated with the high pitch point of the phrase (regardless of expected stress category), but in neither of these instances is the degree of tap clustering the greatest, leading us to conclude that this is not the reason for its popularity. However, the reverse may be true: in all cases, the expected fall or lower pitch at the end of the phrase is present, and this may be what caused the clustering. The finality of -ngi may also be a factor in another way, due to the contrast with following silence, but there is no indication of final lengthening and the vowel, /i/, is subject to devoicing in some of the stimuli. Similarly, there is a spread in the results on the word hōtoke ("winter"): while clusters fall on the first and second syllables, there are also enough taps on the third to be interesting. -Ke is phrase final: it appears before a narrative pause, which in Māori can result in lengthening (Biggs, 1961, p. 11)). However, again, when the stimuli where -ke has a cluster and those where it does not are compared, there is
no pattern to the relative durations of syllable and pause; it is more likely that contrast with silence is what participants noticed. Alternatively, we return to the possibility of non-cue-related physical delay effects. For more on this, see the discussion of the English results in the next section.

3.2.5.2. Results for the English Stimuli

Below are the responses to the English stimuli, for which the utterance used is reproduced in (3) from (1) above.

(3) Although the weather is nice at the moment, the forecast is for hail.

Table 3.2 below is equivalent to Table 3.1 in the previous section: the total taps associated with each syllable, out of the possible 36 for the 6 unfiltered (UF) stimuli or the 6 filtered (LP) stimuli, or the possible 72 over the 12 stimuli in both conditions (All). The final column (Tot.) shows the total taps per condition. The benchmark result is, again, indicated by bold italics.

Table 3.2.
Tap Distribution in Responses to the English Stimuli

<table>
<thead>
<tr>
<th>Cond.</th>
<th>#</th>
<th>although</th>
<th>the</th>
<th>weather</th>
<th>is</th>
<th>nice</th>
<th>at</th>
<th>the</th>
<th>moment</th>
<th>the</th>
<th>forecast</th>
<th>is</th>
<th>for</th>
<th>hail</th>
<th>L</th>
<th>#</th>
<th>Tot.</th>
</tr>
</thead>
<tbody>
<tr>
<td>UF</td>
<td>0</td>
<td>7</td>
<td>17</td>
<td>6</td>
<td>2</td>
<td>11</td>
<td>2</td>
<td>35</td>
<td>0</td>
<td>3</td>
<td>11</td>
<td>3</td>
<td>24</td>
<td>11</td>
<td>2</td>
<td>2</td>
<td>27</td>
</tr>
<tr>
<td>LP</td>
<td>1</td>
<td>5</td>
<td>11</td>
<td>11</td>
<td>3</td>
<td>5</td>
<td>4</td>
<td>20</td>
<td>3</td>
<td>0</td>
<td>13</td>
<td>18</td>
<td>21</td>
<td>16</td>
<td>1</td>
<td>1</td>
<td>26</td>
</tr>
<tr>
<td>All</td>
<td>1</td>
<td>12</td>
<td>28</td>
<td>17</td>
<td>5</td>
<td>16</td>
<td>6</td>
<td>55</td>
<td>3</td>
<td>3</td>
<td>24</td>
<td>35</td>
<td>45</td>
<td>27</td>
<td>3</td>
<td>3</td>
<td>53</td>
</tr>
</tbody>
</table>

In this case, the hash marks (#) at the beginning and end of the sentence do not represent narrative pauses, but periods of silence in which taps were recorded by participants. Such pre- and post-utterance taps did not occur in any of the responses to the Māori stimuli.

The results here follow a similar pattern to the responses to the Māori stimuli. There are definite tap clusters, accompanied by a degree of noise. The largest clusters appear on the syllables -though, nice, mo-, -ment, fore-, -cast, and hai-. Nice and hai- have the highest response rates, and are grouped tightly, both in the benchmark and the other participant responses. Although “hail” is
classically a monosyllable, in this case it was given a syllabic /l/ as a second syllable because all of the speakers very clearly produced one. Taps do fall in this final syllable, but close examination of their actual temporal location shows that the majority of these are closer to the preceding syllable than to the end of the word: here is another possible effect of physical delay.²

Where one or both parts of a disyllable feature a tap cluster (al-though, mo-ment, and fore-cast), the tap distribution is wider. In the case of al-though, the stress would be expected on the second syllable. Impressionistically, however, some of the speakers very clearly placed the stress on the first, manifested in either a sharp rise across al- towards the pitch peak on -though, or the pitch peak on al- followed by a relatively sharp fall across -though: whether this was their usual style or a reading-based error or affectation can only be the subject of speculation. However, because examples like this were included in the experiment, it is possible to see whether listeners picked up on the difference, and it appears that they did. Where this shift is present, the taps do align with or towards it, particularly on examination of the exact temporal tap points. In the readings containing no shift in stress, where the pitch peak is found upon -though with no sharp rise beforehand, the tap cluster is firmly on this second syllable, in some cases even leaking into the onset of the following word (the). The sensitivity to pitch is unsurprising, given that the language concerned is English, under consideration by native English speakers.

In the other disyllabic words that feature large clusters, moment and forecast, the stress would be expected on the first syllable, but in both cases, the tap distribution only partly reflects this. Based on the investigator’s impressions, none of the speakers stressed either of these words in an unusual manner: they do not obviously shift the stress to the other syllable, as occurred with although above.

Another possible reason why the taps fall as they do is, again, delay. A look at the actual time points involved revealed that in both cases, if the taps fell in the second syllable, they were much closer to the first syllable than to the end of the word. Often, the taps were marked in the onset of

² Because of physical variables, the exact single time point associated with any given tap is not very useful in itself; it is more useful to look at the overall syllable unit in which the taps occur. However, trends in the exact tap positions within the syllable units can be informative when questioning or speculating on participant intent or factors affecting it.
the second syllable. A similar effect is visible in the word weather, which, despite a lower number of taps, could overall be considered to have a small cluster on the second syllable. None of the speakers stressed it on this syllable, which would be unconventional, stress being expected on the first. Most of the taps recorded on the second syllable, when the timepoints were examined, were in the onset. All of this does imply some reaction time effect.

3.2.5.3. Tap Totals and Filtering: General Discussion

Across the Māori stimuli, there was a tap on or in the vicinity of each cluster from each participant, including the three results from the investigator. There was also noise, but less than that observed in the responses to the English stimuli. However, when the overall totals were calculated, there were more taps from all participants on the English stimuli, both filtered and unfiltered, than on the Māori stimuli. The spacing of taps on the Māori stimuli was also wider and, for want of a better word, more regular: there was less noise and clustering was more defined. Some participants tapped twice on a single syllable for some of the English stimuli, but this did not occur for the Māori.

There were more syllables in the English sentence (18 σ) than the Māori one (16 σ). More syllables in total might imply more possible tap locations and therefore more possible taps, explaining the higher overall totals for English. The proportions of the tap totals were, indeed, very similar to the proportion of syllable totals: the Māori total was between 80-90% of the English total in all cases (see Table 3.3 below).

Table 3.3.

<table>
<thead>
<tr>
<th></th>
<th>Syllables</th>
<th>Taps on Unfiltered (UF)</th>
<th>Taps on Filtered (LP)</th>
<th>Both Conditions (All)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Māori Stimuli</td>
<td>16</td>
<td>153</td>
<td>137</td>
<td>290</td>
</tr>
<tr>
<td>English Stimuli</td>
<td>18</td>
<td>184</td>
<td>168</td>
<td>352</td>
</tr>
<tr>
<td>Proportion (M/E)</td>
<td>89%</td>
<td>88%</td>
<td>84%</td>
<td>82%</td>
</tr>
</tbody>
</table>

It is, however, more likely that the difference is due to greater variety in the English reading styles than to the extra space provided by only two more syllables.
Although, overall, there were fewer taps recorded on the filtered stimuli than on the unfiltered stimuli in both languages, across the individual participant tap totals there was no pattern: neither condition consistently motivated a higher number of taps. The charts in Figures 3.4 and 3.5 below also show that there was also little difference in distribution between the results for filtered and unfiltered stimuli for each language.

The filtered stimuli did not appear to create uncertainty, confusion or lack of consensus. Where there was a lack of difference between responses to the two conditions, it may have been partly due to the multiple readings of the same two sentences, which could have caused an eventual familiarity to develop with the overall pattern of those sentences. Even though the participants did not know which condition or language they would hear next as they were doing the test, it has been shown elsewhere (Maclagan et al., 2009b) that Māori and English are distinguishable, above chance, in filtered speech. After hearing the stimuli once or twice for practice, the participants would have known which language they were hearing.

Where the patterns of the utterance in the unfiltered version were noticeably different (owing to speaker effects, as discussed earlier), the participants still responded to this difference in the
filtered version. This, combined with the number of taps and their overall distribution, suggested that in an experiment of this type, whose purpose is to establish clustering, filtering did not have enough of an effect on perception of standout syllables to make it worth pursuing post-pilot.

3.2.6. Evaluation of the Methodology and Proposed Modifications

“Your work,” one informal reviewer wrote, when aspects of this experiment (Thompson, 2009) were presented at an academic gathering, “nicely illustrates the importance of undertaking a pilot study first.” They were quite right. The pilot methodology fulfilled its chief aim, which was to show that if Māori stimuli were presented in an experiment of this type, listeners were capable of identifying elements that stood out in those (and the English) stimuli, and that a degree of consistency and agreement across participants would be visible in the elements identified. The taps from different participants did cluster on certain syllables. The results also raised relevant questions and possibilities for what might causing those syllables to stand out for participants, and showed that the inclusion of filtered stimuli did not make an appreciable difference to participant responses with respect to the purpose of the study.

The results suggested that the experiment would be worth pursuing, with some modifications that would both expand the project and stimulus set, and address some of the issues encountered in the pilot. The key points for review were efficiency, accessibility and practicality; participant frustration and fatigue; and a solution for the variable of possible delay effects in the results. Most of these problems were solved by changes to the delivery method. They are discussed in more detail below.

3.2.6.1. Efficiency and Accessibility

Tapping, as a method of eliciting perceived prominence, takes a lot of time. The length of the test process in the pilot varied widely from participant to participant, even within the initial small set of five (including the investigator), depending on how many practices they felt they needed and whether or not they required a repeat recording of results. Not including setup, explanation, and pre-recording practices, the actual response recording periods ranged from just under 20 minutes for the whole stimulus set to well over 45 minutes for only half the stimulus set: as mentioned above, one participant could not finish the exercise, having to leave before the process was
complete. This extreme variation meant that it was impossible to predict how long any individual would take; it is quite likely that participants unfamiliar with such experimental situations or who were not linguists and required further explanation or discussion would require even more time. The risk of unfinished results needed to be minimised for the integrity of the study.

The requirements of investigator presence and reservation of the soundproof booth, as well as not being able to predict how many people could be tested within a certain period, meant that arranging test times with participants was difficult. These factors also made the study untransportable, limiting accessibility. This would pose a particular problem in trying to gain fluent Māori-speaking participants, many of whom were unlikely to be available within range of the primary recording location.

A further issue concerned efficiency in result processing. Emu database markup of the tap locations for each participant and addition of those results to the master tap file was technically effective and suitable for a small number of participants but it was also inefficient and susceptible to human error. In a study with more participants, the tap level in the database would become overcrowded and unwieldy. A more automated process using a spreadsheet to collect the tap totals would speed up the process, make the data much more usable, and reduce the potential for human error.

3.2.6.2. Practicality: Participant Difficulties and Variables in the Results

Participant reaction to any test environment is never completely predictable, and this experiment had several factors which caused difficulties. Firstly, some people do not enjoy the soundproof booth environment, which is not an uncommon reaction to enclosed and muffled spaces of that type. On top of that, the tapping method is, in reality, a test of both mental and physical response. Feedback suggested that some participants felt it measured their physical coordination skills rather than their true intent in showing which part they felt stood out. While they knew where they wanted the tap to fall, they sensed that it was too late or too early when they moved their hands. This had the potential to cause a kind of frustration or performance anxiety. Depending on background and personal circumstances, participants in any further tapping experiment might also be concerned by this, while others might not. However, since a comfortable participant
produces better results, and no experiment gains participants through bad press, this was an important issue to be addressed before the project went further.

The possible evidence of delay in reaction time in the recorded results was introduced above. This could be caused by fatigue or coordination issues, which might be natural or due to age, health, nerves, or other negative responses to the test environment. Delay might be different on any given tap, or it might not even occur at all. Tapping is a complex task in a simple disguise: the demand on attention, fine motor control, and auditory processing is very high. Individual skills also play more of a role than might be expected. Differing performances in a tapping task are linked to structural differences in the motor areas of the brain and further differences in the auditory system’s ability to extract the exact onset time of a sound (Tierney & Kraus, 2013). To determine whether or not delay really was occurring, or attempt to eliminate the effect, participants would either have to engage in exhaustive pre-test practice, or possibly repeat the test more than once in order to produce an average result similar to the benchmark used here. This would be time-consuming, its effectiveness would not be guaranteed, particularly if the participant were naturally ill-disposed to the task. Additionally, because the delay duration itself is unpredictable, the process of correcting for it, mathematically or otherwise, complex. A better solution was to implement a different method that permitted participants to be specific about where they heard the standout elements.

3.2.6.3. Expanding and Altering the Stimulus Set

Several changes were also planned where the stimuli were concerned. The first was enlargement of the set, a logical development between a pilot and a full-scale study. There were many more different sentences and recordings from the letter and the story available for use. These were not only from the speakers used in the pilot, but from other speakers as well: both male and female.

Responses to the filtered and unfiltered stimuli suggested it was a distinction that did not need further testing. Removing the filtered stimuli allowed for the inclusion of more stimuli overall, providing a larger selection of potential standout syllables for analysis, occurring in a larger selection of grammatical contexts. The lack of filtered utterances also worked far more effectively
with the major planned alteration to the delivery method, which relied upon the visibility of syllables in the task.

Moving forward, the goal was to allow the participants to feel more comfortable and in control; to increase portability and efficiency of the experiment; to reduce potential errors in the results; and to gain more information from responses to the stimuli used. The change to the delivery method that provided the answer to most of this was the creation of a web-based interface for the second experiment. Putting the test online would allow it to reach a far wider, perhaps international, audience. The web-based test would contain stimuli that were written out and readable for the participants, with the utterances divided into syllables and presented in that way on the screen. Participants would simply check a box for a syllable they considered to stand out. This was intended to remove any mystery from the concept of “syllable” for those unfamiliar with it, or who might be familiar with it but not know how it applied to Māori or English specifically. This was a further reason for removing the filtered stimuli: in the web format, with the syllables separated and assigned to checkboxes, the filtered utterances could not have been used, as it would not have been possible for the participants to associate the sound directly with the visible text. The checkbox method was intended to allow them to be as specific as they wished, while avoiding physical coordination issues or other variables that arose from tapping. To put the participants further at ease, they would be able to listen to the stimuli as many times as they deemed necessary by pressing a button. This would let them be in control of familiarizing themselves with the exercise and avoid any embarrassment or awkwardness about asking for repeat playback. Allowing multiple chances to hear the utterances also had the potential to provide better results over a more varied set of stimuli. Timing would be up to the participant, without the pressure of others waiting, an investigator observing, or of needing to leave partway through. This should result in a majority of completed and viable test results.

Behind the scenes, this progression of the study required coding and development of the test interface for the web and instalment of more effective methods of result processing to manage the larger number of participants. In addition to all of the above, there were not yet any formal criteria used to define “perceived prominences” within the experiment: in the analysis, standout points were judged on clustering only, and assigned to the nearest syllable. As the number of
results grew, became more reliable, and became candidates for further empirical and acoustic analysis, finding a set of criteria and a definition for prominences in that analysis was an essential development that occurred across the next phases of the study.

3.3. The Second Experiment: Prominence on the Web

The chief benefits of the web-based survey used in the second incarnation of the prominence perception experiment were that it enabled far more stimuli and participants to be involved, and it took away many of the impracticalities of the pilot study.

It had the same basic goals as the pilot study: to assess the effectiveness of the new delivery method; to elicit perceived prominence locations from participants; and to determine whether these locations were consistent across participants. In addition, it was to analyze what the perceived prominences had in common, and as the study progressed, a further question was included. With more participants involved, and with these participants having a wider range of proficiency in Māori, it was hoped that the results would show up any differences in prominence identification related to familiarity with the language.

3.3.1. Stimuli

In this second version of the experiment, the read speech from the first experiment was retained, with more sentences taken from the same Māori story and English letter used in the pilot. The chief reason for retaining the read speech at this stage was that it still allowed different versions of the same utterances to be taken from speakers of different ages and genders, to provide variety and reduce individual speaker effects. However, the effectiveness of this proved to be debatable (§ 3.3.4.3). The retention of the English sentences was to confirm that, as suggested by the pilot study results, participants could do the prominence perception task in both a language that they knew, and one with which they might be less familiar or completely unfamiliar.

This time, there were 50 stimuli in total for participants to hear. These included a core set of 48, from which the results for analysis were taken, plus two additional readings of one of the Māori sentences used as the first and last stimuli in the survey. They were intended as a buffer only: for
practice at the outset and to avoid possible last-task effects, wherein the participant, having reached the end of the exercise, does not give the final task or question their full effort or attention.

The core 48 stimuli were made up as follows: 12 different sentences, six from the Māori story and six from the English letter. Two versions of these sentences were taken from different speakers within each of four speaker groups: male and female present-day elders (born ca.1920-1940), and male and female present-day young L1 Māori (born ca.1970-1990). This is summarized in Figure 3.6 below.

![Figure 3.6. Stimuli used in the second experiment.](image)

Female speakers were included in this version of the experiment to provide a larger pool from which to extract the different sentences, and to ensure that there were enough recordings from which to find excerpts that were clear and free of interruptions. At least three different speakers from each speaker group were used to create the stimulus set, again with the intention of keeping the effects of individual speaker style to a minimum.

The specific sentences that were used in the experiment are shown below. The Māori stimuli are shown here in (4), with translations. Full glossing is available in Appendix B2. The English sentences are shown in (5) over the page. The full text of the English letter, along with the Māori story text and translation, can be found in Appendix B1.
(4)  *I tīmata te tau hou ki te karakia.*  
“The New Year began with a service.”  
*I kitea ko Matariki e rere atu ana i te pae.*  
“Matariki was seen ascending from the horizon.”  
*Ahakoa ko te hōtoke, he tino mao te rangi.*  
“Although it was winter, the sky was very clear.”  
*Ka tūtaki ngā whānau ki ērā atu o ngā whānau.*  
“The families met the other families.”  
*Ko tōku teina te toa o Nūhaka.*  
“My sister is the champion of Nūhaka.”  
*Ka pau haere te rā, kotahi atu te hoki a te iwi ki te marae.*  
“The day draws to a close [and] the people return (straight) to the marae.”

(5)  Well, here I am in the big city.  
Although the weather is nice at the moment, the forecast is for hail.  
I bought a new coat, because they say it gets really cold.  
I have to stay at Auntie Deb’s house for now.  
I’m hoping to get a flat soon.  
The trip up was great, even though it took about ten hours.

These sentences have an average length of 14 syllables. Some of them are rather longer than others: such is the nature of language. Longer utterances allow the listener to become more familiar with the speaker more quickly. Shorter utterances give the participant a break. The relative length of the utterances is not thought to have had any significant effect on the responses. See Chapter Five, Section 5.3 for further discussion on this point as it relates to the utterances in the later experiment.

### 3.3.2. Development of the Survey for the Web

#### 3.3.2.1. Design: Motives and Technical Detail

The web survey had a set of basic requirements: a media player and a selection of web form elements (including radio buttons, checkboxes, and text entry fields). The layout style was very
specific and straightforward; intended to be accessible to as many people on as many different types of connection as possible, while not aesthetically unappealing.

Currently, there are a number of web-based and free tools available online for the purpose of creating a virtually fully customizable web survey from scratch. These tools will write the code for the survey based on the information provided by the user. One example is LimeSurvey (http://www.limesurvey.org/). At the time when the survey for this study was to be produced, however, there were no external tools available that provided precisely the features and the appearance required. The investigator was kindly granted access to the code used in another, similar survey (Teutenberg & Watson, 2010). With permission, elements of this code were used here, with developments and additions to suit the specific needs of the study.

The survey was produced for the web using PHP, HTML, CSS and JavaScript. The media player functionality was provided by the open-source jPlayer (http://www.jplayer.org/), which is written in jQuery. One may alter this media player code in order that only certain features are available to the participant; for example, simply a play button and volume controls, as in the present study. The questionnaire and survey were contained in an HTML form, while the overall appearance was controlled by CSS and jPlayer’s provided skin. The submission and server result file writing functionality were provided by PHP.

Survey creation and preliminary testing occurred on a private web server. When the survey was released to the public, hosting on a secure server was provided by the University of Auckland. Performance was tested in five different web browsers to ensure maximum usability. The original sound files from the MAONZE database were in WAVE format (.wav), but as this format is far too large to allow easy access from the web, compressed formats were used. Browser quirks meant that the files had to be available in two different formats: OGG Vorbis (.ogg) and MP3. The sound levels were not altered at all in either the original files or the web sound files: it was not considered necessary. However, participant feedback suggested that even though none of the stimuli were too quiet to be heard, one or two were, comparatively, over-loud. As perceived loudness is highly subject to environmental factors in both recording and listening, this could either have been due to individual circumstances or to original differences in the stimuli and
possibly both. On the basis of the comments, the variable was addressed in the revised version of the web survey used in the main experiment (see Chapter Four).

3.3.2.2. Survey Structure and Participant Access

As the first incarnation of the exercise to step outside the research group for its participant pool, this study required, and was granted, ethics approval by the University of Auckland’s Human Participants Ethics Committee. This approval was granted in 2010 for a period of three years and also extended to the main experiment (the revised web survey).³

Participants were sourced through networks (friends of friends) and, in some cases, advertising on social media such as Facebook. The provided URL linked to the participant information sheet and consent form, where the purpose and procedure of the project and survey were explained. The participants were assured of their anonymity, unless they should choose to reveal themselves by contacting the investigators for any reason, such as feedback, questions, problems, or requests to view the eventual results. It was also explained to the participants that they did not require any Māori proficiency in order to complete the exercise. They could not proceed with the survey without checking the consent box. When they did, they reached the questionnaire and survey page.

The questionnaire asked for basic information such as age group, gender and ethnic group identification, as well as birthplace and location at the time of completing the test. It also contained questions about language background, language use in daily life, and asked participants to self-rate their proficiency in English, Māori and other languages they might speak.

In the perception test itself, after some minimal instruction on how to use the interface, each of the 50 stimuli was laid out on the screen next to its own play button and volume control. A partial screencapture from the test is provided in Figure 3.7 below.

³ The participant information sheet, including the ethics approval, consent form and questionnaire, are available in Appendix A.
Figure 3.7. The first web survey interface, with controls, syllabification, word-grouping, and checkboxes.

Each participant heard the stimuli in the same partially randomised order. This randomisation was different from that used in the pilot. There, it was entirely artificial: the stimuli appeared random to the participants, but the ordering was intentionally arranged (§ 3.2.3). Here, it is called “partial”, because the initial order was computer-generated by a randomisation script used on the codes for the core 48 stimuli. As there were four recordings of each sentence, the order was checked to make sure that different recordings of the same sentence did not occur in succession, and one or two stimuli were moved as a result. Finally, the two buffer stimuli were placed at the beginning and the end of the list.

First, each sentence was printed just as one would read it in a book, with all punctuation and capitalization included. Below that, the sentence was split into syllables. Punctuation and capitalization were removed, but the syllables remained grouped into words by a darker background. Each syllable had a checkbox underneath it. Participants were instructed to check the box for each syllable that “stood out” to them. They could check as many boxes as they wanted, but were encouraged to check at least one for each utterance.4

The instruction was extremely minimal, and it was, to some degree, open to participant interpretation. Did “stand out” mean “can hear” vs. “can’t hear”; “understand” or “recognise”? Or did it mean “sounds louder, higher, lower, longer or just somehow stronger”? Different people

4 Checking of no boxes for a stimulus recording only occurred four times in all of the submitted results. Three of these were from the same participant, and two of those three were on readings of a sentence that, in all four of its readings, received a very low number of checks overall. Generally speaking, participants followed the instruction to check at least one box per utterance. Most checked more than one.
would inevitably have different interpretations, but just as in the first experiment, the vagueness was necessary in order to avoid leading participants toward one particular acoustic cue over another.

The controls for playing the recordings were very simple: only a play button and volume control were provided. Participants were able to play each recording in full as many times as they wanted before or while checking boxes, but they could not pause the playback partway through. Since prominence is relative, disruption to the flow of the utterances could not be permitted.

3.3.2.3. Result Processing

The result processing in this version of the experiment was an improvement on the pilot in terms of the ability to handle larger participant numbers. When a participant clicked on the button marked “Submit” at the bottom of the web form, the information they had entered in the questionnaire and test was written to a text file on the server. An example section of the raw result text file is given in Figure 3.8 below.

![Figure 3.8. Example of the raw result text file generated by the web form.](image)

In the text file, the information is visible exactly as it came from the web form. Names (L1, EnglishType, and so on) were sent with each of the responses from the questionnaire so that the result file would be readable at a glance. At the bottom of the figure are the codes for syllables that were checked in the perception test. Each checkbox in the perception test had its own unique code that specified the utterance, the speaker group, and the syllable number within the utterance.
The “1” simply means “checked”. This is only the very beginning of the list of checkbox responses for this participant.

The results in the file were then copied into a spreadsheet, where the participant data were counted and analysed as a set. An example of part of the result spreadsheet is in Figure 3.9 below. In the spreadsheet, which shows the checkbox results from the first five participants (numbered across the top) for two of the stimuli, the spacing of checked syllables to allow for the unchecked syllables can be seen. The numbers in the left column are the syllable numbers in the utterance, to which the numbers in the checkbox codes correspond.

![Figure 3.9. Example of a section of the result spreadsheet.](image)

3.3.2.4. Prominence: A First Definition for Analysis

As mentioned in the discussion of the pilot, in that first study, there was no formal experimental definition or set of criteria for “prominences”. The pilot looked only for the appearance or non-appearance of tap clusters on “standout syllables”.

While unquestionably essential overall in the development of the methodology, a formal definition was also required in this new study from a technical point of view, in order for the result processing to be developed appropriately. The spreadsheets that held the results and produced the relevant charts and analysis were created with such a definition in mind. The
intended empirical analysis included assessing levels of agreement on prominence locations across participants and Māori proficiency levels, and how often the perceived prominences aligned with existing predictions made about Māori stress locations (Biggs 1969).

For the purposes of this first web experiment, a syllable was defined as having been identified as “prominent” by a relatively simple criterion. It is referred to as the “60% threshold”, meaning more than 60% participant agreement. In the responses from all the participants in the study, the number of checks for a syllable across four readings had to amount to more than 60% of the total possible checks for that syllable. This 60% threshold, while effectively arbitrary, was not without reason: in the absence, at this stage of the methodological development, of statistical prominence validation measures, it was intended to ensure the agreement of more than half the participant pool on a syllable’s salience before it would be called a prominence and subjected to analysis. In the final experiment, described in Chapter Four, this threshold was abandoned in favour of the statistically-based verification measures mentioned above (see § 4.5, particularly § 4.5.3 - § 4.5.6).

The agreement percentage for a syllable was calculated as follows: \( x/108 \) responses (4 readings x 27 participants) for each syllable in a sentence. The minimum number of votes that would qualify as being over the 60% threshold was, therefore, 65/108. These could be on any of the four instances of the syllable concerned, and may not always have been evenly distributed across those four instances.

With the 60% threshold in mind, the results were presented in percentage-based column charts, which effectively demonstrated the participant agreement. The result charts used percentage values rather than original frequencies to keep the scale of all the plots the same, even where the participants were split into smaller groups (for example, by Māori proficiency levels), whose numbers were not identical. Examples of these charts are in Section 3.3.3 below.

### 3.3.3. Survey Participants

Here, the 27 participants involved in the first web experiment are introduced. There was no specific targeting of Māori participants, but all of them, as New Zealand residents of various durations, would have had some exposure to the Māori language. Ability in the Māori language
ranged from virtually none to high, on a six-point self-rated scale. This scale is the same one that is used in collecting official New Zealand Census data (http://www.stats.govt.nz), and is based on a question that asks about day-to-day conversational ability in the language concerned. Thirteen males and fourteen females participated in the study. Their ages ranged from 20 to 85 years, and the majority (16/27) identified with the New Zealand European ethnic group. Six identified as Māori. Five identified as “other” (e.g. Chinese, Arab, other European). Their language proficiency, as they evaluated it, was as follows. For English, all participants rated their proficiency highly. For Māori, 7/27 rated themselves between fluent and good; 14/27 rated their proficiency from low to “words & phrases”; and 6/27 said they had no ability in the language at all. This information about the participants was all gained from the questionnaire at the beginning of the exercise.

The next section looks at the response of the survey participants to the stimuli. This includes how they responded on prominence identification in general; how they responded to the older and younger speakers used in the study; how their perceived prominences aligned with Biggs’ expected stresses (see Chapter Two) in the Māori stimuli; and outlines the project’s first examination of how differences in participant Māori proficiency might affect prominence perception.

3.3.4. Results from the First Web Experiment

3.3.4.1. Prominence Identification Results from All Participants

Percentage agreement responses in this section are, again, calculated based on x/108 responses (4 readings x 27 participants) for each syllable in a sentence. Numbers of prominences are based on the 60% threshold in the definition above (§ 3.3.2.4). There is not a great deal of difference between the responses to the Māori and English stimuli here. Neither set had any syllables with full agreement at either end of the scale; that is, where either every participant checked all four readings of a given syllable (100%), or no participants checked any readings of a given syllable (0%).
In the Māori stimuli, there were 14 prominences identified altogether, across a possible 98 syllables in the 6 sentences. This is a rate of 14%. All of the six sentences contained at least one of these prominences. Most had two or more. The highest agreement levels were above 80%, found on three syllables in three different sentences. The average agreement level within the prominent zone of 60-100% was 73% ($SD = 7.9\%$).

For the English stimuli, the results were similar. Eleven prominences were identified across a possible 73 syllables: a rate of 15%. Again, all of the six sentences contained at least one prominence. Five sentences contained two prominences each. The highest agreement level for the English sentences was 84%, found on one syllable. The average agreement level within the prominent zone of 60-100% was 78% ($SD = 5\%$).

Participants did, in addition to their agreement on what was most prominent, have some agreement on sections of the utterances that were less prominent. An example from one of the Māori sentences is below in Figure 3.10. Grammatical phrase divisions are shown by vertical bars. It is possible to see that the third phrase, *ki ērā atu* (a prepositional phrase indicating direction; syllabified as | ki ē rā a tu | ), contains relatively low responses compared to the pattern in the other phrases, each of which contains a syllable above the 60% threshold.

*Ka tūtaki | ngā whānau | ki ērā atu | o ngā whānau.*

“The families met the other families.”

---

**Figure 3.10.** Utterance result chart showing more and less prominent syllables and sequences.
3.3.4.2. Prominence Identification in Response to Older and Younger Speakers

Several participants who provided feedback commented that the younger speakers had “less rhythm” than the older speakers (Thompson et al., 2010a). All of them used this same description of the phenomenon, but did so independently. The participants who made these comments were not trained linguists. Neither were they told which speakers were from which age group: any assumptions they made about this were their own, based only on the voices in the recordings. As with the older Māori speakers who have remarked on the change in mita (Chapter Two), these participants were not able to pinpoint exactly what had triggered this thought regarding what they called “rhythm”. They could not isolate a feature: it was more of a general impression, but clearly there was sensitivity to some form of contrast. In response to this, the analysis included a comparison of the prominences identified for the older and younger speaker groups.

When the results from all twelve utterances in the test were taken together (that is, both Māori and English), very slightly more prominences were identified in the speech of the younger group. The contribution to this of the prominences in the two sets of stimuli is roughly equal. This is an interesting observation: an informal listener description along the lines of “less rhythm” might imply less of a perceptible pattern across the utterance. Since the patterns (and prominence) are caused by perceived contrast, “less rhythm” might create an expectation of fewer prominences for those speakers, owing to less contrast. Table 3.4 below shows the prominence results for all of the stimuli together, in response to older (present-day elder) and younger speakers.

<table>
<thead>
<tr>
<th>Speaker Group</th>
<th>Number of Prominences</th>
<th>Average % Agreement in Prominent Zone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Older (PE)</td>
<td>23</td>
<td>77% (SD = 11%)</td>
</tr>
<tr>
<td>Younger (PY)</td>
<td>26</td>
<td>71% (SD = 7%)</td>
</tr>
</tbody>
</table>

When the responses to the older and younger speakers are split into the two languages, there is a small difference. Percentage agreement values for the split results that follow are calculated
differently from those for the overall set. Here, they are based on \(x/54\) responses (2 readings x 27 participants) for each syllable in a sentence. This is because there were only two readings from each speaker group: two older, two younger, and therefore there are half the number of participant responses. Numbers of prominences are still based on the same definition: syllables that had more than 60% agreement. These results are shown in Table 3.5 below.

Table 3.5.

<table>
<thead>
<tr>
<th>Speaker Group</th>
<th>Number of Prominences</th>
<th>Average % Agreement in Prominent Zone</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Máori</td>
<td>English</td>
</tr>
<tr>
<td>Older (PE)</td>
<td>14</td>
<td>9</td>
</tr>
<tr>
<td>Younger (PY)</td>
<td>12</td>
<td>14</td>
</tr>
</tbody>
</table>

For the Māori stimuli, there were a few more prominences identified for the older speakers. In the English stimuli, it was the younger speakers for whom the greater number of prominences were identified, and the difference was greater between the two groups than for the Māori. The average agreement within the prominent zone (60-100%) was slightly higher for the older speakers in both language sets, by approximately the same margin. However, while there appears to be somewhat greater agreement on perceived prominences in older speech (regardless of language), statistical tests (Fisher’s exact) showed that none of the observed differences, either in response to language or speaker age, were significant. The results suggest, then, that based on this dataset, it is not possible to establish a particular pattern in participant responses to older or younger speakers.

