PROSODIC CLUES IN LANGUAGE RECOGNITION:

HOW MUCH INFORMATION DO LISTENERS NEED TO IDENTIFY MĀORI AND ENGLISH?[1]

Catherine I. Watson: *Department of Electrical and Computer Engineering, The University of Auckland <c.watson@auckland.ac.nz>*

Jeanette King: *Aotahi School of Māori and Indigenous Studies, University of Canterbury, <j.king@canterbury.ac.nz>*

Stephen Bier: *Department of Electrical and Computer Engineering, The University of Auckland <s.bier@auckland.ac.nz>*

Margaret Maclagan: *Department of Communication Disorders, University of Canterbury <margaret.maclagan@canterbury.ac.nz>*

Ray Harlow: *24 Marama St, Hamilton <rharlow@waikato.ac.nz>*

Laura Thompson: *Department of Electrical and Computer Engineering, The University of Auckland <la.thompson@auckland.ac.nz>*

Peter Keegan: *Faculty of Education, The University of Auckland <p.keegan@auckland.ac.nz>*

Abstract

As part of an investigation into the mutual influence of Māori and New Zealand English phonologies into the 21st century, we report on three experiments designed to test whether the languages can be identified from purely prosodic cues. In the first experiment, listeners with varying degrees of exposure to the Māori language heard 15-second stretches of low-pass filtered (400Hz) natural speech from three speaker groups (Historical Elders, Present-day Elders and Young speakers). Listeners were able to identify English and Māori significantly above chance level. In the second experiment, listeners were able to identify the three groups of speakers when the stimuli were filtered to parallel a land-line telephone and to a lesser extent when the stimuli were low-pass filtered. Those listeners with greater exposure to Māori performed better. The third experiment used synthetic stimuli that retained only the pitch and/or loudness information from the original speech. Listeners were able to identify the two languages above chance level when pitch information was included.

1. Introduction

The MAONZE project (Harlow et al. 2009; http://www.ece.auckland.ac.nz/~cwat057/MAONZE/MAONZE.html) is investigating sound change over time in Māori, the indigenous language of New Zealand, which has been spoken in this country for some 800 years (Anderson 2009: 25–27). We have compiled the MAONZE database, which allows us to compare archival materials of Māori recorded mainly in the mid 1940s, with present day recordings (King et al. 2011).

The speakers from the database can be divided into three groups: Historical Elders (born mainly in the 1880s), Present-day Elders (born mainly in the mid 1930s) and Young speakers (born mainly in the 1980s). The three groups of speakers are born approximately 50 years apart so that overall they provide a depth of 100 years of apparent time. In total, we have investigated the speech of fifty-eight speakers, both men and women, with roughly equal numbers in each of the three groups (see King et al. 2011 for details). We have Māori and English recordings for most of the archival speakers and all the present-day speakers.

We have already demonstrated that there has been considerable change in the quality and quantity of Māori vowels through acoustic analysis of both the archival and present day recordings (e.g. Harlow et al. 2009, King et al. 2010). Although some of the changes (such as the fronting of /u:/ and /u/) are also compatible with Labov's internally motivated sound changes (Labov 1994), much of the sound change has been influenced by English, and we note that the vowels in the speakers' English have undergone the same vowel shifts as other NZE speakers (Watson, Maclagan, King, and Harlow 2008). The influence from English is perhaps to be expected, as although English and Māori are both official languages in New Zealand,[2] the use of the Māori language has declined considerably since the mid-1900s. Māori has been subject to massive revitalisation efforts since the mid 1980s (see e.g., Benton 1991a, b).

Recently the project has turned its attention to changes in prosodic aspects of Māori (Maclagan et al. 2009, Thompson et al. 2010). There is anecdotal evidence that the rhythm of Māori is changing. Recently, a Māori language commentator stated (without explicit phonetic analysis) that the 'euphony' of the language was changing. In addition various Māori elders have noted that 'the *mita* of the language has changed'. Whilst the precise meaning of 'euphony' in this context and the precise definition of *mita* are unclear, contextual information suggests that both include prosodic features such as rhythm, stress and pitch. In addition, *mita* particularly refers to dialectal differences (Waitangi Tribunal 2010:3). The quantitative changes in Māori vowels over time include a lessening of the length difference between the traditional long/short vowel pairs (Harlow et al. 2009), which must affect the timing of the language. Further, English rhythm is usually identified as being based on stress whereas Māori rhythm has been postulated to be based on the mora, with the mora being defined as one short vowel together with any preceding consonant (Bauer 1981). However, the role of the mora in Māori rhythm is not fully understood. No quantitative research has yet been done on prosodic aspects of Māori, so we do not yet know whether the changes we have observed have affected the intonation and timing of the language including its rhythm. All of this indicates that a focus on the prosody of Māori is timely.

A complexity in the situation is the existence of the variety of English called Māori English. This is currently the most rapidly expanding variety of New Zealand English (NZE) and has received considerable attention over the last 20 years (see Bell 2000; Holmes 1996, 1997, 2005; Maclagan, King and Jones 2003; Warren and Bauer 2004; Maclagan, King and Gillon 2008; Szakay 2008). Although Māori English started as an ethnic variety of English influenced by the Māori language (Benton 1966: 93; Mitcalfe 1967: 20), Richards (1970: 126) questioned whether such was still the case. With the extent of revitalisation of the Māori language since 1980 (see Benton 1991a, b) mutual influence between Māori and Māori English is again a possibility. Māori English is distinguished from general NZE quantitatively rather than qualitatively, in that the features which characterise Māori English such as use of high rising terminals in intonation, the pragmatic particle *eh*, fronting of the vowel /u/ and a distinctive rhythm, are found to a lesser degree in general NZE (on rhythm see Holmes and Ainsworth 1996; Warren 1998 and Warren and Britain 2000). Although some of the young speakers recorded by the MAONZE project spoke Māori English on occasion, none of the MAONZE recordings are typical examples of Māori English, and none of the examples in the experiments described here contained features of Māori English. In addition Watson et al. (2008) showed that the vowels for the English of these speakers were comparable to values found in other studies on NZE. Māori English is therefore not further considered in this paper, which will focus on the Māori language and non-Māori accented NZE.[3]

We next provide background on Māori phonology and on some of the ways in which rhythm has been investigated. We then present three experiments designed as an initial investigation into prosodic features of Māori. In the first experiment we ask whether listeners, both Māori and Pākehā with greater or less familiarity with the Māori language, can identify Māori and English when segmental information is removed by low-pass filtering. In the second experiment we ask whether listeners can identify whether a Māori speaker belongs to the MAONZE Historical Elders group, the Present-day Elders or the Young group using the low-pass filtered stimuli and stimuli filtered to match a land-line telephone. Experiments one and two both use filtered natural speech. Following on from the results of these experiments, in the third experiment we modify the natural stimuli from experiment one by using an artificial sound source so that there is no trace of voice quality left in the stimuli. Only the *F0* pitch contours and/or the RMS intensity contours[4] of the original speech extracts remain. We ask whether listeners can still identify which language is being spoken. There is an increasing influence of English on Māori, and many modern speakers of Māori are actually learners of it as a second language. Because of this, we speculated that listeners would find it easier to separate the two languages in the speech of the Present-day Elders than the Historical Elders, whose English may be influenced by their Māori. We expected the opposite to happen for the speech of the younger speakers and that their Māori could be identified as English.

