THE MAONZE CORPUS: TRANSCRIBING AND ANALYSING MĀORI SPEECH

Jeanette King[1], Margaret Maclagan[1], Ray Harlow[2], Peter Keegan[3] & Catherine Watson[3]

[1]University of Canterbury, [2]University of Waikato, [3]University of Auckland

Abstract

The MAONZE project investigates change over time in the pronunciation of the Māori language by comparing archival recordings of older Māori speakers born in the late 19th century with present-day recordings of both older and younger speakers. The background to the project and details on how the corpus of recordings was compiled are described in an earlier companion piece. This article describes the transcription and analysis protocols that have been employed in the project and gives an overview of some of the results of the analysis of vowels and consonants and the perception of prosodic cues.

Introduction

The aim of the MAONZE (Māori and New Zealand English) project is to analyse changes in the pronunciation of the Māori language over time by comparing archival materials of Māori born in the late 19th century and recorded, for the most part, in the mid 1940s, with present-day recordings of both elders and young adult speakers. The male and female speakers from the database can be divided into three groups: Historical Elders (born mainly in the 1880s), Elders (born mainly in the mid-1930s) and Young speakers (born mainly in the 1980s). In total, we have investigated the speech of 58 speakers, with roughly equal numbers in each group (see King, Maclagan, Harlow, Keegan, & Watson, 2011, for details). We have Māori and English recordings for many of the archival speakers and all the present-day speakers.

An earlier companion paper published in this journal gives details of the methodology and design of the MAONZE corpus (King, Maclagan, Harlow, Keegan, & Watson, 2010). The current paper gives details of how the recordings were transcribed and some of the research methods employed. This is followed by a brief overview of the results of a number of analyses from the MAONZE corpus. As well as analysis of changes over time in the pronunciation of vowels (both monophthongs and diphthongs) and some consonants, recent investigation focusses on changes in the rhythm of Māori.

Transcription

The previously published companion paper gave details on how the historical recordings were collected and how the present day recordings were made. Before any acoustic or auditory analysis of the speech of speakers could be undertaken it was necessary to make transcriptions of what was said on the recordings. Most of the transcriptions of the MAONZE database and the associated Tūhoe and Māori English databases were made by research assistants. Some recordings have also been transcribed by the project team, in particular the Māori recordings.

The transcriptions are made using the program Transcriber, a shareware computer program available in PC and Mac formats, which can be downloaded from http://trans.sourceforge.net/en/presentation.php. The advantage of using Transcriber is that not only is a transcript of the sound file produced, but the transcript is time aligned to the sound file. The time aligning can be seen in Figure 1, which shows a sample screen shot from one of the Transcriber files from this project. Below the waveform in the bottom part of the screen shot it can be seen how each line in the transcript is time aligned to the sound file. This means that it is easy to move to precise places within a sound file, a useful feature when dealing with the textgrids that are produced from the transcriber files (see following section). Time aligning is the pivotal feature which makes very powerful searches and analyses possible over large corpora using the LaBB-Cat software (http://onzeminer.sourceforge.net/). This is explained in more detail below.

Figure 1: Sample screenshot in Transcriber from transcription of a historical female elder.

For transcription, the MAONZE project follows protocols developed for the ONZE (Origins of New Zealand English) project (Gordon, Campbell, Hay, Maclagan, Sudbury, & Trudgill, 2004). For linguistic research purposes, transcriptions should include everything the speakers say, including hesitations, repetitions and false starts. Maclagan and Hay (2011) give a detailed account of different types of transcriptions and their uses. As can be seen in Figure 1, conventional punctuation is not used, with capitals only being used for proper names and the pronoun *I* in English. Various lengths of pauses are indicated with a full-stop, hyphen or double hyphen. Because of difficulties between programs and across PC and Mac computer platforms, with vowels with macrons becoming assorted strange symbols, the decision was made to represent the macrons of long vowels in Māori with a colon after the vowel, *Ma:ori*.

Acoustic analysis

We used Praat (Boersma & Weenink 2010, version 4.125 or later) to carry out the acoustic analysis. Where Transcriber only allows one level of textual notation, Praat allows many tiers to be added for recording analyses. The Transcriber files can be converted to Praat textgrids, and appropriate tiers added so that different analyses can be carried out. Appropriate utilities for converting Transcriber .trs files to Praat textgrids can be found on the web, including on the MAONZE website, http://www.ece.auckland.ac.nz/~cwat057/MAONZE/MAONZE.html. Praat Acoustic Analysis Software is a shareware computer program available in PC and Mac formats. Both research assistants and the research team were involved in the analysis of the sound files in Praat.