3.3.4.3. Relationship Between Identified Prominences and Biggs’ Expected Stress Locations

The analysis in this section is relevant only to the Māori stimuli, and based on the expected word and phrase stress locations described in the grammar by Bruce Biggs (1969). A summary of these rules, including grammatical phrase definitions, may be found in Chapter Two. The purpose of the comparison here was simply to examine the degree of alignment between the expected stresses and the identified prominences.
There are, in the Māori stimulus set, 24 grammatical phrases, and therefore 24 expected phrase stresses (PS). Of these, 17 are also expected word stresses, where phrase and word stress coincide (PWS). A total of 14 prominences were identified by all of the participants as a group. Their distribution across the stress categories is in Figure 3.11 below. There were 24 expected phrase stresses, and eight of these were identified as prominent. The PWS type (both phrase and word stress together) is clearly the most prominent category for these listeners. In addition, five syllables that are word stress (WS) only were selected, but no syllable that was phrase stress only was considered prominent, despite Biggs’ comments that the phrase stress syllable should be the most prominent part of the phrase.

Of the 16 non-prominent phrase stress syllables, some of which were PWS and some PS-only, five had an expected word stress syllable in the same phrase that was considered prominent. This suggests that the qualities of word stress syllables may have a stronger effect on perceived prominence in this task than those of phrase stress syllables, but clearly, where the two coincide, the effect is strongest. It is possible that the speakers were somehow emphasising the word stress syllables, particularly as the stimuli were from read speech, and the participants picked up on that. It is also possible that the larger number of heavy syllables in the WS category attracted attention, if duration and syllable weight were a strong cues for these participants. Further indications of this pattern are readily observable in the results from the main experiment in the project (see Chapters Four to Six and the review in Chapter Seven).

The same investigation of expected stress alignment was made for the older and younger speaker groups. It showed that the stimuli from the older speakers produce exactly the same result as just
described for the whole set above, while those from the younger speakers have just one fewer PWS and one fewer WS identified as prominent: not a notable difference.

It is worth noting here, and this will be elaborated upon in the evaluation that follows, that of the 16 phrase stresses not identified as prominences under the current definition, eight were what could be called “peaks”, which is to say, they had a higher agreement value than the syllables either side of them. The size of these peaks varies, from only a few percent to about 20 percent higher than their adjacent syllables, but their percentage value was simply lower than the 60% threshold. Were they to have been included, five PWS and three PS syllables would have been added to the totals above, which has implications for the confirmation of expectations regarding phrase stress syllables. The question lies in deciding whether or not a peak should have a minimum relative height in order to be counted. This observation led to a modification of the criteria for prominence definition in the subsequent phase of the study.

3.3.4.4. Effect of Māori Proficiency Differences in Participant Responses to Māori Stimuli

This experiment contained the project’s first look (Thompson et al., 2010) at whether proficiency in Māori made a difference to prominence identification in the Māori stimuli, based on observations in other studies that native or higher proficiency speakers respond differently from others to stimuli from their own language, and particularly Bauer’s (1993) related comments regarding Māori stress (see Chapter Two). This comparison was not made for the English stimuli: since all of the participants rated their English proficiency within the “high” range, it was not possible to split them into subgroups. Neither was their relative Māori proficiency likely to make a difference to their response to the English stimuli. For this analysis, then, the 27 participants were divided into three groups based on their self-rating of Māori language proficiency from the questionnaire. The groups and the number of participants included were as shown in Table 3.6. The names of the groups used here are based on self-rating of proficiency level and are relative only.
Table 3.6.

Māori Language Proficiency Groups Based on Self-Rating (SR)

<table>
<thead>
<tr>
<th>Proficiency Group</th>
<th>“High”</th>
<th>“Limited”</th>
<th>“Zero”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Approximate SR Level</td>
<td>fluent to good</td>
<td>low to words &amp; phrases</td>
<td>none</td>
</tr>
<tr>
<td>Number of Participants</td>
<td>7</td>
<td>14</td>
<td>6</td>
</tr>
</tbody>
</table>

The responses from this part of the analysis continued to show that in general, people will identify prominences with a degree of consistency. This supports the idea that proficiency in or understanding of a language does not affect the basic ability of humans to perceive contrasts and identify prominences. Essentially, almost everyone could do it for any language, leaving aside, for the moment, the question of how they are identifying the prominences, which is where proficiency, exposure, and native language affect whether or not people identify the same prominence as others on listening to the same stimuli. What was found in this particular test was a suggestion that proficiency level might affect the number of prominences identified.

Again, while numbers of prominences were still based on the same definition (syllables that had more than 60% agreement), percentage agreement values for these proficiency groups were calculated differently from the overall set, because there were different numbers of participants. The calculations were as follows:

**High proficiency group:** \( x/28 \) responses (4 readings x 7 participants)

**Limited proficiency group:** \( x/56 \) responses (4 readings x 14 participants)

**Zero proficiency group:** \( x/24 \) responses (4 readings x 6 participants)

Comparative percentage result charts, two of which are shown in Figure 3.12 below, showed a difference between the responses of the proficiency groups across the Māori stimulus set.
In some of the stimuli, this difference was more evident than in others, and it was not always one particular group who were consistently different in their behaviour. Generally speaking, there is a trend toward more prominence identification (that is, checking of boxes) in the higher proficiency group (see further discussion below). In some cases, those participants checked a syllable rather obviously more than the lower proficiency groups. For example, the syllables ki- and -a in the last phrase on the top chart, and several syllables on the bottom chart, including a-[bakoa], ko, bō-[toke], and mao. In other, fewer, cases, the lower proficiency groups checked the box for a syllable more than the high proficiency group, or checked it when the higher proficiency speakers did not respond at all: see i in the first phrase; tau in the second phrase; and ki and te in the third phrase of the top chart. The zero proficiency group also picked up on something on the syllable [a]-ba-[koa] in the first phrase on the bottom chart.

Some of the differences are, naturally, larger than others, but even though there are spikes in the various proficiency groups’ results, the actual values on these spikes do not always fall within the prominence zone of 60-100% agreement. This brought up an interesting point related to the threshold that was being used for qualifying prominences. A breakdown of the different proficiency groups’ prominence results, based on the 60% threshold, is shown in Table 3.7 below.
Table 3.7.

Prominence Identification by Proficiency Group at 60% and 50% Thresholds

<table>
<thead>
<tr>
<th>Proficiency Group</th>
<th>Number of Participants</th>
<th>Prominences Identified in x/6 Sentences</th>
<th>Prominences Identified in x/6 Sentences</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>7</td>
<td>4</td>
<td>2/6</td>
</tr>
<tr>
<td>Limited</td>
<td>14</td>
<td>2</td>
<td>1/6</td>
</tr>
<tr>
<td>Zero</td>
<td>6</td>
<td>4</td>
<td>1/6</td>
</tr>
</tbody>
</table>

Owing to the size of the groups, the numbers were very small: both the prominence counts and the number of sentences in which prominences validated by the threshold were identified. No proficiency group stood out. However, closer observation of the charts for all of the sentences revealed that lowering of the threshold to 50%\(^5\) caused more valid prominences to appear, along with a difference between the proficiency groups. At the right of Table 3.7 are the figures for the 50% threshold.

Under the lower threshold, the high proficiency group had a much larger number of prominences, across nearly all of the sentences. The results for the other two groups changed little, suggesting that the behaviour of the high proficiency group was indeed different. As a group, not only did they appear to identify more syllables as standing out, they also appeared to identify them in more environments. This finding, along with the fact that the threshold had to be altered to make it visible, was an indication that the prominence definition criteria would need revision in later versions of the study (\$3.3.5.2. below, and Chapter Four).

The possible effect of proficiency level on the number of prominences identified on syllables that were expected word or phrase stresses in the Māori stimuli was also investigated. Again, the numbers were very small, but the results from this analysis followed a similar pattern to both the overall results for prominence and phrase/word stress alignment as presented in the previous section, and the proficiency-based results just seen above. The higher proficiency participants do identify more expected-stress syllables as prominent than do the other groups, but in order to see

---

\(^5\) Because of the small numbers of participants in the proficiency groups, lowering the threshold by 10% only means a reduction of between 3 and 6 actual votes.
this clearly, once again, the threshold had to be lowered to 50%. The comparison is illustrated in Table 3.8 below.

Table 3.8.

Prominence Identification by Proficiency Group on Expected Stress Syllables at 60% and 50% Thresholds

<table>
<thead>
<tr>
<th>Proficiency Group</th>
<th>No. of Ptcpts.</th>
<th>Prominences &gt; 60% Threshold</th>
<th>Prominences &gt; 50% Threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>PWS</td>
<td>PS</td>
</tr>
<tr>
<td>High</td>
<td>7</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Limited</td>
<td>14</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Zero</td>
<td>6</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

As in the overall alignment results between expected stress and prominences, the syllables where phrase and word stress are expected to coincide (PWS) were most likely to be identified as prominent by all groups, followed by expected word stress (WS) syllables. Again, results for phrase stress only (PS) syllables were very low: here, zero. In the results with the lower threshold, there was a syllable with no expected stress (OS) that appeared as prominent: the same syllable was identified by both the high and zero groups. It was, in fact, the final syllable -ngi, in the final phrase of the _ahakoa_ stimulus shown in the bottom chart in Figure 3.12 above.

Thus, despite low numbers, there was a suggestion that higher proficiency participants were more likely to check a PWS or WS-only syllable as prominent. Something about the quality of word stress syllables, either alone or with phrase stress, appears to attract attention, even under the proficiency split. It is important to note that despite comments about change in the language, a good rate of alignment between rule-based stresses and participant-identified prominences is not unexpected. There are obscuring factors, such as awareness of the expected stress rules in some higher proficiency participants, which may have led them to check certain boxes because they felt it was “correct” to do so, or because they truly heard prominence there, or a combination of the two. With regard to this, it is worth noting, again, the complete lack of PS-only syllable prominences above. If higher proficiency listeners were simply following rules when checking boxes, we would expect at least some of these PS-only syllables to appear in their results, if no-one else’s. However, none appear, and therefore these listeners are either not aware of, or not using,
the (complete) phrase stress placement rule, or they are expecting a cue that is not strong or not present at all. Alternatively, it is possible that they are not aware that PS may occur separately from WS, given that PS-only is the minority case, or their sense of stress and prominence has now become focused on WS-related syllables (i.e. WS and PWS). Correcting for a possible rule-following variable in this kind of study would be virtually impossible without manipulating the signal, but given the evidence here, it may not be a significant variable at all.

The potential effects of participant age, gender or ethnic group on prominence identification were also explored in the process of this study, but none were apparent (Thompson et al., 2010b, 2010c). If such effects do exist, they were likely obscured here by both the small number of participants and the variety of proficiency levels involved: for example, most of the participants with higher Māori proficiency were in the older age group.

3.3.5. Evaluation of the Methodology and Proposed Modifications

The change to the web format successfully resolved most of the issues identified in the pilot. Taking away the physical aspect of tapping, removing the presence of the investigator, and providing self-controlled playback and interaction was intended to reduce the potential for participant frustration or fatigue. Such proved to be the case: some participants chose to provide feedback that they had rather enjoyed the web-based task.

Further feedback suggested that the whole exercise took most participants about half an hour. This information indicated that the web-based delivery method was much more efficient and predictable than the method used in the pilot. It also suggested that the number of stimuli was acceptable to most. The actual time required for each participant would have been dependent on the number of playbacks and decision time taken. However, there was no negative feedback about the overall length of the process.

The efficiency of the study, in time, portability, and result processing, was vastly improved. Participants were gained from all around New Zealand, as opposed to locally to the research location. The larger number of stimuli included meant that the results obtained were more reliable and gave more information. Participants did identify prominences consistently, and the
method was capable of eliciting that information effectively. Moving forward, however, there were still remaining adjustments to make.

3.3.5.1. Changes to the Web Survey

The web survey code and result processing were made still more efficient. Firstly, the manner in which the stimuli themselves were handled by the code was streamlined and simplified. This meant that the utterances, and their syllables, could be added to or changed with less effort on the part of the survey programmer. The result processing was also improved by altering the code that took the data from the web form. The checkbox results were made binary: simply a 1 for “checked” or a 0 for “unchecked”. All of this was written to the result file, removing the need for manual spacing of the checked syllables in the spreadsheet. By allowing a direct copy-paste action, the potential for human error in the manual spacing was eliminated, and both totals and updated result charts were available instantly.

3.3.5.2. Changes to the Definition of Perceived Prominence

As a theoretical parallel to the changes to result processing, revision to the criteria for the definition of perceived prominence was also required. In the discussion of the results above, it was mentioned that the 60% agreement threshold occasionally obscured a genuine difference between proficiency groups. However, the more important observation is that the results in the 50-60% agreement zone that were excluded were still classifiable as “peaks”, meaning that those syllables were considered more prominent than the syllables either side of them, even if the overall agreement value was below the 60% threshold. The size of these peaks varied, but the circumstances suggested that since some of them were quite large, and prominence is relative rather than based on an absolute value, they should not be ignored. As discussed in § 3.3.2.4 earlier, the threshold was arbitrary, if logical and even effective, to a degree; however, in the next experiment, new criteria were formulated for the definition of prominences based on these “peaks”. Because of the varying sizes, statistical analysis was included in order to determine which peaks were significant and which were not. A method was devised which had its own complications and process of refinement, and these will be discussed in detail in Chapter Four. This approach was intended to provide a more robust list of prominences on which to do further analysis.
3.3.5.3. Changes to the Stimuli and Adjustment to the Scale of the Study

Based on the findings from the first web experiment, the nature of the stimulus set was entirely re-evaluated. The major alterations were a change from read to continuous speech; the inclusion of another, older, speaker group, and removal of the English stimuli.

The results from this first experiment did not strongly bear out the comments about differences in “rhythm” made by participants, or the comments about change in the language perceived by older Māori, if that is based on differences in response to older and younger speakers, or alignment of perceived prominences with expected stress locations. The pilot and first web survey stimulus sets used read speech, with the goal of taking multiple versions of each utterance. However, the story and letter from which they were taken were only available in the modern recordings, and could not include the third, oldest vintage of speakers available in the MAONZE database: the historical elders, recorded in the 1940s (see Chapter Two). Since continuous, conversational speech recordings were available from all three speaker vintages, and continuous speech was likely to give a better indication of natural prominence locations than the read speech, the set of stimuli was altered accordingly. As there were, at the time, no recordings of female speakers from the historical era in the database, it was decided to return to recordings from male speakers only. This would also simplify any acoustic analysis by avoiding the need to normalise for intrinsic pitch differences between male and female speakers.

A further reason for the change to continuous speech was that there were occasional “performance effects” in evidence in both the Māori and English recordings, particularly from some of the older speakers; for example, theatrically-styled additional volume or expression. There were also sentences in the English letter that were ambiguous: they could be, and sometimes were, stressed differently according to speaker interpretation. In a reverse type of performance effect, read speech situations can cause self-consciousness on the part of the reader, leading, perhaps, to errors, unnatural speech patterns, pauses, hesitations, or even possibly excessive focus on correctness and rules, such as Biggs’ stress rules in Māori. Having sets of readings of the same sentences in the pilot and first web survey was intended to counter or mitigate speaker effects such as these, but the effectiveness of this strategy is in question, especially where it was evident that participants did
pick up on the different patterns. Finally, since Biggs’ rules were based on conversational speech (Biggs, 1961, pp. 5-6), then the part of the analysis that dealt with the alignment of perceived prominences with expected stress locations was best served by a list of prominences that had been identified in that type of speech.

The second major change to the set of stimuli was that English utterances would no longer be included. Both English and Māori continuous speech recordings were available for all three speaker vintages, but once it had been determined by the first web survey that participants would, in fact, identify prominences with an acceptable degree of consistency in both languages and in this test format, the comparison between English and Māori had outlived its usefulness. It was more potentially informative and relevant to revise the primary focus of the study specifically towards the location and nature of prominence in Māori. By leaving out English stimuli, more Māori stimuli could be added. This provided a larger dataset with more variety, eliminating the need to find multiple clean recordings of each stimulus sentence.

With the intention of major changes to the nature of the stimuli, the scale of the study was also re-evaluated after the first web experiment. Perception tests in general have an enormous size range, from very few tasks, finished in minutes, to literally hundreds of tasks that must be done over a course of days or weeks. For this project, it was essential to keep the size appropriate to the delivery method. Replacing tapping with checkboxes had meant that more stimuli could be included without participants suffering fatigue. The fact that there was a mixture of English and Māori stimuli in the first web test, and that there were repeats, also meant that a set of 50 was not an unreasonable demand. Participants handled that volume without trouble, which did suggest that, moving forward, the number could have been increased further.

The test in this study should ideally be completed in one sitting, however, and it is much easier to gain participation in a study that can be advertised as requiring a relatively minimal time commitment. When the final phase of the experiment was being planned, this was taken into account, and the number of stimuli was ultimately smaller (32 utterances instead of 50). This was for several reasons.
Firstly, since the English stimuli were to be removed, most participants would be dealing exclusively with a language with which they were less familiar or not familiar at all. Secondly, all of the stimuli would be different, which would mean participants did not build familiarity with the utterance patterns. Thirdly, when the new stimuli were being chosen, it was found that their average length (in syllables) was longer, at 21 instead of the 14 in the first web experiment. Though none of these factors were likely to cause true difficulty, they would raise the level of concentration required by the participants, and therefore the potential for fatigue. Ultimately, the 32-stimulus format proved to be appropriate for keeping the average amount of time required for the test to about half an hour. Results and feedback from the second web survey suggested that in these circumstances, particularly where participants who are language-unfamiliar were included in the test, the 30-utterance size was adequate (see the evaluation of the second web test in Chapter Seven for more detail).

3.3.5.4. Projections for Further Analysis of the Results & a New Proficiency Study

Both empirical and acoustic analyses were originally intended for the results from the first web experiment. As the results above have shown, however, the responses available from the small numbers of participants make drawing conclusions difficult. The same problem extends to acoustic analysis, which is why it was ultimately not pursued here, in favour of developing the test and stimulus set further. There was, however, a suggestion in results from both preliminary studies that some acoustic features might be common to the syllables identified as prominent, such as duration (heavy syllables) or pitch (as predicted by Biggs). In order to control the scope of the investigation, and taking into account the existing description that was to be examined for possible change, as the study moved into its third phase, it was decided that the acoustic investigation was to be confined to pitch, duration and vowel quality. The question of proficiency effects on perceived prominence were also to be explored in more detail, under the same analysis method.

In this first web experiment, the numbers in the different participant Māori proficiency groups were quite uneven, with the high and zero groups having only half what the limited group had. With evidence of different behaviour in the groups having appeared, however, it was desirable to even them out for a more accurate comparison. In a New Zealand-based study, finding
participants in the limited category was much easier than finding high or zero proficiency participants. Māori speakers, particularly fluent ones, are not as easy to acquire as might be expected. However, one of the advantages of the web survey method is its accessibility, which aided in fixing this difficulty. It allowed specific targeting of Māori-speaking participants from around the country, as well as participants from overseas who had no exposure to Māori at all. The effects of proficiency group behaviour on elements of the methodology can be found in Chapter Four, and the results of the expanded proficiency study that was part of the main experiment are presented in Chapter Six.

In the final phase, the study was revised with the changes as outlined above. The prominence elicitation methodology (that is, the web survey method) remained largely the same, however, as it had fixed the problems encountered in the pilot study and proven itself effective. The final phase is described in detail in Chapter Four.
4

THE PROMINENCE ELICITATION TOOL:

MAIN STUDY METHODOLOGY AND DATASET

4.1. Introduction

This chapter introduces the third phase of the experimental development in the project, which came in the form of the second web survey. The principal goals of this main experiment were as follows: to use the web survey methodology, tested and proven effective (see Chapter Three, Section 3.3), to elicit participants’ perceived prominences based on Māori conversational speech stimuli, and to determine through analysis where those prominences were heard; their nature, number and frequency; and whether there was any pattern to them. Finally, it would explore what effect, if any, the participants’ Māori language proficiency levels had on their perception of prominence.

The experiment had a very similar shape to the first web survey, in terms of the task and interface presented to the participants. However, this version used an entirely new set of stimuli, and implemented some changes to the web code and result handling. It also increased the robustness and detail of the criteria for defining perceived prominences in the study. A database was created and a set of analysis approaches was formulated and then carried out on the results from the participants. These approaches were not only empirical but acoustic. They documented the distribution of the prominences, including alignment with existing descriptions of Māori stress, as in the previous versions of the study, and other characteristics such as duration, vowel quality, and F0. For the motivations behind the changes made between the two web-based studies, see the evaluation of the first web survey in the previous chapter (§ 3.3.5).
The second web experiment was a large study, and is presented across three chapters. This chapter covers the setup and development of the study, including the changes mentioned above. It introduces the stimuli and the new participant sets, known as Group Z (all 92 participants) and Group A, B, and C (subsets of Group Z according to Māori language proficiency). Changes to the data handling that were made to accommodate the different groups’ behaviour are described, and an overview of the dataset on which the study is based is provided. A discussion of the prominence identification results based on speaker vintage is also included, prior to the main results and analysis from Group Z in Chapter Five, and from Group A, B, and C in Chapter Six.

4.2. Stimuli

In this version of the experiment, a completely new set of stimuli was used. Where previously there had been read speech stimuli from both Māori and English, here there were Māori stimuli only, from continuous speech. They were, as before, taken from the MAONZE database, but this time from interview recordings.

These new stimuli included three different speaker vintages from the MAONZE database: historical elders (HE), present-day elders (PE) and present-day young L1 Māori speakers (PY). Broadly, these speakers’ birthdates span more than 100 years. All of the speakers used in this study were male, because restricting speaker gender in the stimulus set helps control variables in F0 and vowel quality analysis: there is no need to normalise for intrinsic gender-based pitch differences. Additionally, at the time when the stimuli were being selected, recordings of female speakers of a comparable age to those of the male historical elders were not available in the database. (See Chapter Two for more detail on the MAONZE speaker groups and their birth and recording dates.) The use of the three different vintages was also intended to allow a diachronic comparison based on the PPMs. However, given the results of the preliminary investigation (see Section 4.8 for an explanation), this line of analysis was discontinued.

Participants heard 32 stimulus sentences in total. This included a core of 30, made up of 10 from each of the three speaker vintages (HE, PE, PY). In addition, there were two buffer sentences: one each for the beginning and end of the exercise. These were extra sentences from the PY and PE groups respectively. Just as in the first web survey, the buffer sentences were used to provide
practice for the participants, and to avoid any last-task effects, but the participants were given no indication that this was their purpose.\textsuperscript{1} The results from these two buffer sentences were not included in the analysis. A summary of the stimuli is in Figure 4.1.

\[
\begin{array}{ccc}
3 & \times & 10 \\
\text{speaker vintages} & \quad & \text{sentences each} \\
(\text{HE, PE, PY}) & + & 2 \\
& & \text{buffer} \\
\end{array} \\
= 32 \\
\text{different Māori sentences}
\]

Figure 4.1. Summary of sentence stimuli used in the second web survey.

All of the sentences used in the study were different, though some did contain similar phrases. This was intended to prompt responses to a more varied and useful set of Māori syllables of different types, in different grammatical environments, than had been available in the previous experiment.

The stimuli were selected based on several criteria, including content, clarity, and length. Topics in the source interviews had a wide range, so sentences were checked for suitability and neutrality of subject matter. Given that they were out of context, this was generally not a problem. Obvious proper nouns were avoided, with the exception of the word Māori. This was allowed on the grounds that it provided tokens of the long diphthong /aao/, long diphthongs being rare. It might also indicate whether prominence was extra likely to be perceived in such a familiar word, or by participants whose only exposure to Māori thus far was the name of the language itself.

Some of the sentences contained discourse particles such as nā “and so” or similar. There were both simple and complex sentences and one question, averaging 21 syllables in length. This was an increase on the average 14 syllables used in the first web survey. The stimuli were kept to total durations of between two and six seconds, with an average of 3.5 seconds. There is no evidence that the variation in length of the stimuli affected the participant’s responses. Longer utterances help the listener familiarise themselves with the speaker, while shorter ones help them keep their focus by providing a break, of sorts. Including both provides variety, which is realistic and

\textsuperscript{1} For example, the first sentence in the test, intended for practice, was not labelled “practice”: it was simply labelled “1”.

85
characteristic of language. Expected stress and intonation patterns would still be expected in both shorter and longer utterances. See Chapter Five, Section 5.3 for further comment on this.

The speaker sources for the recordings were kept as varied as possible, including several speakers from each of the speaker vintage groups. There were sentences from five different historical elders, four different present-day elders, and four different present-day young L1 speakers. Some of the full source recordings, particularly those from the older speakers, do contain noise and interruptions such as coughing. Because of this, the stimuli that were used in the test were generally selected based on their smoothness, though there are a few natural pauses. These included hesitations, both narrative and error, and two restarts. They were allowed to remain in order to see whether they would have any effect on perceived prominence. For discussion of this, see Sections 4.8.4 and 4.9.3.

Some examples of the sentences the participants heard are below. (1) is from a historical elder, (2) from a present-day elder and (3) from a present-day young L1 speaker. A list of all 32 sentences used in the study, with full glosses and translations, may be found in Appendix C.

(1)  *Me tīmata atu taku kōrero i ngā rā ki muri rā.*
    “My story begins in days long gone by.”

(2)  *Kua tīmata au te tuhituhi i te pukapuka tuarima.*
    “I have begun writing the fifth book.”

(3)  *He tuakana teina ōhau kei konei?*
    “Do you have siblings here?”

It was considered that these 32 stimuli would provide sufficient syllables for the analysis, without putting undue strain on participant stamina. The nature of the experiment made it difficult to control for the former, but most participants completed the test without any trouble, suggesting that the latter was true.
4.3. Participants

The group that provided the results discussed Chapter Five, known as Group Z, includes all of the participants who completed the test: 92 in total. This was over three times the number who participated in the first web survey. Overall, the process of participant acquisition for this second survey spanned more than a year, and required repeated recruitment drives, but two of the main reasons for the successful increase in participant numbers were better advertising and specific targeting. For this experiment, participants were acquired via personal networks, as in the previous experiment, and also by several other means. Advertisements containing the link to the survey were placed on various forms of social media, such as Facebook; put up with posters at conferences and handed out on cards (see the example in Figure 4.2 below); and distributed via relevant listservs.

As before, participants were provided with a URL at which they could access the test, but this time, as the figure shows, the URL was customized to make it more memorable. There was also, in this version of the experiment, deliberate sourcing of Māori-speaking participants in order to increase their numbers, and deliberate targeting of overseas participants who genuinely had no ability in, or prior exposure to, the Māori language.

There were 29 males and 61 females who took part, with 2 unspecified.² The age range of the participants was from 16 years to over 65. Almost half of the participants fell within the 25-45

² A flaw existed in the survey code that allowed participants to submit unfinished results, meaning that in a very few cases, there was some information missing from the questionnaire. Where possible, this information was inferred from examination of the participant’s answers to other questions. For example, a participant who did not specify their English proficiency, but who was born in New Zealand and had always been resident here, was likely to have English proficiency of a high level. The one participant who did not specify their age range did state their length of residence in New Zealand, which allowed a lower boundary to be put on their age group. Where it was not possible to infer the information, as with gender, the results were recorded as “unspecified”. This issue shows the usefulness of
Sixty-five participants listed New Zealand as their current location. Fifty-one listed New Zealand as their birthplace. Forty-eight said they were both born and resident in New Zealand. This made New Zealand residents dominant in the survey. Of the participants resident overseas, 15 were in Australia, 6 in the USA, 3 in Europe, and one each in Asia, South Africa and South America. Some of these had been born in New Zealand, while others had not. Despite the deliberate targeting, overseas participation was partly a matter of chance.

General language exposure and background were quite varied for these participants. Some of them had formal linguistic training; others had language learning experience but no linguistic training; others still had no such experience but may have been exposed to other languages in daily life. Some of the participants were native English speakers but some were not. Some spoke second and third languages to different levels, but there were also participants whose questionnaire responses suggested that they had little or no familiarity with or ability in a language other than English. This information is based on questions (see Appendix A2) that were asked about a participant’s first language; other languages; languages spoken at home and with friends; birthplace; and country of residence at the time of completing the survey.

In New Zealand, participants with minimal exposure to the Māori language are easy to come by. However, there are not as many highly proficient Māori speakers available as might be expected (see Chapter Two). Low or zero proficiency speakers may be, and were, sourced from overseas. For this study, a particular effort was made to gain both participants who had no exposure to the Māori language at all, and those with very high proficiency. As a result of this targeting, the 92 participants in the main experiment could be divided into three relatively even subgroups based on their Māori proficiency level, which was, as before, self-rated. The groups were equivalent to those used in the first web test (“zero”, “exposed”, “high”), and the division criteria are detailed below.
4.3.1 Background to the Proficiency Group Divisions

A pair of questions in the pre-test questionnaire related specifically to the participants’ familiarity with Māori and English. As in the first web survey, they were asked to self-rate their proficiency in the two languages, based on a question in the New Zealand national Census forms that asks about ability in day-to-day communication in the language concerned. The proficiency rating scale used here had six points, ranging from “very well” to “not at all”, which can be seen in Tables 4.1 and 4.2 below. Unsurprisingly, English proficiency (Table 4.1) was rated highly. The lowest rating level (“not at all”) was not provided as an option for English, since in order to participate in the survey, a proficiency greater than zero in English was obviously required.

Table 4.1.

*English Proficiency Self-Ratings from the Second Web Survey Questionnaire*

<table>
<thead>
<tr>
<th>English Proficiency Rating</th>
<th>Total</th>
<th>very well</th>
<th>well</th>
<th>fairly well</th>
<th>not very well</th>
<th>words &amp; phrases</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of Participants</td>
<td>92</td>
<td>84</td>
<td>6</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

The purpose of retaining the English self-assessment was to provide a benchmark both for the participants when answering the same question about their Māori proficiency, and for the investigator when it came to splitting participants into Māori proficiency groups. These ratings, which did include the “not at all” option, contained much more variation, as shown in Table 4.2.

Table 4.2.

*Māori Proficiency Self-Ratings from the Second Web Survey Questionnaire*

<table>
<thead>
<tr>
<th>Māori Proficiency Rating</th>
<th>Total</th>
<th>very well</th>
<th>well</th>
<th>fairly well</th>
<th>not very well</th>
<th>words &amp; phrases</th>
<th>not at all</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of Participants</td>
<td>92</td>
<td>14</td>
<td>7</td>
<td>3</td>
<td>4</td>
<td>30</td>
<td>34</td>
</tr>
</tbody>
</table>

The Māori proficiency ratings were used to group the participants into four different cohorts for use in the analysis: one master group and three proficiency-based subgroups. The ratings divided...
the participants quite naturally into evenly-sized categories, and additional information provided in the questionnaire, such as length of residency in New Zealand, helped to confirm that the groupings were as appropriate as possible. The four groups were named alphabetically, for simplicity, as “Z”, “A”, “B” and “C”. The exact numbers of participants in each group may be found in Table 4.3 below. The reasoning behind the divisions is explained in (1) to (4), which follow.

Table 4.3.

<table>
<thead>
<tr>
<th>Proficiency Group</th>
<th>Group Z “All”</th>
<th>Group A “High”</th>
<th>Group B “Exposed”</th>
<th>Group C “Zero”</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of Participants</td>
<td>92</td>
<td>28</td>
<td>30</td>
<td>34</td>
</tr>
</tbody>
</table>

(1) Group Z: “All”

Group Z included all of the participants, considered as a single set, regardless of Māori proficiency rating. See Chapter Five for their results.

(2) Group A: “High”

Participants were included here if they rated their Māori ability in the top four categories: from “very well” to “not very well”. The latter may sound like a negative rating; however, a participant’s selection of that level as opposed to the next, “words and phrases”, suggests that they considered their ability to include sentence formation or basic conversation, simply not at a confident level. Additionally or alternatively, there may be an element of modesty in self-rating. Group A certainly has the widest range of ability included, but relatively speaking, for the lowest rating in this category, the ability to form actual sentences would still imply higher proficiency than that of the participants in Group B and C, as described below.

(3) Group B: “Exposed”

The Group B participants were those who rated their Māori proficiency at the “words and phrases” level. This would be the expected proficiency level for most
New Zealand residents of any duration who have not actively studied or learned Māori. They are certainly exposed to the language; they can parse it to varying degrees; and they may recognize, understand or use some lexical items or phrases in daily life; for example, when reading the newspaper, watching television, or in conversation with other people. However, they would be very unlikely to construct a sentence or hold a full conversation in Māori, and their knowledge of the grammar would range from limited to non-existent. They may or may not use or hear the Anglicized pronunciations of Māori words and place names, depending on education, awareness and source of exposure. This point is only relevant to the study here because the Anglicized versions often do not exhibit the vowel length distinctions that exist in the Māori pronunciations; they are sometimes missing syllables or are phonetically incorrect; and they are often subject to English stress patterns. Regular exposure to any of this could affect prominence detection responses to L1 Māori stimuli as required by the task in this experiment. Given the numbers and circumstances involved, Group B formed itself naturally, but it should be noted that it does not correspond directly to participants who were New Zealand residents. Some of the overseas participants were New Zealanders who had relocated for only a short period at the time of completing the survey, and are included if they rated their proficiency at this level.

(4) Group C “Zero”

Group C included participants who rated their day-to-day Māori ability as non-existent (“not at all”). This group also formed itself on numbers, but the reasons the participants rated themselves in this manner varied more from person to person than in the other groups. Some of the overseas participants had definitely never heard, or even heard of, the Māori language before. Others knew of the language but had very little or no exposure. It is possible that this group has some small overlap with Group B, because of a variation on the modesty factor mentioned in the description of Group A. Some of the participants claiming to have no ability at all in Māori may have been rating themselves negatively because they did not consider themselves able to have a conversation; they may not have taken into account passive understanding of words and phrases as included in “ability to use” a language. However, on examination of the participants’ overall responses to the questionnaire, there were only a few instances where this might have been the case. The evidence was not strong enough to justify moving them into another group.
In Chapter Five, all the participant responses are considered together. In Chapter Six is found the study of proficiency effects on PPM identification. For now, we return to the description of the survey.

4.4. Further Development of the Web Survey

The web survey construction for this experiment was very similar to that of the first version, as described in detail in the previous chapter. The principal changes were behind the scenes, however: a participant who completed both tests would have noticed little difference in the nature of the task.