2. Background

2.1 *Māori phonology*

The Māori vowel system is usually analysed in terms of five short vowels /i, e, a, o, u/, which can occur alone or in sequences, with sequences of like vowels, at least within morphemes, being usually realised as long monophthongs. We have shown that both the qualitative and quantitative distinctions between the short and long vowels are being lost, with the exception of /a:/~/a/ (e.g. Harlow et al. 2009, King et al. 2010). In terms of quality, the short vowels are becoming more peripheral in the first versus second formant vowel space so that they are totally within the space of their paired long vowels. In terms of quantity, the duration of the long vowels has reduced markedly while the duration of the short vowels has hardly changed, so there is little duration difference between some short versus long vowel pairs. Figure 1 plots the overall average durations of the short vowels and the individual durations of the five long vowels for the three speaker groups and shows the marked

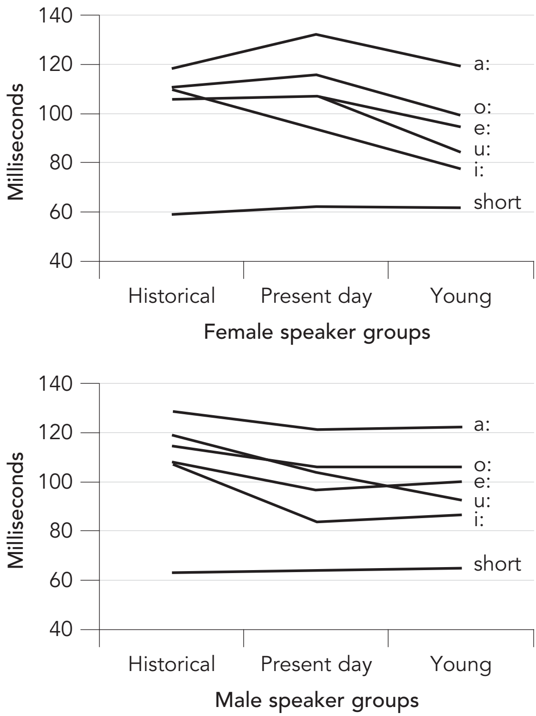


Figure 1: Changes in the duration of the long vowels over time for male and female speaker groups compared with the relatively unchanging mean short vowel duration.

An overall mean is provided for the short vowels because the individual differences among these vowels are much smaller than the differences among the long vowels.

reduction in the mean durations of each of the long vowels over time (for full details of changes to the short vowels as well as the long vowels, see Harlow et al. 2009:140 figure 4). For the present day Young speakers, the order of long vowel durations corresponds exactly with vowel height as it does in English, with the lower vowels longer than the higher vowels. This is not the case for the older speakers. These durational changes must inevitably be leading to changes in the rhythmic aspect of Māori prosody.

The syllable structure of Māori is described as (C)V(V(V)), since in addition to the short and long monophthongs and to the diphthongs, there is a set of long diphthongs consisting of a long /a:/, /e:/ or /o:/ followed by a short higher vowel, thus for instance *tāu* 'your sg.' contrasts with *tau* 'year'. As already noted, the rhythm of Māori is usually described as based on patterns of morae (Bauer 1981). However, whilst morae clearly played a part historically in Māori poetry (Biggs 1980; McLean 1981: 53–63) and still play a grammatical role in modern Māori, their role in stress and rhythm is somewhat less clear. In contrast, English rhythm is based on the patterns of stresses. Some vowels are stressed, and others reduced, leading to a distinctive rhythmic pattern. Traditionally, Māori unstressed vowels are not reduced, in either quality or quantity (Bauer 1993). Current MAONZE analyses, however, are indicating greater centralisation of unstressed vowels for modern speakers (Kaefer et al. 2010), which again will lead to changes in the prosody of Māori.

2.2 *Analysis approach*

We started our analysis of Māori prosody with a consideration of rhythm, which has been analysed by various means. Our initial focus was on the Pairwise Variability Index (PVI) (Grabe and Low 2002), a popular 'rhythm metric', which compares the durations of adjacent pairs of vowels in spoken text. Languages with stress induced reduction of vowels such as English tend to have a higher PVI value, while languages which do not have such vowel reduction, like Spanish, tend to have lower PVI values. Japanese, a language in which morae may play a part in rhythm (see for example Port et al. 1987; Beckman 1982; Warner and Arai 2001), lies closer to Spanish than to English. PVI is a strictly phonetic measure, which is applied to a language without any phonological analysis. In addition, most analysis has involved read texts (Grabe and Low 2002) or has compared different varieties of the same language (Warren 1998; Szakay 2008).

PVI proved ineffective for analysing Māori because the large number of vowel clusters in Māori led to a great deal of variation in the duration of adjacent vocalic segments. Consequently this led to Māori having a high PVI, which aligned it with languages like English (Maclagan et al. 2009) whereas the expectation had been that Māori would have a low PVI, like Japanese. PVI also did not pick up any changes in Māori over time, though the progressive shortening of long vowels would have been expected to have had an effect on this type of measurement. Any duration based rhythm metric is likely to be as ineffective for Māori as PVI. Dauer (1987) points out that duration is only one component of rhythm as a whole. Kohler (2009b: 6) reemphasises this and states, 'The various duration metrics have … been applied to language data without a clear conception of what speech rhythm is and how it may differ between languages in the way they bundle its physical exponents.' Listener perception is crucial in any rhythm analysis (see for example Kohler (2009b)) and alternative approaches to investigating rhythm involve perception as well as production. Kohler (2009a: 35) comments that judgments of rhythm in a language should be made by what he calls 'the competent language user' meaning, of course, the native speaker. This point is further emphasized by Arvaniti (2009). The reason for involving native speakers is that the listener's perception of linguistic rhythm is language-dependent. Different languages use different acoustic cues to encode rhythm or prominence, and whatever the language being heard, a listener will perceive rhythm or prominence according to the cues with which they are most familiar – that is, their own native cues (see for example studies by Beckman (1986) and de Jong (1994)).

It was because of the difficulty in identifying precisely what the acoustic correlates of rhythm are, that the MAONZE group moved to a more general consideration of prosody. Because of the close relationship of Māori with English over the last 170 years, it seemed important to start by seeing how well the two languages could still be identified by prosodic cues alone. The three experiments presented in this paper were designed as an initial investigation of this focusing on pitch and intensity which, as well as being basic aspects of prosody, are also usually important factors in rhythm. Other current work on prosody-related areas focuses on the identification of prominences in spoken Māori (Thompson et al. 2010) and on the acoustic analysis of unstressed vowels (Kaefer et al. 2010).

There have been many perception studies that have assessed listeners' ability to identify languages using prosodic features of speech. Komatsu (2007) provides a very good summary of these perception studies. There are a number of different ways that speech can be modified for these studies, such as low-pass filtering, use of laryngograph signals, resynthesised pulsetrains, linear-predictive filtering, and segmental resynthesis. These approaches vary in complexity and applicability for our data set (for instance, the use of laryngograph signals recorded from the larynx is not possible, since such data is not obtainable for the Historical Elders). In this study we focus on two approaches: low-pass filtering of speech and harmonic sinusoidal resynthesis.

3. Experiments with natural speech, Experiments 1 and 2

3.1 Aims

Experiment 1 was designed to test the hypothesis that Māori and English could be identified when most of the segmental information in the speech signal was removed by low-pass filtering. We hypothesised that the two languages should still be identifiable, and that greater exposure to the Māori language would result in greater accuracy of identification. Experiment 2 was designed to test the extent to which listeners were aware of change over time in the Māori language. We hypothesised that listeners would be able to identify the three speaker groups from unfiltered Māori speech. We also tested whether such identification would be possible with the filtered Māori speech used in Experiment 1.