Often speech that was unclear when transcribed in Transcriber was easier to interpret when performing the analysis in Praat. For this reason those doing the analysis were given both the Praat textgrids and Transcriber (.trs) files and asked to alter text in both formats if mistakes were found or unclear speech could be deciphered. This was to ensure that the Transcriber files were as correct as possible, because ultimately it is these files, and not the textgrids, which are uploaded to LaBB-Cat and are available for searching.

The acoustic analysis in Praat for the English and Māori speech of each speaker included the following:

1. monophthongs – at least 30 stressed tokens for each vowel (where possible), from environments with surrounding consonants and not in word final positions. It is important that the tokens analysed are stressed, so that they are unambiguously tokens of the intended vowel, and there are no effects from unstressed vowels. Values for the first three formants and fundamental frequency (F_0) as well as length were taken. Unless unavoidable, no more than five tokens of any one word were used.
2. diphthongs – as per the monophthong analysis with measurements taken for both the first and second target.

Figure 2 shows a sample screen shot showing marking up of part of the Māori acoustic analysis for a present day female elder. The top part of the screen shot shows the sound waves with the dots showing the formant placements which are generated automatically by Praat. The figure shows how the data can be labelled on many levels in Praat. The first tier underneath the sound waves contains the time-aligned transcript. The phrase tier, here shown empty, was used on some textgrids to conduct an analysis in order to investigate rhythm.

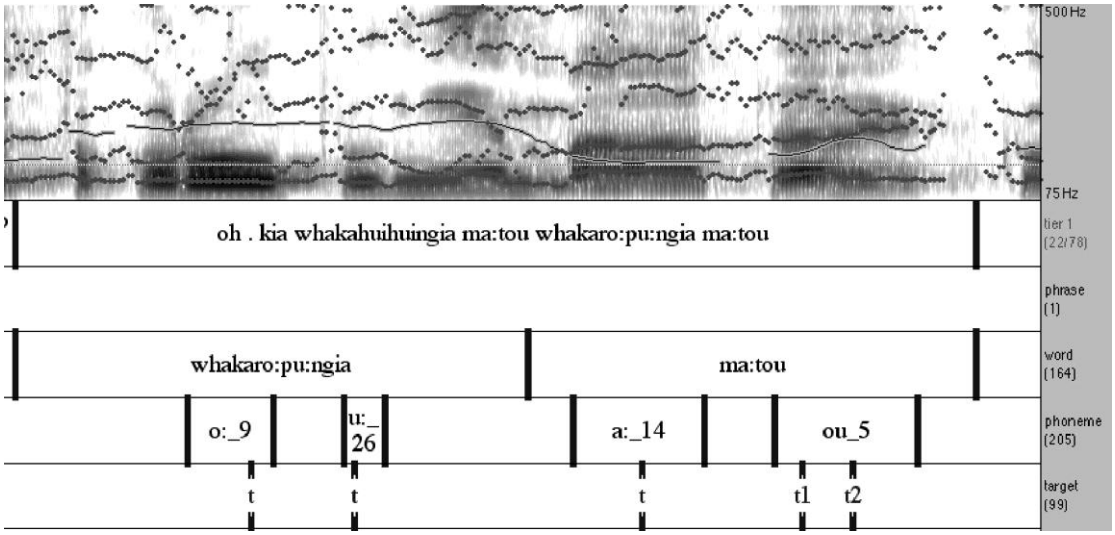


Figure 2: A sample screen shot in Praat showing part of the acoustic analysis for a present-day female elder.

The vowels for analysis are shown on the phoneme tier with the beginning and end of each analysed phoneme marked. Each vowel is labelled, here the /u:/ in the word *whakarōpūngia* has been analysed with this token being number 26. The target tier marks the point in the vowel where the formant and F_0 readings were taken. Similarly the /ou/ diphthong is the fifth to be analysed and t1 and t2 mark where the formant readings for the two target points in the diphthong were taken.