4.4.1. Structure and Participant Access

Access to the test was as before: a URL was provided via a number of means, as described above, which brought the participants to the information sheet and consent form. After checking the consent box, they could proceed to the questionnaire and survey. Minimal changes were made to the interface itself. Firstly, the wording of the instructions was altered to emphasize the usefulness of headphones in the task. It is impossible, in a web-based, unsupervised task such as this, to control background noise that might distract the participants. However, this encouragement was included in an attempt to minimize the possible effects. The other wording, including the core instruction of “check the boxes underneath the syllables which stand out most to you”, remained unchanged. There was one visual alteration to the presentation of the stimuli. Word groupings for syllables were no longer indicated by a darker background, as it was considered that the provision of the sentence written out normally above the checkboxes made this unnecessary. A partial screencapture from the second web survey is shown in Figure 4.3 below.

![Figure 4.3. Partial screencapture showing two stimuli from the second web survey.](image-url)
The player and its controls remained the same in appearance and functionality. As before, participants could play the recordings as many times as they wished, but they could not pause or stop the utterances partway through. This was so that the natural flow and any potential perceived contrasts within the utterance would not be interrupted. Participants could proceed at their own pace, and informal feedback suggested that the average time for completion amounted to around half an hour.

When the participants were satisfied that they had completed the task, they clicked on the button marked “Submit” at the bottom of the test screen. This sent the results to the server, and the participants were shown a screen notifying them of successful submission. In case of errors, they were advised simply to go back and resubmit. The survey was tested in five of the most commonly-used web browsers in use at the time: Firefox, Google Chrome, Internet Explorer, Opera, and Safari. In all of these browsers, the contents of the survey form would have been retained in case of error, and it would not have been necessary to re-do the entire test.

Behind the scenes, changes were made to the code in order to improve efficiency. Where previously the stimuli (as both sentences and syllables) had been individually coded into the HTML form that displayed the test, here they were stored in arrays and then written to the screen, with their checkboxes, in each test question. This was a simple change, but made the code much more streamlined and easier to view and edit. It was also intended to allow for easier re-use with other stimuli. The player controls for each utterance were added to the interface using the same process, which linked them to the correct sound file.

A sound issue that had been discovered in the first web survey was also addressed. The recordings used in the experiment were originally in .wav format. This format was, as before, converted to both .mp3 and .ogg formats for use in the web survey, in order to reduce loading times and allow for different web browsers being used by the participants. After feedback that one or two of the recordings in the first survey were relatively loud, an effort was made to ensure that the recordings in the present experiment had more even volume. Some of the oldest recordings included in this study were very quiet. Even though volume controls were provided, it was necessary to prevent the base volume of one recording being much louder than another and perhaps unexpectedly deafening a participant wearing headphones. Before the survey was released to the public, overall
volume was regulated on the recordings using the volume correction function in Adobe
Soundbooth. No other adjustments to the sound were made.

In order to produce a set of usable tokens for further analysis, the raw submitted results needed to be processed in several ways, from checking the submissions for completeness to finding the perceived prominences. These methods are outlined in the following sections.

4.4.2. Managing Submissions

The results continued to be written to a file on the server, but the format was changed in a series of small ways to make it easier to read and search the file, and to process the data that came out of the survey. Each participant submission was allocated a timestamp and a unique, randomly generated participant ID, where none had existed before. Code was also inserted to notify the investigator via email when a new response was submitted, to avoid the need for constant checking of the file. This functionality was useful because the survey remained open for a long period, and participants did not submit results at regular or scheduled times.

The questionnaire and test results in each participant submission were written into the result file in a horizontal string instead of the vertical structure that had been used in the first survey. Line breaks were inserted between each of these strings, in order to make the file more readable at a glance. The structure of the result string was also changed. Previously, there had been labels on each of the answers to indicate the question or checkbox involved. For example, “gender = female” (from the questionnaire). These labels had to be removed when the data was prepared for analysis, and because only the checkbox results with a “checked” value appeared, the responses had to be spaced manually to allow for the non-checked syllables. In the revised version, the labels were not included in the result string; instead, generic placeholders (usually a simple “X”) were assigned by the code to any blank form fields (as, for example, in the case of participants who did not need to specify a length of residency in New Zealand). Most importantly for the data processing, zeros were assigned for the unchecked syllables in the test, so that when each result string was transferred to the result spreadsheet, no reformatting was required: all of the responses lined up accurately with the headings in the table. Because of the numerical format of the checkbox results, addition of vote totals was instant, and any charts produced from the totals were updated automatically.
4.4.3. Cleaning of the Raw Data

Ninety-two complete participant responses were used in the analysis in the study, out of a total of 104 submitted tests. The responses were collected over a period of more than a year. The process of gaining participants for an exercise such as this is unpredictable and time-consuming when they are random rather than a captive group. The responses tended to come in waves, which allowed for refinement of the data-cleaning process and some preliminary examination of responses, but did not permit full analysis until all of the results were in. A further side-effect of random participants (or semi-random, as some targeting did prove necessary) is that control over their manner of participation is limited, and this can include whether or not they finish an exercise. In this case, the perception test data for twelve of the responses were incomplete, and had to be omitted in order not to create discrepancies in total possible votes for the syllables in the dataset. Examination of the omitted responses showed that there were a number of reasons for the incomplete results. In some cases, participants had stopped at a certain point after completing several utterances successfully. In others, the participants had completed all or part of the questionnaire and no information at all for the perception test, but submitted the result anyway. This may have been due to interruption or lack of time. There were also instances of reported connection failure, where the entire submission was empty. This remains unexplained: there was no discernible pattern due to type of connection (Wi-Fi, wired, dial-up, etc.), time of day, location, or any evidence of submission clash, but the reports were consistent with the timestamps on the empty submissions.

Once the unusable responses had been removed and the basic vote counts for each syllable had been established, the first indication of which syllables stood out to the participants was provided. It was then necessary to examine the results further to discover which of the syllables could be called the perceived prominences and used in the empirical and acoustic analysis.

4.5. Preparing the PROM Dataset for Analysis

In the 30 core stimuli included in this experiment, there were a total of 636 syllables. These 636 syllables make up what will be known as the PROM Dataset in the analysis. Within the PROM Dataset, there are, based on participant group responses, different sets of both perceived prominences (to be known as PPMs), and syllables not perceived as prominent (non-PPMs). For
example, the overall group (Group Z) has a list of PPMs based on votes from all the participants in the study. Similar lists of PPMs were made for each of the proficiency groups A, B and C, based on the votes from only the participants in those groups (see Chapter Five). In this chapter, however, we will be concerned with the nature of the dataset as a whole, along with the creation of the PPM list for Group Z, and the results based on it.

Several steps were required in order to process the cleaned results gained from the perception test, find the PPMs for any given participant grouping, and prepare them for analysis, both empirical and acoustic. The criteria for perceived prominences were closely examined and refined, both to strengthen the definition of the concept used in this version of the experiment and to assess the possibility of including some potentially relevant syllables that the previous experiment’s definition would not allow. The lists of PPMs were then created based on these criteria. In addition, the stimuli were entered into a database in the Emu speech analysis software (http://emu.sourceforge.net). This will be referred to as the PROM Database. It was prepared with relevant information for the analysis, including segmentation of the sound files into units such as phrases, syllables and phonemes; hierarchical linking of these units; PPM marking; and F0 and formant tracking. Following this preparation, the database could be queried efficiently and was ready for analysis. More detail about the process will be provided throughout this section.

### 4.5.1. Refining and Defining the Criteria for Perceived Prominences (PPMs)

As discussed in Chapter Three, in order to provide a solid base for analysis, a clear definition of perceived prominences was required. In the first web experiment, this definition was based on what was referred to as the **60% threshold**:

A syllable is *prominent* if there is more than 60% participant agreement that it stands out.

This meant that across the responses from all participants the study, the number of checks for a given syllable was required to amount to more than 60% of the total possible checks for that syllable. For example, in that first study, there were 27 participants and four readings of each

---

3 See also Chapter Three, Section 3.2.4 and this chapter, Section 4.6.1, for more on Emu databases as used in these studies.
sylable, bringing the possible checks for any syllable to x/108, of which the minimum requirement for the 60% threshold was 65/108.

Under this criterion, any syllable with over 60% participant agreement was included in the pool of perceived prominences for analysis, regardless of the vote counts of its adjacent syllables. This did, in theory, allow for two adjacent prominences of different levels above 60%. In some respects, this clashes with the inherent relativity in the concept of prominence, but in reality, the first web experiment’s participant set was small enough that their vote totals meant two adjacent syllables with more than 60% agreement did not occur in the overall result set, and only very rarely in any of the analysis subsets such as proficiency group. This can be seen below in a result chart from one of the sentences used in the first web experiment (Figure 4.4). Phrase divisions are indicated by vertical bars.

*Ka pau haere | te rā, | kotahi atu | te hoki | a te iwi | ki te marae.*

“The day draws to a close [and] the people return (straight) to the marae.”

![Figure 4.4. Result chart from the first web survey: non-adjacent distribution of >60% syllables; excluded <60% peaks.](image)

The two prominences (syllables >60%), *pau* and *rā*, are not adjacent to other syllables with greater than 60% agreement. However, what is also visible in the same chart is that there are a number of other clear peaks of varying sizes (the syllables *ta, ho, i*, and *ma*), but these were left out of the prominent list because the vote counts for those peak syllables came in below the 60% threshold.

Why should these syllables not be included, when there was clearly some perceived contrast for the listeners? In the analysis to the first survey, lowering the threshold, even by a few votes, had
shown up differences in the participant proficiency subgroups: more evidence that some of these potential but excluded prominences also merited attention. This showed the 60% threshold to be inadequate, and to accommodate these observations, a modified approach to the definition of perceived prominence was formulated. The new definition, explained below, uses the linked concepts of **triplet** and **peak**.

### 4.5.2. Triplets, Peaks and the Peak Criterion

A **triplet**, here, is *a unit consisting of any syllable together with its two directly adjacent syllables, left and right*. Like their musical counterparts, prominence triplets may have any shape: they can be peaks, valleys, upward or downward slopes, bent (two the same, one different; fall or rise) or level (all the same). The shapes (Figure 4.5) are produced by the vote counts for the included syllables.

<table>
<thead>
<tr>
<th>(a) peak</th>
<th>(b) valley</th>
<th>(c) upward &amp; downward slopes</th>
<th>(d) bent rise &amp; bent fall</th>
<th>(e) level</th>
</tr>
</thead>
</table>

Figure 4.5. Triplet shapes.

It is important to note that in this analysis, unlike their musical counterparts, prominence triplets *overlap*. This is so that every syllable in an utterance may be considered in relation to its two directly adjacent syllables. Where a syllable is initial or final and therefore has only one adjacent syllable, the other side is simply treated as silence, with a zero value. Figure 4.6 shows a medial phrase in one of the stimuli from the main experiment, *e toru* (“[there were] three”), and all of the overlapping triplets connected with the syllables in it.

*Ko aua pukapuka | e toru | ngā wāhanga.*

“Those books **had three** parts.”

![Figure 4.6. Phrase with partial context. The shaded areas show overlapping syllable triplets.](image)
Each row in the figure contains one triplet. The bold type shows all triplet-centres in the sequence, while the shading shows how each of the three syllables in the phrase e toru are related to their two adjacent syllables, contributing to the triplets around them. The chart in Figure 4.7 below has the same phrase (in bold), with the percentage votes for each of the syllables, which shows the shapes of all the triplets involved. These percentage vote counts are from Group Z (all 92 participants), calculated by \((x/92) \times 100\).

![Figure 4.7. Phrase e toru with partial context, showing overlapping syllable triplets.](image)

In this case, some of the triplets are slopes; some are peaks and some are valleys. When searching for prominence, the triplet shapes in any utterance that are of interest are the peaks. The others are part of the landscape, providing the context and contrast. Peaks must be obtained using the vote counts, meaning that they can only be determined once all of the results are in hand. Peaks can be defined simply under what will be called the Peak Criterion:

A peak is a triplet-centre syllable that received more participant votes than the syllables either side of it in the triplet. A peak may consist of two or more adjacent syllables if those syllables each received exactly the same number of votes.

The Peak Criterion allows isolation of all and only the syllables which were considered more prominent than their triplet neighbours, without relying on a semi-arbitrary percentage vote count threshold. A basic example of this is, of course, e toru above. The criterion also allows for instances of what are called, in this study, pair-tops and tri-tops, where the peak of the triplet was actually two or three syllables, in sequence, with the same number of votes. An example can be seen in the phrase whakaaro ai au (“why I thought”), from another of the stimuli, in Figure 4.8 below.

Here, in the sequence wha-ka, which is part of the word whakaaro (“think”), both of the syllables (shown in bold) received the same percentage of votes (37%) from Group Z participants. This
gave the triplets of which the syllables are the centres “bent” shapes (both rise and fall). Generally speaking, bent triplets would be excluded from the peak list; however, these two syllables together still received more votes than the syllables either side of them (*i* and *a*). It is not possible to say that either *wha* or *ka* is the more prominent, and therefore they must be considered as constituting the top of a peak *together*. Thus: pair-top. A tri-top is simply a version of this that is one syllable greater, but these were very rare in the data. There were no peaks that spanned more than three syllables.

*Ko ēnei tonu tētahi o ngā take nui whakaaro ai au ki te haere ki reira.*

“These, indeed, are one of the main reasons why I thought to go there.”

Peaks are clear numerically, but are not all the same. They come in different sizes, and also, like triplets, different shapes. In some cases, the two outer syllables in the triplet received very similar or the same number of votes to each other, producing an even, caret-like peak shape. In other cases, either the left or right syllable had fewer votes than the other, producing an uneven, hockey stick-like peak shape. All of the shapes in the results varied in their dimensions; that is, some had a very high triplet-centre value and very low adjacent values, giving a very pointy appearance. In other cases, the difference between the centre value and the outer values was very small, producing a flatter shape. These variations and could also occur on higher or lower values overall. For example, a caret of either type could occur whose three values were all below 50%. General examples of peak shapes are shown in Figure 4.9 below.
The variation identified in peak size and shape is essentially relative, and based on observation. In the course of examining the peak shape data for the purposes of developing the assessment of peak validity, some numerical criteria were initially established for peak size in general and what constituted pointy versus flat. The criteria were, however, arbitrary. Ultimately, statistical processing, still based on the triplet unit, provided a better, more objective, and more efficient assessment solution.

4.5.3. Triplet Mode: A Triplet-Based Statistical Approach to Refinement of the Peak List

The approach here began with the entire PROM Dataset of 636 syllables (not just the peaks). In addition to these 636 triplet-centres with actual syllable content, there were 60 zero-centres that provided the edge-padding in the triplets for the initial and final syllables in each utterance, making 696 triplet-centres in total. A slice of the dataset is provided in Table 4.4 below to illustrate this.

Table 4.4.
Slice of Dataset Showing Part of Two Utterances, With Syllables, Triplet Context, and Added Zero-Centres

<table>
<thead>
<tr>
<th>Utterance</th>
<th>Syllable</th>
<th>Triplet</th>
<th>Group Z Vote Count (x/92)</th>
</tr>
</thead>
<tbody>
<tr>
<td>he_01</td>
<td>taa</td>
<td>ngaataanga</td>
<td>37</td>
</tr>
<tr>
<td>he_01</td>
<td>nga</td>
<td>taangata</td>
<td>23</td>
</tr>
<tr>
<td>he_01</td>
<td>ta</td>
<td>ngata_</td>
<td>40</td>
</tr>
<tr>
<td>he_01</td>
<td>_</td>
<td>ta_</td>
<td>0</td>
</tr>
<tr>
<td>he_02</td>
<td>_</td>
<td>_me</td>
<td>0</td>
</tr>
<tr>
<td>he_02</td>
<td>me</td>
<td>_metii</td>
<td>27</td>
</tr>
<tr>
<td>he_02</td>
<td>tii</td>
<td>metiima</td>
<td>54</td>
</tr>
</tbody>
</table>

On the left is the utterance code (this slice shows the end of utterance “he_01” and the beginning of utterance “he_02”; both are from historical elders). The next two columns show each syllable and then the triplet of which it is the centre. The double vowels here represent long vowels,
which would usually be marked with macrons: this was for simplicity, in order to remove special characters from the data processing. Note the two zero-centres, marked with the underscore (_) in the table, at the transition between the two utterances. As mentioned above, this was both to account for the edge of the sentences, where there is indeed silence for a listener, and to allow for easier application of the statistical testing. The right column is an example of vote count, where the silences have been assigned a placeholder zero vote count.

To this master list of 696 triplet-centres, two statistical tests were applied. Chi square tests were used on the original vote frequencies to find p-values for the number of votes for each of the triplet-centres, in order to show whether or not the difference between the numbers of votes for a peak and for its neighbours was statistically significant. Anywhere there was a zero value contained within a triplet, Fisher’s exact test was used instead. In all cases, the alpha level was 0.05.

Next, any non-peaks (for example, slopes or valleys) were removed using a spreadsheet formula version of the Peak Criterion. This formula compared each vote count total (again, the original frequencies) to those either side of it, and isolated all of the peaks: that is, the potential prominences. The formula was also designed to allow those special cases mentioned above, the pair-tops and tri-tops, to remain in the list.

There were not many of these pair- and tri-tops (only three in the Group Z results), and they were dealt with in the following way. They were isolated and manually turned into a triplet by treating the identical syllables as a single syllable with the existing value. The following tables (4.5 and 4.6 below) explain the pair-top treatment. The example pair-top used here comes from the phrase *i ērā whakahātere* (“those circumstances”), in an utterance from one of the present-day elders (pe_07), where the adjacent syllables *rā* and *wha* each received the same number of participant votes (30/92). Table 4.5 is a dataset slice showing the relevant sequence (shaded) and its triplet context.

---

4 The query method used in Emu/R for extracting syllables and/or vowels for use in the acoustic analysis (§ 4.6.2) still allows the individual syllables to be examined separately, even though they are treated as a unit for statistical purposes here.
Table 4.5.

*Slice of Dataset Showing Adjacent Syllables with Identical Vote Counts*

<table>
<thead>
<tr>
<th>Utterance</th>
<th>Syllable</th>
<th>Triplet</th>
<th>Vote Count</th>
<th>%</th>
<th>$X^2$</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>pe_07</td>
<td>ee</td>
<td>iecraa</td>
<td>18</td>
<td>20</td>
<td>11.68421</td>
<td>0.002903</td>
</tr>
<tr>
<td>pe_07</td>
<td>raa</td>
<td>eeraawha</td>
<td>30</td>
<td>33</td>
<td>3.692308</td>
<td>0.157843</td>
</tr>
<tr>
<td>pe_07</td>
<td>wha</td>
<td>raawhaka</td>
<td>30</td>
<td>33</td>
<td>5.157895</td>
<td>0.075854</td>
</tr>
<tr>
<td>pe_07</td>
<td>ka</td>
<td>whakahae</td>
<td>16</td>
<td>17</td>
<td>36.86726</td>
<td>9.87E-09</td>
</tr>
</tbody>
</table>

The chi square and p-value results, on the right, are different for each syllable, since at first they were calculated on different triplets. However, because these syllables, *rā* and *wha*, had the same score, and it was not possible to say that one or the other was more prominent or the “true” peak top, they were concatenated and treated as a single unit, appearing as *raa.wha* in Table 4.6 below, with a score of 30/92. The chi square statistic and p-value were recalculated based on the new triplet. The concatenated unit and its new statistical values were then returned to the list of peaks as a single line.

Table 4.6.

*Concatenated Pair-Top in New Triplet, with Recalculated $X^2$, p-Value, and Adjusted p-Value*

<table>
<thead>
<tr>
<th>Utterance</th>
<th>Syllable</th>
<th>Triplet</th>
<th>Vote Count</th>
<th>%</th>
<th>$X^2$</th>
<th>p</th>
<th>Adjusted p</th>
</tr>
</thead>
<tbody>
<tr>
<td>pe_07</td>
<td>raa.wha</td>
<td>eecraawhaka</td>
<td>30</td>
<td>33</td>
<td>5.375</td>
<td>0.06805</td>
<td>0.094564</td>
</tr>
</tbody>
</table>

The final column in Table 4.6 is new and shows an adjusted p-value. This adjustment was not specific to the pair-tops: it was part of the overall process, and was done on the entire list of p-values, in order to correct for the very large number of statistical tests involved. The adjustment procedure used was Benjamini & Hochberg (1995). Following this, the list of peaks was sorted, on the adjusted p-values, from highest to lowest significance level. As mentioned above, the alpha value used was 0.05. This left a list of what could be considered the statistically significant peaks. These would be the perceived prominences (PPMs) for further analysis. This gave a revised definition for a **perceived prominence (PPM)** in this experiment:

A PPM is a syllable which constitutes the centre of a peak-shaped triplet, and whose difference from the other syllables in that triplet is statistically significant at $p < 0.05$. 

103
This PPM identification method, named Triplet Mode, was carried out separately on the results from all of the different participant groups established in the study. It was primarily this method that produced the lists of perceived prominences for the whole group of participants (Group Z), as well as the proficiency group subsets (Groups A, B and C). However, the three proficiency groups exhibited differences in behaviour that required some adjustments to be made to the Triplet Mode process in order to accommodate them. These are described in the following sections.

4.5.4. Response of the Proficiency Groups to the Perception Task

The main point of difference between the proficiency groups is their behaviour in the perception task. It concerns how likely they were to check a box for any syllable in the first place; that is, to identify a syllable as prominent or “standing out” to them. The chart in Figure 4.10 below illustrates this for all three groups and all of the syllables in the PROM Dataset.

![Figure 4.10. Rate of syllable selection by proficiency group.](chart)

On the x axis are all 636 PROM syllables. There is a data point for every syllable, and they are ordered across the chart by the percentage of participants in a given group who checked the box for that syllable. On the y axis is the percentage value. There are three lines, one for each group. The wider a plateau on the line for that group, the more syllables were checked by that number of participants. For example, following the line for Group A (blue), there is an obvious wide plateau at around 32% on the y axis. This represents 84 syllables (out of the 636) that were selected by 32% of Group A participants (about 9 people). These would not necessarily be the same nine people for all 84 of the syllables, but there were nine each time. If Group A is a valid group in
PPM selection, then the motivations for PPM selection ought to be the same or similar for any nine members of that group. As we will see, there is evidence of Group A behaving differently to the other groups that reinforces this.

With regard to differences in group response to the task, the section of interest in the chart begins at the left side. Group A (the blue line) has a steeper initial rise in the number of checks, and overall higher percentages (selection rate), until about the 60% mark. This shows that Group A participants were more likely to check any syllable as prominent, but not as close to unanimity on the syllables with higher percentages. Mild foreshadowing of this behaviour was visible in the previous web experiment (Chapter Three, Section 3.3.4.4), and the present experiment has confirmed the observation. Above the 60% mark, it is Group B (the red line) with the steeper curve, showing stronger agreement within the group on those syllables. Group C’s results (the black line) are very close to Group B’s, though it is worth noting that C’s percentage selection rate stops at 88%, while both other groups reach 100% for at least one syllable.5

The average percentages of checks for the groups, which could be restated in terms of the likelihood that any syllable would be checked by a participant from that group, were as follows: Group A had 40%, Group B had 32%, and Group C had 28%. Group A was clearly more likely to check syllables in the first place. This was foreshadowed in the earlier results (Chapter Three, Section 3.3.3.3), and might be for a number of reasons, but is almost certainly related to proficiency, interpretation of the instructions, or more likely, both. It might be expected that Group A participants would, owing to their greater parsing ability, select fewer syllables overall. However, as discussed in previous chapters, because the test instructions were intentionally minimal, there was a good deal of room for interpretation. On being told to select syllables that “stood out” to them, it is possible that some Group A participants took this to mean that they should check all syllables of which they understood the meaning, or perhaps those that they could hear clearly. Some of them occasionally checked whole words or even phrases, which in turn caused noise at the lower end of the percentage agreement range. This noise, visible as the lump on the left side of the Group A line on the chart, has the effect of reducing the contrast between

---

5 The Group Z line, not shown on this chart, follows very close to the Group B and C lines, which is likely due to the majority contribution from those two groups to the Group Z vote counts.
the scores for these syllables and the scores of the syllables that were checked by many more of the participants in Group A: it makes the peaks smaller.

Another way of describing the noise and reduced contrast in the Group A results is to say that those results have a **higher floor**. In its own right, this is an interesting behavioural observation on Group A’s response to the exercise they were given, but further effects of the higher floor became obvious when the Triplet Mode statistical PPM list refinement method was applied. These effects, the problem they created, and the solution that was implemented are described below.

4.5.5. Proficiency Group Effects on the Triplet Mode Statistics

When the initial lists of PPMs for the different proficiency groups were generated using Triplet Mode, there was an obvious difference in the number of PPM results between Group A and Group B/C. This is summarised in Table 4.7 below, where Group Z’s result is also included for reference.6

<table>
<thead>
<tr>
<th>Participant Group</th>
<th>Group Z “All”</th>
<th>Group A “High”</th>
<th>Group B “Exposed”</th>
<th>Group C “Zero”</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of Participants</td>
<td>92</td>
<td>28</td>
<td>30</td>
<td>34</td>
</tr>
<tr>
<td>No. of PPMs (Triplet Mode)</td>
<td>148</td>
<td>41</td>
<td>100</td>
<td>90</td>
</tr>
</tbody>
</table>

Note the similar numbers for Group B and C, and the much lower number of PPMs for Group A. This is the effect of the Group A higher floor. For the Triplet Mode method, it meant that the differences between the numbers in the triplets (that is, between the peaks and their adjacent syllables) were generally smaller. When the statistical tests were applied to the triplets, it created more non-significant results, excluding those peaks. With only 41 PPM results, the small size of the Group A list, as opposed to the Group B and C lists, created problems with both comparability and token numbers for other parts of the analysis, particularly the vowel spaces. Owing to the use of averages, low numbers of tokens reduce the accuracy of a vowel space plot

---

6 There is partial overlap in the PPM lists: some of the PPMs for different groups are the same syllables from the same words; others are not. For a full explanation and discussion of this, see Section 5.4.3.
Besides this, inspection of the non-significant peaks in the Triplet Mode Group A list revealed a total of 63 peaks with over 60% of the participant vote (that is, the threshold used in the first web experiment) that would not be included in the analysis. Some of these peaks had agreement scores of up to 93%. This was an echo of the problem discovered in the earlier experiment (see the evaluation in Chapter Three (§ 3.3.5.2) and the development of Triplet Mode in § 4.5.2 and § 4.5.3 for more detail), where the 60% participant vote threshold had also eliminated some peaks that appeared worthy of investigation. It had been established that the 60% threshold was insufficient when used on its own. However, in a study encompassing the concept of proficiency effects and looking at possible features that cause prominence, it seemed counter-intuitive to ignore syllables that up to 93% of the Group A participants indicated they perceived as prominent. Figure 4.11 below contains two of the more extreme examples of these excluded but potentially useful peaks. These are single phrases extracted from two different utterances.

Figure 4.11. Examples of non-significant Triplet Mode peaks with very high selection rates.

On the left hand side is the phrase me tīmata atu (“[it] begins”). The triplet of interest is me- tī- ma, where the peak syllable -tī- has a vote count of 82%, but was not significant. On the right hand side is the phrase ki ngā whakatauāki (“to (the) proverbs”), where the triplet of interest is tau- ā- ki. In this triplet, the peak syllable -ā- has a vote count of 93%, but under Triplet Mode, this was not significant either.
One possible solution to this problem would have been to discard Triplet Mode in favour of some other method entirely. However, the general principle behind Triplet Mode had proven itself adequate in most cases. Alternatively, the alpha level used in Triplet Mode could have been raised, in order to include a few more syllables. This would only have amounted to a small increase in PPM syllables, and any chosen significance level, even the standard 0.05, is, in a way, as arbitrary as the 60% threshold used in the previous study. In fact, there is a better solution, which allows the retention of the peak content comparison and statistical validation (at a standard level) provided by Triplet Mode, while accounting for the different prominence identification behaviour displayed by the proficiency groups and including more of the extra peaks. This solution is named Quinlet Mode.7

4.5.6. Quinlet Mode

As the name suggests, Quinlet Mode is a modification to Triplet Mode. The difference is that the unit on which this mode operates is not a triplet, or sequence of three syllables, but a quinlet, or sequence of five syllables. These five include the target syllable, or quinlet-centre, and its four adjacent syllables: two on either side. This is illustrated in Figure 4.12 below. The example is contained within a single phrase from one of the utterances (as before, phrase boundaries are shown by vertical bars): i [sic] tētahi kāinga noho (“a place to live”).

The peak syllable concerned is kāi-, which had a vote count of 68%. The smaller, inner, grey brackets show this peak within its triplet, -hi-kāi-nga-. The larger, outer, black brackets show it as the centre of its quinlet: -ta-hi-kāi-nga-no-.

---

7 Quinlet is intentionally incorrect: it was named facetiously and proved too popular to change.
The unit of five syllables is primarily intended to correct for the higher floor in Group A’s results. By extending the floor area over which the statistical analysis is applied and creating a greater potential contrast for the central peaks concerned, it draws out the syllables on which there was real consensus and allows a better view of the overall shape of Group A’s prominence identification in particular. However, in order to keep the analysis consistent, Quinlet Mode was applied across the board to Groups A, B, and C.

The process involved in Quinlet Mode is identical to that of Triplet Mode, across five syllables instead of three. All quinlet units in the database were identified and, as in Triplet Mode, there were zero placeholders added where the syllables concerned were near an edge (initial or final). They were then tested using a chi-square or Fisher’s exact test, depending on whether or not there were zeros in the quinlet. Because of the wider unit, about a third of the quinlets required the Fisher’s exact test to allow for the edge-zeros. The p-values resulting from the statistical tests were adjusted in the same way as in Triplet Mode, using Benjamini & Hochberg (1995). They were then sorted by significance level. For consistency, the cutoff point was, once again, 0.05. All of the syllables near edges turned out to be significant in the final list, indicating that the sensitivity to zeros and edges exhibited by the Triplet Mode process is not reduced by widening the unit. The Quinlet Mode results were added to the Emu database on the QuinMode level in the hierarchy.

After application of Quinlet Mode, the PPM lists of all three proficiency groups were larger, which improved the set sizes for analysis. A summary of the increases in the groups’ PPMs may be found below in Table 4.8. It is clear that the Group A list was effectively doubled in size, while the Group B and C lists are about half as large again.

Table 4.8.

| Summary of Proficiency Group PPMs: from Triplet Mode to Quinlet Mode |
|-----------------------------|-----------------------------|-----------------------------|-----------------------------|
| No. of PPMS                 | Group A                     | Group B                     | Group C                     |
| Triplet Mode                | 41                          | 100                         | 90                          |
| Quinlet Mode                | 84                          | 155                         | 155                         |

109
There are two further factors to consider. The first concerns the effect of edge-adjacent syllables on the statistical tests used. The statistically-refined lists of PPMs for all participant groups, both overall and according to proficiency split, featured a large number of edge syllables. These were both initial and final, though initial syllables were dominant. In all cases, the triplet tested by the statistics contained at least one zero value, to which it would appear that the test is particularly sensitive, regardless of the height of the actual peak. When the list was sorted, it was the edge syllables that had the highest significance. A correction for this was not found; therefore, the edge syllables remain in the final lists. For discussion of the contribution of these edge syllables to the PPM lists, see Chapter Five, Section 5.4, and Chapter Six, Section 6.5.

The results and analysis that now follow are concerned with both Group Z and the three proficiency groups. Presentation and discussion of Group Z’s results are in Chapter Five, and use the Triplet Mode list created for that group. The proficiency groups’ (ABC) results are presented and compared in Chapter Six, and will refer only to the Quinlet Mode lists described above.

4.6. The PROM Database: Construction and Analysis Methods

There were a number of processes used in the analysis of the PPM list, once that list had been obtained. This section describes those methods and the reasons for their inclusion. It also discusses the building of the Emu database that facilitated most of the analysis.

4.6.1. Creating the PROM Database in Emu

All of the recordings in the study, in .wav format, were prepared for access by the speech analysis software Emu. To expand on the brief description in Chapter Three (§ 3.2.4), in this software, one may create hierarchically structured databases that appear to “divide” a sound file into sections known as segments: for example, phonemes, syllables, words, phrases or any other unit specified by the user in the database template. Events (single time points) may also be marked: for example, vowel targets. These marked sections are given user-defined labels as required, such as maaori, mao, ri, m, aao, and so on. The segmentation, as it is called, does not compromise the sound file, which remains intact. In Emu, one may also view a spectrogram and waveform representation of the utterance; the F0 contour; and formant tracks, all of which were used in this
study. Again, the reader is encouraged to consult Harrington (2010b) for a comprehensive description of database creation and related factors.

In the Emu database used to manage the recordings in the PROM Dataset, there were eight levels of labels, designed to encompass all of the segments and events that might be useful in queries. Some of the segments and events were given specific sets of codes as labels; the others had simply the words or syllables involved. The initial labelling of the database included Speaker Reference, Word, Syllable, Stress and Phoneme levels.

The Emu Pitch and Formant Tool was employed, at appropriate settings for recordings of male speakers, to calculate the F0 and formant tracks. For the formants, F1-F4 were calculated, though practically speaking, only F1-F3 were of use in the present investigation for vowel target marking and vowel space plots. Once the F0 and formants had been calculated and examined, the Phrase Contour and Target/PitchPeak levels (for pitch peak only) were added. When the refined PPM lists had been created, the Perceived Prominence level was marked and the Target/Pitch Peak level had the vowel targets added for PPM vowels. The eight levels of the PROM Database are shown below, in hierarchical order from top to bottom.

**Speaker Reference**

Codes for each whole utterance, including a speaker group designation and an utterance number from 1-10.

**Phrase Contour**

A dual-purpose level, containing the observed contour type (e.g., fall or rise), corresponding to grammatical phrase divisions for visual reference.

**Word**

Word divisions.

**Syllable**

Division of the words into syllables. Syllabification was according to the rules for Māori as outlined in Chapter Two (§ 2.2.1). The syllable is the principal unit on which the analysis is based. The following two levels are also syllable-sized and are based on the same segmentation.
Stress

The Biggs expected stress category (phrase stress, word stress, etc.) for each syllable.

Perceived Prominence (Triplets)

The perceived prominences (PPMs), as identified by the participants in Group Z and categorised as significant by the triplet-based statistical analysis. The syllables were labelled with Z (prominent) or NP (not prominent), and position in the utterance: initial, medial or final. Thus, a label on this level might appear as, for example, “Zm” or “NPi”.