3.2 Methodology

The first two experiments used natural speech stimuli that were taken from the MAONZE male interviews. Continuous extracts of approximately 15 seconds duration in both English and Māori were chosen from five speakers in each of the three groups (Historical Elders, Present-day Elders and Young speakers). One extract was chosen per speaker. The stimuli were free from hesitations or inappropriate pauses and were chosen so that the topic material would not automatically point to one of the three speaker groups. The same speakers from each group supplied the English and Māori extracts. The English extracts were all typical NZE speech. None had marked Māori English features. The recordings of the historical speakers were made by the Mobile Broadcasting Unit of the New Zealand Broadcasting Service, were band-limited and had an audible hiss due to the recording equipment of the day (King et al. 2011). In contrast, the recording methodology for the present day speakers was standardised and digital (op. cit.). The present day recordings had a bandwidth of 0–10 kHz, but the historical recordings had an effective bandwidth of

0–4 kHz. To ensure listeners were not making a judgment based on the nature of the recordings, we first removed the audible hiss from the historical recordings via the audio tool Audacity (http://audacity.sourceforge.net/).

Low-pass filtering of speech is considered one of the most convenient ways of creating stimuli that remove segmental features of speech but retain its prosodic features (Komatsu 2007). We followed Szakay's (2008) approach and low-pass filtered the extracts at 400 Hz (with 50 Hz smoothing) in Praat (Boersma and Weenink, version 4.125 or later). The original recording levels varied, and low-pass filtering reduces the loudness of the resulting stimuli. The filtered extracts were therefore intensity scaled in Praat to an average intensity of 70dB. In Experiment 2 listeners were asked to identify the speaker groups. To minimise the recording differences, we band-pass filtered all the excerpts from 300–3000 Hz to match the bandwidth of a land-line telephone, and again intensity scaled the resulting stimuli in Praat to an average intensity of 70dB. We refer to these stimuli as 'telephone' stimuli. The stimuli for both experiments were concatenated in a random order, with each extract preceded by a non-filtered number. In Experiment 1 we used both the Māori and the English extracts for a total of 30 stimuli. Experiment 2 used only the Māori extracts, but also had 30 stimuli because we used both telephone and low-pass filtered extracts. Practice extracts from speakers not included in the experiment preceded each experiment so listeners knew what the stimuli would sound like. The low-pass filtered stimuli were presented to the listeners first to avoid any transfer of learning from the telephone speech.

There were three different listener groups with different levels of exposure to Māori (see Table 1). The group with the least exposure to Māori (YP in Table 1) were students in a speech-language therapy programme. These students were mainly Pākehā (New Zealanders of predominantly European extraction), and while they were familiar with many Māori lexical items, they had limited Māori language skills. The second set of young listeners were students of Māori, many of whom were ethnically Māori (YM in Table 1). They were competent second language users of the language. The final group of listeners were Māori elders (EM in Table 1), whose first language was Māori. This group would have heard speech similar to that of the Historical Elders in their childhood, as the grandparents of these EM participants would have belonged to the same age group as the Historical Elders. These participants were also contemporaries of the Present-day Elder speakers and of the same generation as the grandparents of the Young speakers. The smaller cohort of the older listeners reflects the fact that it is hard to find participants in that age

Table 1: Listener Details for Experiment 1

LISTENERS	#	MEAN AGE	SD	RANGE
YP	40	23.5	9.23	17–58
YM	31	24.9	7.86	19–49
EM	9	66.1	7.12	55–75

YP = Speech-language therapy students, mainly Pākehā. The median age of participants in this group was 19. We therefore decided to retain the name 'young' even though the group included 3 older students.

YM = Students of Māori. The median age of participants in this group was 22. We therefore decided to retain the name 'young' even though the group included 3 older speakers.

EM = Māori elders. The median age of participants in this group was 69.

group. The same three categories of listeners participated in both Experiment 1 and Experiment 2, with Experiment 2 being conducted several months after Experiment 1. Because of availability, the listeners in YM category changed between Experiment 1 and 2, but no changes were made for the listeners in the YP and EM categories. A preliminary version of Experiment 1 was reported in Maclagan et al. (2009) without the EM group results.

The WAVE PCM file containing the stimuli was presented via loudspeakers to the younger listeners in their usual university lecture rooms. Because there were only small numbers of suitable Māori elders who were in different parts of NZ, it was not possible to gather them together as we did for the young listeners. Nor was it possible to eliminate listeners with some degree of hearing loss. The stimuli were therefore played via headphones to the Māori elders to help compensate for their known hearing problems. Although headphones would have provided better listening conditions than loudspeakers in lecture rooms, we believe this would have been counteracted by the elders' sometimes marked hearing loss.

The listeners filled out background information based on Szakay (2008). This included questions asking for age, sex, ethnicity and familiarity with the Māori language, together with questions designed to gauge the listeners' Māori integration. These questions asked for the ethnicity of any partner and asked whether/how often the listeners participated in activities such as watching Māori TV or listening to Māori radio or visiting marae (Māori meeting houses and associated facilities). Responses from the non-New Zealanders in the YP group are not included in the results presented here. Listeners were asked to make forced choice responses to the experimental stimuli. For Experiment 1, they were asked to tick one of two boxes, Māori or English, for each stimulus.

For Experiment 2, they were given three choices: Historical Elders, Modern *Kaumātua* (elders) or Young speakers.

For our statistical analysis we followed an overall approach similar to those of Barkat, Ohala, and Pellegrino (1999) and Komatsu, Arai, and Sugawara (2004) in that we first checked whether the overall results were different from chance and then examined details and interactions among the factors. We used a one-sample test of a binomial proportion to see whether the correct responses were greater than could be expected by chance.[5] In cases where the results were significantly different from chance, we calculated, for each listener group, the number of correct responses per stimulus then used ANOVAs to investigate significant interactions between the various factors involved. If post hoc tests were required we used Tukey HSD tests. Because the number of listeners in the various groups was different for both Experiments 1 and 2, we expressed the correct response data as proportions for these experiments before conducting the ANOVAs.

3.3 Results: Experiment 1

Overall 70% of the responses correctly identified the language type from the low-pass filtered stimuli. Because there were only two choices, Māori or English, listeners could have achieved 50% correct responses by chance. We used a one-sample test of a binomial proportion to assess the significance of the overall results ($p<0.001$).

Overall, all three listener groups were able to identify both languages significantly correctly above the rate expected by chance ($p<0.01$, see Table 2) as shown in Figure 2. In order to tease out the details, a three-way ANOVA was performed, with the proportion of correct responses per stimulus as the dependent variable, and listener group, speaker group and language as the

Table 2: Correct identification of language by the three listener groups in Experiment 1

LISTENER GROUP	# LISTENERS	# CORRECT	TOTAL	% CORRECT	SIGNIFICANCE
YP	40	798	1200	67%	** p<0.01
YM	31	669	930	72%	** p<0.01
EM	9	219	270	81%	** p<0.01

** the languages are identified at a rate significantly above chance level.

Statistical significance was assessed by a one-sample test of a binomial proportion (for details see note 5).

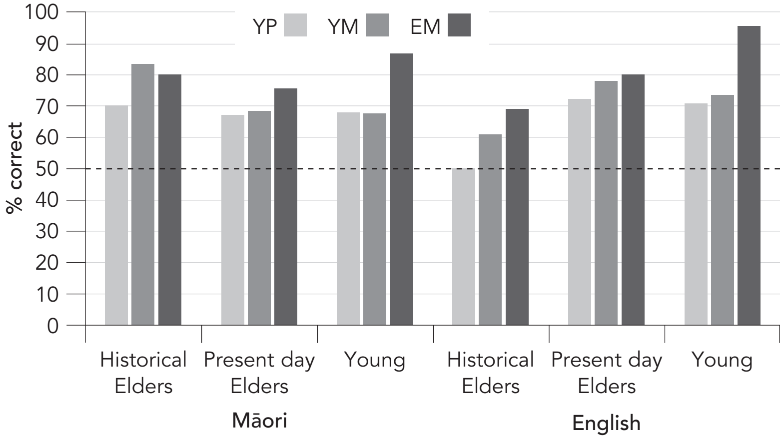


Figure 2: Results of Experiment 1 presented according to listener group and speaker group

The dashed line shows the rate that could be expected by chance.

three independent variables. We looked at main effects and then at two and three factor interactions.