Formants were calculated using the default Praat settings (25 ms analysis frame, Gaussian window, 10 pole LPC filter). The formant positions were visually checked and corrections made to the analysis parameters as necessary. Measurements were taken during the steady state portion of the vowel. If there was no steady state, formant readings were taken at the F2 maximum (and F1 minimum) for front vowels, the F1 maximum (and F2 minimum) for central vowels and the F2 minimum (and F1 minimum) for back vowels. Two target measurements were taken for the diphthongs. Consonant transitions were included within vowel length measurements so long as vowel formants could be seen (that is, so long as there was voicing). For English, we avoided analysing tokens adjacent to /w/, /l/ and /r/ because of co-articulation effects. For Māori, we avoided tokens adjacent to /w/ and /h/ (which was often voiced and vowel-like). /r/ in Māori is flapped and did not affect the adjacent vowels to the same extent as the English approximant /r/.

The Māori analysis for all speakers also included the following analyses:

/t/ analysis – as part of an investigation into increasing rates of aspiration/ affrication the voice onset time (VOT) of up to 30 tokens of word initial /t/ were taken in the following six contexts: /ta:/, /ta/, /ti:/, /ti/, /tu:/, /tu/, where /a:/ and /a/ do not facilitate aspiration and the other contexts do.

Ka analysis – as part of an investigation into changes in the length of the verbal particle *ka*, formant values and length measurements were taken of up to 30 tokens of the /a/ vowel in this particle in contexts where the verbal material consisted of either two morae or more than two morae (a mora being defined as a short vowel plus any preceding consonant).

All formant and length measurements taken from the Praat textgrids were recorded in Excel. Statistical analysis has been performed with SPSS, SYSTAT 12 and R (http://www.r-project.org/).

MAONZE Miner

Once the files have been transcribed, the transcriber files are uploaded onto the MAONZE Miner server, which allows them to be easily searched and interacted with. MAONZE Miner is software which has been adapted from the ONZE Miner software (now renamed as LaBB-Cat) designed for the ONZE project by Robert Fromont and Jen Hay (http://onzeminer.sourceforge.net/ and Fromont & Hay, 2008).

It is possible to upload the sound files to the server with the transcriber files. However, because members of the MAONZE team are scattered round the country, we do not do this. Even though we downsample them (see King Maclagan, et al., 2010), the files are still large and they would take too long to access over the internet. Instead, we have produced DVDs of the sound files. Because of the terms of the University of Canterbury's agreement with Sound Archives / Ngā Taonga Kōrero we are unable to make copies of recordings obtained from them available to people outside the MAONZE team (this includes all the historical male recordings and many of the historical female recordings).

Figure 3 shows a sample screen shot from the selection page of MAONZE Miner where male speakers from Ngāti Porou have been selected. At this point speakers can be selected using the tick boxes on the left hand side and various types of searches can be performed for words or phrases.

The results of such a search show the context in which the word appears. Users then have the option of listening to the relevant part of the corresponding sound file and also exporting the results of the search to an Excel spreadsheet. The MAONZE miner software was integral in an analysis of changing uses of the verbal particle *ka* (see below).