Perceived Prominence (Quinlets)

The perceived prominences (PPMs), as identified by the Group A, B, and C participants and categorised as significant by the quinlet-based statistical analysis. The syllables were labelled according to participant group: A, B, C, or any combination of these, or NP (not prominent for any group), and position in the utterance: initial, medial or final, for example, “BCi” or “NPf”.

Phoneme

Individual phonemes, whose labels were allocated according to the Māori phonemic inventory (see Chapter Two). Segmentation of all phonemes, including non-diphthong vowels in sequence, was based on a set of criteria derived from careful examination of spectrogram and waveform, as well as the signal itself.

Target / Pitch Peak

Another dual-purpose level, this time containing two different events:

Vowel targets. The targets for the monophthongs in syllables identified as PPMs were marked by hand according to standard criteria on F1, F2 and occasionally F3 (Harrington, 2010b). Diphthong nuclei were marked with a first and a second target by this method. Long vowels produced across a syllable boundary were also marked with two targets, again using the same criteria. Vowel targets for monophthong nuclei of all other (non-PPM) syllables were taken automatically at the vowel midpoint during later analysis in R. The purpose of the vowel targets was to provide formant values for vowel space plots. Vowels not marked manually with targets (that is, the non-PPMs), had their target values extracted at the vowel midpoint.

Pitch peaks, labelled H (after H*), were marked according to the F0 contour as produced by the Emu Pitch & Formant Tool. Known articulation effects on F0 were taken into account when marking these events.
A screencapture of the Emu signal view for an utterance in the database is shown in Figure 4.13 below. Not all of the nine levels of markup are shown in the figure: they can be toggled on and off so as to show only those relevant to the current task or to provide more space on the screen.

![Image](image1)

Figure 4.13. Emu signal view: selected levels, waveform, spectrogram, and formant tracks.

The levels were also connected in a hierarchy. Figure 4.14 below is a screencapture of Emu’s hierarchy view, showing the same utterance as in Figure 4.13.

![Image](image2)

Figure 4.14. Emu hierarchy view, showing linked levels.

Hierarchies may be linked up either manually using a mouse or automatically by use of a function. In this case, a function was used, after which each of the utterance hierarchies was checked for assignment errors and corrected as necessary. The purpose of creating the hierarchy is to enable complex queries when searching the database (§ 4.6.2). Queries may be performed within the Emu software itself, or, as in this study, by using the statistical analysis software R, v.2.14.1 (R
Development Core Team, 2011), with the package that enables it to interact with Emu (see Harrington, 2010b).

The information contained in the levels, which consists of the start times for segments (here, all units except target and pitch peak), the time point for events (targets or pitch peaks) and the label given to that segment or event, is kept in separate files in a database folder accessed by the software. These files are connected to the .wav file, the Emu-generated hierarchy file, and the files that store the track data (formants and F0) for each utterance by their filename and by instructions in the database template.

4.6.2. Use of R to Query the Emu Database

Using R, one may execute simple searches; for example, find all the tokens of the vowel /a/ in the dataset. Depending on the level of detail in the Emu database, one may also execute very complex queries. For example, in the PROM Database, it would be possible to find all the monophthongs which are preceded by /k/ and occur in expected phrase stress syllables and are also PPMs as identified by proficiency Group B.

The query results are stored in “objects”, such as vectors (a kind of list), or matrices or data frames (tables), for later refinement or use. With further tools in the Emu-R package, it is possible to produce F1-F2 vowel space plots (§ 5.7) and acquire duration data for segments, among many other functions.

An example of an Emu-R query, with results, is below. This is a complex query, searching the entire dataset for all of the PPMs identified by Group Z that are also phrase stresses and contain the phoneme /aː/. This query creates an object “x” of the type matrix. Object names are one word, specified by the user, and can be much longer and more informative than “x”.

R commands in the example below are preceded by the prompt sign > and are bolded in these examples. In the query syntax, first, the database to search is specified ("prom_expt": the PROM Database), then the utterances within the database (here, * simply means “search all”). The remaining section contains the levels to search, in their hierarchical order from top to bottom. The caret (^) shows ranking. The hash mark (#) indicates the level on which to focus, which here is the Phoneme level.
R found 7 records, which means there are 7 PPMs in the database that match all of the criteria, and they are now stored in “x”. Displaying the contents of “x” brings up the matrix that was created.

```
> x<-emu.query("prom_expt","*","[Stress=P|PW^[PPTriplet=ppz^[Phoneme=a:]]]")
moving data from Tcl to R
Read 7 records
```

```
> x
segment  list from database:  prom_expt
query was:  [Stress=P|PW^[PPTriplet=ppz^[Phoneme=a:]]] 
labels    start      end utts
1     a: 1834.793 1904.912 pe_02
2     a: 2194.446 2373.091 pe_03
3     a: 1273.336 1378.823 pe_07
4     a: 3594.451 3752.318 he_02
5     a: 4006.366 4150.186 py_03
6     a:  646.594  801.637 py_04
7     a: 1137.833 1266.659 py_07
```

Working from right to left, the “utts” column shows which utterance the tokens come from: this is useful when tracking errors or investigating specific tokens, for example. The “start” and “end” columns are start and end times (in milliseconds) for the segments concerned. These can be used to calculate durations or to locate segments in an utterance where there may be, for example, two tokens of a certain vowel. The “labels” column, on the left, is the focus of the query. It was mentioned above that the Phoneme level is in focus here, which is why the column is filled with /a://. If an additional level is added to the query, R will also show the words in which the phonemes found are contained. Note the addition of #Word!=x to the query below.

```
> x<-emu.query("prom_expt","*","[#Word!=x^[Stress=P|PW^[PPTriplet=ppz^[Phoneme=a:]]]]")
moving data from Tcl to R
Read 7 records
> x
segment  list from database:  prom_expt
query was:  [#Word!=x^[Stress=P|PW^[PPTriplet=ppz^[Phoneme=a:]]]] 
labels    start      end utts
1      waahanga 1706.241 2134.748 pe_02
2      whakatauaakii 1738.175 2698.999 pe_03
3       maatou 1175.963 1511.774 pe_07
4        raan 3565.748 3752.384 he_02
5      raanei 3965.414 4429.972 py_03
6       naana  546.486  960.093 py_04
7   taangata 1051.121 1633.348 py_07
```

The hash mark has shifted the focus to the Word level. !=x means “is not equal to x”, which allows any label from the Word level to be included in the query. This is necessary because the Word level does not have any defined codes: the labels are entirely dependent on the words in the
utterance. When the contents of the object $x$ are displayed once more, the labels column is now filled with words, all of which contain /a:/.

The start and end times have also changed to reflect the larger units.

In addition to basic list queries and token-counting, in this study, R and Emu together provided much more information about the dataset and the objects made from it. This included duration data, F0 readings, formant data, means and standard deviations, and production of plots, including vowel spaces. R, along with Excel and some online statistical tools, also provided the ability to assess the results that were being produced.

4.6.3. Statistical Assessment of Results

As a separate exercise from the statistical methods used in Triplet Mode as described above, four different statistical tests were employed in assessing the significance of the results. These tests were applied to both the Group Z results presented in this chapter, and the proficiency split results presented in Chapter Five.

The tests used were the chi square test, Fisher’s exact test, t-test and ANOVA. The chi square test was used, in Excel, for the majority of the counted results, particularly in comparing the three proficiency groups’ results in Chapter Five. Where values were lower than 5, R was used to perform Fisher’s exact test in place of the chi square. The t-test was used for the numerous results involving means (for example, mean duration or mean F0); calculated with the aid of an online statistical calculation tool: GraphPad QuickCalcs (http://www.graphpad.com). ANOVA was used, in R, to assess the significance of differences in nuclear durations and formant values between various categories including syllable weight and Biggs stress, and across the proficiency groups. For all tests, the alpha level was 0.05.

The remainder of this chapter contains the results of the main study. These will be presented in sections, beginning with a PROM Dataset profile and moving on to a perceived prominence (PPM) profile for Group Z. The PPM profile covers general characteristics, including distribution, and alignment with the Biggs stress categories, along with some acoustic
characteristics: duration, vowel quality and F0 behaviour. These angles of investigation do occasionally overlap. Where this is the case, it is explained in the discussion.

4.7. Results & Analysis: PROM Dataset Overview

A profile of the characteristics of the PROM Dataset as a whole will be presented in the following section. It is intended as a reference point for the later discussion. There are 30 utterances in the PROM Dataset, these being the 30 core stimulus sentences used in the perception task. They do not include the two buffer sentences, as explained earlier. Within these 30 utterances are 147 grammatical phrases, with an average of 5 phrases per utterance. The following subsections summarise the main characteristics of the dataset.

4.7.1. Syllable Weight, Type and Content

There are 636 content syllables, at an average of 21 syllables per utterance. Onsets are found on 499 of these 636 syllables. All of the Māori consonant inventory (see Chapter Two, Section 2.2.1) is represented in these onsets, but the distribution is uneven: just over a third are either /t/ or /k/. The syllable nuclei contain 564 monophthongs (long and short) and 72 diphthongs (long and short). All of the monophthongs in the Māori vowel inventory are represented; however, one diphthong (/eu/) and four long diphthongs (/aae, aau, oou, eei/) are missing. The most commonly occurring vowel in the dataset is /a/, which accounts for approximately one third of all vowel tokens. This is not an accidental effect of the chosen utterances: /a/ is, in fact, the most frequent Māori vowel (Harlow, 2001; Harlow et al., 2011). Table 4.9 below summarizes the syllable characteristics in the dataset.

Table 4.9.
Summary of PROM Dataset Syllable Characteristics

<table>
<thead>
<tr>
<th>Total Syllables</th>
<th>Syllable Weight</th>
<th>Syllable Type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Heavy</td>
<td>Light</td>
</tr>
<tr>
<td>636</td>
<td>168</td>
<td>468</td>
</tr>
</tbody>
</table>

The syllables in the dataset were classified as either heavy or light according to criteria based on moraic weight, as described in Chapter Two (§ 2.2.1). Light syllables are monomoraic: (C)V, and account for 468 of the 636. Heavy syllables, of which there are 168, are those with nuclei
containing long monophthongs, diphthongs, or long diphthongs. Long monophthongs and
diphthongs are both bimoraic. While syllables containing long diphthongs are technically
trimoraic and therefore could constitute their own super-heavy category, here they are included in
the heavy category. This is for two reasons. Firstly, there are so few long diphthong tokens in the
dataset that they cannot be analyzed as a group in their own right. Secondly, it is a categorisation
supported by mean duration figures: the long diphthongs in this dataset are within or close to the
durational range of long vowels and diphthongs: there is no significant durational difference
between the categories.

There is also durational support for the light-heavy distinction: the light (1μ) syllables’ mean
nuclear duration is significantly shorter than that of the heavy (2-3μ) syllables (ANOVA,
F(1,630) = 221.97, p = 0). Considering the individual vowel categories, the short monophthongs
(M = 66.1ms, SD = 43.3ms) in the dataset are approximately half the length of the long
monophthongs (M = 118.6ms, SD = 54.1ms): this difference is significant, t(562) = 10.34, p =
0.0001. There is also, as it happens, a significant difference between the mean durations of the
long monophthongs (M = 118.6ms, SD = 54.1ms) and the diphthongs (M = 139.8ms, SD =
58.6ms): the diphthongs are roughly 1.5 times the length of the long monophthongs, t(166) =
2.43, p = 0.01). However, given the durational characteristics of the long diphthongs mentioned
above; and considering the relative lack of heavy nuclei in the dataset overall (only 168/636), all
of the non-light syllables have been grouped into one category for simplicity. This is true for all of
the analysis in the principal experiment presented in this thesis (Chapters Five and Six).

Onsets do make a difference to both whole syllable and nuclear duration. In the dataset as a
whole, syllables with onsets (M = 210.3ms, SD = 71.8ms) are significantly longer than those
without (M = 145.6ms, SD = 58.6ms), regardless of syllable weight classification, t(166) = 3.78, p
= 0.0002. Syllables with onsets have significantly shorter nuclear durations than those without
(ANOVA, F(1,630) = 33.39, p < 0.0001), but this does not affect the established heavy-light
nuclear duration contrast.

4.7.2. Biggs Stress Categories

Within the 147 grammatical phrases in the dataset, under the stress rules devised by Bruce Biggs
(1969) and summarised in Chapter Two (§ 2.4.1), there are 153 expected phrase stress syllables.
Generally speaking, there is only one phrase stress per grammatical phrase, and this is still true here. However, it is possible for that phrase stress to span a morpheme or word boundary and cover two syllables in certain circumstances (e.g. *e pā ana*, “about, concerning”), so when the expected phrase stresses are counted in syllables, they total 153 instead of 147. This is particularly important with regard to prominence identification, especially in this test, because both of the two syllables had checkboxes and so either (or both) could be selected by the participants. Figure 4.15 below shows syllable distribution in the PROM Dataset across all of the stress categories.

One hundred and twenty-one expected phrase stresses coincide with expected word stress (PWS). Thirty-three are phrase stress only (PS). Expected word stress only (WS) accounts for 91 syllables, and the remaining 391 are not expected word or phrase stresses (OS).

![Figure 4.15. Syllable distribution in the PROM Dataset by Biggs stress category.](image)

4.7.3. Phrases, Phrase Contours, and $H^*$

Existing description of Māori mentions a fall contour in the majority of phrases (see Chapter Two, Section 2.4.2), with rise being the other option. From this description, one might expect a binary classification of phrase contour categories in this dataset: fall and rise. The 147 phrases in the dataset were, however, classified into three categories of contour, according to observation of F0 in each phrase.

The dominant category is “fall”, with 122 phrases, while “rise” accounts for 15 phrases. During the analysis, when listening to the utterances and viewing the F0 contour as calculated by Emu, it became apparent that 10 phrases very definitely belonged in a third category, in which there was sufficiently “minimal” change in F0 to make them different from the fall or rise contours, hence the name. This difference can be seen in Table 4.10 below.
Table 4.10.

Average Mean F0\(^8\) and Mean Range of F0 for Phrase Contour Categories in PROM Dataset

<table>
<thead>
<tr>
<th>Contour Category</th>
<th>No. of Phrases</th>
<th>Mean Range of F0 (Hz)</th>
<th>SD (Hz)</th>
<th>Average Mean F0 (Hz)</th>
<th>SD (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall</td>
<td>122</td>
<td>151.4</td>
<td>52.7</td>
<td>97.4</td>
<td>30.4</td>
</tr>
<tr>
<td>Rise</td>
<td>15</td>
<td>172.4</td>
<td>50.3</td>
<td>116.6</td>
<td>32.2</td>
</tr>
<tr>
<td>Minimal</td>
<td>10</td>
<td>87.9</td>
<td>51.6</td>
<td>85.6</td>
<td>29.9</td>
</tr>
</tbody>
</table>

The mean range of F0 for phrases categorized as “minimal” is significantly smaller than that of the phrases in both the fall, \(t(130) = 3.67, p = 0.0004\), and rise, \(t(23) = 4.07, p = 0.0005\) categories. A graphical example is provided in Figure 4.16, which appears over the page.

These are not extreme examples of the phrase contour categories, but rather an excerpt from one of the stimuli in which all three phrase contour types occur in sequence. The three relevant phrases are shown in bold in the sentence below. Full glossing for the example is available in Appendix C. The F0 contour is from Emu; word and phrase segmentation has been added.

Nō reira | korekore noa | ngā Mahi | a Rēhia | ka tae | ki te tau | iwa tekau | mā whā.

“So, there were no kapahaka competitions until ‘94.”

Figure 4.16. Examples of Emu-generated F0 contour for minimal, fall, and rise type phrases.

In this figure, there are also two points marked H*: this is the high pitch point mentioned in the literature (see Chapter Two, Section 2.4.2). The H* is expected to occur on the phrase stress, in Biggs’ stress category terminology, after which the pitch in the phrase is generally expected to fall.

---

\(^8\) This term “average mean” refers to the fact that the F0 reading for a single unit, such as a syllable or phrase, is in fact the mean of several readings taken across the unit. When a group of such units is considered together, this creates a mean of means, or average mean. For events, such as H*, there is a single F0 reading and so for a group of events, the mean is a single mean. Some further discussion of this can be found in Section 4.9.7.
In a phrase with a rising contour, this fall would not occur. In the PROM Database, 137 instances of H* were marked: one in each fall or rise type phrase (§ 4.6.1). No H* events were marked in the phrases classified as minimal, as it was not possible to discern a clear F0 maximum.

The mean F0 of the H* events is significantly higher than the average mean phrase F0 in both fall type, \( t(242) = 10.87, p < 0.0001 \), and rise type, \( t(28) = 3.23, p = 0.0031 \) phrase contour categories (Table 4.11). This is to be expected, since the targets for those events are the highest points in the phrase.

Table 4.11.

<table>
<thead>
<tr>
<th>Contour Category</th>
<th>Phrases &amp; H*s ((x/137))</th>
<th>Average Mean F0 (\text{Hz})</th>
<th>SD (\text{Hz})</th>
<th>Mean F0 (\text{Hz})</th>
<th>SD (\text{Hz})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall</td>
<td>122</td>
<td>97.4</td>
<td>30.4</td>
<td>143.6</td>
<td>35.8</td>
</tr>
<tr>
<td>Rise</td>
<td>15</td>
<td>116.6</td>
<td>50.3</td>
<td>169.2</td>
<td>37.9</td>
</tr>
</tbody>
</table>

According to Biggs’ descriptions, as mentioned above, the high pitch point is expected to align with the phrase stress. The chart in Figure 4.17 below shows how the H*’s aligned with stress category types in the whole dataset.

Figure 4.17. Distribution of H* across stress categories in the PROM Dataset.

Over half of the 137 H* events occurred on PS-Group syllables (72/137, or 53%), but slightly more occurred on WS-Group syllables (76/137, or 55%). As elsewhere, PWS (59/137) has the highest result individually, but it is also the largest available category: most of the phrase stresses in the dataset are PWS type. Otherwise, word stress (WS) syllables (17/137) are better represented than PS-only syllables (13/137).
This particular dataset appears to support comments about the coincidence of H* and expected phrase stress. As regards Bauer’s observation (see Chapter Two, Section 2.4.2) that H* may also be heard on the syllable preceding PS, in this dataset, pre-PS H* (18/137) does not occur any more frequently than post-PS H* (22/137). Just over a third of the H* in the dataset (48/137) were also found on non-expected-stress (OS) syllables. Further H*-related analysis, concerning the participants’ perceived prominences and their locations relative to it, can be found in Chapter Five, Section 5.8.2, and Chapter Six, Section 6.9.

4.7.4. Pauses, Errors, and Edges

Pauses and errors in speech can affect the perception of prominence. These were marked in the database where they occurred, to facilitate their measurement and the examination of where and how PPMs were identified around them. There are 20 noticeable pauses and 2 errors in the dataset. Both of the errors are restarts, and the majority of the pauses are narrative. A few are obvious hesitations or thinking pauses, and one is for breath. There is one super-long narrative pause (1469ms); the next longest has only about half the duration, at 712ms. The shortest pause measures 78ms. The mean duration, excluding the super-long pause, is about 315ms (SD = 174ms). Apart from two phrase-internal exceptions, the pauses and errors are mostly found between grammatical phrases, as predicted by Biggs, who defines his phrases, both grammatical and phonological, by pause-bounding and silence (Biggs, 1961, 1969, 1973). Pauses break the flow of an utterance in the same way as the silence found before and after an utterance given in isolation, as in this task. The contrast with silence may increase the likelihood of prominence detection on the initial or final syllable. Since there are 30 utterances in the dataset, there are 60 edge-positions for syllables: 30 initial and 30 final. When the location of selected PPMs was analysed, any that appeared in these positions were counted to assess their contribution to the total: as mentioned in the description of Triplet Mode, the statistical process appeared particularly sensitive to edge-zeros.

It was mentioned in Section 4.2 that the utterances used in the survey spanned three speaker vintages, in order to allow for diachronic analysis, and also that this line of investigation had been discontinued. The following section contains a discussion of the preliminary analysis and explains why it was set aside.
4.8. Speaker Vintages

In order to assess the possibility of perceived change in Māori being the result of a shift in prominence (or their perception) over time, three different speaker vintages from the MAONZE database were used. As explained in Section 4.2, all speakers were male, and the three vintages were as follows: Historical Elders (HE; b. ca. 1880; rec. 1940s), Present-Day Elders (PE; b. ca. 1920; rec. 2000s), and Present-Day Young (PY; b. ca. 1970; rec. 2000s). Biggs’ (1969) description, based on speakers who would fall into the HE or PE categories, tells us that phrase stress (PS) syllables should be the most prominent.

We might therefore expect that if something has changed, PS syllables would be identified as a PPM more often in the speech of the older speakers (HE, PE) than the younger speakers (PY). In the previous experiment, described in Chapter Three, it was observed that there was no observable pattern in responses to older (PE) and younger (PY) speakers (§ 3.3.4.2). With a wider range of speaker vintages, and more stimuli from those vintages, it was possible that a difference might be observed. In fact, as discussed below, this was only true to a very limited extent.

This section will be concerned with the PPM response, in terms of their number and distribution across stress categories, from all of the participant groups across the speaker vintages, and what the observations may or may not allow us to speculate about change.

Table 4.12 on the next page shows how the groups responded to the different speaker vintages in terms of how many PPMs they identified. None of the differences between any of the totals here are significant. Group Z, who, given their number, their varied Māori language exposure and proficiency, and their own language experience, might be considered a general group of listeners, do not show a significant difference in rate of PPM identification for any of the vintages. Even when the listeners are divided into Māori proficiency groups, however, the result is effectively the same: no speaker vintage consistently inspires more or fewer PPMs to be identified than the others. This is also true when the totals are considered as a proportion of the available syllables from that vintage.
Table 4.12.

**PPM Selection by All Participant Groups for Three Speaker Vintages**

<table>
<thead>
<tr>
<th>Participant Group (participant)</th>
<th>Total PPMs</th>
<th>HE</th>
<th>PE</th>
<th>PY</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Z (92)</strong></td>
<td>148</td>
<td>48</td>
<td>44</td>
<td>51</td>
</tr>
<tr>
<td><strong>A (28)</strong></td>
<td>84</td>
<td>28</td>
<td>25</td>
<td>28</td>
</tr>
<tr>
<td><strong>B (30)</strong></td>
<td>155</td>
<td>50</td>
<td>50</td>
<td>47</td>
</tr>
<tr>
<td><strong>C (34)</strong></td>
<td>155</td>
<td>47</td>
<td>50</td>
<td>47</td>
</tr>
</tbody>
</table>

Looking one level deeper, into the distribution of Biggs stress category PPMs, the results are little different. Tables 4.13 to 4.16 below show the distribution of stress category PPMs for the different vintages and participant groups. There are minor patterns which suggest that the participants may be responding to a degree of hyper-correctness in speakers from the present-day elder (PE) vintage. Interestingly, this appears most obvious in the word stress (WS; Table 4.13) and non-expected stress (OS; Table 4.14) categories, which in itself suggests that those speakers could be making an effort to stress syllables that they know by rule belong to the WS category, and not to stress syllables they know by rule are not meant to bear it. However, in neither case is the difference in results from the other vintages significant.

Table 4.13.

**Distribution of Word Stress (WS) PPMs by Speaker Vintage and Participant Group**

<table>
<thead>
<tr>
<th>Speaker Vintage</th>
<th>Participant Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Z</td>
</tr>
<tr>
<td><strong>HE</strong></td>
<td>14</td>
</tr>
<tr>
<td><strong>PE</strong></td>
<td>17</td>
</tr>
<tr>
<td><strong>PY</strong></td>
<td>13</td>
</tr>
</tbody>
</table>

Table 4.14.

**Distribution of Non-Stress (OS) PPMs by Speaker Vintage and Participant Group**

<table>
<thead>
<tr>
<th>Speaker Vintage</th>
<th>Participant Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Z</td>
</tr>
<tr>
<td><strong>HE</strong></td>
<td>14</td>
</tr>
<tr>
<td><strong>PE</strong></td>
<td>8</td>
</tr>
<tr>
<td><strong>PY</strong></td>
<td>13</td>
</tr>
</tbody>
</table>

The results for WS (Table 4.13, above) and PWS (Table 4.15, below) contain no significant differences across the vintages or groups for either category. They rate similarly, at about 50% of
PPMs for all groups in each vintage. The PWS and WS results are, as will be echoed in the analysis in later chapters, indicative of the role of WS in bringing PS syllables out: the importance of overlap.

Table 4.15.

*Distribution of Phrase+Word Stress (PWS) PPMs by Speaker Vintage and Participant Group*

<table>
<thead>
<tr>
<th>Speaker Vintage</th>
<th>Participant Group</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>HE</td>
<td>Z</td>
<td>18</td>
<td>14</td>
<td>20</td>
</tr>
<tr>
<td>PE</td>
<td></td>
<td>17</td>
<td>8</td>
<td>19</td>
</tr>
<tr>
<td>PY</td>
<td></td>
<td>26</td>
<td>11</td>
<td>27</td>
</tr>
</tbody>
</table>

The truly interesting category in these results is that of the phrase stress only (PS) syllables (Table 4.16).

Table 4.16.

*Distribution of Phrase Stress (PS) PPMs by Speaker Vintage and Participant Group*

<table>
<thead>
<tr>
<th>Speaker Vintage</th>
<th>Participant Group</th>
<th>Z</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>HE</td>
<td></td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>PE</td>
<td></td>
<td>3</td>
<td>0</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>PY</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

While the PS PPM totals are low for all vintages and all groups, the present-day young (PY) vintage utterances elicited no PPMs at all from any group. This is significant for Group Z (two-sided Fisher’s exact test, $p = 0.03$) and also for Group A (two-sided Fisher’s exact test, $p = 0.004$). However, it is important to note that Group A also selected no PS PPMs from the PE vintage, which elsewhere (Group B and C) elicits the most, and selected only two from the HE vintage.

PS-only syllables, the smallest category in the dataset as a whole, have a low prominence selection rate in most all of the stress category analysis in this study, as was foreshadowed in Chapter Three (§ 3.3.4.3) and will become more apparent in later chapters. In light of this, it would be overstating the case to single out the PY vintage as particularly different.

If we were to extrapolate these results into the realm of pure speculation, however, there is a possibility that the young (PY) speakers may not be cueing PS-only syllables at all, or so slightly
that they are less likely to be heard as prominent, even by proficient speakers. The possibility would merit further investigation, but with a much larger dataset than the one in the present study, owing to the scarcity of PS-only syllables in general.

Although there are some small differences here, then, none of the observations can be said to have such certain bearing on participant response that the vintages need to be separated in subsequent analysis. Therefore, for the remainder of the discussion, the dataset will be treated as a whole.

4.9. Summary

This chapter has presented the fullest description of the prominence-elicitation survey tool developed over the course of the project, including the web-based perception test itself, the data processing, and the adjustments made to allow the methodology to cope with different behaviour from participant groups and produce results more consistent with participant intent.

The participants and their groupings, along with the stimuli used in the survey have been presented in this chapter. While the Triplet Mode method of validating perceived prominences (PPMs) was very effective for the results from Group Z (all of the participants together), the three proficiency groups (A, B, and C) exhibited different behaviour in the prominence identification task, with Group A more likely to check any syllable as prominent than the other two groups. This resulted in fewer significant peaks on the Group A PPM list after the statistical analysis (Triplet Mode) was applied. The probable cause of this different behaviour in Group A is that individual syllables stood out less to those higher proficiency participants because of their greater understanding of the content of the utterance. Group B and C, the lower proficiency groups, were both less likely, in the first instance, to check a box than Group A, which led to larger “spikes” in their data on standout syllables, and consequently more significant peaks when the statistical analysis was applied. Group B and C, without the advantage (or distraction, or blurring) provided by increased semantic ability, may have been more likely to grab onto acoustic cues when asked to identify prominences.

The different behaviour of Group A required a methodological modification. The higher floor, or reduced contrast, in their results caused the Triplet Mode statistical refinement to ignore some
peaks that should have been included, based on the high number of participants in Group A who actually checked the box for those syllables. This reduced the Group A PPM list to below a useful size, creating a problem with token numbers for further analysis and inaccurately rendering participant intent. The problem was addressed by increasing the unit used in the statistical tests from a triplet to a quinlet, or sequence of five syllables, which allowed more of the highly-selected but previously non-significant peaks to be included. This Quinlet Mode was applied across all three participant groups’ responses, resulting in PPM sets that were more readily comparable in size. A complete evaluation of the study, as well as some future directions for the project, may be found in Chapter Seven.

This chapter has also provided an outline of the characteristics of the PROM Dataset itself, as well as the main analysis areas of syllable weight and duration, Biggs stress categories, and F0 behaviour, to provide a point of comparison for the results in the coming chapters. The preliminary investigation into possible differences in response to different speaker vintages was also discussed, and it was found that despite small indications that phrase stress syllables are considered less prominent in the speech of young speakers, there is very little difference between the responses of any group to any speaker vintage, in any of the other stress categories. With this avenue of investigation set aside, we now move into the next chapter, and the results from Group Z, followed by the proficiency groups’ results in Chapter Six.
THE MAIN STUDY

RESULTS FROM ALL PARTICIPANTS

5.1. Introduction

This chapter presents the results from the participant set known as Group Z, which was comprised of all of the 92 participants from the study. It begins with an overview of the Group Z PPMs and their general characteristics. Following that are further sections covering in more detail aspects of the PPM distribution and acoustic characteristics, including alignment of PPMs with Biggs’ expected stress categories; duration; vowel quality; and the relationship of PPMs to elements of F0 behaviour such as the high pitch point and phrase contour.

5.2. Group Z PPM Profile: Overview

From the 636 syllables in the dataset, Group Z identified 214 peaks overall, three of which were pair-tops. From these 214 peaks, the statistical list refinement (Triplet Mode, § 4.5.3) allowed 145 as significantly prominent. This included all three Group Z pair-tops, meaning that the list of Group Z PPM syllables for analysis totalled 148 (or 23% of the syllables in the dataset). All of the characteristics in Table 5.1 on the next page, along with others, are discussed in more detail in the sections that follow.

---

1 When compared to the 14% rate observed under the previous definition of prominence used in the first web survey, this suggests that the question of potentially excluded prominences raised in that study was a valid one.
Table 5.1.

Summary of Group Z PPM and Non-PPM Characteristics

<table>
<thead>
<tr>
<th>Syllables</th>
<th>PPMs</th>
<th>Non-PPMs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>148</td>
<td>488</td>
</tr>
<tr>
<td>Syllable Weight</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heavy</td>
<td>83</td>
<td>85</td>
</tr>
<tr>
<td>Light</td>
<td>65</td>
<td>403</td>
</tr>
<tr>
<td>Syllable Type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Onset</td>
<td>139</td>
<td>360</td>
</tr>
<tr>
<td>No Onset</td>
<td>9</td>
<td>128</td>
</tr>
<tr>
<td>Biggs Stress</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PWS</td>
<td>60</td>
<td>54</td>
</tr>
<tr>
<td>PS</td>
<td>7</td>
<td>32</td>
</tr>
<tr>
<td>WS</td>
<td>45</td>
<td>39</td>
</tr>
<tr>
<td>OS</td>
<td>36</td>
<td>363</td>
</tr>
<tr>
<td>Phrase Contour</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fall</td>
<td>99</td>
<td>23</td>
</tr>
<tr>
<td>Rise</td>
<td>12</td>
<td>3</td>
</tr>
<tr>
<td>H* Events</td>
<td>54</td>
<td>83</td>
</tr>
</tbody>
</table>

The next three sections (§§ 5.3 – 5.5) will discuss aspects of the Group Z PPM distribution: their rate of selection in utterances and phrases, their relationship to pauses and edges, and their alignment with the Biggs expected stress category syllables.

5.3. PPM Rate of Selection in Utterances and Phrases

Group Z identified more than one PPM in each of the 30 utterances. The average was 5 PPMs per utterance, from a range of between 2 and 7. The distribution is shown in Table 5.2 below.

Table 5.2.

Group Z PPM Distribution Across Utterances

<table>
<thead>
<tr>
<th>PPMs per Utterance</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of Utterances</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>7</td>
<td>10*</td>
<td>4**</td>
<td>5</td>
<td>30</td>
</tr>
</tbody>
</table>

An asterisk (*) indicates that the number includes an utterance that contained one of the three pair-top PPMs (see § 4.5.3). The pair-tops must be counted as two PPMs each, because each syllable was identified separately by the participants.
With the exception of some longer stretches, there were generally about 2-3 syllables between each of Biggs’ expected stresses (including PWS, PS-only and WS-only). In an utterance with the average 21 syllables from this set, one might expect about 7 PPMs, if the rate of PPMs were consistent with the rate of any expected stresses. In fact, with an average of 5, the actual rate of PPMs per utterance was the same as the average number of expected phrase stresses (PS-Group) per utterance in the dataset. If the distribution were entirely even, this would mean, effectively, an average of one PPM per phrase. However, the reality of the distribution of PPMs across the phrases (Table 5.3 below) and the alignment with expected stresses (see § 4.7.2) is not so clean-cut.

Table 5.3. Group Z PPM Distribution Across Phrases

<table>
<thead>
<tr>
<th>PPMs per Phrase</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>33</td>
</tr>
<tr>
<td>1</td>
<td>86</td>
</tr>
<tr>
<td>2</td>
<td>22</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
</tr>
</tbody>
</table>

For these participants collectively, there was not a PPM in every grammatical phrase, but the most common number of PPMs per phrase was indeed one. The maximum was three. Some utterances contained up to three phrases with no PPM, and these were sometimes sequential, leaving a long gap where no PPM was identified. However, phrase length itself does not appear to be connected to lack of PPM: phrases without PPMs ranged from 2 to 7 syllables in length, representing almost the full range of possible phrase length. There was also no apparent pattern to the intervals between PPMs in or across utterances, either in millisecond duration or number of syllables.