The responses of the listener groups were significantly different (F(2,72) = 6.31, p<0.01), and post hoc Tukey HSD tests revealed that the EM listener group was significantly better at language identification than the YP group (p<0.01), it was also almost significantly better than the YM group (p=0.08). The listeners familiar with Māori were more successful than those who were not, and the listeners from the elders group (EM) were the most successful at the task. Across all listeners there was no significant difference between the correct identification rates for either language, nor did the speaker group have any significant impact on the correct identification rate.

There was a significant interaction between speaker group and language (F(2,72) = 5.46, p<0.01). Post hoc Tukey tests revealed that for all listener groups the English from the Historical Elders was not as well identified as that from the Māori Youth (p<0.01). The post hoc tests also revealed that all listener groups found the Māori from the Historical Elders easier to identify than their English (p<0.05). Here and in other statistical analyses, interactions other than those reported were not statistically significant.

The results from Experiment 1 thus show that all three listener groups could identify both Māori and English when segmental cues were removed.

Prosodic cues are therefore important for identification of NZE versus Māori. In our next experiment we sought to investigate whether listeners could identify the Māori spoken by the three different speaker groups. Our acoustic studies had demonstrated that there was a measurable sound change in Māori in terms of vowel quality and duration. We now wanted to see how salient this change was.

3.4 Results Experiment 2

It was more difficult identifying which speaker group the stimuli came from than identifying the language of the stimuli. For the 'telephone' condition, 61.8% of the responses correctly identified the speaker group and this was at a rate greater than chance ($p<0.01$). The low-pass filtered condition also allowed the speaker group to be identified above the rate of chance ($p<0.01$) even though only 41.8% of the responses were correct. Table 3 and Figure 3 give the mean correct responses for the three listener groups for the 'telephone' and low-pass filtered speech. All three listener groups were able to identify the speaker groups above chance levels for the 'telephone' speech, with the Māori listeners achieving higher results than the Pākehā. The Māori elders again achieved the highest results for the low-pass filtered speech, but

Table 3: Overall correct identification of speaker group by each of the listener groups from telephone speech in Experiment 2

LISTENER GROUP	YP	YM	EM	TOTAL
# Listeners	28	26	11	65
'Telephone' speech				
# correct	221	267	115	603
total	420	390	166	975
% correct	53%	68%	70%	62%
Significance	** p<0.01	** p<0.01	** p<0.01	** p<0.01
Low-pass filtered speech				
# correct	180	144	82	401
total	420	390	166	975
% correct	43%	37%	50%	43%
Significance	** p<0.01	ns	** p<0.01	** p<0.01

** the speaker groups are identified at a rate significantly above chance level.

Statistical significance was assessed by a one-sample test of a binomial proportion (for details see note 5).

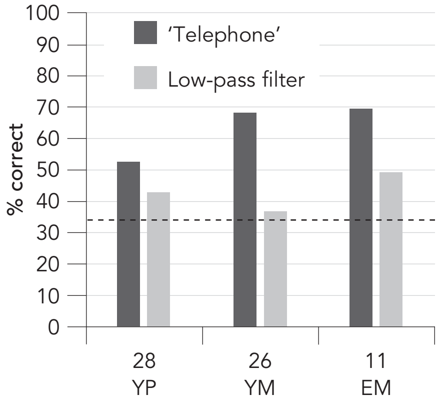


Figure 3: Results of Experiment 2 presented according to listener group and condition

The dashed line shows the rate that could be expected by chance.

in this condition, the YM group were not able to identify the speaker groups above the level expected by chance.

To investigate whether there were differences in the way the participants responded to the languages or speakers in the two different conditions ('telephone' or low-pass filtered) a three-way ANOVA was performed, with stimulus condition included as the third independent variable. As expected, the two conditions came up as significantly different ($F(1,72) = 25.83$, $p<0.001$) but nothing else was significant, nor were there significant interactions between conditions, language and listener or speaker groups. Further investigation of the low-pass filtered condition revealed no significant differences between listener group responses or between the speaker groups.

A two-way ANOVA was then performed on the 'telephone' condition responses, with the proportion of correct responses per stimulus as the dependent variable, and listener group and speaker group the independent variables. We looked at both main effects and at two factor interactions. There was a significant difference between the responses of the three listener groups ($F(2,36) = 5.23$, $p<0.01$). Post hoc Tukey tests revealed that YP listeners performed significantly worse than both the YM and EM listeners (YP-EM $p<0.05$, YP-YM $p<0.05$), however responses between the listener groups YM and EM were not significantly different.

Turning now to speaker group differences, and still focusing on the 'telephone' condition, when all the listeners were combined, there was little difference between the identification of the Historical Elders (at 56% correct) and the Elders (at 59% correct). However the Young speakers were correctly identified 71% of the time. The ANOVA revealed the responses to the three speaker groups were significantly different ($F(2,36) = 3.29$, $p<0.05$), and the Post hoc Tukey HSD tests confirmed that the Historical Elders were harder to identify than the Young group ($p<0.05$). This could indicate that the listeners, especially the Māori listeners, were sensitive to the identified changes over time in the Māori language. However, human listeners are good at identifying the age of speakers, using perceptual cues such as pitch, speech rate and voice quality (Schötz 2007). In this experiment, listeners might have been identifying a younger and older voice quality rather than responding to changes in prosody. In addition, the average pitch of the Historical Elders at 165Hz[6] was higher than the pitch of the other groups (127 Hz), and this could have made these speakers stand out and thus have aided listener identification.

4. Experiment with synthetic stimuli: Experiment 3

Experiment 1 suggested that listeners could identify Māori and English based on non-segmental cues. However there was a chance that some residual segmental information existed in the low-pass filtered stimuli. Although low-pass filtering to 'remove' the segmental aspects of speech has been a popular approach in perception studies (for example van Bezooijen and Gooskens 1999; Frota, Vigario and Martins 2002; Maclagan, et al. 2009), there is some debate about whether low-pass filtering is the most appropriate approach. As Ramus and Mehler (1999: 513) point out, low-pass filtering 'does not allow one to know which properties of the signal are eliminated and which are preserved'. They go on to acknowledge that whilst most segmental information is removed, it 'is only an approximation' (op. cit.: 513). In our own data there definitely is segmental information below 400 Hz, for example, the first formants of the high vowels in both languages (/iː/, /uː/, and /ɔː/ in New Zealand English and /iː/ and /uː/ in Māori). Further it is possible that a difference in recording conditions could be detected in the low-pass filtered speech between the historical mobile unit recordings and the present day recordings. Although the audio tool Audacity was used to reduce the audible hiss in the historical recordings (see Section 3.2), it is possible that the historical recordings still sounded different from the modern ones.

To overcome these problems, we created stimuli based solely on the pitch and intensity contours of the original speech utterances. In this third experiment we investigated whether listeners could identify Māori and English from stimuli which retained only the *F0* and/or the intensity of the original speech segments. This approach is similar to that taken by Ramus and Mehler (1999), Barkat, Ohala and Pellegrino (1999), and Komatsu, Arai, and Sugawara (2004), although the stimuli used in the present experiment were created in a different manner.