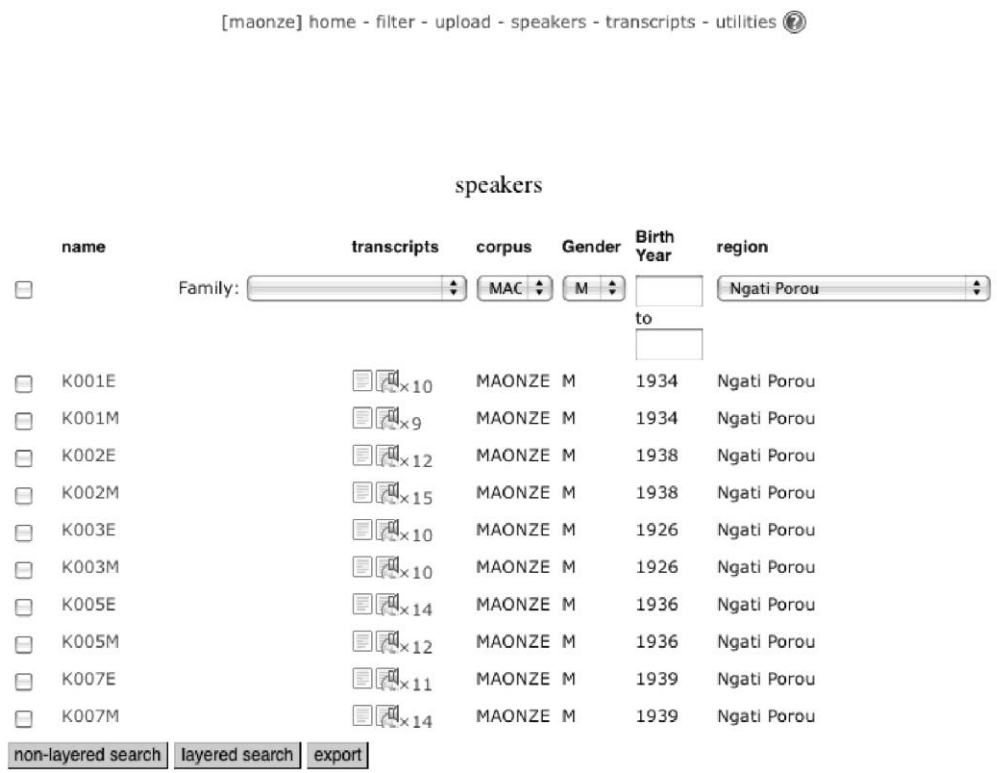


Figure 3: Screen shot of selection page in MAONZE Miner.

Once the orthographic transcription has been uploaded, LaBB-Cat and MAONZE Miner use information from the CELEX database (Baayen, Piepenbrock, & van Rijn, 1993) to create an automatic phonemic transcription for English. The HTK toolkit (http://htk.eng.cam.ac.uk/) can then be used to automatically force align the sound to the phonemes, so that each phoneme symbol is graphically aligned with the appropriate section of the sound file. This forced alignment is done individually for each speaker, and needs a minimum of about 1,000 words of spoken text. The final analysis usually needs to be hand-corrected, but it is much faster than doing the entire analysis by hand. Phoneme alignment facilitates very powerful analysis over the whole corpus or subsets of it. Because Māori is much less studied than English, a phonemically transcribed electronic database like CELEX is not available. The MAONZE team has developed letter to sound rules that allow successful phoneme alignment of Māori words within an English text. We are still working on rules and a dictionary for Māori text.

Results

In this section we briefly describe the results of the various analyses conducted by the MAONZE project. The main focus of the project has been an acoustic analysis of the vowel space in Māori and change over time in both vowel quality and quantity (duration) between the three sets of speakers. In addition the role of women in the attested sound changes has been examined, and some of the consonants have been analysed both auditorily and acoustically. The English speech of the speakers has been analysed as well as their Māori. More recently, the

MAONZE group has been focussing on changes to the rhythm of te reo Māori. The following section explains the use of KEYWORDS for the vowels and diphthongs analysed. Subsequent sections give some detail on the findings of each of the various analyses.

Keywords

Because the primary focus of the MAONZE project is phonetic analysis, the team decided to adopt the use of KEYWORDS to name the vowel phonemes of Māori, following Wells' (1982) example for English. Vowels in KEYWORDS are paired with unique consonants, so that the vowel phoneme may be identified even if speakers' accents vary. For Māori vowels, KEYWORDS will be particularly helpful in clarifying whether the long or short vowel is intended. Table 1 lists the KEYWORDS for the 5 short and 5 long Māori monophthongs.