Variation in utterance length, between 12 and 34 syllables and two and six seconds in duration, did not affect the results. The average number of PPMs per phrase is not different for phrases in shorter or longer utterances. The overall number of checks per utterance, when considered as a percentage of the number of possible checks for that utterance (i.e. the number of syllables in the utterance), is in fact no larger for the longest utterances than for the shortest, or vice versa. It sits at an average of 30% across all utterances, and the two utterances with the highest check ratings (around 40%), come from opposite ends of the length spectrum.
5.4. Edges and Pauses: PPM Relationship to Silence

There are, as described earlier, 20 pauses in the PROM Dataset. Group Z identified PPMs adjacent to 11 of those 20 pauses. Generally speaking, if a PPM was identified on a syllable adjacent to a pause, it was most often preceding that pause (7/11 cases here) and the pause was longer in duration than the PPM syllable (9/11 cases). However, owing to variation and the occurrence of multiple sets of similar circumstances where no PPM was identified, this cannot be called a rule. Neither is there any clear relationship between the length of a pause and whether a PPM was identified adjacent to it, preceded it, or followed it. Therefore, for the meantime, it appears that presence of pause does not have any obvious influence on PPM identification for these participants: if a pause occurs, a PPM is about as likely to be identified adjacent to it as not, and if this does occur, it is most likely to be under the circumstances above. The suggestion that PPMs are more likely before a pause (silence) is also given the lie by the fact that sentence-initial syllables (following edge-silence) are more likely to be identified as PPMs than sentence-final syllables (preceding edge-silence). More instances of pause (both with and without adjacent PPMs) than are available in this dataset would be needed in order to draw proper conclusions.

In Group Z’s PPM list, the PPMs with the highest significance levels included 26 utterance edge-adjacent syllables, 15 initial and 11 final. Because the statistics used to create the PPM lists (Triplet Mode, § 4.5.3) appeared sensitive to the presence of a zero value (no votes, or pre-initial/post-final placeholder) in a triplet, it was necessary to count these instances of edge-adjacent PPMs in the list to see if they made a significant contribution to the PPM total. In fact, they do not. From this result, it is possible to say that the apparent statistical sensitivity in the result processing did not actually allow an excessive number of this type of syllable, which is methodologically useful. Additionally, it can be assumed that the participants are not biased towards these syllables in PPM selection, meaning that they do not need to be analyzed separately for prominence-causing features, at least in the first instance.

5.5. PPM Alignment with Biggs Stress Categories

Biggs’ stress categories are, of course, not a prominence-causing feature in themselves. They are a construct based on Biggs’ impressions. These impressions were, in their turn, based on the
perceptual cues Biggs was using as a proficient Māori speaker with linguistic training, listening to a certain dataset. Belonging to a certain Biggs stress category is not what makes a PPM a PPM: being an expected phrase stress, for example, is not the inherent quality of a syllable that makes a listener hear it as prominent. However, since Biggs (1961, 1969, 1973) does use the term “prominent” in his description, and since one of the central questions in this project concerns whether or not there has been some change in the language, existing expectations regarding stress and prominence locations are an essential point of comparison when examining current PPM distribution. A high rate of alignment of Group Z PPMs with Biggs’ expected stress syllables might suggest that these participants were tuned to the same cues as he was.

When considering this particular alignment analysis, there are some points to keep in mind. Firstly, where Biggs used one corpus of recorded speech; this experiment uses entirely another dataset. Both datasets contain continuous speech, making the comparison valid, but where his original analysis and rules were based largely on older speakers or those contemporary to himself (see Chapter Two), the PROM Dataset contains younger speakers as well. Where expected alignment, or degree of alignment, between PPMs and expected stress syllables does not occur, it may be due to production differences between older and present-day Māori language. In the half-century since Biggs’ rules were formulated, something might well have changed, particularly owing to the influence of English.

While it was pointed out that stress category does not make a PPM a PPM, it is also possible that formal or informal knowledge of the stress categories could have affected which syllables were checked by some of the participants, particularly those with higher proficiency. For example, if one knows (by learning) that a syllable belongs to the phrase stress category, then one might be tempted to check the “correct” box. Additionally, even though the participants were told there were no right or wrong answers, test-like environments and language exercises in particular can sometimes produce undue focus on the desire to “get it right” rather than respond to what is heard, if these are perceived to be different. However, the influence of this effect in Group Z’s results would likely have been cancelled out by the lower proficiency level of the majority of the

---

2 Feedback from one or two participants (personal communication, 2012) indicated that they had, in fact, made a conscious effort to put this out of their minds.
participants, to whom the categories would not have been known at all. It is probably a safe assumption that any participant knowledge of the stress categories does not have an effect on the overall PPM results. See Chapter Six for some further support for this assumption.

Methodologically, the alignment of PPMs with Biggs’ expected phrase and word stresses was immediately examinable by eye, using the percentage-based charts generated by the result spreadsheet. However, using the levels for Stress and Perceived Prominence in the Emu database, the alignment of the stress categories with a number of other dataset units could be queried more efficiently, from PPMs to phonemes to H* events. The distribution and alignment results are presented below.

PPM distribution across the Biggs stress categories was examined in two ways. First, if perceived prominence and certain stress categories do align strongly for these listeners, more of their PPMs would be expected to belong to a Biggs stress category (PWS, PS, WS) than to the other category (OS). Of the 148 PPM syllables, 112, or 76%, were Biggs’ expected stresses of various types (that is, not OS). Figure 5.1 below shows the PPM distribution across the categories.

![Figure 5.1. PPM distribution across Biggs stress categories.](image)

The PS-Group (PS+PWS) might have been expected to dominate, owing to the descriptive connection made between phrase stress and greatest prominence. However, while the PWS category accounted for the largest number of PPMs (60/148), WS-only syllables (45/148) made up a much larger proportion of the results than PS-only syllables (7/148), meaning that the “WS-Group” (PWS+WS) was, in fact, the majority. This suggests that (a) the overlap syllables (PWS) carry the stronger cues, and (b) PPMs are most likely to be identified on WS syllables of any sort.
When looking at the same distribution across the 488 syllables that Group Z did not identify as prominent ("non-PPMs"), we would have expected to see more OS category syllables than any other, and this was the case (363 syllables, or 74% of the non-PPM set). There were, however, also 54 PWS, 32 PS-only, and 32 WS-only syllables that were not selected as PPMs.

This is relevant to the second way of looking at PPM distribution across the stress categories: the rate of selection of a given category out of all the available syllables of that category in the whole dataset. Again, it would have been expected that PS-Group syllables would be the most likely to be identified as PPMs out of their set, and OS syllables the least likely. As Table 5.4 below illustrates, only the prediction regarding OS syllables was correct.

Table 5.4.

Selection of Biggs Stress Category Syllables as PPMs Out of All Available in Category in Dataset

<table>
<thead>
<tr>
<th>Stress Category</th>
<th>PWS</th>
<th>PS</th>
<th>WS</th>
<th>OS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whole PROM Dataset</td>
<td>121</td>
<td>33</td>
<td>91</td>
<td>391</td>
</tr>
<tr>
<td>Group Z PPMs</td>
<td>60</td>
<td>7</td>
<td>45</td>
<td>36</td>
</tr>
<tr>
<td>PPMS as % of Available</td>
<td>50%</td>
<td>21%</td>
<td>49%</td>
<td>9%</td>
</tr>
</tbody>
</table>

The categories with the highest selection rates were, again, PWS and WS-only, and the rate of selection for PS-only was comparatively very low. The WS-Group accounted for 105/212 available, a rate of 50%, which was greater than that of PS-Group syllables (67/154 available, or 44%). Just as before, this suggests that word stress syllables of both types (PWS and WS-only) are more likely to be selected. The correlates related to one level of stress or prominence may be magnified by those related to another level, where they co-occur. Therefore, the rate of selection being higher for PWS syllables than PS-only, or even WS-only, is not surprising. What is more interesting is that the WS-only syllables had approximately the same rate of selection out of their set as the PWS syllables, which may mean that, for these participants at least, what causes the perceived prominence on WS syllables in general is itself a very strong cue or collection of cues, perhaps stronger than the cue(s) for phrase stress.
In the next sections (§§ 5.6 – 5.8), some of those possible cues, in the form of aspects of PPM duration, vowel quality and F0 behaviour, will be examined.

5.6. Duration

This part of the analysis was concerned with the rate of selection as PPMs of heavy and light syllables, whose classification is based on moraic weight and connected with nuclear duration (§ 4.7.1), and the possible influence of the presence or lack of an onset (whole syllable duration). The durations of any segments (units) labelled in an Emu database may be calculated from the start and end times. Events, such as the H* or vowel target, have no duration because they are single time points.

Biggs’ descriptions and rules associate heavy syllables with stress (see Chapter Two), so if there is alignment between what is described and what the participants perceive as prominent, then more heavy syllables (2-3μ) ought to be selected as PPMs than light syllables (1μ). In the whole PROM Dataset, there are more light syllables (468) than heavy (168). However, in the Group Z PPMs, more heavy syllables (83) were identified than light syllables (65), which confirms the prediction above. As in the whole dataset, the mean nuclear duration of Group Z’s heavy PPMs ($M = 131.0\text{ms}$, $SD = 53.6\text{ms}$) is significantly longer than that of their light PPMs ($M = 63.9\text{ms}$, $SD = 38.4\text{ms}$), $F(1,142) = 77.26$, $p < 0.0001$.

Since there is a heavy-light distinction, PPM syllable nuclei, or indeed whole syllable PPMs, might also be longer on average than their non-PPM counterparts. The PPM vowel categories (short monophthongs, long monophthongs, diphthongs and long diphthongs) behave, again, the same way as in the whole dataset, with the long diphthongs not durationally separable from the long vowels and diphthongs, and the short monophthongs ($M = 63.9\text{ms}$, $SD = 38.4\text{ms}$) significantly shorter than the long monophthongs ($M = 116.6\text{ms}$, $SD = 42.2\text{ms}$), $t(110) = 6.86$, $p = 0.0001$. This is also true of the non-PPMs: short ($M = 66.4\text{ms}$, $SD = 44.1\text{ms}$) and long ($M = 120.5\text{ms}$, $SD = 63.9\text{ms}$), $t(450) = 7.67$, $p < 0.0001$). However, despite the significant difference between long and short monophthongs, the difference between PPM and non-PPM nuclear
durations in either vowel category (short or long) was not significant. A table of the mean nuclear durations can be found in Appendix D.

In Table 5.5 below are whole syllable mean duration data. The difference in mean duration of the whole syllable between heavy and light, where heavy are longer, was highly significant for both PPMs, \( t(146) = 8.18, p < 0.0001 \), and non-PPMs, \( t(486) = 11.00, p < 0.0001 \). However, unlike nuclear duration, the difference in mean whole syllable duration between PPMs and non-PPMs was also significant in both the heavy category, \( t(166) = 2.66, p = 0.0087 \), and the light category, \( t(466) = 4.18, p < 0.0001 \). In both cases, the PPM syllables had greater mean duration than the non-PPMs.

Table 5.5.
Mean Whole Syllable Durations of Heavy and Light Syllables

<table>
<thead>
<tr>
<th>Syllable Type</th>
<th>PROM Dataset</th>
<th>PPMs</th>
<th>Non-PPMs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No.</td>
<td>Mean Dur. (ms)</td>
<td>SD (ms)</td>
</tr>
<tr>
<td>Heavy</td>
<td>168</td>
<td>203.0</td>
<td>73.0</td>
</tr>
<tr>
<td>Light</td>
<td>468</td>
<td>117.7</td>
<td>51.2</td>
</tr>
</tbody>
</table>

Since the presence of an onset causes greater mean whole syllable duration (and shorter mean nuclear duration, though this does not affect the heavy-light nuclear durational contrast), syllables with onsets might also account for a greater proportion of the PPMs than onsetless syllables.

Table 5.6.
Mean Whole Syllable Durations of Heavy and Light Syllables by Presence or Absence of Onset

<table>
<thead>
<tr>
<th>Syllable Type</th>
<th>PROM Dataset</th>
<th>PPMs</th>
<th>Non-PPMs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No.</td>
<td>Mean Dur. (ms)</td>
<td>SD (ms)</td>
</tr>
<tr>
<td>Heavy Onset</td>
<td>149</td>
<td>210.3</td>
<td>71.8</td>
</tr>
<tr>
<td>Heavy No Onset</td>
<td>19</td>
<td>145.6</td>
<td>56.6</td>
</tr>
<tr>
<td>Light Onset</td>
<td>350</td>
<td>128.8</td>
<td>41.1</td>
</tr>
<tr>
<td>Light No Onset</td>
<td>118</td>
<td>84.7</td>
<td>62.8</td>
</tr>
</tbody>
</table>

In the PROM Dataset, syllables with onsets are in the majority (Table 5.6). This was also true of the PPMs, but in the PPMs the percentage was greater: in the whole dataset, onset syllables
constitute 88% of heavy syllables and 75% of light syllables. In the PPMs, onset syllables accounted for 93% of heavy PPM syllables, and 94% of light PPM syllables. It appears that syllables with onsets were more likely to be selected as PPMs, and those that were selected had a greater whole syllable mean duration than those that were not.

All of this suggests that these listeners may be tuned to basic syllable weight, in that they did choose more heavy syllables than light, despite there being more light syllables in the dataset overall. However, they also appear to have taken longer duration of the whole syllable into account when identifying PPMs in either weight category. This may or may not be connected to the presence of onsets, since while these contribute to whole syllable duration, they do not alter the heavy/light distinction. The proportions of the dataset may also have affected the onset result.

When the PPMs’ mean nuclear durations were compared across Biggs stress categories, it was found that the PWS category had longer durations than the PS, WS or OS categories, all of whose durations were remarkably similar. This suggests that any connection between duration and the perceived prominence of expected stress syllables is additive: only where expected phrase and word stress coincide do the PPMs exhibit the extra length. However, ultimately, there was no significant difference between any of the mean durations across the categories.

Durational contrasts can also be connected to vowel quality, in that greater duration in a vowel allows more time to reach and sustain the vowel target, and therefore fuller articulation. This, in turn, can be visible as greater peripherality in the vowel space. This will be explored in the next section, on vowel quality.

5.7. Vowel Quality

This part of the PPM profile examined the PPMs in terms of their positions in the vowel space, and compared that with the vowel space of non-PPMs. As part of the database creation process, the formant tracks were calculated for each utterance, and vowel targets were marked on the PPM syllables, as described earlier in the chapter. Using the Emu-R function `eplot()`, the vowel targets were used to create plots of the F1-F2 vowel space.
The vowel spaces presented here compare only the monophthongs, long and short. Each vowel label that appears on the plots is a centroid, representative of all of the tokens of that vowel in the set concerned. It is possible to plot all of the points individually, but to do so in this case would make the plot illegible given the numbers involved. The centroid is a two-dimensional point, obtained from the mean F1 and mean F2 value, and like any average, by itself, it does not give a sense of the variance. This can be caused by incorrect automatic formant tracking, which is manually alterable in Emu, as well as phonetic differences that may be caused by anything from coarticulation effects or speaker error to regional variation or physiology. A further cause may be the difference between tokens from read and continuous speech. Any outliers must be checked, and altered or removed where necessary, before conclusions can be drawn from the vowel space shown. This was done for the PROM Dataset when the vowel space plots were created for analysis here.

5.7.1. Vowel Spaces: MAONZE Database vs. PROM Dataset

The first part of the vowel space analysis compared a set of monophthongs in the PROM Dataset (both PPMs and non-PPMs) with a set of tokens from the MAONZE database. The MAONZE set is comparable to PROM in that it includes male speakers from the same three vintages, but it is on a much larger scale. There are 26 speakers: 11 historical elders, 10 present-day elders, and 5 young L1 speakers. The MAONZE set includes 8132 monophthong tokens.

The vowel tokens identified in the MAONZE set are different from the PROM set in that they include only the tokens that fall into the Biggs stress categories referred to in the previous sections as the PS-Group (PWS and PS) and WS-only. They are not based on perception test results; not prominences identified externally by a group of participants: they are vowels from syllables identified as stressed purely on the basis of Biggs’ rules. Other tokens (those which fall into the category known in this project as OS), are not included, because their targets were not marked in the original database. For this reason, the first two vowel space comparisons here were between (a) the MAONZE set and the equivalent PS-Group and WS tokens from the whole PROM set,

---

3 It is also possible to plot diphthong trajectories, but owing to the relatively small number of diphthong tokens in the PROM Dataset, and the even smaller number in the Group Z PPM set, a diphthong analysis was not included here. Low token numbers can make the vowel space plots unreliable; the reasons for this will be explained later in the present section.

4 For more detailed comment on this subject, the reader is referred to Chapter Six of Harrington (2010).
and (b) the MAONZE set and the PROM PPM tokens. This was to see how the PROM set compared to a much larger set (that is, whether or not PROM was representative of the larger set from which it was taken), and to find out whether the participant-identified prominences in PROM were different from a very large set of theoretically stressed syllables, which, owing to its size, ought to provide a very stable point of reference. The number of vowel tokens in each of the sets is shown in Table 5.7 below. The first two plots are presented in the following figures.

### Table 5.7.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Total Tokens</th>
<th>Short Monophthongs</th>
<th>Long Monophthongs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>a  e  i  o  u  a: e: i: o: u:</td>
<td></td>
</tr>
<tr>
<td>MAONZE</td>
<td>8132</td>
<td>925 810 880 839 830</td>
<td>926 854 659 794 615</td>
</tr>
<tr>
<td>PROM All</td>
<td>564</td>
<td>191 68 109 39 61</td>
<td>58 11 6 13 8</td>
</tr>
<tr>
<td>PROM PS-Group+WS</td>
<td>188 62 9 16 15 30</td>
<td>25 11 5 8 7</td>
<td></td>
</tr>
<tr>
<td>Z-PPM</td>
<td>112</td>
<td>30 6 7 7 15</td>
<td>22 7 5 6 7</td>
</tr>
<tr>
<td>Z Non-PPM</td>
<td>452</td>
<td>161 62 102 32 46</td>
<td>36 4 1 7 1</td>
</tr>
</tbody>
</table>

Figure 5.2 shows the MAONZE set of stressed monophthongs (PS-Group and WS) in a comparison plot with the equivalent PS-Group and WS syllables from the PROM Dataset.

![Figure 5.2. PS-Group+WS monophthongs: PROM vs. MAONZE.](image)

The groupings for each vowel here are, in most cases, fairly tight, which means that the PROM Dataset can be called representative of the larger set from which it comes. While there is no real
peripherality difference between the two sets, it is possible to see a slight difference within both sets in some of the long-short vowel pairs: for example, /i - i:/, /a - a:/ and /o - o:/. The long vowels are the more peripheral, which is what would be expected given the extra time to reach the vowel targets.

The same is true in the second plot (Figure 5.3 below), which compares the same MAONZE set with the PROM PPMs identified by Group Z.

![Figure 5.3. PROM Group Z PPM monophthongs vs. MAONZE PS-Group+WS monophthongs.](image)

The two plots are almost identical, suggesting that the PROM PPMs are, on average, little different in vowel quality from a large set of syllables allocated by rule to the Biggs stress categories, and little different from the equivalent PROM syllables (with which, it must be noted, some of the PPMs would overlap: recall the (collective) dominance of PS-Group and WS-only category syllables in the PROM PPM set that was seen in Section 5.5 above).

There is only one obvious difference between the two plots, and it concerns the vowel pair /a - a:/.

The long and short versions are further apart for the PROM PPMs here than they were for the PROM PS-Group+WS set in the first plot, owing to the long version being very slightly higher in F1 and therefore more peripheral. This is worth noting, despite the difference being small, because /a - a:/, being the most frequently-occurring vowels in Māori and having a high functional load, are also the best represented here in terms of token numbers, making them the most reliable centroids.
5.7.2. PPM vs. Non-PPM Vowel Space Comparisons

The next comparison is specific to the PPM monophthongs and their non-PPM counterparts. Figure 5.4 below compares the Group Z PPMs with the non-PPMs. For the token numbers relevant here, see the last two rows of Table 5.7 above. In this plot, the difference in number of tokens between the two categories (PPM and non-PPM) is especially wide, and some of the numbers involved are quite low, particularly for the long vowels. This can affect the centroids, and as before, /a - a:/ are the most reliable owing to the greater numbers.

Figure 5.4. Vowel space comparison of Group Z PPMs and non-PPMs.

Visually, overall, there are differences in most of the long-short vowel pairs, with the long vowels more peripheral. This is supported by a significant long-short mean durational difference (§ 5.6) and is therefore not an unexpected result. However, what is more important to note is that the differences in formant values within these vowel pairs are not significant. This applies to both PPMs and non-PPMs, so it cannot be said that this contrast is relevant for PPM selection from the data available here.

Additionally, in comparing the overall PPM space (black) with the non-PPM space (red), there is a visual suggestion of greater centrality in the non-PPMs, which is entirely logical given the much higher proportion of light syllables included in that set. However, an ANOVA showed that there is, in fact, no significant difference in the formant values between PPMs and non-PPMs. While one reason for this might be the large discrepancy in token numbers between the two categories,
even where the difference appears clearest, in the most reliable tokens, /a – a:/, it is not significant. There was also no significant durational distinction between PPMs and non-PPMs in either the short or long vowel category.

From these data, then, owing to the size of the set, it is not possible to draw strong conclusions about vowel quality differences between PPMs and non-PPMs as a cue for these participants. It is also not possible to filter the PPMs into smaller categories for comparison, such as Biggs stress. This is a very tricky variable for which to correct: it is impossible to predict exactly which syllables participants in a test such as this will choose as PPMs, and equally impossible to predict which of those syllables will then be statistically significant. The best potential solution would be to increase the pool of possible PPM syllables by using a larger dataset, with an attempt to balance the vowel tokens in that stimulus set. However, in a language such as Māori, where there is a high occurrence of a certain vowel, it still may not be possible to even out the token set appropriately. There are some indications, even if not significant, that the categories in the dataset investigated here (long vs. short; PPM vs. non-PPM) do display some of the expected peripherality differences. Where they do or do not, there is a connection with a durational distinction, or lack thereof.

5.8. F0 Behaviour

The final part of the PPM profile included in the present study concerns selected aspects of the behaviour of F0 and its relationship to PPM identification. The points for discussion will be phrase contour, H* events, and the connection between H*, PPMs and Biggs stress categories.

5.8.1. PPM Selection by Phrase Contour Category

The main question concerning phrase contour categories is how they relate to the identification of PPMs. There are three types of phrase in the dataset, according to contour (or lack thereof): 122 fall, 15 rise and 10 minimal. Overall, 114 of the 147 phrases in the dataset contained PPMs. As discussed earlier, there is a relative lack of F0 contrast in the minimal type phrases, whose mean F0 range was found to be significantly smaller than the mean F0 ranges of the other phrase types (§ 4.7.3). Since prominence is relative, if these participants were using contrast in F0 as a cue in prominence identification, it would be expected that fewer PPMs would be identified in
these minimal type phrases. This was indeed the case, and significantly so (two-sided Fisher's exact, $p = 0.002$). One or more PPMs were identified in 81% of the fall type phrases and 80% of the rise type phrases. Of the minimal type phrases, 33% contained a PPM, but never more than one. Table 5.8 below illustrates the distribution of PPMs across phrase contour types.

Table 5.8.

<table>
<thead>
<tr>
<th>Contour Category</th>
<th>Total Phrases ($x/147$)</th>
<th>Phrases with PPM(s) ($x/114$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall</td>
<td>122</td>
<td>99</td>
</tr>
<tr>
<td>Rise</td>
<td>15</td>
<td>12</td>
</tr>
<tr>
<td>Minimal</td>
<td>10</td>
<td>3</td>
</tr>
</tbody>
</table>

5.8.2. Relationship of PPMs to H* Events

The H* is the main point of comparison for the PPMs, owing to the descriptions associating the high pitch point with the most prominent part of a phrase. An H* event was identified in the 137 fall and rise type phrases in the dataset, but not in the 10 minimal type phrases.

It is possible to take a mean F0 measurement for a syllable unit, just as for a phrase unit. However, because this mean is obtained from several readings across the unit, once a group of units (for example, the PPM syllables) is involved, the mean F0 for that group is in fact a mean of means (here referred to as average mean), which is less reliable. In addition, across a whole unit of any type, there is a good deal of potential for F0 fluctuation that can affect the mean. For example, this might be caused by the intrinsic F0 of the vowels concerned; consonantal to vocalic proportions within the unit (relating to whether or not there is F0 to measure at a given time point); or segmental effects that adjacent consonants can have on vowel F0.

The marked H*s in this study are, as mentioned previously, events rather than units: their F0 is taken as a single reading. This means that, firstly, the mean of a set of H* events, such as those that were found in syllables that were PPMs, is a single mean, not an average mean. Secondly, since the H* events were marked manually, it was possible to “correct” for some of the variables described above that affect mean F0 across a larger unit. When searching for the true highest
pitch point in a phrase, the H*s are a more reliable measurement or point of contrast than a syllable average mean F0. However, given the relativity involved in prominence, it is also more likely that the contrast will be noticed in a larger context than in fluctuation within a single syllable. That is, if (higher) F0, F0 movement, or any F0 contrast is a cue to prominence, it is more likely that a syllable will be noticed and selected as a PPM because it contains the high pitch point, rather than that the high pitch point will be noticed within the syllable.

Earlier, it was observed that the mean F0 for all H* events in the dataset is higher than the average mean F0 of the phrases in which they occur, and that this is significant in the fall and rise type phrases (§ 4.7.3). Therefore, contrast does exist. It was also observed that according to the literature, H* events should coincide with greatest perceived prominence. Based on these points, we would expect to find PPMs on the majority of the 137 H*-bearing syllables in the dataset. However, there were only 54 syllables in the dataset where PPMs coincided with the H*: 39%. Additionally, when the mean F0 of the H* events on PPM syllables was compared with the mean F0 of H* events on non-PPM syllables, while the PPM H* mean was slightly higher (see Table 5.9), the difference was not significant.

Table 5.9.

<table>
<thead>
<tr>
<th></th>
<th>No. of Syllables</th>
<th>Mean F0 (Hz)</th>
<th>SD (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPM H*s</td>
<td>54</td>
<td>148.7</td>
<td>38.8</td>
</tr>
<tr>
<td>Non-PPM H*s</td>
<td>83</td>
<td>140.7</td>
<td>35.8</td>
</tr>
</tbody>
</table>

Thus, even though PPMs were found to be proportionally more common in fall or rise type phrases, and syllables in these phrases do contain significant contrast in pitch, at first, there does not seem to be a corresponding strong relation of PPM to H*.

Since prominence may be perceived where there is movement of F0 rather than an absolute value (see Chapter Two), the locations, relative to H*, of the 83 cases where PPMs were not identified on the H* syllable were investigated. The result was that 43% (36/83) of these non-H* PPMs were directly adjacent to the H* syllable. If the “on” (54) and “adjacent” (36) results are then taken together, 90/137 H*s were directly connected in some way with a PPM (66%). This is a
better rate, and more consistent with expectations. Further to this, of the H*-adjacent PPMs, 69% (25/36) preceded the H*. This makes an interesting suggestion: given the roughly 2:1 ratio of precede-H* to follow-H*, it may often be the rise towards the H* (and the H* itself) that the participants notice: a movement rather than a single contrasting value.

There is a further prediction that not only will H*s and perceived prominence align, but that these will also coincide with the phrase stress (here, the PS-Group syllables). However, a common thread in the results so far has been that the WS category of syllables, whether as WS-only or the WS-Group (PWS and WS), has a high selection rate – higher than the PS category. The pattern of distribution of the 54 H*-PPMs across the stress categories is as shown in Figure 5.5 below.

![Figure 5.5. Distribution of H*-PPMs by Biggs stress category.](image)

The PWS category was the largest, and the WS category accounted for more H* PPMs than the PS category, as in previous results, though the margin is smaller here. As a proportion of the available OS H* syllables, those selected as PPMs accounted for only 16%. For PS, this was 29%; for WS, 62%; and for PWS, 56%. If layering of cues is indeed important, then it is not surprising that the OS syllable rate is low. Add that result to the fact that PS also has a comparatively low rate, and it appears that F0 alone is unlikely to be responsible for selection of a PPM: these are the two categories which are least likely to have the support of syllable weight and duration.

Put together, the preceding results indicate that there is a connection between the H* events and the PPMs selected by these participants. The impression is that when PPMs are selected by these participants, higher F0 (and certainly contrast in F0) play a role, but these may not always necessarily be directly linked with the H* event in a phrase; rather, they may be part of movement
around the H*. However, the interaction of the H* PPMs with stress categories suggests that F0 is not the only cue in use, and not very strong as a cue alone.

5.9. Summary

The PROM Dataset, containing the stimuli used in this phase of the study, has been introduced. The results from the entire group of 92 participants have been described, as well as the analysis approaches taken where the perceived prominences (PPMs) were concerned. These included examination of both general characteristics and distribution, and selected acoustic characteristics. The main points may be summarised as follows.

The PROM Dataset contained 636 syllables that were candidates for selection as PPMs by the participants. These syllables were distributed across 30 utterances containing a total of 147 grammatical phrases. Distributionally, the characteristics of the PPMs identified by Group Z in the dataset were as follows. There were 148 significant PPMs in total, once the statistical list refinement (Triplet Mode) had been applied: at least two in every utterance, at an average of 5 from a range of 2 to 7.

PPMs were not identified in all of the 147 phrases. The average number of PPMs per phrase was 1, from a range of 0 to 3. Selection of PPMs by Group Z participants does not appear to have been strongly influenced by the presence of pauses and initial or final silence (edges) in the stimuli. There was also no observable pattern to the intervals between PPMs, either in millisecond duration or number of syllables.

Examination of the alignment of PPM syllables with Biggs’ expected stress category syllables showed the expected result: more of the PPMs belonged to the PWS, WS or PS categories than to the OS category. It was also clear that WS-Group syllables (PWS and WS) were more likely to be selected out of their respective sets than other categories, and together account for the largest proportion of PPMs, suggesting that there is a strong connection between the features of the word stress category and PPMs. Stress category itself is not likely to be a factor influencing selection of syllables as PPMs unless some participants were aware of the classification and were
choosing accordingly. This scenario would probably only have been possible for about a third of the participants.

Acoustically, the PPMs were measured for duration, vowel quality and F0 behaviour. The selection as PPMs of heavy syllables was more likely than light, and the difference in mean nuclear duration between the weight categories was highly significant. Mean nuclear duration was not significantly different between PPMs and non-PPMs; however, mean whole syllable duration was. Far more syllables with onsets were selected than syllables without, though this will most likely have been a reflection of the dataset proportions. The suggestion is that participants are sensitive to durational contrast in the form of general syllable weight (that is, heavy or light), and greater whole syllable duration in their perception of prominence. This may partially account for the high selection rate of word stress group (WS and PWS) syllables described above, as the rule for the designation of WS is in favour of heavy syllables. The selection rate of heavy syllables in those categories as PPMs was much greater than that of light syllables.

In terms of vowel quality, using F1-F2 vowel space plots of the long and short monophthongs in the dataset, it was shown that the PROM Dataset is representative of the much larger MAONZE set from which it is taken. The vowel spaces of both the PROM Dataset PS-Group and WS tokens and the Group Z PPM tokens were particularly closely aligned with the MAONZE space, which is also based on tokens that align with Biggs' expected stress syllables (PS, PWS and WS). When the PPMs and non-PPMs were compared, their vowel spaces exhibited, visually, some of the predicted peripherality differences, where the PPMs were more peripheral. However, the difference in formant values between the PPM and non-PPM vowels was not significant, nor was it supported by a significant difference in nuclear duration. The analysis here did suffer from low token numbers; attempts can be made to correct this, but not within the scope of the present study. For that reason, analyses that filtered the results further, such as stress category, were not explored. It is not possible to say, from these results, that vowel quality is a cue for these participants. However, there is potential for stronger results with more data.

Where F0 behaviour was concerned, the rate of selection of PPMs in various phrase contour types showed that there were significantly fewer PPMs in minimal type phrases than in fall or rise type
phrases, likely due to reduced contrast in F0 in the latter. The basic alignment of PPMs with syllables containing H* events was relatively low; however, when PPMs on and adjacent to the H* were taken together, there was a suggestion that movement and higher F0 around the H* may be a cue for the participants. As far as the connection between H*, PPMs, and stress categories is concerned, the expected link between the PS-Group and H* is, in fact, fractionally exceeded by the link between the WS-Group and H*, providing further evidence of the prominence of word stress. The comparatively low rates of selection of H* PPMs out of those available in the PS and OS categories, which are less likely to have durational support, strongly suggest that F0 alone is not a strong cue for PPM selection.

With these results based on all of the participants in hand, the question remains of whether or not dividing the participants by their Māori proficiency level will make a difference to any of the findings. Do the different groups identify PPMs in the same ways, and at the same rate? We have already seen, in Chapter Three, some possible indications that they do not. An exploration of these questions can be found in Chapter Six.
6

MĀORI LANGUAGE PROFICIENCY AND PROMINENCE IDENTIFICATION

6.1. Introduction

In the analysis of the results from the first web survey, discussed in Chapter Three (Section 3.3.4.4), there was an examination of whether or not participants’ familiarity with the Māori language affected their responses to the stimuli. It was found that higher proficiency participants appeared to check a greater number of boxes in the task overall, and more of the prominences they identified were on expected stress syllables (particularly PWS (phrase and word stress) and WS (word stress only) syllables) than in the results from the lower proficiency groups.

The number of participants in the first web survey was relatively low, at only 27 people, and the proficiency groups into which they were divided were not even in size. The results also raised questions about the definition of prominence used: while there were clearly differences between the groups, the “60% threshold” had to be reduced to 50% in order to make many of those differences visible. This made drawing proficiency-related conclusions difficult based on those results.

The main study, described in Chapter Four, and some results from which have already been presented in Chapter Five, had many more participants and took a different, statistical approach to prominence validation, which eliminated some of those difficulties. However, as we saw in Chapter Four (Section 4.5.4), the three proficiency groups also behaved differently in this perception task, in the same way as before. Methodologically, the result processing method
Triplet Mode) that had been developed for Group Z did not handle this efficiently, and a modification was made to it, in the form of Quinlet Mode (§ 4.5.6), to address the problem.

This chapter presents the results of the investigation into the sets of perceived prominences (PPMs) generated by Quinlet Mode for the three proficiency groups. The central questions in this investigation were whether any of the proficiency groups would identify more or fewer PPMs, and whether or not the PPMs they identified would share characteristics. These characteristics, both distributional and acoustic, will be presented in Sections 6.4 to 6.6, and a summary of the proficiency split sub-study can be found in Section 6.7.

6.2. Participants: A Review of the Three Proficiency Groups

In New Zealand, participants with some exposure to the Māori language are easily found, while highly proficient Māori speakers are more difficult to find (see Chapter Two, Section 2.2). Low or zero proficiency speakers were found overseas. A particular effort was made to gain participants from both ends of the proficiency spectrum. As a result, the 92 participants (“Group Z”) in the study could be divided into three relatively even subgroups based on their Māori proficiency level, which was self-rated. The proficiency rating scale (of conversational ability) had six points, ranging from “very well” to “not at all”. For the detail of the group divisions, see Chapter Four, Section 4.3; a short recap is provided here. The distribution of participants in each group may be found in Table 6.1 below, followed by an explanation of the reasoning behind the divisions.