4.1 Method

Type of Stimuli

The stimuli generated were produced as a sum of harmonically related sinusoids, with the frequency of the lowest sinusoid being defined by the pitch contour of the original stimuli. The original pitch contour, *F0*, and intensity contour (calculated by the RMS method), were extracted using the ESPS algorithm via EMU speech tools (http://emu.sourceforge.net/). To produce the sinusoidal sequence, the pitch and intensity contours were first interpolated to the desired sampling frequency of the final stimuli (in this case, 22050Hz). Next sawtooth-like stimuli (*s[n]*) were generated with five harmonically related sinusoids by:

$$s[n] = A[n]\sum_{j=1}^{5} H_j \sin(2\pi \sum_{k=1}^{n} jF_0[k]),$$

where *A[n]* represents the time varying amplitude of the sinusoidal sequence and is derived from the interpolated RMS contour, *F0 [k]* is derived from the interpolated *F0* contour and *Hj* represents the proportion of the fundamental and each of the next 4 harmonics. For the stimuli that were generated it was decided to have a gradual roll off of the higher frequency components, so values of 1, 0.5, 0.4, 0.3 and 0.2 were chosen for *H1 (=F0)* through to *H5* respectively. This was used to model normal speech where the high frequencies are lower in intensity than the lower frequencies.

In an earlier version of this experiment we used stimuli that were frequency modulated sinusoids which contained the fundamental frequency only. However since the frequencies of these pure tones were always below 500Hz, the participants found the stimuli hard to hear, and the results were inconclusive. In retrospect, we should have realised the low frequency pure tones would be an issue. Pure tone audiometry has shown that auditory thresholds are much higher at frequencies below 500 Hz (British Standard

ISO 226 2003), and therefore adding harmonics within the 500-1000Hz range significantly increases audibility of the stimuli (Moore and Peters 1992). The complex tone used in Experiment 3 is very audible, and sounds a lot like muffled speech, as heard through a thick wall.

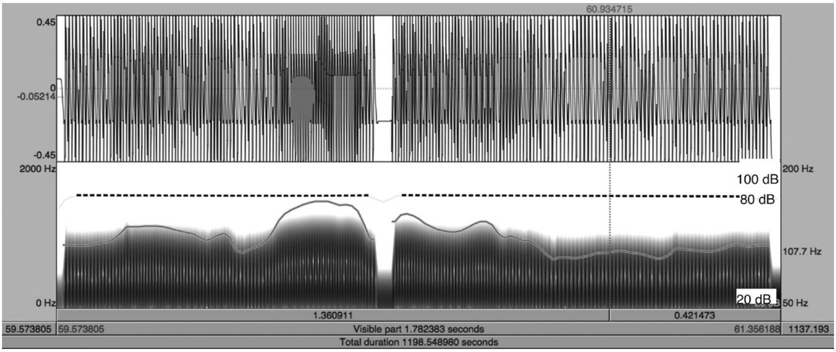
For Experiment 3 the stimuli were derived under three different conditions:

- with *F0* contours the same as the original speech extracts but with a constant amplitude, which was set at approximately 80 dB in order for the stimuli which still lacked high frequency components to be clearly audible.
- with the RMS (intensity) contours the same as the original speech extracts but with a constant *F0* (set at 120Hz which was close to the mean pitch in both English and Māori for the Young speakers and Present-day Elders used in Experiments 1 and 2. The original pitch for the Historical Elders was higher in both languages. See note 6.)
- both the *F0* and RMS contours of the original speech.

These three conditions will be referred to as the pitch (P), intensity (RMS), and pitch plus intensity (PRMS) conditions respectively. An extract from one of the stimuli produced for the P condition is given in Figure 4, where Figure 4a shows that the stimulus has constant amplitude and the *F0* contour is speech like (copied from the original extract used in Experiment 1). It can be seen that the waveform is sawtooth like (Figure 4b). The bandwidth of the signal is around 0-1500Hz. Figure 4b is an enlargement of the point around the cursor in Figure 4a.

Choice of Materials

The stimuli used in Experiment 3 were based on a subset of the utterances used in Experiment 1 and 2. Because there were three conditions in Experiment 3, we only used three utterances for each speaker group (in contrast to five in the earlier experiments) in order to keep the time to approximately 20 minutes. We chose stimuli from Experiment 1, using those speakers who were best identified as Māori when speaking Māori, and English when speaking English, with the proviso that we wanted the same speakers for both English and Māori. Since all the speech was taken from conversations, occasionally there were audible background sounds in the extracts. Since these sounds could inadvertently have been conflated with the actual speech in the calculation of the *F0* contours, care was taken to ensure that no extraneous sounds were a)



b)

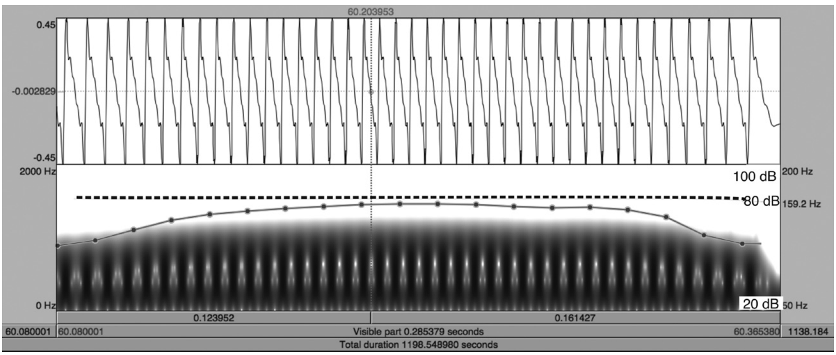


Figure 4: Extract from a synthetic stimulus with varying pitch and constant intensity used in Experiment 3

Extract a) shows a sample of just under 2 seconds. The lower pane shows that the pitch varies as in the original stimulus and the intensity line is straight showing that the intensity is constant. The frequency scale on the left is for the spectrogram and is from 0-2000 Hz. The frequency scale on the right is for the pitch and is from 0-200 Hz.

Extract b) is a magnified section of extract a), with the cursor at the same point in the wave. The top pane in this extract shows the sawtooth nature of the signal.

audible in the excerpts used in Experiment 3. Because we needed to avoid any background sounds, and because some speakers were identified very well in one language but not the other, we had to compromise and choose some speakers who were not unambiguously correctly identified in each language in Experiment 1.

All stimuli were around 15 seconds long. As with Experiment 1 and 2 the stimuli were concatenated in a random order, with each extract preceded by non-filtered numbers. In total, the participants listened to 54 stimuli. As in Experiments 1 and 2, the WAVE PCM file was played over loudspeakers to the listeners in their usual University lecture room. The form the participants filled out was identical to that used in Experiment 1, except for the number of items. As with the first two experiments the participants first filled out the questionnaire gauging their familiarity and exposure to spoken Māori. Then they listened to the stimuli and identified the language of each excerpt as being either English or Māori, ticking a box to indicate which language.

Participants

The participants in this study comprised 38 speech-language therapy students. This group is the same demographic as the YP group in Experiments 1 and 2, however it was drawn from a different cohort.

4.2 Results

Overall, listeners correctly identified the language of the stimuli above chance levels for the P (pitch only) and PRMS (pitch plus intensity) conditions (59% correct and 61% correct respectively, $p<0.01$, tested using a one-sample test of a binomial proportion (see note 5)). However for the RMS (intensity only) condition, the correct identification rates at 48% were not above chance (see Table 4 and Figure 5 for details). To investigate the matter further, and to establish whether there was any difference in responses between the three conditions of Experiment 3, and the low-pass filtered extracts from Experiment 1, we performed a three-way ANOVA. Here we had the correct responses of

Table 4 : Overall correct identification of the language from each of the stimulus conditions in Experiment 3

STIMULUS TYPE	# STIMULI	# CORRECT	TOTAL*	% CORRECT	SIGNIFICANCE
P	18	399	682	59%	** p<0.01
RMS	18	328	684	48%	ns
PRMS	18	412	680	61%	** p<0.01

* the total number of stimuli responded to differs in the three conditions because 2 listeners arrived late and missed the first 3 stimuli in the experiment. Because these were not heard, they were not counted as errors in the analysis.

** the languages are identified at a rate significantly above chance level.

Statistical significance was assessed by a one-sample test of a binomial proportion (for details see note 5).

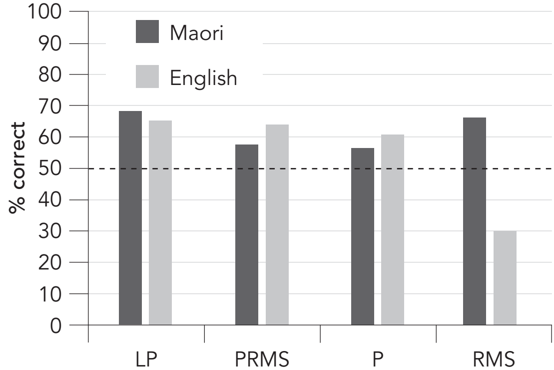


Figure 5: Results of Experiment 3 together with the YP results from Experiment 1 according to language and experimental condition

The dashed line shows the rate that could be expected by chance.

the YP group as the dependent variable, and speaker group, language, and stimulus type as the independent variables.[7]

Considering the main effects first, there were significant differences for the correct responses between the four different stimulus types, and between the two different languages ($F(3,60) = 5.52$, $p<0.01$, and $F(1,60) = 4.37$, $p<0.05$ respectively). Post hoc Tukey HSD tests revealed there was no difference in the rate of correct responses between the P and PRMS stimuli from Experiment 3 and low-pass filtered stimuli from Experiment 1 (henceforth referred to as the LP condition). However the rate of correct responses for the RMS condition was significantly worse than that for the LP condition ($p<0.01$), and it was almost significantly worse than for the PRMS condition ($p=0.09$). In addition the tests revealed that the participants were significant better overall at identifying Māori over English ($p<0.05$). However this is very likely due to their performance in the RMS condition, and will be discussed below.

When the interactions between the main effects were considered, the ANOVA revealed a significant interaction between stimulus condition and language ($F(3,60) = 7.44$, $p<0.01$). Post hoc Tukey HSD tests revealed that for English the RMS condition yielded significantly fewer correct responses than the other three conditions ($p<0.01$ for all). However there were no significant differences in the numbers of correct responses for the other three conditions for English (P, PRMS and LP), and for Māori there were no significant differences in the number of correct responses for any of the four conditions. The results also revealed that for the RMS condition Māori was correctly identified significantly better than English, but for the other three conditions there were no significant differences in the way the two languages were identified.

It is important to note that the fact that Māori was significantly better identified than English for the RMS condition is unlikely to indicate that RMS is a useful cue for identifying the two languages. The P and PRMS stimuli sounded like muffled speech, as did the LP condition, but the RMS stimuli with their constant pitch sounded very strange. It is far more likely that the participants found the RMS stimuli so strange that they attributed them to the language they were less familiar with, that is Māori. Extending this experiment to listeners who are also speakers of Māori will enable us to clarify this point.[8]

In summary the results from Experiment 3 suggest that for this set of listeners, pitch is the only cue necessary to separate English and Māori. The additional RMS information does not significantly increase the rate of correct language identification. The results for the P and PRMS stimuli were in fact not significantly different from those for the LP condition of Experiment 1.

5. Discussion

In this study we investigated whether three groups of listeners with varying exposure to the Māori language could identify Māori and English when segmental cues were removed from speech. We also investigated whether listeners could detect change in Māori over time. In Experiment 1 we established that Māori and English were identifiable on the basis of non-segmental cues, and listeners with more exposure to Māori did the task better.

In Experiment 2 we established that it is harder to identify speaker group than language. When segmental information was available in the ‘telephone’ speech, all listener groups could identify the speaker groups above chance level. Without segmental information in the low-pass filtered condition, younger Māori listeners could not identify the groups above chance level. The Pākehā and older Māori were able to do this, though their results were not as good as for the ‘telephone’ speech, with the older Māori listeners again performing best. When segmental information was present in the ‘telephone’ speech, we found that the Elder and Young speakers were identifiable, but

Historical and Present-day Elders were mistaken for each other. The older Māori listeners again performed best.

Of all our groups of listeners, it is the older Māori listeners who had most direct familiarity with all three speaker groups. They grew up listening to speakers of the same age group as our Historical Elders, their peers are the Present-day Elder speakers and their grandchildren are the Young speakers. In Experiment 2 they were best at identifying all three speaker groups in both telephone and filtered speech. Because of this, we suspect that the Elder listeners were able to use something other than just age differences in voice quality to identify the Historical Elders and the Present-day Elder speaker groups. They did appear to be using prosodic clues. Nevertheless it was possible that the participants, including the older Māori listeners, might have been reacting to age related voice quality changes rather than prosodic clues.

Although the results in Experiment 1 were very promising, low-pass filtering the stimuli is a very inexact method of removing segmental information (Ramus and Mehler 1999). We therefore adopted a different approach in Experiment 3 which enabled us to generate stimuli based only on the *F0* and/or intensity contours of the original utterances. We found that the listeners were still able to identify Māori and English at a rate greater than chance for the P and PRMS conditions, although there was no significant difference in the stimuli identified as correct across the three speaker groups. We also found that stimuli based on the RMS contour alone were not useful for identifying the two languages. However, since all listeners in Experiment 3 have so far come from the YP group, this needs to be further tested.

In the introduction we noted the increasing influence of English on Māori, and the fact that many modern speakers of Māori are actually second-language learners of Māori. We speculated that listeners would find it easier to separate the two languages in the speech of the Present-day Elders than the Historical Elders whose English may be influenced by their Māori, with the opposite happening for the younger speakers, whose Māori is likely to be influenced by their English. However our results only partially confirmed this. It did not prove to be the case for the Present-day Elders or for the Young speakers in any of the experiments. However, in Experiment 1, the Māori of the Historical Elders was identified correctly significantly more often than their English was, and their English was identified significantly less correctly than that of the Young speakers, demonstrating the effect of their first language Māori on their English.

We turn now to a comparison of Experiments 1 and 3. It is clear that listeners could use P, PRMS and LP stimuli to identify English and Māori speech. However we found no significant difference between the three stimulus types. This is important, because it shows that there is no bias towards either Māori or English for these three stimulus conditions (as there was a bias towards Māori in the RMS condition). Further it shows that the synthetic stimuli, which allowed us to focus on the importance of specific prosodic features, did enable identification of the two languages. From these experiments it is also quite clear that *F0* contour differences play a role in identifying the two languages. Preliminary investigations into the pitch contour of fluent Māori speakers suggest some different *F0* contour patterns between English and Māori. This is currently being further investigated.

In other language discrimination studies, stimuli based on *F0* only were not found to be good cues for discrimination. Ramus and Mehler (1999) found that listeners could not discriminate between English and Japanese from stimuli based on the vowel /a/ and intonation. Komatsu et al. (2004) found that *F0* and intensity played varying roles in the discrimination between English, Spanish, Chinese and Japanese, with stimuli containing pitch alone not being adequate to discriminate English and Spanish. Barkat et al. (1999) found both *F0* and intensity played a role in the discrimination between two Arabic dialects. Our findings are somewhat different. Experiment 3 showed listeners could identify Māori versus English with stimuli based on the *F0* contour and that the additional information that the intensity contour provided did not cause a significant increase in the number of correct responses. Differences between our findings and those of the other researchers could well be explained by differences in the pertinent perceptual features of the languages.