Table 6.1.

Breakdown of Participant Numbers in Māori Proficiency Groups

<table>
<thead>
<tr>
<th>Proficiency Group</th>
<th>Group Z “All”</th>
<th>Group A “High”</th>
<th>Group B “Exposed”</th>
<th>Group C “Zero”</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of Participants</td>
<td>92</td>
<td>28</td>
<td>30</td>
<td>34</td>
</tr>
</tbody>
</table>

Group Z, containing all of the participants, was covered in Chapters Four and Five. Group A, the high proficiency group, contained the 28 participants who rated their Māori ability in the top four categories: “very well”, “fairly well”, “well”, and “not very well”. This was taken to mean that
they felt they could produce at least a sentence, as contrasted with Group B, the “exposed” group (30 participants), who rated their Māori proficiency at the “words and phrases” level. This would be the expected level for most New Zealand residents who have not actively studied or learned Māori. They are regularly exposed to the language, have some parsing ability and may use or recognise some lexical items in daily life, but have minimal to no knowledge of the grammar. They would be very unlikely to hold a full conversation in Māori. Group C, the “zero” proficiency group, included the 34 participants who rated their day-to-day Māori ability as non-existent (“not at all”). Some of the overseas participants had neither heard, nor heard of, the Māori language before. Others knew of the language but had very little or no exposure.

As we will see in the remainder of this chapter, these three groups responded differently to the perception task, and their results do show some differences that could be attributed to proficiency-related factors. The main point of difference between the proficiency groups is their behaviour in the perception task: Group A were more likely to check any syllable as prominent than the other two groups. This might be for a number of reasons, but most likely either their proficiency or their interpretation of the instructions. The Quinlet Mode adjustment in the result processing was made to cope with this, but there are places in which it may have had some small bearing on the results. These instances are discussed in the text.

The rest of this chapter, is modelled after the description of the Group Z results in Chapter Five, and presents the results of the empirical and acoustic analysis carried out on the three groups’ PPM lists, looking for similarities and contrasts in the nature of the PPM sets. It is divided broadly into three sections. First, some general characteristics of the groups’ PPMs are presented (§ 6.3). The next section (§ 6.4) includes observations about how the PPMs are distributed across the utterances: their rate of selection; their position relative to elements such as silence; and their alignment with Biggs’ expected stress locations. The final section discusses some acoustic characteristics of the PPMs in more detail: duration (including syllable weight), vowel quality, and relationship to F0 and the marked H* (§ 6.5).
6.3. The Proficiency Group PPMs: General Characteristics

6.3.1. Overview

A summary of the PPMs from the three different proficiency groups is presented in Table 6.2 below. All of these characteristics will be discussed further later in the chapter.

<table>
<thead>
<tr>
<th></th>
<th>Group A PPMs</th>
<th>Group B PPMs</th>
<th>Group C PPMs</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of PPM Syllables</td>
<td>84</td>
<td>155</td>
<td>155</td>
</tr>
<tr>
<td>Weight</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heavy</td>
<td>52</td>
<td>74</td>
<td>76</td>
</tr>
<tr>
<td>Light</td>
<td>32</td>
<td>81</td>
<td>79</td>
</tr>
<tr>
<td>Type</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Onset</td>
<td>76</td>
<td>148</td>
<td>144</td>
</tr>
<tr>
<td>No Onset</td>
<td>8</td>
<td>7</td>
<td>11</td>
</tr>
<tr>
<td>Biggs Stress</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PWS</td>
<td>33</td>
<td>66</td>
<td>64</td>
</tr>
<tr>
<td>PS</td>
<td>2</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>WS</td>
<td>27</td>
<td>44</td>
<td>41</td>
</tr>
<tr>
<td>OS</td>
<td>22</td>
<td>36</td>
<td>42</td>
</tr>
<tr>
<td>H*-PPMs</td>
<td>26</td>
<td>62</td>
<td>60</td>
</tr>
</tbody>
</table>

Even here, we can see that the three groups are similar in some areas and different in others. Firstly, as discussed above (and in detail in Chapter Four), Group A had a lower overall number of PPMs than the other groups. They also selected obviously fewer light syllable PPMs than Group B and C. However, for all groups, syllables with onsets were much more likely to be selected than those without, and the response to the different stress category syllables was much the same. The differences between the groups were not always as wide as might have been expected, and where they were present, they were sometimes not significant. Statistical assessment of these results was as for the Group Z results in the previous chapter: detail is given in Chapter Four, Section 4.6.3). The alpha level was 0.05.

6.3.2. Overlap in PPM Lists

The three groups’ different responses to the perception task and the ensuing PPM list refinement process created three separate PPM lists. An important point to keep in mind when reading the
following sections and observations on the different groups’ responses is that while the PPM lists are based on exclusive groups of participants, some of the syllables in the lists do overlap. That is, for Groups A, B and C, the participants who selected the PPMs for each list did not overlap, but they did select (as three groups) some of the same syllables as PPMs. The tables below illustrate the nature of the overlap.

Table 6.3.

Overlapping PPMs by Participant Group

<table>
<thead>
<tr>
<th>Proficiency Groups</th>
<th>Shared PPMs</th>
</tr>
</thead>
<tbody>
<tr>
<td>A + B + C</td>
<td>56</td>
</tr>
<tr>
<td>A + B</td>
<td>5</td>
</tr>
<tr>
<td>A + C</td>
<td>5</td>
</tr>
<tr>
<td>B + C</td>
<td>64</td>
</tr>
</tbody>
</table>

As the top row of Table 6.3 shows, there are 56 syllables that appear in the PPM lists of each proficiency group individually. However, Group B and C clearly have many more PPMs in common with each other than Group A does with either of them on their own. This parallels the language proficiency situation: Group B and C are closer to each other in proficiency than Group A is to either of them. The observations here also parallel the test behaviour response: A is different from B and C. It should also be noted that the Group A, B and C PPM lists are not subsets of the 148 Group Z PPMs analyzed in Chapter Four. Even though the responses of the participants in the subgroups collectively contributed to Group Z’s (everybody’s) PPM results, when A, B and C were considered as separate subgroups, the responses were different, and there are some PPMs that belong to each group alone. The A vs. B/C trend is also visible in the following table (6.4), which shows the PPMs belonging exclusively to each group.

Table 6.4.

Exclusive PPMs by Participant Group

<table>
<thead>
<tr>
<th>Proficiency Group</th>
<th>Exclusive PPMs</th>
<th>Total PPMs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group A</td>
<td>18</td>
<td>84</td>
</tr>
<tr>
<td>Group B</td>
<td>30</td>
<td>155</td>
</tr>
<tr>
<td>Group C</td>
<td>30</td>
<td>155</td>
</tr>
</tbody>
</table>
None of the proficiency groups has significantly more or fewer exclusive PPMs than the others. When the actual numbers of exclusive PPMs are considered as a subset of the total PPMs identified by any group, they are relatively low, at about 20%, which shows a degree of consensus across the groups.

Looking briefly at the characteristics of the exclusive PPMs for Groups A, B, and C, some patterns appear. For all groups, most of the exclusive PPM syllables have onsets. There are more light syllables than heavy, but Group A has closer to an even number of each, while Group B and C have about three times as many light as heavy. Very few of the exclusive PPMs are found on or adjacent to an H*. Non-expected stress (OS) syllables make up the majority of the exclusive PPMs for all groups, but not significantly so. Most of the exclusive PPMs are medial rather than near edges (sentence-initial or -final syllables). The patterns (and lack thereof) observed in the overlap and exclusive PPM values and characteristics, then, do suggest that proficiency affects PPM selection in some way, but for these participants, the effect may not be as great as expected. The rest of this chapter will explore the extent of the patterns in these data, beginning with the distribution of PPMs for the different groups.

6.4. PPM Distribution: Rate of Selection in Utterances and Phrases

In Chapter Five (Section 5.3), we saw that Group Z (all participants) identified an average of 5 PPMs per utterance. This was approximately equivalent to the average number of expected phrase stresses per utterance in the dataset, even if the PPMs did not always align with the phrase stress syllables. The results for the three proficiency groups are somewhat different. Group A selected, on average, 2 to 3 PPMs per utterance, while Group B and C selected an average of 5 PPMs per utterance. Group A were inclined to select fewer overall: as a unit, they did not select more than 6 PPMs per utterance across the set of stimuli, while Group B and C allowed more, selecting up to 8 (C) or 12 (B) PPMs in some utterances. The difference is significant between Group A and B (A: $M = 2.8$, $SD = 1.1$; B: $M = 5.1$, $SD = 2.0$), $t(58) = 5.55$, $p < 0.0001$), and between Group A and C (A: $M = 2.8$, $SD = 1.1$; C: $M = 5.2$, $SD = 1.7$), $t(58) = 6.42$, $p < 0.0001$). The difference between Group B and C is not significant. This grouping of A vs. B/C is consistent with other results. The distribution is shown in Table 6.5 below.
Table 6.5.

*Group A, B and C Rate of Selection: PPMs per Utterance*

<table>
<thead>
<tr>
<th>Prof. Group</th>
<th>Total PPMs</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7+</th>
<th>Total Utts.</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>84</td>
<td>0</td>
<td>1</td>
<td>13</td>
<td>10</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>155</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>10</td>
<td>5</td>
<td>7</td>
<td>4</td>
<td>30</td>
</tr>
<tr>
<td>C</td>
<td>155</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>6</td>
<td>3</td>
<td>7</td>
<td>4</td>
<td>9</td>
<td></td>
</tr>
</tbody>
</table>

Because Group A were more likely to check syllables overall in the task, as we saw in the earlier discussion, their PPM list is ultimately smaller than those of Group B or C. Consequently, they often have fewer PPMs in any result category. Here, this is shown by their average number of PPMs per utterance being much lower than those of Group B and Group C. The pattern is visible again in the rate of selection of PPMs per phrase.

The Group Z identification rate was an average of one PPM per phrase, and the phrases in which they identified one or more PPMs accounted for 78% of the 147 phrases in the dataset. For the three proficiency groups, the average is also one PPM per phrase, though the spread in the table (6.6, below) shows that there was a difference in the groups’ behaviour. Again, Group A were less likely to have one or more PPMs in a phrase, selecting a maximum of three in any phrase in the dataset. Of the 147 phrases available, those containing Group A PPMs accounted for 46%. Group B and C were different, selecting zero or one in the majority of phrases, but allowing up to four (B) or five (C) in some cases.

Table 6.6.

*Group A, B and C Rate of Selection: PPMs per Phrase*

<table>
<thead>
<tr>
<th>Proficiency Group</th>
<th>Total PPMs</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3+</th>
<th>Total Phrases in Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>84</td>
<td>80</td>
<td>53</td>
<td>13</td>
<td>1</td>
<td>147</td>
</tr>
<tr>
<td>B</td>
<td>155</td>
<td>40</td>
<td>67</td>
<td>31</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>155</td>
<td>40</td>
<td>70</td>
<td>30</td>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>

PPM-containing phrases accounted for 73% of the 147 available for both groups. Once again, this is the effect of the higher floor in Group A’s results and their smaller PPM list. The difference
between Group A and the others here is significant, \( \chi^2 (2, N = 441) = 21.85, p < 0.0001 \). It is also worth noting that in the difference between whether they selected either zero or one PPM(s) per phrase (i.e. whether they identified any PPM or not), Group B and C behaved in the opposite way to Group A: While Group B and C selected one more often than zero, Group A selected zero more often than one.

6.5. Edges and Pauses: PPM Relationship to Silence

When the PPM lists were generated for all of the proficiency groups, as for Group Z, among the PPMs with the highest significance level were a number of edge syllables, both initial and final. Owing to the fact that the statistics used to create the PPM lists appeared particularly sensitive to the presence of a zero value (either no votes, or a pre-initial/post-final placeholder) in a triplet or quinlet, the contribution of these edge syllables to the overall PPM totals was investigated. Position of a syllable adjacent to silence can have an effect on its perceived prominence, and Biggs, for example, makes specific reference to pause as a factor in phrasing and prominence perception in Māori (1973, p. 1-2). The suggestion from Group Z’s results was that pause did not strongly influence their selection of PPMs. However, on closer examination of the 11 pause-adjacent PPMs that were identified by Group Z, over half of the set was shared by Group B and C: the lower-proficiency groups. That is, 8 of those 11 syllables were also pause-adjacent PPMs for Groups B and C. There was very little apparent overlap with Group A, the higher proficiency group: only 1/11. It was therefore possible that the three subgroups’ results would pattern differently from each other where selection of pause-adjacent PPMs were concerned. Table 6.7 shows PPM selection under these circumstances for the three proficiency groups.

Table 6.7.

<table>
<thead>
<tr>
<th>Proficiency Group</th>
<th>Total PPMs</th>
<th>Edge PPMs</th>
<th>Pause-Adjacent PPMs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Initial</td>
<td>Final</td>
</tr>
<tr>
<td>A</td>
<td>84</td>
<td>13</td>
<td>7</td>
</tr>
<tr>
<td>B</td>
<td>155</td>
<td>12</td>
<td>9</td>
</tr>
<tr>
<td>C</td>
<td>155</td>
<td>16</td>
<td>12</td>
</tr>
</tbody>
</table>
Although the usual pattern of Group A identifying fewer PPMs in all categories is apparent, it turns out that Group A did select significantly fewer pause-adjacent syllables than expected, (two-sided Fisher's exact test, $p < 0.05$). This seems to indicate that Group A participants were less sensitive to pause when identifying PPMs, which would be logical in participants with greater semantic ability: the contrast between syllable and pause-silence would not be as salient for them. Among the pause-adjacent syllables selected by Group B and C, about half were expected stresses (phrase+word (PWS) or word only (WS)), as was the case for the sole syllable selected by Group A in this category, and without contrasting examples, it is difficult to prove that it was the pause that caused the perceived prominence rather than other cues such as duration or pitch that would have been expected regardless. The other half of Group B and C’s sets, however, are syllables which are not expected stresses (OS). In all cases, the trend is very much in favour of syllables that precede the pause.

For those participants who are more sensitive to pause-silence, however, edge-silence does not have the same effect. There was no significant difference between the groups in their selection of any-edge-adjacent, initial, or final syllables, and though initial syllables were identified slightly more often than final ones, this is the opposite effect to the trend towards pre-pausal PPM identification, just as was observed with Group Z’s results (§ 5.4).

Thus, the implication is that in prominence identification in Māori, proficiency level is inversely proportional to sensitivity to pause-silence but not edge-silence, and where pause-silence has an effect on perceived prominence, this is most likely to be on the syllable preceding the pause. This is logical on both counts: Biggs (1961, 1973) identified pause-bounded contours, and therefore it would be expected that more proficient speakers would be less concerned with pauses that occur out of turn. The location of what is expected (by rule) to be the most prominent part of the phrase, the phrase stress syllable and high pitch point, generally occurs towards the end of the phrase, which would precede a pause boundary.
6.6. PPM Alignment with Biggs Stress Categories

This part of the analysis, as in Chapter Five, examined the rate of alignment between the syllables in the utterances that are predicted to be stressed, based on Biggs' (1969) rules, and those syllables the listeners in this study heard and identified as prominent (the PPMs).

The rules on which the stress categories used in this analysis are based were described in detail in Chapter Two. To provide a quick recap, the categories are phrase stress (PS), word stress (WS), and two other categories not specifically named so by Biggs, but used here for convenient reference: phrase and word stress coinciding (PWS), and other (unstressed) syllables (OS). Together, PWS and PS are referred to as the PS-Group; PWS and WS as the WS-Group. According to Biggs' other commentary, the phrase stress syllables are meant to be heard as the most prominent. When examining alignment here, again as before, we would therefore expect a good match (that is, a high rate of PPM selection) on PS-Group syllables. However, given Group Z's results in Chapter Five, in which WS was very strong, a high rate might also be predicted for WS and the WS-Group. A much lower result would be expected for OS syllables.

Proficiency levels could affect the results. Higher proficiency listeners (Group A) might select more expected stress syllables as PPMs through greater knowledge, selecting based on what they feel or know the answer “should” be. Their responses might depend on whether they have studied the language formally or learned it by immersion. However, since Biggs' stress categories are not an acoustic feature but a representation of a possible collection of these as heard by Biggs, higher proficiency speakers may also be closely tuned to what Biggs (a native, high proficiency speaker) was hearing, and will select the same prominences based on the same cues, on syllables produced with the same features. The examination of category selection here was intended to see if the different groups were selecting more syllables as PPMs from certain categories over others. Figure 6.1 below is a comparison of the category distribution across the groups' PPM sets, expressed as percentages of the total number of PPMs for each group.
Collectively, selection of stress category syllables (PWS, PS, and WS) as a PPM was, as predicted, much higher than the selection rate for OS syllables. Across the categories, just as in the results for Group Z, PWS syllables accounted for the majority of PPMs for all of the groups. The WS-Group far outweighed the PS-Group for all of the proficiency levels. However, while word stress syllables, particularly WS-only, accounted for a larger proportion of Group A’s PPMs than those of the other groups, both chi-square and Fisher Exact tests showed that there was not a significant difference between the groups. PS-only syllables had a very low rate for all groups. Table 6.8 provides the original values.

Table 6.8.

Alignment of PPMs with Biggs Stress Categories and as % of Available Syllables of Category Type

<table>
<thead>
<tr>
<th>Sylls. Available in Dataset</th>
<th>Total PPMs</th>
<th>PWS</th>
<th>PS</th>
<th>WS</th>
<th>OS</th>
<th>PS-Group</th>
<th>WS-Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>84</td>
<td>33</td>
<td>2</td>
<td>27</td>
<td>22</td>
<td>35</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>27%</td>
<td>6%</td>
<td>30%</td>
<td>6%</td>
<td></td>
<td>23%</td>
<td>28%</td>
<td></td>
</tr>
<tr>
<td>155</td>
<td>66</td>
<td>9</td>
<td>44</td>
<td>36</td>
<td>75</td>
<td>110</td>
<td></td>
</tr>
<tr>
<td>55%</td>
<td>27%</td>
<td>48%</td>
<td>9%</td>
<td></td>
<td>49%</td>
<td>52%</td>
<td></td>
</tr>
<tr>
<td>155</td>
<td>64</td>
<td>8</td>
<td>41</td>
<td>42</td>
<td>72</td>
<td>105</td>
<td></td>
</tr>
<tr>
<td>53%</td>
<td>24%</td>
<td>45%</td>
<td>11%</td>
<td></td>
<td>47%</td>
<td>49%</td>
<td></td>
</tr>
</tbody>
</table>

No group had, as a proportion of their PPM set, significantly more or fewer of any category of syllable. However, as a proportion of the available syllables of that category type in the database,
the picture is different but consistent. Significantly fewer syllables than expected were selected by Group A, and more by Group B and C, in the categories of PS (two-sided Fisher’s exact test, $p = 0.009$) and PWS, $X^2 (1, N = 363) = 11.15$, $p < 0.001$, and accordingly also in the PS-Group category, $X^2 (1, N = 462) = 18.20$, $p < 0.0001$. Owing to the effects of the PWS category result, the WS-Group category showed the same pattern: significantly fewer syllables for Group A, $X^2 (1, N = 636) = 20.12$, $p < 0.0001$. There was no significant difference in the results for the WS category itself, however, even though the same pattern of a lower percentage selected by Group A was in evidence. The OS category also showed the pattern, with no significant difference between the groups, and is the category with the lowest and most consistent rate of selection, despite having the most syllables available. This is entirely unsurprising, given that these syllables are, by rule at least, not expected to be prominent. Since Group A demonstrates fewer selections than expected essentially across the board, it is difficult to say that proficiency makes the predicted difference, here: one might anticipate that higher proficiency speakers would identify a relatively higher rate of expected stress (PWS, PS, WS) category syllables, but they do not. What can be taken from this is that there is a stronger response to WS-only syllables than to PS-only syllables by all groups, and that all groups selected roughly the same proportion of each category, with a relative dispreference for OS syllables.

The pattern here is, therefore, consistent with the A vs. B/C behaviour seen so far. It will continue to be visible in most of the other analysis that follows. The next section examines some acoustic characteristics of the three groups’ PPMs (and non-PPMs): duration, vowel quality and F0 behaviour.

6.7. Duration

The possible effects of duration in PPM selection were considered here, in terms of both syllable weight (heavy and light syllables, or nuclear duration), and whole syllable duration (considering the presence or absence of an onset, since these do contribute to the total duration of a syllable). The central question concerned the selection rate of these syllable types and whether or not the durational effects of either the basic heavy/light distinction, or the presence or absence of onsets in those categories, affected PPM selection for any of the proficiency groups.
As a background, mean nuclear duration was significantly different between heavy and light syllables in the PPMs of all the proficiency groups: Group A (ANOVA, $F(1,76) = 49.90, p < 0.0001$), Group B (ANOVA, $F(1,147) = 94.95, p = 0$) and Group C (ANOVA, $F(1,147) = 80.85, p < 0.0001$), just as in the dataset as a whole (Chapter Four, Section 4.8.1). The long monophthongs were also significantly longer than the short, in all groups' PPMs: Group A, $t(61) = 5.49, p < 0.0001$; Group B, $t(117) = 6.86, p < 0.0001$; Group C, $t(112) = 5.77, p < 0.0001$.

Tables containing the mean durations and standard deviations for the different vowel categories for each group can be viewed in Appendix D.

### 6.7.1. Syllable Weight and Rate of Selection

The predicted outcome here would be that all groups would select more heavy syllables as PPMs than light ones, since many of the expected stress syllables in Māori are heavy, particularly word stress syllables, and this was the expected stress category in which the groups' results were closest in the previous section. There was also an apparent dispreference for OS syllables by all groups, and since most of the OS syllables in the dataset are light, the pattern could be expected to continue here. In fact, not all of these predictions are borne out. Table 6.9 below shows the groups' PPMs divided into heavy and light syllables.

<table>
<thead>
<tr>
<th></th>
<th>Total PPMs</th>
<th>Heavy Syllables</th>
<th>Light Syllables</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group A</strong></td>
<td>84</td>
<td>52</td>
<td>32</td>
</tr>
<tr>
<td><strong>Group B</strong></td>
<td>155</td>
<td>74</td>
<td>81</td>
</tr>
<tr>
<td><strong>Group C</strong></td>
<td>155</td>
<td>76</td>
<td>79</td>
</tr>
<tr>
<td>PROM Dataset</td>
<td>n/a</td>
<td>168</td>
<td>468</td>
</tr>
</tbody>
</table>

Within the subsets of their total PPMs, Group B and C each selected more light syllables than expected. Group A selected fewer, but not significantly so. When PPM selection by syllable weight is considered as a proportion of the total heavy or light syllables available in the dataset, however, Group A responded as expected, selecting more heavy syllables than light, if still significantly fewer than expected, $X^2 (1, N = 504) = 6.26, p = 0.01$. Group B and C were the
opposite, selecting fewer heavy syllables than light, but the margin was smaller and not significant. Rather than a preference for heavy syllables, however, Group A’s result appears, in fact, to be caused by a dispreference for light syllables: when the Group A light PPMs are considered as a subset of all 468 light syllables available in the dataset, again, there are fewer than expected, but the significance level is very high, $X^2 (1, N = 1404) = 24.13, p < 0.0001$.

This is a response that could be called consistent with expectations regarding higher versus lower proficiency: given that syllable weight is the basis for a rule, the higher proficiency speakers could be expected to have some sensitivity to it (whether they are following learned rules or instinct). It is also reminiscent of Group A’s lower OS PPM selection in the Biggs stress category analysis. The two may be connected, but it cannot be said that one entirely explains the other, as there are more light syllables in the OS category to begin with.

As far as whole syllable duration is concerned, it does not appear that presence or absence of an onset is related to any group’s selection of PPMs (Table 6.10, below). All of the groups behaved the same way in their response to presence or lack of onset: just as for Group Z, in both weight categories, they selected many more syllables with onsets than without. This is, as before, most likely to be a reflection of the proportions in the dataset.

Table 6.10.

<table>
<thead>
<tr>
<th>Proficiency Group</th>
<th>Total PPMs</th>
<th>Heavy Syllables</th>
<th>Light Syllables</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Onset</td>
<td>No Onset</td>
</tr>
<tr>
<td>A</td>
<td>84</td>
<td>48</td>
<td>4</td>
</tr>
<tr>
<td>B</td>
<td>155</td>
<td>70</td>
<td>4</td>
</tr>
<tr>
<td>C</td>
<td>155</td>
<td>70</td>
<td>6</td>
</tr>
</tbody>
</table>

It might be asked whether or not there was a durational contribution to Group A’s dispreference for light syllable PPMs, but the data suggest that this is unlikely: the PPM light syllables chosen by Group A were not significantly longer than the light syllables they did not select. The mean nuclear duration of the light syllable PPMs is not significantly different from that of the light non-PPMs for any of the groups. However, when the same comparison is made using whole
syllable duration (including onsets where present), the mean duration of the light PPMs is significantly longer than that of the light non-PPMs for Group B (light PPMs: $M = 140.7\text{ms}$, $SD = 39.8\text{ms}$; light non-PPMs: $M = 112.9\text{ms}$, $SD = 52.0\text{ms}$), $t(466) = 4.54$, $p < 0.0001$. This is also true for Group C (light PPMs: $M = 141.7\text{ms}$, $SD = 44.7\text{ms}$; light non-PPMs: $M = 112.9\text{ms}$, $SD = 51.1\text{ms}$), $t(466) = 4.66$, $p < 0.0001$), but not for Group A. For all groups, however, the PPMs are still longer than the non-PPMs, which is what would be predicted.

It is possible that use of duration as a cue to PPM selection, at least where light syllables are concerned, relies more on the duration of a whole syllable for Group B and C than it does for Group A, who appear to be more influenced by weight. However, as the heavy-light nuclear weight distinction is not altered by the presence of an onset, it is not possible to be certain that this is the case. Additionally, since prominence identification is extremely unlikely to be limited to a single cue, there are other factors that can be considered. Besides the chance that there is semantic knowledge and implicit recognition of syllable weight involved in Group A’s selection process, a further intriguing possibility for the selection of light PPM syllables, particularly given that the majority of them have onsets and whole syllable weight appears to make a difference for Group B and C, is the “false diphthong” effect.

6.7.2. False Diphthong Effect

This potential selection effect refers to circumstances where participants identified a light onset syllable as a PPM, and that PPM is followed by an onsetless syllable of either weight category. In these cases, especially to a lower-proficiency listener, the resulting sequence may have sounded like a diphthong where in fact there is none in Māori, either because the vowel sequence does not form one, or there is a word or morpheme boundary. This could have resulted in the participant checking the preceding syllable (the one with the onset), thinking that it and the following syllable were in fact a single heavy syllable.

The present experiment was not designed to test for this, and determining the exact effects on each individual syllable has the scope of an investigation in its own right. However, it was possible to do a small comparison of the different proficiency groups in this respect.
The whole PROM Dataset contains 69 sequences where the \( \sigma[\text{light with onset}] \rightarrow \sigma[\text{onsetless}] \) circumstances are present and the effect could occur. All of the proficiency groups did identify some of these 69 sequences as PPMs, including Group A. Many of the sequences where a PPM was identified were the same ones across all of the groups' PPM sets, and there is a pattern to them. In the majority of the cases, the second syllable in the vowel sequence is /a/, making false “falling” diphthongs (for example, */ia/, */ua/). This kind of diphthong does not occur in Māori. However, similar-sounding sequences (for example: fear, seal, fewer, tour) are present as diphthongs in English (New Zealand and some other varieties), with which all participants were familiar, as well as in other languages. Given that Group A were most familiar with the Māori language, they might have been expected to be more likely than the other two groups to know that the sequences were not diphthongs. Even if they were not consciously aware of the phonological rule, they might be less likely to hear the sequence as a single heavy syllable.

The number of false diphthong sequences selected by Group A should therefore account for a smaller proportion of the 69 available sequences in the dataset than those selected by Group B or C. We would also expect to find that these sequences account for comparatively fewer of Group A’s light onset PPM total than they do for Group B or C’s totals. Table 6.11 shows the results.

Table 6.11.  
Possible False Diphthong Sequence PPMs by Proficiency Group  

<table>
<thead>
<tr>
<th>Proficiency Group</th>
<th>False Diphthong PPMs</th>
<th>Light Onset PPMs</th>
<th>Total PPMs</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>10</td>
<td>28</td>
<td>84</td>
</tr>
<tr>
<td>B</td>
<td>29</td>
<td>78</td>
<td>155</td>
</tr>
<tr>
<td>C</td>
<td>25</td>
<td>74</td>
<td>155</td>
</tr>
</tbody>
</table>

As percentages of the 69 possible cases in the dataset (A = 14%, B = 42%, C = 36%), all of the groups’ results are relatively low. However, Group A were indeed significantly less likely to identify PPMs under these false diphthong circumstances, \( X^2 (1, N = 207) = 10.64, p = 0.001 \).

Considering the contribution of the false diphthong PPMs to the total light onset PPMs selected by any group, however, the result was about one third in all cases (A = 36%, B = 37%, C = 34%).
Statistically, Group A selected slightly fewer than expected; Group B and C slightly more, but in this case the differences between the groups were not significant.

Since, overall, Group A did select significantly fewer of these sequences out of all those available than did the other two groups, the high proficiency listeners may indeed be less sensitive to the effect than Group B and C listeners. However, as the potential false diphthong sequences contribute approximately equally to all of the proficiency groups’ light onset PPMs, it is not certain that Group A’s lower number of these syllables is not simply a result of their lower overall light PPM count. It is not possible to say whether, were the count higher, the potential false diphthong sequences would still account for the same proportion.

As with all PPMs, these light syllables will have other features that could be influencing their selection instead of, or in addition to, the false diphthong effect. A specifically designed experiment would be necessary to investigate this properly, but a brief examination here suggested that neither the consonant in the onset of the PPM syllables, nor the Biggs stress category of either syllable in the sequence, appeared to make a difference to selection for any of the groups.

On the other hand, there is a suggestion that the PPM syllable’s position relative to an H* event (adjacent to it or aligning with it), may make a difference to one or more of the groups. Circumstances in which the PPM (the first syllable of the sequence) bears H*, or the H* is on an adjacent syllable (usually the second of the sequence) account for 60% (A), 83% (B) and 76% (C) of the false diphthong PPM syllables selected. However, as we will see later (§ 6.9), this particular pattern with regard to PPMs and H* events applies more broadly than the false diphthong set, so it is likely to be a matter of coincidence here.

The findings of this small investigation are inconclusive. While they do show some of the differences in behaviour with regard to these sequences that would be expected from the different proficiency groups (that is, Group A selecting fewer such sequences), there is some obscuring of the results by patterns involving F0, which is often a very strong cue to prominence, and has support elsewhere in the analysis for its use by these participants. A more specifically designed investigation into the effect may provide clearer results, but from the data available here, it is not
possible to determine whether or not the false diphthong effect is a real factor in light syllable PPM selection, either for any of the proficiency groups or at all.

6.8. Vowel Quality

This section of the analysis investigated whether or not there were differences in peripherality between long and short monophthongs, both in general and in PPMs and non-PPMs. This was examined for each proficiency group and across the three groups.

The long and short categories equate approximately to the heavy/light syllable weight category distinction. Diphthongs are not included here, but the reader may remember from the overview in Chapter Four that the long monophthongs in the PROM dataset are significantly longer than the short monophthongs: on average, twice the length (Section 4.7.1).

It was mentioned above (§ 6.7) that this is also true for all of the proficiency groups’ PPMs and non-PPMs. If greater duration is generally linked to greater peripherality in the vowel space, owing to greater time allowed to reach the vowel target, then long monophthongs should be more peripheral than short monophthongs. Since the long-short durational distinction was significant across the board, then we might also expect the formant values to show the same difference for all groups’ PPM spaces. As for the PPM vs. non-PPM distinction, PPM syllables might be expected to be more peripheral than non-PPMs. This, however, is not supported by a nuclear durational distinction: no group has a significant difference in mean duration of either long or short monophthongs between PPMs and non-PPMs. This might also be reflected in the vowel space shape when PPMs are compared with non-PPMs for the different groups.

The proficiency split vowel quality analysis suffers from a similar lack of tokens for some vowels in the PPMs to that shown by Group Z in Chapter Five (Section 5.7). The table opposite (6.12) shows the numbers for each vowel in each group.
Table 6.12.

**PPM and Non-PPM Vowel Token Sets for All Proficiency Groups**

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Total Monoph. Tokens</th>
<th>Short Monophthongs</th>
<th>Long Monophthongs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>a:</td>
<td>e:</td>
<td>i:</td>
</tr>
<tr>
<td>PROM</td>
<td>564</td>
<td>191</td>
<td>68</td>
</tr>
<tr>
<td>A-PPM</td>
<td>63</td>
<td>13</td>
<td>3</td>
</tr>
<tr>
<td>A non-PPM</td>
<td>501</td>
<td>178</td>
<td>65</td>
</tr>
<tr>
<td>B-PPM</td>
<td>119</td>
<td>39</td>
<td>6</td>
</tr>
<tr>
<td>B non-PPM</td>
<td>483</td>
<td>152</td>
<td>62</td>
</tr>
<tr>
<td>C-PPM</td>
<td>114</td>
<td>34</td>
<td>7</td>
</tr>
<tr>
<td>C non-PPM</td>
<td>450</td>
<td>157</td>
<td>61</td>
</tr>
</tbody>
</table>

Figure 6.2 contains two separate F1-F2 vowel space plots of the monophthong PPMs for Groups A, B and C: one for short monophthongs (on the left) and one for long (on the right). Owing to the difficulty with token numbers, the reader is advised that when viewing the plots, the vowels with the most reliable centroids (points) are generally /a/ and /a:/.

![Figure 6.2. Short (left) and long (right) monophthong PPM vowel spaces by proficiency group.](image)

When the two plots are compared, the first observation is that there is a peripherality difference between the short and long monophthongs, as expected. The short monophthongs have a smaller, more central and less uniform vowel space than the long monophthongs. This is also true for all three proficiency groups, whose long and short vowel spaces are very similar. The appearance of
the vowel spaces (that is, the visual peripherality difference) is supported by the significant durational difference between long and short monophthongs in all of the groups’ PPMs mentioned above. However, the peripherality difference itself, in terms of formant values, is not significant for any group: none of these long monophthong PPMs are significantly more peripheral than their equivalent short monophthong PPMs.