However it may well be that listener exposure to the different languages is also an important variable to consider. In Ramus and Mehler (1999) not much listener information was provided, but the participants were not first-language users of either language (English and Japanese). There was even less participant information in Komatsu et al. (2004). It can be inferred that the listeners were Japanese, but no details were given about the language(s) spoken by the participants. The listeners for Experiment 3 came from the YP category (speech-language therapy students, see 3.2), and, though not proficient in Māori, they had been exposed to it throughout their life. In order to increase our understanding of the differences in prosodic features between Māori and English, we intend to repeat Experiment 3 with listeners with a greater exposure to Māori. We anticipate the rates of correct language identification will increase with listeners with greater exposure to Māori, as they did in Experiment 1. Further, the identification rates may vary more across the three different speaker groups and between the two languages as the number of prosodic features in the stimuli increases. Barkat et al. (1999) did an Arabic dialect discrimination task with stimuli based on the *F0* and amplitude contours of the original speech segments. They found that Western Arabic listeners were able to correctly distinguish between two Arabic dialects at a rate greater than chance, but listeners with no exposure to Arabic could not.

6. Conclusions

By manipulating the information available in spoken stimuli, we isolated some of the prosodic clues that listeners need to identify extracts of English and Māori speech and showed that these languages can indeed be identified from prosodic features. We first tested low-pass filtered natural stimuli and then synthetic stimuli based on the *F0* contour and/or the intensity contour of a subset of the natural stimuli. Listeners could separate utterances in English from utterances in Māori, for low-pass filtered natural stimuli, and the synthetic stimuli that included the *F0* contour. In contrast the synthetic stimuli based only on the intensity contour were not sufficient. The *F0* contour, therefore, plays a role in identifying English and Māori. At the more subtle level of detecting language change between our three speaker groups, we found natural stimuli filtered to match speech heard over a landline telephone were sufficient for speaker group identification, but low-pass filtered stimuli were not as good.

We predicted that the language spoken by the Young speakers would be identified with less accuracy than that spoken by the Historical Elders and the Present-day Elders. This was not the case. This is a positive sign and indicates that the prosodic features of the two languages still differ, though precisely how they differ has not yet been established. However the English of the Historical Elders was identified least well in Experiments 1 and 3, presumably because it was influenced by their first language, Māori. We also predicted that listeners with greater exposure to the Māori language would perform best in Experiments 1 and 2 and this was the case.

Our study has shown that Māori and English are identifiable at a supra-segmental level in spite of over 150 years of contact and increasing changes in the pronunciation of Māori. We are now embarking on acoustic studies to quantify these differences. Our study is also a reaffirmation of the importance of native speaker perception in speech perception studies. The responses of the EM and YM listeners will be crucial in identifying the distinguishing prosodic cues between English and Te Reo Māori.

Notes

1 This research was supported by grants from the Marsden Fund of the Royal Society of NZ and the University of Canterbury. We also thank the participants in our various perception studies, and the students and staff of the Speech Pathology programs at the University of Canterbury and Auckland for supporting our study. We also acknowledge our many research assistants, and particularly Rosie Lamb who helped with coding and analysis of Experiment 3. We also wish to thank the reviewers who made helpful comments on earlier versions of this paper.

2 New Zealand has two legally defined official languages, Māori and New Zealand Sign language. English is *de facto* an official language and does not need such legal protection. See http://nzcurriculum.tki.org.nz/Curriculum-documents/The-New-Zealand-Curriculum/Official-languages (accessed 16 June, 2011).

3 The Historical and Present-day Elders are L1 speakers of Māori and L2 speakers of English. Their English contains L2 features such as stopping or affrication of dental fricatives and grammatical errors such as pronoun disagreements (*he* instead of *she* because Māori *ia* is gender neutral). These are not features of present day Māori English.

4 RMS stands for Root Mean Square, the standard method of calculating intensity.

5 A one-sample test of a binomial proportion gives the expected mean of the results as np, where n is the number of responses and p is the probability of a correct response. The standard deviation of the mean is the square root of npq, where $q = 1 - p$. For Experiments 1 and 3, there were two choices, Māori or English, so the probability was 0.5. For Experiment 2, there were three choices, Historical Elder, Present-day Elder or Young speaker, so the probability was 0.33. The experimental results were then compared against the normal distribution where values more than 1.96 sd above or below the mean are significant at the $p<0.05$ level and values more than 2.58 sd above or below the mean are significant at the $p<0.01$ level.

6 The average pitch of 165 Hz was derived from the extracts used in these experiments. Later analysis, using longer extracts of speech, found a somewhat lower pitch of 149 Hz for the Historical Elders. The longer extracts provided similar pitches (123 Hz) for the Present-day Elders and the Young speakers, so the Historical Elders still used a higher pitch, even when longer extracts were analysed. The pitch levels of the speaker groups were similar in Māori and English.

7 We considered that it was appropriate to compare responses from Experiments 1

and 3. Only the responses from the YP group in Experiment 1 were included in the comparison. The listeners in Experiment 3 were from the same demographic group as the listeners in Experiment 1 but from a different cohort. In addition, the stimuli in Experiment 3 matched a subset of those used in Experiment 1.

8 Because of the situation in Christchurch after the 2010 and 2011 earthquakes, we have not yet been able to carry out further listening tests. We plan to do so as soon as we can and will report the results.

References

Anderson, Atholl. 2009. 'Origins, Settlement and Society of pre-European South Polynesia.' In Giselle Byrnes (ed.) *The New Oxford History of New Zealand*. Melbourne: Oxford University Press. 21–46.

Arvaniti, Amalia. 2009. 'Rhythm, timing and the timing of rhythm.' *Phonetica* 66: 46-63.

Barkat, Melissa, John Ohala, and Francois Pellegrino, 1999, 'Prosody as a distinctive feature for the discrimination of Arabic dialects.' In the *Proceedings of the Sixth European Conference on Speech Communication and Technology (Eurospeech'99)*, 395–398.

Bauer, Winifred. 1981. 'Hae.re vs. Ha.e.re: A note.' *Te Reo* 24:31–36.

Bauer, Winifred. 1993. *Māori*. London and New York: Routledge.

Beckman, Mary E. 1982. 'Segment duration and the 'mora' in Japanese.' *Phonetica* 39:113–135.

Beckman, Mary E. 1986. 'Stress and non-stress accent.' *Netherlands Phonetics Archives 7*. Dordrecht: Foris.

Bell, Allan. 2000. 'Māori and Pākehā English: a case study.' In A. Bell and K. Kuiper (eds.), *New Zealand English*. Wellington: Victoria University of Wellington. 221–248.

Benton, Richard. 1966. *Research into the English Language Difficulties of Māori School Children 1963–1964.* Wellington: Māori Education Foundation.

Benton, Richard. 1991a. 'Māori English: a New Zealand myth?' In J. Cheshire (ed.), *English Around the World*. Cambridge: Cambridge University Press. 187–199.

Benton, Richard. 1991b. *The Māori Language: Dying or Reviving*. Hawaii: East-West Center.

Biggs, Bruce. 1980. 'Traditional Māori Song Texts and the "Rule of Eight".' *Pānui* 3: 48–50. Anthropology Dept, University of Auckland.

Boersma Paul, Weenink David. 'Praat: Doing phonetics by computer.' Last downloaded 4 August 2010. http://www.fon.hum.uva.nl/praat/.