In terms of comparison between PPMs and non-PPMs, even though there was no significant difference found in the formant values between PPMs and non-PPMs for Group Z, it was possible that any of the three proficiency groups could show such a difference: Group A PPM vowels might be found to be less peripheral than those of Group B and C, if the lower proficiency groups were attracted by more fully articulated vowels and greater contrast. In fact, PPM and non-PPM vowels are remarkably similar in peripherality characteristics across the three groups, and are also similar to what has been seen so far: the long monophthong spaces are larger than the short, and no group shows a difference in overall vowel space size or shape. An ANOVA showed that there is no significant difference in formant values, either between PPMs and non-PPMs or across the groups. Again, however, the sample sizes are very small. In most cases, the difference in token numbers between PPMs (few) and non-PPMs (many) for any vowel in any group’s results puts the variation at a level that is too high to allow proper comparison between the two categories.

Overall, these results give the impression that while the speakers may be producing contrasts in the vowels as might be expected, no participant proficiency group appears more likely than the others to use vowel quality contrast in selection of PPMs. If it is a factor, it affects all groups equally. However, low token numbers may have affected this result: with a larger number of tokens, further significant differences could become apparent. Additionally, because the goal of this particular experiment is not to control which syllables are selected as PPMs by participants, it is not possible to balance the PPM tokens for phonetic environment. Therefore, with those caveats, the suggestion from these data is that no proficiency group specifically uses vowel quality contrast in selection of PPMs any more than the others do.
6.9. F0 Behaviour

This final section of the proficiency group comparison includes a discussion of the groups' PPMs and their relationship to various aspects of F0 behaviour. The three groups' responses are considered in three ways: whole phrases (in the form of phrase contour type); H* events (both rate of PPM alignment and mean F0); and stress category. We will see if there is a difference in response at these levels across the groups or between their PPMs and non-PPMs, in case any of them appear to affect PPM selection.

The first question concerns PPM identification as it relates to phrase contour type for each proficiency group. If the A vs. B/C pattern continues, Group B and C might be expected to be less likely to identify PPMs in minimal-type phrases; that is, the phrases with less contrast in F0. Table 6.13 below shows the number of phrases of each contour type where PPMs were identified by each group, along with the total number of available phrases of that contour type.

Table 6.13.

<table>
<thead>
<tr>
<th>Contour Category</th>
<th>Phrases Available</th>
<th>Phrases Containing PPMs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Group A</td>
<td>Group B</td>
</tr>
<tr>
<td>Fall</td>
<td>122</td>
<td>58</td>
</tr>
<tr>
<td>Rise</td>
<td>15</td>
<td>9</td>
</tr>
<tr>
<td>Minimal</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>147</td>
<td>68</td>
</tr>
</tbody>
</table>

All groups identified more PPMs in fall- and rise-type phrases than minimal, but in no case was this significantly more than expected. Group B and C identified more PPMs in minimal type phrases than Group A, which was not what was predicted, but all the groups identified significantly fewer than expected in that case (two-sided Fisher’s exact test, p<0.0001 for each of A, B and C).

The next comparison concerns PPMs and non-PPMs and their relationship to the H*, or high pitch point, in each phrase. Each H* is, of course, associated with a syllable. Description by Biggs and others suggests that this ought to coincide with greatest prominence (that is, a PPM) and also
with the expected phrase stress syllable (PS or PWS). The H*, being an event, is more specific than a syllable and allows a single reading of F0, rather than an average of several readings across the syllable unit.

Not every syllable in the dataset that is associated with an H* event was selected as a PPM by any proficiency group: For Group A, H* PPMs amounted to only 19% of the 137 available H*-bearing syllables in the dataset. Group B had 45% and Group C 44%. The difference between A and B/C is significant: Group A selected fewer than expected, $X^2 (2, N = 441) = 9.11, p = 0.01$.

Group B and C’s selection rates were slightly higher than expected. The original values can be seen in the first line of Table 6.14 below. This A vs, B/C pattern in group behaviour is the same one we have seen before. When considering the selected H*-bearing syllables as a subset of any group’s PPMs, however, no group has significantly more or fewer than expected.

Table 6.14.

<table>
<thead>
<tr>
<th>PPM Location</th>
<th>Number of PPMs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Group A</td>
</tr>
<tr>
<td>On H*</td>
<td>26</td>
</tr>
<tr>
<td>_ H*</td>
<td>21</td>
</tr>
<tr>
<td>H* _</td>
<td>7</td>
</tr>
<tr>
<td>Elsewhere</td>
<td>30</td>
</tr>
</tbody>
</table>

Since the selection rate for H*-bearing syllables as PPMs is low for Group A and under 50% for Group B and C, the finding can be tested in the same way as in Chapter Five: by examining the syllables surrounding the H* to see if F0 movement over a larger area is relevant. For Group Z, it was found that there were a number of PPMs that did not occur on the H* syllable but were adjacent to it (most commonly preceding), and that this increased the rate of PPM selection in relation to H* (§ 5.8.2). Earlier in the present chapter, while investigating possible false diphthongs, it was also discovered that many of those disyllabic sequences featured a PPM on the syllable preceding an H* syllable (§ 6.7.2). When the same investigation into H* adjacency was made for the three proficiency groups’ PPMs, the result was, predictably, similar (Table 6.14).
Across the three groups, of the PPMs that did not align with H*, between 35% and 50% were adjacent to it. In all cases, of those H*-adjacent PPMs, more preceded the H* syllable than followed it. When the PPMs that fell on the H* and all of the H*-adjacent PPMs are taken together for each group, the rate of PPM-H* association is 64% of PPMs for Group A, 61% of PPMs for Group B, and 63% of PPMs for Group C. The fact that the groups pattern so similarly here provides support for the interpretation that they are all influenced by contrast in F0, or a movement in F0, that occurs around the H*: particularly, what precedes it, or the H* itself. This is the same result as for Group Z, but shows that use of this cue in this way is neither special to, nor dispreferred by, any group individually.

Concerning the issue of contrast, when the difference between PPMs and non-PPMs is examined further, the mean F0 for H* events in PPM syllables can be compared with that of the H* events in non-PPM syllables. Table 6.15 below contains the measurements. The H* events in the syllables that were identified as PPMs do have, for all groups, a slightly higher mean F0 than the H* events in the syllables that were not identified as PPMs. However, the difference is not significant for any group. In Chapter Four, it was observed that across the whole dataset, the mean F0 for H* events is significantly higher than the phrase average mean F0 in fall- and rise-type phrases (§ 4.7.3). Therefore, the contrast does exist, but it does not appear to be used as a particular cue by any participant group.

Table 6.15.

<table>
<thead>
<tr>
<th>Proficiency Group</th>
<th>H* Event Type</th>
<th>No. of H* Events</th>
<th>H* Mean F0 (Hz)</th>
<th>SD (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A</strong></td>
<td>PPM 26</td>
<td>149.0</td>
<td>45.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Non-PPM 111</td>
<td>142.8</td>
<td>36.0</td>
<td></td>
</tr>
<tr>
<td><strong>B</strong></td>
<td>PPM 62</td>
<td>154.8</td>
<td>41.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Non-PPM 75</td>
<td>139.3</td>
<td>35.0</td>
<td></td>
</tr>
<tr>
<td><strong>C</strong></td>
<td>PPM 60</td>
<td>154.9</td>
<td>41.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Non-PPM 77</td>
<td>140.3</td>
<td>35.5</td>
<td></td>
</tr>
</tbody>
</table>

Finally, based on Biggs’ comments about the connection between higher pitch, the most prominent part of a phrase, and the phrase stress syllable, we might also expect to see a link
between H*, PPMs and PS-Group syllables. In Chapter Five, for Group Z, the connection was present to an extent. The results were dominated by the PWS category, but the WS-Group (PWS and WS) rather than the expected PS-Group (Section 5.8.2). The same was true here, for all proficiency groups, and the results were remarkably consistent: selection of H*-bearing syllables of any stress category was not significant for any group. Figure 6.3 below shows the H* PPMs for each stress category as a percentage of the available H* syllables in that category, for the three proficiency groups. PS-Group and WS-Group results are also provided for reference.

Figure 6.3. H* PPMs for the stress categories as % of the available H* syllables in that category.

Here again, as in Group Z’s results, the salience of word stress (WS-Group) and the importance of overlapping cues in prominence identification (PWS) are apparent, and affect all proficiency groups in almost the same manner.

6.10. Summary

In this investigation, the participants from the main experiment were divided into three groups based on Māori language proficiency: Group A (high), Group B (exposed) and Group C (zero). Their response to the perception task, and the resulting perceived prominences (PPMs), were compared. There was a difference in the groups’ response to the task, with Group A checking more syllables overall in the survey than the other two groups, resulting in a smaller overall PPM list for Group A after the statistical validation was applied (see Chapter Four, Section 4.5.5 – 4.5.6). The A vs. B/C pattern is echoed in various ways throughout the PPM results of the different groups here.
The three groups’ PPMs were measured and compared distributionally and acoustically. Overall, 84 significant PPMs were identified by Group A, 155 by Group B, and 155 by Group C, and the overlap between the PPMs selected by Group B and C was much greater than that of Group A with either of the others, which is consistent with relative proficiency levels. There were not PPMs in every utterance or phrase for any group. The average number of PPMs per utterance was 2-3, and the average number of PPMs per phrase was 0-1 for all groups; Group A were more likely to have none, while Group B and C were more likely to have at least one, and the difference between the groups was significant.

The rate of alignment with Biggs’ expected stress category syllables was generally as expected given previous results: the PWS and WS categories had the highest rate of selection across all of the groups and, together as the WS-Group, accounted for the largest proportion of stress category selection. No category made up a significantly larger proportion of any group’s PPMs (thus, it is not possible to say that Group A selected more PS syllables than Group B, for example). When the PPMs of a given category are considered as a proportion of all the available category type syllables in the dataset, however, Group A selected significantly fewer than expected in the categories of PS and PWS, and accordingly also in the PS-Group. Group B and C, in fact, selected more. This is contrary to expectation: higher-proficiency speakers would be the group theoretically most likely to be sensitive to cues on PS syllables, unless, as we might speculate, the cues are not present or too weak to be detected, or are not the cues that are expected by those speakers. There is some evidence that the cues on PS-only syllables in particular may be weakening, for want of a better term (see also § 4.10 on speaker vintages, § 5.5 on stress category alignment, and § 5.8.2 and below on H* alignment). In the category of non-expected stress (OS) syllables, however, was found the lowest rate of selection for all groups, which is an expected result.

In response to syllable weight, within the PPMs of each group, none selected significantly more or fewer syllables of either weight category: heavy or light, though Group A selected many more heavy than light, and Group B and C selected fractionally more light than heavy. When the PPMs of each weight category were taken as a proportion of the total available syllables of that category in the dataset, Group A showed a clear and very highly significant dispreference for light
syllables, while Group B and C did not. However, there did not appear to be a durational motivation for this: the light syllables that were chosen by Group A were not significantly longer than those they did not select. In fact, the mean nuclear duration of light PPMs was not different from that of light non-PPM syllables for any group. Presence or absence of onset made no difference to PPM selection for any group, but when light PPMs and light non-PPMs are compared using whole syllable (including the onset where present) rather than nuclear duration, the light PPMs were significantly longer than the non-PPMs for Group B and C, but not Group A. Overall, when comparing PPMs and non-PPMs of any weight, this pattern still applies: mean nuclear duration is not significantly different for any group, but in whole syllable duration, PPMs are significantly longer for Group B and C, but not Group A. This suggests that all groups, but particularly Group A, are tuned to weight to a degree. However, actual nuclear duration may make less of a difference than predicted. In fact, it is likely that as might be expected given the inherent contrast involved in prominence, it is relative duration that will be noticed. It is possible that whole syllable duration plays more of a role in PPM selection for Group B and C than it does for Group A, but the fact that presence or absence of an onset (whole syllable duration) does not affect the basic heavy-light proportions (and therefore contrast), even though it may shorten both long and short nuclei, is an obscuring factor in these results. Despite this, it is still certainly possible to say that all groups are tuned to weight in some form.

Vowel quality of the groups’ PPMs was also examined. However, there were no significant differences between the PPM vowel spaces of any of the groups. Neither were there any significant differences in formant values between vowels in PPMs and non-PPMs selected by any group. A reason for this result may be the very low token numbers for all vowels other than /a/, and more data would be needed to test efficiently, but it appears that even though speakers are producing contrasts in the vowels as expected (both in quality and quantity), none of the proficiency groups are using this cue to a greater or lesser degree than the others.

The rate of PPM selection in phrases of different contour types (fall, rise and minimal) was consistent across the groups, with all groups identifying significantly fewer PPMs than expected in minimal type phrases. Group A were significantly less likely to select an H* (high pitch point) syllable as a PPM than Group B or C, who identified more than expected out of the 137 available
in the dataset. However, no group has a higher proportion of H* syllables in their PPM list than the others. For all Groups, the H* PPM mean F0 was higher than the H* non-PPM mean F0, but this was not significant for any group, which may simply mean that they use the cue equally. Additionally, when the PPMs from each group that did not coincide with H* but were adjacent to it (preceding or following) are added to those that did coincide with H*, the rate of H* association is very similar for all groups: just over 60%. These adjacent PPMs were most often found preceding the H* syllable, echoing the finding in Chapter Five for Group Z. This suggests that participants of all proficiency levels are sensitive to (a) higher F0, and (b) some movement on or around a point of high F0, particularly preceding it. This is certainly not an unexpected result. Where the further expected association of H* and PPM with the phrase stress category was concerned, no significant difference was found between the groups: all exhibited the same tendency towards PWS and WS category syllables (rather than PS or PS-Group syllables) seen elsewhere, underlining the importance of overlapping cues that has become apparent across the study.

Overall, there is some evidence from this investigation that participant language proficiency does affect both response to a perception task and the results that come out of it. Group A, with greater knowledge of the language, responded to the stimuli in a different way to Group B and C, the lower proficiency speakers. This pattern is quite clear and recurs throughout, even though the difference is not always significant. Though the pattern often shows Group A selecting fewer PPMs of a given type, this is partly due to their lower overall PPM total, and this in turn is a result of the way they responded to the exercise: by selecting more syllables to begin with. There is not, however, strong evidence from the results of this experiment that “low” proficiency is greatly different from “zero” proficiency (Group B and C) in task behaviour and when selecting PPMs. The two groups behaved in a very similar manner, and they were certainly capable of responding to the task. Where some cues are concerned, such as in the observations on vowel quality or on F0 in the final section, the results for the three different proficiency levels can be very consistent: there is evidence that they are using the cues, and to roughly the same degree as each other. However, it is also clear that syllable weight (and to some extent duration) have what can probably be considered to be the most contrasting influence on PPM selection by the
different groups, and in that case the pattern is definitely higher (A) versus lower (B/C) proficiency.

Ideally, as ever, an investigation such as this would be conducted on a larger scale. If the proficiency groups were of a comparable size to the entire participant set, that is, with 90-100 participants each, then the results would be clearer still. This provides plenty of scope for further investigation, some of which will be discussed in the evaluation and conclusions in Chapter Seven.
CONCLUSION

DISCUSSION, EVALUATION, AND FUTURE DIRECTIONS

7.1. Introduction

There were six goals at the core of this project, motivating its development. The aim was to develop an approach that would discover and analyse the current locations of perceived prosodic prominence in Māori, amid anecdotal descriptions of change in the language, evidence of change already discussed in other studies, and existing descriptions of stress, prominence, and connected acoustic features such as F0 (§ 2.4). By asking what contemporary listeners heard, it could be determined whether their perceived prominences were consistent with the existing descriptions of stress locations formulated in the 1960s, and if not, what had changed. The MAONZE database, containing read and continuous speech recordings from Māori speakers, provided a base from which to launch an investigation: participants’ prominence perceptions in response to the speech of speakers from three different time periods spanning roughly a century were compared. In order to examine the perceived prominences further, selected acoustic features, namely, duration, vowel quality, and F0, were examined, to determine, again, whether or not this was consistent with the existing impressionistic descriptions. Considering the known effects of native versus non-native proficiency in perception of cues to prominence (§ 2.3.3.3), there was the question of whether or not exposure to or proficiency in the Māori language on the part of the participant listeners made a difference to the prominences they identified, and if so, what that difference might be.
The project’s first task was to develop a suitable method for eliciting the perceived prominences. In principle, this was a straightforward goal, but there are a number of different ways to have people communicate what they hear. The progression of the study and development of the methodology, from the pilot study to the main experiment, explored two of these possibilities. Initially, listeners were asked to tap along with a recorded utterance (§ 3.2). While the results confirmed the ability of participants to identify prominences consistently and established that filtering the stimuli did not affect this outcome, the elicitation method itself was not especially efficient, and was replaced with a process requiring direct identification of units (here, syllables) using a specific web-based checkbox method (§ 3.3). This method proved a great deal more efficient and practical; solved some participant confidence problems; and allowed the study to expand, in both participants and stimuli. A much fuller discussion and evaluation of these first two phases of the project can be found in the summary and conclusion to Chapter Three.

The methodology used in the main study (Chapter Four) retained the web-based format, but continued to expand, with a more robust technical definition of prominences for use in the analysis; statistical refinement of the results; the creation of a detailed speech analysis database (the PROM Database) based on the stimuli; and the empirical and acoustic analysis of the prominences obtained. Overall, the adjusted web survey methodology and analysis methods used in the main experiment were successful. They achieved the goals of effective prominence elicitation and analysis of the results. However, as in any exercise of this nature, there were still a few matters to address. There is always room for change, improvement, and exciting new directions. These include the scale of the study, the web survey itself, and its possible applications, both within the Māori language and without. We will return to further evaluation and discussion of these later on, but for now, there is the question of what was found out in the course of the project. The observations on which the following discussion is based come chiefly from the main study (Chapters Four to Six), with additional references to the experiments described in Chapter Three where relevant.
7.2. Issues in Participant Prominence Identification

Participants were, as expected, capable of identifying prominences (PPMs) consistently, though there was early evidence, confirmed throughout the experiments, that speakers with different proficiency levels in Māori responded differently to the task. Higher-proficiency speakers had a tendency to check more boxes for syllables in the perception test, perhaps because they understood the meaning of the words, or perhaps owing to their interpretation of the instructions (of which more below). This created noise in their results and fewer significant prominences following statistical validation using Triplet Mode (§ 4.5.2 - § 4.5.3), despite a high degree of consensus. The methodological issues associated with this were addressed with reasonable success in the form of a modification to the statistical prominence validation measures: Quinlet Mode (§ 4.5.4 - § 4.5.6) widened the area within which an agreement peak’s significance was assessed, countering the effect of the noise.

The deliberately minimal nature of the instructions in the perception task has been a constant through all of the experiments in the project. As explained in the descriptions of the experiments themselves (§ 3.2.3, § 3.3.2.2, and some discussion in § 4.5.4), the reason that participants were told no more than to identify syllables that “stood out” to them was to avoid steering the listeners toward any particular cue that might be present, and thereby skewing the results in favour of that cue. It was desirable that the responses be as instinctive as possible. That said, such minimal instructions obviously provide scope for a rather wide range of interpretation, not limited to possible perceptual cues such as pitch or duration, but including semantic knowledge, recognition of lexical items, or even simply the ability to hear the utterance. It is entirely possible that any or all of these factors had some effect on the results. However, examination of the tapping patterns from the pilot study and the box-checking patterns from both the first web experiment and the main experiment shows that checking of whole utterances was extremely rare, and no participant did so all the time. These experiments did contain different participants, and it was clear that broadly speaking, there was some common understanding of “prominence” (that is, contrast). Checking of larger blocks of utterance was also relatively rare and, interestingly enough, occurred at both ends of the proficiency spectrum. Thus, there is an indication that semantic knowledge or familiarity was the interpretation of “standing out” for some participants, but this was not
consistent and not exclusive to a proficiency level. The fact that participants with very low proficiency are included in this group suggests that a reaction to audibility may have been involved. However, these observations are based upon the raw results and, as mentioned above, the statistical validation measures employed in the main experiment were designed to counter any effects this behaviour may have had: most especially Quinlet Mode (§ 4.5.6), which was applied to results from participants of all proficiency levels. Analysis of the statistically validated results with regard to the stress category alignment has shown that what might be called excessive checking has not affected the results greatly. Some evidence of this is found in the relatively low alignment rate of perceived prominences with phrase stress only (PS) syllables despite the number available in the dataset, which occurred in both the early (§ 3.3.4.3 - § 3.3.4.4) and main (§ 4.8; § 5.5; § 6.6) experiments. If participants were checking whole words, phrases or utterances more than expected, and the effect of this filtered through into the results, the number should be much larger. The rate of alignment with non-expected stress syllables (OS) should also be much higher. Since the participants who showed a tendency to check more boxes were in the minority, it is also likely that the larger the group of participants, the smaller the effect such behaviour can have on the results.

Prominence is based on contrast; it is relative salience. This theoretically entails a type of recognisability that is different from (semantic) understanding, but that may still be linked to it, particularly for listeners who are also familiar with English, in which stress, which is based on prominence, is contrastive. All participants in this study rated their English proficiency highly. Therefore, it is likely that the two interpretations of the instruction, that is, reaction to perceptual cues versus basic recognition, are also linked. The word Māori, in which the first syllable bears stress, appeared three times in the stimuli. The first syllable is heavy, and therefore a candidate for the duration or weight cue, and in two of three cases, it is associated with the high pitch point. In all cases, it was selected by the majority. Being the name of the language under investigation, for some of the participants, this word would have been the only one in the exercise that they recognised. For the rest of the participants, it would have been extremely distinctive. Under the current investigation methodology, it is impossible to tell what prompted each individual participant who checked the box for the syllable māo to do so: it may have been recognition, it may have been a cue, or it may have been both. The best method of isolating the two would be to
filter the stimuli, removing segmental information and therefore the possibility of understanding; this would require the use of a different methodology for prominence elicitation and analysis. As far as the present project is concerned, it cannot be said that all participants interpreted the instructions in the same way, but the results of all of the experiments presented here suggest that the interpretations were, for the vast majority of participants, similar to each other and as intended by the task.

Statistically validated (i.e. significant) prominences (PPMs) identified by the 92 participants in Group Z (all participants in the study), as well as those of the three Māori proficiency based participant groups (Group A (high), Group B (exposed), and Group C (zero)) are included in the following discussion.

7.3. Perceived Prominences: Distribution

The rate of selection of PPMs per utterance was approximately 5 for Group Z as well as Group B and C; while only 2-3 for Group A. This is not surprising since B and C make up the majority of Group Z. While the rate of 5 per utterance is consistent with the number of expected phrase stresses per utterance in the dataset, the reality of PPM distribution across phrases in the utterances is different. The selection rate was an average of one per phrase for Group Z, but while Group B and C were more likely to have one PPM in a phrase, Group A were significantly more likely to have no PPM at all. No group identified PPMs in every phrase, however, and there is no evidence that utterance length, in either syllables or seconds, had any effect on the prominence identification rate. There was also no observable pattern to the intervals between PPMs, either in millisecond duration or number of syllables.

Biggs makes specific reference to pause as a factor in phrasing and prominence perception in Māori (1973, p. 1-2). The stimuli used in the main experiment all had initial and final silence (edges), and a few also contained pauses either within or between phrases. However, selection of PPMs by Group Z participants does not appear to have been strongly influenced by the presence of these pauses or silences in the stimuli. Where the proficiency split is concerned, Group A selected significantly fewer pause-adjacent syllables than expected; this seems to indicate that the
Group A participants were less sensitive to pause when identifying PPMs, and would be logical in participants with greater semantic ability: the contrast between syllable and pause-silence would not be as salient for them. Group B and C selected more pause-adjacent syllables, but it is difficult to prove that in such cases, it was the particular contrast of pause that caused the perceived prominence rather than other cues such as duration or pitch: coincidence was high between the selected pause-adjacent syllables for all groups and expected stress (phrase+word (PWS) or word only (WS): the two most salient categories – see next section). Overall, the trend is very much in favour of syllables that precede the pause. For the participants who are sensitive to pause-silence, however, edge-silence (sentence-initial or -final) does not have the same effect, with no significant difference observed between the groups in their selection of any-edge-adjacent, initial, or final syllables. Although initial syllables were identified fractionally more often than final ones, this contradicts the above trend towards PPM identification on syllables preceding a pause. It seems, therefore, that in prominence identification in Māori, the higher the proficiency level, the lower the sensitivity to pause-silence but not to edge-silence. Where pause-silence affects perceived prominence, it will most likely be in favour of the syllable preceding the pause. This is logical on both counts, and consistent with description in the literature, to a degree: Biggs (1961, 1973) identified pause-bounded contours, and therefore it would be expected that more proficient speakers would be less concerned with pauses that occur out of turn. The most prominent part of the phrase, as it is generally described, that is, the phrase stress syllable and high pitch point, is generally towards the end of the phrase; this would precede a pause boundary.

7.4. Perceived Prominences: Stress Category Analysis

The stress categories discussed here are phrase stress (PS), word stress (WS), and two other categories that were not named so in Biggs’ (1969) description, but are used in this study: the coincidence of phrase and word stress (PWS), and other (unstressed) syllables (OS). Together, PWS and PS are referred to as the PS-Group; PWS and WS as the WS-Group. As mentioned above, the phrase stress syllables are, according to the description, meant to be heard as the most prominent. Belonging to a certain stress category is, of course, not an inherent quality of a syllable that makes a listener hear it as prominent. The categories are a construct based on Biggs’ impressions, which were, in their turn, based on perceptual cues that Biggs, a proficient Māori
speaker with linguistic training, was using when listening to a dataset containing continuous speech. The expectation in the results was that if perceived prominence aligned strongly with certain stress categories for the participants in the study, more PPMs would be expected to belong to a stress category (PWS, PS, WS) than to the other category (OS), and this effect would be strongest in the listeners with the highest proficiency level (Group A). However, this was not exactly so.

Throughout the study, in all of the experiments, the PS-Group (PS+PWS), which would have been expected to dominate the PPM results owing to the descriptive connection made between phrase stress and greatest prominence, was in fact overshadowed by the WS-Group (PWS+WS) for all of the participant groups. The rate of selection of PS-only syllables, both out of those available in the dataset and as a proportion of any group’s PPMs, was very low. By contrast, the response to word stress (WS) syllables was very strong for all groups, and the PWS overlap category was the strongest in all cases. While there was potential for higher Māori proficiency in the participants to have skewed the results, by virtue of extra knowledge or rule-based selection of syllables, the results do not suggest any such influence. For example, if this were the case, the selection rate of PS-only syllables in particular by Group A participants ought to have been much higher, but in fact, their rate was the lowest of the three proficiency groups for that category.

Looking at the relative proportions of PS-only to PWS category syllables in the PROM dataset, which, at its size, could be called representative, the PWS category far outweighs the other. This was also true of the smaller dataset used in the first web experiment in Chapter Three. When Biggs formulated his stress rules, he did separate WS and PS, acknowledging that they can overlap, but that they can also be separate in non-final phrases (§ 2.4.1.2). He did not specify anything more than syllable weight and position in the word as a defining characteristic for WS syllables, but he associated PS syllables with pitch and prominence specifically. What he did not do was conduct a specific examination of possible differences between PS and PWS syllables, or consider relative prominence in this respect; perhaps it was to be taken as read that PWS would be more prominent – we cannot know. While the present study did not examine relative prominence for individual syllables specifically either, the selection rate of PWS syllables compared to that of PS syllables out of those available in the dataset speaks for itself: clearly the former has the greater
salience, regardless of proficiency level, and this is true in the results from both of the web experiments. This observation, along with the extremely strong response to WS syllables (sometimes, to the extent of apparently overruling PS-only in the same phrase, as in the first web experiment (§3.3.4.3); similarly observable in the PPM distribution from the second web experiment) and the resultant overall dominance of the WS-Group in these results, suggest two main conclusions. First, that overlapping cues (in this case, syllable weight for WS and F0 for PS) create a much stronger impression than single cues, which is entirely reasonable and even expected, and second, that either current listeners are more strongly tuned to the cue associated with WS, or the PS cue has grown weaker. Further support for these conclusions was found in the acoustic analysis.

7.5. Perceived Prominences: Acoustic Analysis

The three acoustic features investigated were duration, vowel quality, and the behaviour of F0. It was clear that all participants were using syllable weight as a cue, with a clear dispreference for light syllables on the part of Group A; this was to be expected, at least in their selection of WS-Group syllables, which form the majority of PPMs in all cases. However, while the expected difference in nuclear duration was observable; that is, heavy syllable nuclei were significantly longer than light syllable nuclei, nuclear duration measurements did not provide support for Group A’s dispreference. Indeed, the light syllables chosen as PPMs by all groups were not significantly longer than those that were not chosen. The results from measurement of whole syllable duration (including onsets where present) show that in all cases, heavy syllables were significantly longer than light, but when the groups’ light PPMs were compared with the light non-PPMs, only for Group B and C were the PPMs significantly longer. All of this points to sensitivity to syllable weight, possibly almost in the abstract, in all participants, but greater sensitivity to whole syllable duration on the part of the lower proficiency groups. Differences in nuclear duration make a difference on a general level, but apparently do not influence the choice of some light syllables over others as prominent. The results from the vowel quality analysis were consistent with this, in that where increased duration generally corresponds to increased peripherality in a vowel space, this was observable in the comparison of the PPM syllables with
the non-PPMs. The difference in formant values, however, despite being visually obvious, was not significant for any proficiency group.

The duration results also link back to the stress categories. It was observed that the overlapping stress category syllables (PWS) were by far the most commonly selected as prominent out of those available, as were word stress syllables (WS). As word stress is linked with (heavy) syllable weight, this is unsurprising given what has just been discussed with regard to duration. Extra support for the PWS overlap category as the most prominent may be found here, too: in all cases, the PPM syllables from the PWS category are significantly longer than those from the other categories, including WS itself: only where expected phrase and word stress coincide do the PPMs exhibit the extra length. If we take into account that PWS in final phrases are often the penultimate or ultimate syllable in a sentence, this extra length may be some form of final strengthening. While it is true that such PWS have a better response rate than PWS found further away from the end, a similar pattern is also observable in non-final phrases. However, sentence-internal PWS are not necessarily shorter than the sentence-final PWS, which would merit closer examination.

PWS syllables had longer duration, then, and as it happens, also the best rate of association with the high pitch point, H*. The results show that F0 was used as a cue by participants of all proficiency levels. All groups selected very few PPMs in minimal type phrases (those which fell into neither fall nor rise type categories owing to insufficient change in F0 across the phrase). It is possible that these ten phrases were instances of what Biggs (1961, 1998) and Bauer (1993) have described as multiple grammatical phrases (XP) being included in a phonological phrase (PPh); see §2.4.1.2 for more detail. In the present investigation, most of these cases were sequences of a minimal type phrase followed by a fall type phrase, and in the second (fall) phrase, the PPM identification was consistent with the expected pattern in terms of coincidence with H* and or PWS category across all proficiency levels.

Overall, the response to H* was also consistent: H*-PPMs made up the same proportion of PPMs for all of the groups, and the mean F0 of the H* in those PPM syllables was higher than that of the H* in the non-PPM syllables. It is highly likely that all participants use the F0 cue equally and in a similar way; this is unsurprising given that English, with which language all of the lower
proficiency participants were familiar, also has pitch as one of its principal cues. There was a further suggestion in the results that what participants were most sensitive to was both higher F0 and the movement on or around, but most particularly preceding, the H* itself. These are, as mentioned above, not unexpected results, and they are essentially supportive of the existing description.

The predicted coincidence of the H* with PS-Group syllables, however, is another matter: the expected association is outweighed, if only marginally, by association of the H* with the WS-Group. This has been a pattern throughout the results, although it is true that here, the contribution of PS and WS to their respective groups’ results is closer than it has been in other analyses so far, and this is, in fact, the best showing of PS-only category syllables in any results in the study. This indicates that the association of PS with F0 is not gone, and it should be noted that in both the PS- and WS-Groups, PWS syllables, which are still phrase stress, account for the vast majority of the scores. Though the differences between the proficiency groups’ responses were not significant, an interesting observation was that Group A did have the highest response to WS-only syllables in terms of H* association; this would not be expected and suggests that they are more sensitive to H* if it occurs on another stress category than Group B and C. Group A’s OS-H* PPM response was the lowest of the three groups, however; this is likely to be a result of their dispreference for light syllables, which make up the majority of the OS category, and may indicate that Group A’s weight sensitivity outweighs their F0 sensitivity. If true, this would provide further support and explanation for the apparent emergence of word stress (and, by extension, the overlapping category) as stronger than phrase stress alone. A further indication of weakening in PS-only syllables comes from the comparison of prominence identification in the speech of speakers from different time periods (§ 4.10).

7.6. Speaker Vintage Comparison

The three vintages used in the study were Historical Elders (HE; b. ca. 1880; rec. 1940s), Present-Day Elders (PE; b. ca. 1920; rec. 2000s), and Present-Day Young (PY; b. ca. 1970; rec. 2000s). Biggs’ (1969) description was based on speakers who would fall into the HE or PE categories. In the face of expectation that phrase stress syllables, or the PS-Group, would be
identified more often in the older vintages, and that there would be a difference in PPM rate of selection between the different vintages, this was only partially the case.

The results from the first web experiment, which compared PE and PY vintages, showed that there was no significant difference between listeners’ responses to the two vintages in terms of either PPM selection or phrase-category PPM selection. In both vintages, PWS and WS dominated, and no PS-only syllables were selected at all. In the main experiment, which compared responses to all three speaker vintages (HE, PE, and PY), the result was essentially the same in terms of rate of selection. All of the participant groups selected approximately the same number of PPMs, and also approximately the same proportion of the syllables available, from each vintage.

When the stress category results from the main experiment, are considered, however, the observations are more useful. The response to PWS and WS syllables was strong in all vintages, as has been a theme throughout the study. This suggests, again, that WS cues were strong in all vintages, as this is the common factor between the two categories. The response to PS-only syllables is the most interesting result in this analysis: where previously there was simply no response to PS-only in any vintage, here, the selection rate is low in all vintages, but it is in the PY (youngest) vintage speech that there is no response at all. Participants from all proficiency levels had not a single PY-vintage PS-only syllable as a PPM.