British Standard ISO 226 2003, 'Acoustics – Normal equal-loudness level contours,' BSI, London, 2003.

Dauer, Rebecca M. 1987. 'Phonetic and phonological components of language rhythm.' *Proceedings of the 11th International Conference of the Phonetic Sciences (ICPhS)*, Tallinn. 447–449.

de Jong, Kenneth. 1994. 'Initial tones and prominence in Seoul Korean.' *OSU Working Papers in Linguistics* 43:1–14.

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

Grabe, Esther and Ee Ling Low. 2002. 'Durational Variability in Speech and the Rhythm Class Hypothesis.' In C. Gussenhoven and N. Warner (eds) *Laboratory Phonology* 7. Berlin: Mouton. 515–546.

Harlow, Ray, Peter Keegan, Jeanette King, Margaret Maclagan and Catherine Watson. 2009. 'The changing sound of the Māori language.' In J. N. Stanford and D. R. Preston (eds) *Quantitative Sociolinguistic Studies of Indigenous Minority Languages*. Amsterdam: John Benjamins Publishing Company. 129–152.

Holmes, Janet. 1996. 'Losing voice: is final /z/ devoicing a feature of Māori English?' *World Englishes* 15(2):193–205.

Holmes, Janet. 1997. 'Māori and Pākehā English: some New Zealand social dialect data.' *Language in Society*, 26(1):65–101.

Holmes, Janet. 2005. 'Using Māori English in New Zealand.' *International Journal of the Sociology of Language,* 172:91–115.

Holmes, Janet and Helen Ainsworth. 1996. 'Syllable-timing and Māori English.' *Te Reo,* 39:75–84.

Kaefer, Stephanie, Jeanette King, Catherine Watson, Margaret Maclagan, Ray Harlow and Peter Keegan. 2010. 'Evidence for the development of a centralised vowel in Māori and its effect on rhythm.' Paper presented at the 12th New Zealand Language and Society Conference, 22–23 Nov 2010, Auckland, New Zealand.

King, Jeanette, Margaret Maclagan, Ray Harlow, Peter Keegan and Catherine Watson. 2011. 'The MAONZE project: changing uses of an indigenous language database.' *Corpus Linguistics and Linguistic Theory* 7 (1):37–57

King, Jeanette, Catherine I. Watson, Margaret Maclagan, Ray Harlow and Peter Keegan. 2010. 'Māori women's role in sound change.' In J. Holmes and M. Marra (eds) *Femininity, Feminism and Gendered Discourse*. Newcastle upon Tyne: Cambridge Scholars Publishing. 191–211.

Kohler, Klaus J. 2009a. 'Rhythm in speech and language: A new research paradigm.' *Phonetica* 66(1-2):29–45.

Kohler, Klaus J. 2009b. 'Whither speech rhythm research?' *Phonetica* 66(1-2):5–14.

Komatsu, Masahiko, Tsutomu Arai, and Takayuki Sugawara, 2004, 'Perceptual discrimination of prosodic types and their preliminary acoustic analysis,' in *Proceedings of the 5th Annual Conference of the International Speech Communication Association (Interspeech 2004)*, 1280–1283.

Komatsu, Masahiko. 2007, 'Reviewing Human Language Identification.' In Müller, Christian. (ed) *Speaker Classification II, Lecture Notes in Artificial Intelligence 4441*, Springer-Verlag, Berlin. 206–228.

Labov, William. 1994. *Principles of Linguistic Change Volume 1: Internal Factors*. Oxford: Blackwell

Maclagan, Margaret, Catherine I. Watson, Jeanette King, Ray Harlow, Laura Thompson and Peter Keegan. 2009. 'Investigating Changes in the Rhythm of Māori over time.' *Proceedings of the 10th Annual Conference of the International Speech Communication Association (Interspeech 2009)*, Brighton. 1531–1534.

Maclagan, Margaret, Jeanette King, and Gail Gillon. 2008. 'Māori English.' *Clinical Phonetics and Linguistics*. 22(8):658–670.

Maclagan, Margaret A., Jeanette King, and Ivan Jones. 2003. 'Devoiced final /z/ in Māori English.' *New Zealand English Journal*, 17:17–27.

McLean, Mervyn. 1981. 'Text and music in "rule of eight" waiata.' In J. Hollyman and A. Pawley (eds). *Studies in Pacific Languages and Cultures in Honour of Bruce Biggs*. Auckland: Linguistic Society of New Zealand. 43–53.

Mitcalfe, Barry. 1967. 'Survivals of Māori in English.' *Education*, 16(8):20–22.

Moore, Brian and Robert Peters. 1992. 'Pitch discrimination and phase sensitivity in young and elderly subjects and its relationship to frequency selectivity,' *The Journal of the Acoustical Society of America*, 91: 2881.

Port, Robert F., Jonathan Dalby, and Michael O'Dell, (1987). 'Evidence for mora-timing in Japanese.' *Journal of the Acoustical Society of America*, 81, 1574–1585.

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

Richards, Jack C. 1970. 'The language factor in Māori schooling.' In J. Ewing and Jack Shallcrass (eds), *Introduction to Māori Education.* Wellington: New Zealand Universities Press. 122–132.

Schötz, Susanne. 2007. 'Acoustic Analysis of Adult Speaker Age.' In Christian Müller, (ed) *Speaker Classification I*, *Lecture Notes in Artificial Intelligence 4343*, Springer-Verlag, Berlin. 88–107.

Szakay, Anita. 2008. *Ethnic Dialect Identification in New Zealand: The Role of Prosodic Cues*. VDM Verlag, Germany.

Thompson, Laura, Catherine I. Watson, Helen Charters, Peter Keegan, Ray Harlow, Jeanette King and Margaret Maclagan. 2010. 'An experiment in *mita*-reading: investigating perception of rhythmic prominence in the Māori language.' In *The Proceedings of 13th Australasian International conference on Speech Science and Technology (SST2010)*, Melbourne.150–153.

van Bezooijen, Renée and Charlotte Gooskens. 1999. 'Identification of language varieties. The contribution of different linguistic levels.' *Journal of Language and Social Psychology*, 18: 31–48.

Waitangi Tribunal. 2010. Prepublication version of Chapter 5 of Waitangi Tribunal Report 262. Wellington: Author. Accessed 25 November 2010 from http://www.waitangitribunal.govt.nz/reports/downloadpdf.asp?ReportID={BF981901-5B55-441C-A93E-8E84B67B76E9}

Warner, N. and Takayuki Arai. 2001. 'Japanese mora-timing: A review.' *Phonetica,* 58:1–25.

Watson, Catherine I., Margaret Maclagan, Jeanette King and Ray Harlow. 2008. 'The English Pronunciation of Successive Groups of Māori Speakers.' *Proceedings of the 9th Annual Conference of the International Speech Communication Association (Interspeech 2008) incorporating the 12th Australasian International Conference on Speech Science and Technology (SST 2008)*. 338–341.

Warren, Paul. 1998. 'Timing patterns in New Zealand English rhythm.' *Te Reo* 41: 80–93.

Warren, Paul and Laurie Bauer. 2004. 'Māori English: phonology.' In Edgar Schneider, Kate Burridge, Bernd Kortmann, Rajend Mesthrie and Clive Upton (eds) *Handbook of Varieties of English*. Berlin and New York: Mouton de Gruyter. 614–624.

Warren, Paul and David Britain. 2000. 'Intonation and prosody in New Zealand English.' In Allan Bell and Konraed Kuiper (eds). *New Zealand English*. Wellington [N.Z.]: Victoria University Press. 146–172.