To speculate from these results, it is possible that the young speakers may not be cueing PS-only syllables at all, or so slightly that they are less likely to be heard as prominent, even by proficient speakers. This would certainly merit further investigation with a much larger dataset than the one in the present study, owing to the relative scarcity of PS-only syllables in general. A further observation from the speaker vintage analysis is that there are mild but non-significant indications of a participant response (at all proficiency levels) to what may be a degree of hyper-correctness in the PE (middle) vintage: this is based on the fact that the HE results (that is, those for the oldest speakers) did not display the same effects. Other than chance, the best explanation for a small spike in expected alignment in the middle category is rule-consciousness on the part of the speakers, and the possible associated emphasis on certain cues.
Given the tendencies that have been observed in the results from this study, it may be that where Biggs had originally associated the effect of greatest prominence with phrase stress, by which we must assume he meant the PS-Group, word stress now plays more of a role, and from that we might conclude that greatest prominence is now associated with PWS rather than the PS-Group. It is not possible to exclude PS-only syllables entirely, though this may become more likely with time. The importance of the response of higher proficiency listeners in this conclusion is shown in the fact that even they did not have any PY phrase stress PPMs: of any group, they would have been the most likely to select such syllables, even if it were only by rule.

7.7. Proficiency Group Comparison

As the previous discussion has shown, there is evidence that participant language proficiency did affect both response to the perception task and the results that came out of it, but that in many cases, in this investigation at least, the differences were not as great as might have been expected. In the results from the first web experiment, there were indications that higher proficiency listeners identified more prominences, and had a slight tendency towards more expected-stress syllables (of PWS and WS type), than the lower proficiency listeners. The results from the main experiment also showed a general pattern of Group A (“high”) versus Group B (“exposed”) and C (“zero”). This was not at all surprising, given the relative proficiency of those involved. The pattern is quite clear and recurs throughout, even though the difference is not always significant. The greatest difference between the groups along this pattern in the main experiment is, as before, in their response to the exercise: Group A checked more syllables in the task to begin with, and through statistical validation, their overall significant PPM count was lower. Consequently, in the main experiment, the analysis often has Group A selecting fewer PPMs of a given type, but when considered proportionally, this does not affect the conclusions. There was, conversely, little evidence from the results of the main experiment that “low” proficiency is greatly different from “zero” proficiency (Group B and C) in task behaviour and when selecting PPMs in response to the cues investigated. What difference there was between Group A and the lower proficiency groups was variable. Occasionally, the difference was very clear, as in syllable weight, where Group A obviously dispreferred light syllables, but Group B and C did not react in the same way. That said, all groups used some form of duration or weight as a cue. Where other cues were
concerned, such as in the observations on vowel quality or on F0, or more generally, as in the responses to the different speaker vintages, the results for the three different proficiency levels could be very consistent: there was evidence that they were each using the cues, and to roughly the same degree as each other. In order to tease out any proficiency related differences in greater detail, further study would be needed, which brings us to a discussion and evaluation of the current study, and its potential for future development.

7.8. Evaluation and Development

As was apparent in some of the discussion in the course of the main experiment, there were occasional effects of the size of the set of responses on the results. While the second web survey featured a much broader set of stimuli than either of the experiments in the first two phases of the study, and while the set of PPM tokens provided was generally adequate for the purpose, a set of responses that was larger still would be ideal. In a study of this type, where the initial number of participants is not precisely controlled, the number of responses cannot be foreseen. It also cannot be predicted how many PPMs a given participant will identify; this is, after all, partially the point of the exercise. The most effective way to correct for the former is to control the participant pool; this has obvious additional benefits in the form of proficiency level control. The only way to attempt to influence the number of selected PPMs, and therefore the token numbers for analysis, is to provide a wider pool of potential syllables: a larger set of stimuli. This, among other developments, has been considered across the course of the project and following its conclusion.

Potential development of the project falls into several categories. First, there is the technical development of the experiment and the methodology. There are some exciting possibilities here, in addition to the practical ones. The test itself, and by extension also the participants, would benefit greatly from increased functionality: for example, the addition of save buttons, spaced throughout the task, or perhaps the division of the exercise into multiple sections. This would not only allow the inclusion of more stimuli, but it might also encourage participants by reducing the pressure to complete the task in one session, since in a web-distributed survey, which does not provide a consistent survey environment, it cannot be guaranteed that they will be free of
interruptions for the entire time required. Additionally, if they found the task challenging, they could take a break.

Based on evaluation of the result processing methods in this study, the efficiency could be improved by having the submitted results written directly into a database, rather than a text file, and some further automation of the early analysis involving result charts and data processing: automatic peak identification, for example, and application of PPM statistical list refinement (Triplet or Quinlet Mode: see Chapter Four) which would change on the fly as results were submitted. From another point of view, the experimental process developed here is, and always was, intended to be more broadly applicable than to the present purpose. With that in mind, it may be possible to create a more streamlined survey creation tool with additional functionality that could be implemented for use with different languages: sliders in addition to or in place of checkboxes, to allow for the inclusion of filtered stimuli where syllables are not visible; a facility for prominence grading as well as identification; the ability to “draw what you hear”, and so on. Such a tool could also be designed to lay out, syllabify or segment stimuli automatically, making additions or changes to a large set much easier.

The second development to the project is expansion. As mentioned above, a larger response set would be very useful. This can be achieved by increasing both the stimulus and participant sets. One is, of course, easier than the other. Where the present project is concerned, the MAONZE database contains vastly more potential stimuli than were used here, from speakers of all vintages and genders: this would provide many further avenues for investigation, but on a very practical level, more syllables mean more potential PPMs. A larger stimulus set would also allow an attempt to correct for the large proportion of /a/ found in Māori, and to balance the vowel token numbers in, for example, vowel space plots. Controlling any of this is, again as mentioned above, very difficult given the unpredictability of PPM selection, but with sufficient potential tokens to begin with, a satisfactory minimum ought to be reached. More stimuli would also permit the investigation of more varied phonetic, grammatical, or intonational environments, or alternatively, the restriction of these to one specific, much larger, set.
The same is true of expansion in the participant set. With random participants, it is very difficult to place a long-term time limit on a study such as this, and participant recruitment may be more or less effective, either overall or at given times. In the present study, more and better marketing for the second web survey did contribute to the increase in participant numbers. In addition, there is the matter of participant type: some groups are easier to populate than others. As has been discussed throughout the description of the studies, finding fluent Māori speakers is not as straightforward as it might be, and this could apply equally to any specific participant set required for an experiment of this kind. The proficiency split used here would benefit greatly from the groups being about three times the size (that is, around 100 people each). Given that the present experiment took over a year to gain the number it did, finding so many more, within a reasonable time frame, would be a challenge without a captive participant group of some sort. This difficulty would be magnified in any language with a smaller number of speakers.

The proficiency split employed in this experiment, and the results from it, did confirm observations in the literature that proficiency makes a difference to perception results. However, some of the results, as mentioned above, also showed no significant difference between the groups. While the importance of native speaker evaluation in perception tasks is not disputed, in some cases, as with extreme minority languages, it may not be possible to obtain enough L1 or proficient speakers for experimentally adequate results. Additionally, the speakers involved in producing the stimuli may end up making judgements on their own or their contemporaries’ language, which can cause discomfort (Bird, 2011). In such circumstances as these, based on the results from this experiment, it is cautiously suggested that non-native or L2 participants may provide results for the perception test, provided that their language background and proficiency levels are known and controlled.

The third development concerns the analysis methods, and the set of cues for investigation. The three cues chosen for analysis in the present study (duration, vowel quality, and F0) were chosen based on several factors, including existing description and results from other studies of the language, as well as characteristics of the acoustic correlates in general (see Chapter Two). The vowel quality analysis would benefit from the inclusion of diphthong trajectory plots. As before, here, the diphthong investigation was precluded by token numbers; however, considering that
they constitute a proportion of the heavy syllables selected as PPMs by the participants, their features are worth exploring. A further correlate, loudness, in the form of intensity, was not included for a number of reasons; chiefly that it is not often considered a reliable cue, owing to the fact that it is highly subject to environmental, speaker, and recording effects. The recordings used in this study were made on different media, were of different ages, and the speakers were not recorded in a controlled environment, the result being that there is some variation in voice quality, along with coughing and background noise, for example. However, several studies have found that spectral tilt, balance or emphasis; that is, the effect of increased energy in the higher regions of the spectrum, is a more reliable measure of loudness and less subject to these effects (Heldner, 2003; Sluijter & van Heuven, 1996). This may be a worthwhile addition to the set of correlates analysed here, in order to attempt to account for PPMs that are unexplainable by acoustic means using the current set.

A further way in which the analysis can be expanded, and one that is particularly important for the investigation of change in the language, is further investigation of participant responses to the different speaker vintages in the MAONZE database. In the investigation presented here, it was seen that while the overall response to the three vintages that were included was rather similar for all participants, there was a definite difference in response to the PS-only stress category in the speech of the PY (young) vintage. Ten utterances from each vintage were available in the current study, but a much larger set, potentially chosen with respect to their PS-only syllable content, would provide the capacity for a thorough analysis of this phenomenon.

Theoretically, the concepts of stress and prominence, as defined for use in the study (see Chapter Two), along with the particular technical definition used for perceived prominences in the results (Chapter Four), proved adequate, even considering the required modification from Triplet to Quinlet Mode to account for proficiency group behaviour (Chapter Four). Methodologically, what has been discovered during this project development process is that participants must be kept comfortable, but it is very easy to underestimate their fortitude. Apparently, they also rather enjoy checking boxes. The scale of the study is a fine balance, because the participation and responses are necessarily unrestricted. However, the scale must unquestionably be expanded in terms of participants and stimuli to facilitate more in-depth analysis and interpretation of results.
Proficiency level can affect not only a participant’s response to stimuli, but also their behaviour in the perception task, and the result processing methods must be able to cope with this efficiently. The essential design of the experiment, given the purpose for which it was intended, ought to remain the same, in that it is free of assumptions.

7.9. Applications

The potential applications of an experiment or study such as this are varied. Its original purpose was motivated by a desire to add to description of the Māori language, and to investigate change that was being observed. However, with the development of the methodology came a hope that it might be useful in a wider sense. The first way in which a study like this is useful is simply illustration, and adding to a field of knowledge. It is a research tool. But with the description it provides, a number of other possibilities arise, not the least of which is helping to preserve or document a language that is vulnerable. It may help with language revitalisation in a number of ways. On a very simple level, it causes participation and interaction with the language on a level other than speaking and listening: it makes people think about speech. That may not always be a comfortable proposition for them, but it is important in keeping a language alive. Both language revitalisation enthusiasts and sticklers emphasise the need for good language role models, and the results from studies like this do have applications in teaching and learning. For example, they help to connect grammatical description and “rules” with the reality of speech, or, possibly, to explain an apparent lack of connection between those same rules and reality. As prosody is just as important as other language components in understanding and achieving naturalness in a second language, greater understanding of the features involved, and where they occur, can benefit both teachers and learners of pronunciation.

Finally, the move towards more electronic and technological involvement in our speech and language is already occurring. Robots speak; voices open doors. Automatic speech recognition is a fast-growing field. While artificial language may not replace real human speech, given its wide range of applications, there is a responsibility on language scientists to work towards making it as natural and intelligible – as useful – as possible. The information gained in studies like this can help: prosodic information increases naturalness and comprehensibility, and the acoustic analysis,
for example, being based in technology, provides results in values that machines can use. While
global languages like English receive a great deal of attention in refining the tools available both
to analyse them and to implement the results of that analysis in artificial environments, minority
languages may not. Applying studies like the one presented here to these minority languages
means that they too might be easily included in the e-linguistic revolution. If it is the way of the
future, they ought not be left behind: in a sense, it is a new form of language preservation and
revitalisation.

Language is a taonga, but examining a treasure does not lessen its value. The more we know about
it, the more it is protected and brought into the light, where we can enjoy it and use it and
maintain the spiritual connection to it felt by so many. This sort of analysis, going in under a
question mark; investigating what people hear; and finding the qualities involved, is an important
part of language maintenance. Languages are not suspended in time and space: because they are
inherently human, they change, and we must keep pace with that. It’s not enough to learn and
replicate – it’s important to ask questions. Curiosity is valuable: it leads to greater appreciation
and understanding, both of which facilitate learning.

The three experiments in this project have developed an approach to discovering how
prominence is perceived in Māori by modern listeners, both proficient and naïve, and what causes
them to perceive that prominence. The results have demonstrated divergence from the existing
descriptions, as discussed above. Initially, the question of a possible change in rhythm in
Māori was part of what prompted the study; inevitably, this led to a focus on prominence, as the
two are connected: rhythm is, at its core, patterns of prominence, and so knowledge of
prominence, both its location and its cues, is essential to any discussion of rhythm in a language.

What has been observed here is the importance of overlapping cues in relative salience; the
tendency (not an absolute) of higher proficiency in a language to affect prominence
identification; and evidence that in Māori, greatest perceived prominence appears to be changing:
shifting in favour of word stress rather than phrase stress and above all, in favour of those
overlapping cues. This study, therefore, is a contribution towards greater discussion of the rhythm
question, in that it moves us closer to describing the rhythm; that is, the *patterns of prominence* in Māori.
Appendices

Appendix A. Ethics and Survey Material

A1. Online Participant Information Sheet with Ethics Approval and Consent Checkbox

Project Title: Perception of Prosodic Prominence in Māori

Participant Information & Consent

Principal Investigator: Dr. Catherine Watson

My name is Laura Thompson, and I am conducting this survey as part of my research towards a PhD degree in the Department of Electrical & Computer Engineering at the University of Auckland, supervised by Dr. Catherine Watson. My research is on the analysis and description of prosody (rhythm, stress and/or intonation) in the Māori. Māori is the language of the indigenous people of New Zealand. It is a Polynesian language with five vowels and ten consonants. It is currently spoken by about 157,000 people.

In particular, I am interested in learning about how people hear prosodic prominence in Māori: in other words, what patterns they hear, and what makes some syllables stand out over others. In this study, I am testing which syllables in a word or sentence stand out to listeners. The results will provide insight into the patterns mentioned above, and allow me to investigate features of speech that are important in creating this effect in Māori specifically.

I would like to invite you to take part in this study. The exercise consists of an online questionnaire and survey. It's designed to take roughly 30 minutes to complete, but this will vary from person to person. There is no time limit.

The questionnaire asks for basic information such as your gender, age group and language proficiency, that will help me interpret the results. In the survey, you will be asked to listen to a number of short recordings in Māori. It doesn't matter if you know the language well or have never heard it before - we are simply interested in what you hear.

You will see the words in each recording on the screen, divided into syllables, and for each recording you will be asked to check boxes for the syllables that stand out to you as you listen. There are no 'right' or 'wrong' answers to this exercise: you may check as many or as few boxes as you feel are necessary, and you may play each recording as many times as you need to.

No information that could identify you personally as a participant is required by this exercise. On our side, no record associating you with your answers will be kept. Upon submission, the results of the exercise will be automatically processed and deleted. However, as with any
data collection or internet-based activity, there is no absolute guarantee of anonymity. Though extremely slight, the possibility of a third party identifying you from your answers does exist.

You can withdraw your participation, including withdrawal of any information you have provided, at any time until you click the 'Continue' button at the end of the survey. After that, because the survey is anonymous, the information cannot be retrieved. If you wish to withdraw from the survey, simply close your browser.

When analysis is complete, the results of this research will be available to all participants. Please send me an email if you would like an electronic copy to be sent to you. You are more than welcome to make this request, but you are advised that doing so will mean that, while your answers may well remain anonymous, the fact that you took part in the study will then be revealed.

On this webpage, there is a consent checkbox. From your checking this box and then submitting the survey electronically, it will be understood that you are 16 years of age or older; that you have read and understood the information supplied describing the aims, content, and terms of the project; that you agree to participate in the project under those terms, and that you consent to publication of the results of the project with the understanding that anonymity will be preserved.

Thank you very much for your time and help in making this study possible. Either I or Dr. Watson will be pleased to discuss any questions or concerns you may have about participation in the project. Our contact details can be found below.

Laura Thompson ([email addresses supplied])
Department of Electrical & Computer Engineering, University of Auckland
Supervisor: Dr. Catherine Watson ([email address supplied])
Head of Department: Professor Allan Williamson

APPROVED BY THE UNIVERSITY OF AUCKLAND HUMAN PARTICIPANTS ETHICS COMMITTEE on April 14, 2010 for (3) years, Reference Number 2010/161. For any queries regarding ethical concerns you may contact the Chair, The University of Auckland Human Participants Ethics Committee, The University of Auckland, Office of the Vice Chancellor, Private Bag 92019, Auckland 1142, New Zealand. Telephone +64 9 373 7599 extn. 83711.

☐ I am 16 years of age or older. I have read and understood the information above describing the aims and content of the project. I understand that, by submitting this survey electronically, I agree to participate in the project under the terms indicated in the information supplied.
A2. Participant Questionnaire Used in the Web Surveys

Questionnaire

Please fill out the form below.

1. Gender:  ○ Male ○ Female

2. Age group:  ○ 16-24  ○ 25-34  ○ 35-44  ○ 45-54  ○ 55-64  ○ 65+

3. Ethnic group: (Please check all that apply.)
   ○ NZ European
   ○ Māori  [ ] iwi:  [ ] Hapu:  [ ]
   ○ Other  Please specify:  [ ]

4. Place of birth:  [ ]

5. If you were not born in New Zealand, how long have you been here?  [ ] years

Please tell us about your language use.

6. What is your first language?  [ ]
7. Which of the following do you think best describes the type of English you speak? Please choose just one.
- New Zealand English
- British English
- Australian English
- American English
- Māori English
- Pasifika English
- South African English
- Something else (e.g. Chinese English, Indian English, Canadian English, etc.)

8. How well can you speak the following languages in day-to-day conversation?

<table>
<thead>
<tr>
<th>Language</th>
<th>very well</th>
<th>well</th>
<th>fairly well</th>
<th>not very well</th>
<th>words &amp; phrases</th>
<th>not at all</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Māori</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

9. Do you speak any other languages?  Yes  No

9a. What other languages can you speak, and how well can you speak them in day-to-day conversation?

<table>
<thead>
<tr>
<th>Language</th>
<th>very well</th>
<th>well</th>
<th>fairly well</th>
<th>not very well</th>
<th>words &amp; phrases</th>
<th>not at all</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

10. What language(s) do you or your family speak at home?  

Notes:

*Iwi / hapu*: These boxes allowed Māori participants to enter tribal affiliations.

*Marae*: Literally, the courtyard in front of a meeting house, but often also refers to the meeting house and associated buildings in the complex. New visitors are welcomed onto the *marae* in a special ceremony called *powhiri*.
Appendix B. Stimuli Used in the Pilot Study and First Web Survey

B1. Full Text of Letter and Story and Full Translation of Story

Notes:
1) The sentences used in the pilot study are underlined and the sentences used in the first web survey are in bold in both the original text and in the translation of the Māori story.
2) For both texts, and for the translation of the Māori story, I am indebted to the MAONZE team.
3) The story text is divided into grammatical phrases by vertical bars ( | ).

Dear Mum and Dad,

Hi. How are you? **Well, here I am in the big city. Although the weather is nice at the moment, the forecast is for hail.** But that should soon clear. I **bought a new coat, because they say it gets really cold. I have to stay at Auntie Deb’s house for now, but I’m hoping to get a flat soon. The trip up was great, even though it took about ten hours.** Well, I must go.

You know how rarely I write, but I will try to do better this year.

Love, Blair

---

I tīmata | te tau hou | ki te karakia. | I kitea | ko Matariki | e rere atu ana | i te pae.
| Ahakoa | ko te hōtoke, | he tino mao | te rangi. | Arā | mātou | e noho ana | i roto | i tō mātou wharenui, | i te taha | o te poutokomanawa. | I mui mai | te katoa | o ngā whānau | nō te haukāinga, | arā, | nō Nūhaka. | I waimarie hoki () mātou | ki te kite | i te whānau | o Rangi | i hau mai | i te moutere | o Tahiti. | I Mutu ana | ngā karakia | mō te tau hou, | i ka hau atu | te māpū | ki roto | ki te wharekai | kai ai. | I kainga | ko ngā kai pai | nō te ūkaipō | o tātou | te tangata, | arā, | nō Papa-tūā-nuku. | Ko te kūmara | nō te māra | o Koro Hōne, | te pūha | nō waho | o te whare | o Nana Mere, | me te maha | o ērā atu kai | i taona | ki roto | i tuku umu, | o tūira | ko ngā kai | nō te tai. | Ka tūtaki | ngā whānau | ki ērā atu | o ngā whānau. | Ka kitea | ko ngā pēpi ihu hūpē | o te kāinga, | rātou | ngā tamariki whakatoi, | ngā tama hoihoi, | ngā kōtiro whakahihi. | Ko irā kē | te pai | o te wāhi pēnei | i a Nūhaka, | he maha | ngā tūmomo rōpū. | Mutu ana | ngā mahi | o te ata, | ka pīrangi | te katoa | ki te haere | ki te awa | hoe ai. | Rere wahangū ana | te wai | me ngā pīpī rakiraki | e whakaakohia ana | e ō rātou mātua | ki te kaukau. | Ka tae atu | ngā whānau | ki te awa, | ka uru | ki ngā waka, | ā, | ka rangoa | ko ngā hoe | e pao ana | ki te mata | o te wai. | Ka haere (()) rātou | ngā tāne mā pū | ki te ngau poaka, | ki te whaiwhai tia. | Ka rangoa | ngā pū | e pūhia ana | i te awa. | Kātahi | ko ngā tāne mātua | ki te whakamahi pū | ko rātou. | I te awa, | i muri | i ngā otaota toetoe | e tipu roa ana, | ka tākaro | ngā tai tamariki | i te mūrere. | Ko tōku teina | te toa | o Nūhaka. | Engari | nā te hē | o tētahi | o ngā nekehanga, | ka hinga | ko ia | i te rā nei, | heoi | ka haere tonu | ngā kēmu. | Ka pau haere | te rā, | kotahi atu | te hoki | a te iwi | ki te marae. | Ka kawea | ko ngā karakia | o te pō | e ngā kaumātua, | ā, | i rangoa | ko te tūi | e
The New Year began with a service. Matariki was seen ascending from the horizon. Although it was winter, the sky was very clear. There we were sitting inside our meeting house, at the side of the main-post. The entirety of the families gathered from home, that is, from Nūhaka. We were fortunate in seeing the family of Rangi, which arrived from the island of Tahiti. When the service for the New Year was over, the group went into the dinner room to eat. We ate the good food from the ‘mother’ of us, human beings, that is, of the Earth. Kūmara from Koro Hōne’s garden, pūhā from outside Nana Mere’s house, with the abundance of other food which was cooked in [my oven, that is, the food from the sea]. The families met the other families. You could see the snotty-nosed babies of home, the cheeky kids, the noisy boys, the show-off girls. That’s the great thing of a place like Nūhaka, there are lots of types of group. When the morning’s activities were over, everyone wanted to go to the river to paddle. The water was flowing silently, as were the ducklings, which were being taught to swim by their parents. The families arrived at the river, they got into the canoes, and there were heard the paddles striking the surface of the water. The men who knew how to use guns went pig-hunting to hunt deer. The guns were heard being fired, from the river. What knowledgeable men at using guns they are! At the river, behind the toetoe reeds growing tall, the children play at mūrere. My sister is the champion of Nūhaka. But because of a mistake in one of the moves, she loses today. Anyway, the games continue. The day draws to a close, [and] the people return straight [back] to the marae. The evening prayers are led by the elders, and one could hear the tui with its beautiful colours, flying from the top of the tree outside the meeting house to his own home, to the Great Forest of Tāne.”

---

1 The Pleiades (star cluster): the sign of Māori New Year.
2 Sweet potato.
3 Grandpa.
4 Sow thistle (gen. Sonchus). A kind of flowering wild herb, the leaves of which are eaten by Māori as a vegetable.
5 A kind of long grass with tall, pale, featherlike flower stalks (gen. Austroderia).
6 A kind of draughts game.
7 See Footnote 2 above.
8 New Zealand wood pigeon: Prosthemadera novaeseelandiae.
9 Refers to Tāne-mahuta, who, in Māori mythology, is god/ancestor of forests and birds, and one of the children of the earth mother (Papa-tū-ā-nuku) and sky father (Rangi-nui) (Moorfield, 2013).
B2. Glossing of Māori Stimuli

Notes:
(1) Stimuli are alphabetically arranged.
(2) Vertical bars indicate grammatical phrase divisions.
(3) Biggs stress categories: Phrase & Word Stress (PWS); Phrase Stress (PS); Word Stress (WS).
(4) Detail of glosses used may be found in the List of Glosses on page xviii.

"Although it was winter, the sky was very clear."

Ahakoa ko te hōtoke, he tino mao te rangi.
although Foc. Det. winter Det. very clear Det. sky

"Although it was winter, the sky was very clear."

I kitea ko Matariki e rere atu ana i te pae.

Matariki new fly Dir. TA P Det. horizon

"Matariki was seen ascending from the horizon."

I tīmata te tau hou ki te karakia.

begin Det. year new P Det. service

"The New Year began with a service."

Ka pau haere te rā, kotahi atu te hoki a te iwi ki te marae.

exhaust go Det. day one Dir. Det. return P Det. people P Det. marae

"The day draws to a close [and] the people return (straight) to the marae."

Ka tūtaki ngā whānau ki ērā atu o ngā whānau.

meet Det. family P Det. Dir. P Det. family

"The families met the other families."

Ko tōku teina te toa o Nūhaka.

Det.of.1Sg. sister Det. champion P Nūhaka

"My sister is the champion of Nūhaka."
Appendix C. Continuous Speech Stimuli from the Main Experiment

Notes:
(1) Stimuli are alphabetically arranged; buffer stimuli (not analysed) are marked.
(2) Vertical bars ( | ) indicate grammatical phrase divisions.
(3) Biggs stress categories are marked on syllables as follows: Phrase & Word Stress (PWS); Phrase Stress (PS); and Word Stress (WS).
(4) Detail of glosses used may be found in the List of Glosses on page xviii.

Anā noa atu ngā mea pai i [sic] i tutuki i a au i te kura ngā mea kino rānei.

Interj.  long ago  Det. thing  good  [err.]  TA done  P Poss. 1Sg.  P Det. school  Det. thing  bad  or

“So those are the good things that I did at school - or the bad things!”

E mihi ana ki te whenua, e tangi ana ki ngā tāngata.

TA greet  TA  P Det. land  TA call  TA  P Det. people

“Greetings to the land, and calling out to the people.”

Ehara i te mea nā te taha nā ki te Māoritanga nei.


“It wasn’t as if it were on the Māori side. (Refers to the ‘positions’ in the previous sentence.)”
Engari ko taku pirangi tērā ki te hoki ki runga i ngā marae.

however Top. my want that Inf. return P on P Det. marae

“However, that’s what I want: to return to the maraes.”

He tuakana teina ēhau kei konei?

Indef. older siblings younger siblings do you have P here

“Do you have siblings here?”

I kīia ai e ia he tangata i tū ai i ngā tūranga nunui.

TA say.Pass. Ana. TA 3Sg. be as person TA stand Ana. 3Sg. P Det. position big

“He is said to have been a man who held important positions.”

I te mea nāna i whakaako mai ngā kupu ki ahau.

because Ag.Emph.3Sg. TA teach Dir. Det. word P 1Sg.

Because it was he who taught the words to me.
I tū ai hei māngai he kanohi mō tō tātou Ariki-nui i runga anō i tana hiāhia.

"I stood as a representative and as a face for our Queen, according to her wishes."

Ka mutu te ao i runga i tērā kōrero.

"[It's not as if] the world will end because of those words."

Ka tangi te kuia nei, kore rawa i tae ki whare kōti.

"This old woman was crying because she just could not get to the courthouse."

Ka tū koe ki tēnā taha, ka karanga atu, ka ō mai ko ia tēnā.

"You stand on that side and call out, and when you hear a reply, that will be him."
Ka whakatūngia hoki he whare mōna ki reira tū ai.

KA build.Pass. also Indef. house for.3Sg. P there stand Ana.

"A house was also built to stand there for him."

Kātahi ka kīia hei tino tangata nui.

Kātahi ka kīia hei tino tangata nui.

"And so, it is said, he was a very important person."

Kei te anga te kūwaha ki te uru, kei konā ngā tāngata i kite ai ahau.

Kei te anga te kūwaha ki te uru, kei konā ngā tāngata i kite ai ahau.

"The doorway faces towards the west, where there were the people that I saw."

Ko aua pukapuka e toru ngā wāhanga.

Ko aua pukapuka e toru ngā wāhanga.

"Those books had three parts."
Ko ēnei tonu tētahi o ngā take nui whakaaro ai au ki te haere ki reira.

“These, indeed, are one of the main reasons why I thought to go there.”

Ko mātou tērā kua uru atu ki a ia tautoko ai he ahakoa te aha.

“It was we, who went in to support her [the Queen], no matter what.”

Koirā taku mahi i roto i te mahi tūruhi he kapahaka.

“And so, my work in the tourist industry was kapahaka.”

Kua piri tonu ngā tāngata te nuinga o ngā Māori ki ngā āhuatanga ki ngā tikanga o te Māori.

“The people; the majority of Māori still cling to the aspects and customs of Māori.”

---

10 Māori cultural performance, with dancing and singing.
Kua tīmata au te tuhituhi i te pukapuka tuarima.

"I have begun writing the fifth book."

Me haere tonu atu koe kia kite i te wai i kite ai abau.

"You should go straight there, to see the water that I discovered."

Me tīmata atu taku kōrero i ngā rā ki muri rā.

"My story begins in days long gone by."

Na, ka haere tonu ngā mahi.

"So, the work goes on."
“And so, it turned out that there wasn’t a competition in ‘93.”

“And so, I alone will find myself a place to live.”

“And so, we were raised under those conditions.”

“So, there were no kapahaka competitions until ‘94.”
Because that is the thing which is called the ‘food cupboard’.” (Refers to a bay for fishing.)

“Of this thing; of the choices; the divisions; the grammatical analysis of related aspects.”

“When it got to the ‘80s, I finished working.”

“The second part relates to proverbs.”
“The first part relates to string games.”
Appendix D. Syllable Nuclear Duration Tables for Group Z and A, B, C

Table D1.

*Group Z Mean Nuclear Durations by Nuclear Vowel Category and Syllable PPM Status*

<table>
<thead>
<tr>
<th>Nucleus Vowel Type</th>
<th>Total Tokens</th>
<th>PPMs</th>
<th>SD (ms)</th>
<th>Non-PPMs</th>
<th>SD (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Total Tokens</td>
<td>Mean Dur. (ms)</td>
<td>SD (ms)</td>
<td>Total Tokens</td>
</tr>
<tr>
<td>Short Vowels</td>
<td>65</td>
<td>63.9</td>
<td>38.4</td>
<td>403</td>
<td>66.4</td>
</tr>
<tr>
<td>/ a e i o u /</td>
<td></td>
<td>47</td>
<td>116.6</td>
<td>42.2</td>
<td>49</td>
</tr>
<tr>
<td>Long Vowels</td>
<td></td>
<td>31</td>
<td>152.7</td>
<td>62.6</td>
<td>36</td>
</tr>
<tr>
<td>/ a: e: i: o: u: /</td>
<td></td>
<td>5</td>
<td>131.5</td>
<td>54.6</td>
<td>0</td>
</tr>
</tbody>
</table>

Table D2.

*Group A Mean Nuclear Durations by Nuclear Vowel Category and Syllable PPM Status*

<table>
<thead>
<tr>
<th>Nucleus Vowel Type</th>
<th>Total Tokens</th>
<th>PPMs</th>
<th>SD (ms)</th>
<th>Non-PPMs</th>
<th>SD (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Total Tokens</td>
<td>Mean Dur. (ms)</td>
<td>SD (ms)</td>
<td>Total Tokens</td>
</tr>
<tr>
<td>Short Vowels</td>
<td>32</td>
<td>58.5</td>
<td>21.2</td>
<td>436</td>
<td>66.7</td>
</tr>
<tr>
<td>/ a e i o u /</td>
<td></td>
<td>31</td>
<td>111.2</td>
<td>49.8</td>
<td>65</td>
</tr>
<tr>
<td>Long Vowels</td>
<td></td>
<td>17</td>
<td>142.3</td>
<td>49.2</td>
<td>50</td>
</tr>
<tr>
<td>/ a: e: i: o: u: /</td>
<td></td>
<td>4</td>
<td>135.7</td>
<td>62.2</td>
<td>1</td>
</tr>
<tr>
<td>Diphthongs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/ ae ai ao au ei oe oi ou /</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Long Diphthongs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/ aai aao /</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table D3.

*Group B Mean Nuclear Durations by Nuclear Vowel Category and Syllable PPM Status*

<table>
<thead>
<tr>
<th>Nucleus Vowel Type</th>
<th>PPMs</th>
<th>Non-PPMs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total Tokens</td>
<td>Mean Dur. (ms)</td>
</tr>
<tr>
<td>Short Vowels /a e i o u/</td>
<td>81</td>
<td>60.5</td>
</tr>
<tr>
<td>Long Vowels /a: e: i: o: u:/</td>
<td>38</td>
<td>112.1</td>
</tr>
<tr>
<td>Diphthongs /ae ai ao au ei oe oi ou/</td>
<td>31</td>
<td>151.7</td>
</tr>
<tr>
<td>Long Diphthongs /aai aao/</td>
<td>5</td>
<td>131.5</td>
</tr>
</tbody>
</table>

Table D4.

*Group C Mean Nuclear Durations by Nuclear Vowel Category and Syllable PPM Status*

<table>
<thead>
<tr>
<th>Nucleus Vowel Type</th>
<th>PPMs</th>
<th>Non-PPMs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total Tokens</td>
<td>Mean Dur. (ms)</td>
</tr>
<tr>
<td>Short Vowels /a e i o u/</td>
<td>79</td>
<td>64.9</td>
</tr>
<tr>
<td>Long Vowels /a: e: i: o: u:/</td>
<td>35</td>
<td>115.8</td>
</tr>
<tr>
<td>Diphthongs /ae ai ao au ei oe oi ou/</td>
<td>36</td>
<td>152.0</td>
</tr>
<tr>
<td>Long Diphthongs /aai aao/</td>
<td>5</td>
<td>131.5</td>
</tr>
</tbody>
</table>