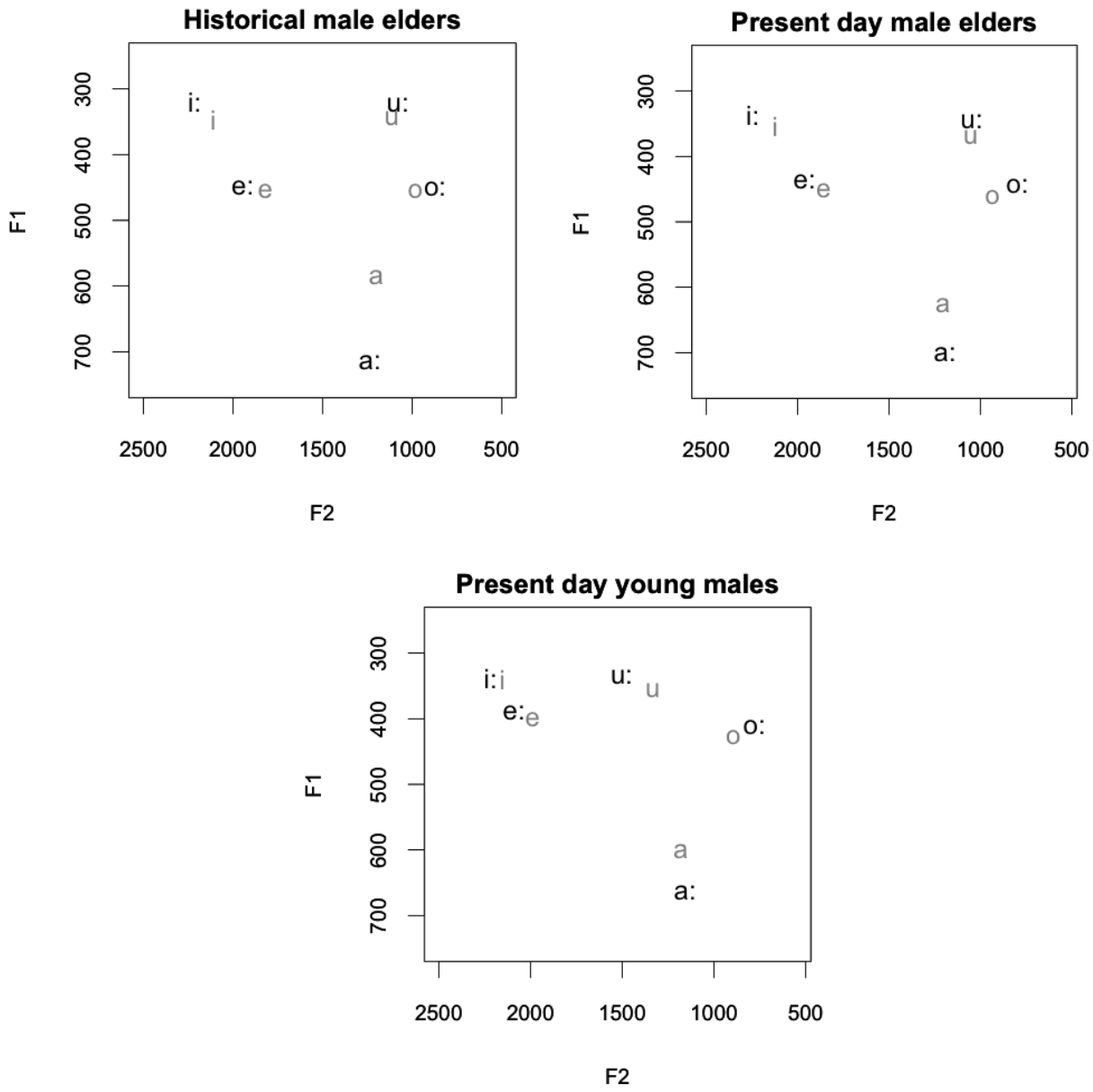
Table 1: KEYWORDS for Māori monophthongs

KEYWORD	phoneme	KEYWORD	phoneme
PĪ	/i:/	PIKI	/i/
KĒ	/e:/	KETE	/e/
WĀ	/a:/	WAKA	/a/
MŌ	/o:/	MOKO	/o/
TŪ	/u:/	TUKU	/u/

KEYWORDS were also coined for five of the most frequent diphthongs in Māori: MAI /ai/, WAE /ae/, RAU /au/, HOU /ou/ and PAO /ao/.

Vowel quality

As mentioned above, the MAONZE project has quantified sound change over time in the Māori language. Changes in the vowel space for the three sets of male speakers are shown in Figure 4. These F1 vs F2 plots show the centroid of the mean F1 and F2 value from each of the three speaker groups in the male data. The centroid is indicated by the IPA symbol of the vowel it represents. Please note that whilst the vowels are referred to by the IPA symbols on the plots, the KEYWORDS are used in the text to facilitate comparison with the relevant vowels for New Zealand English (NZE). From the earliest to the youngest set of speakers we can note the progressive raising of the mid-vowel pairs KĒ/KETE and MŌ/MOKO. Also noticeable in the speech of the young male speakers is the fronting of TŪ/TUKU. Both of these sets of changes parallel changes in vowels which inhabit similar space in NZE, that is, the raising of the DRESS and THOUGHT vowels and the fronting of GOOSE (see Gordon et al., 2004 for details). With regard to the short and long vowel pairs, it can be seen that except for WĀ/WAKA, the long vowels are becoming very like their short vowel counterparts in quality.

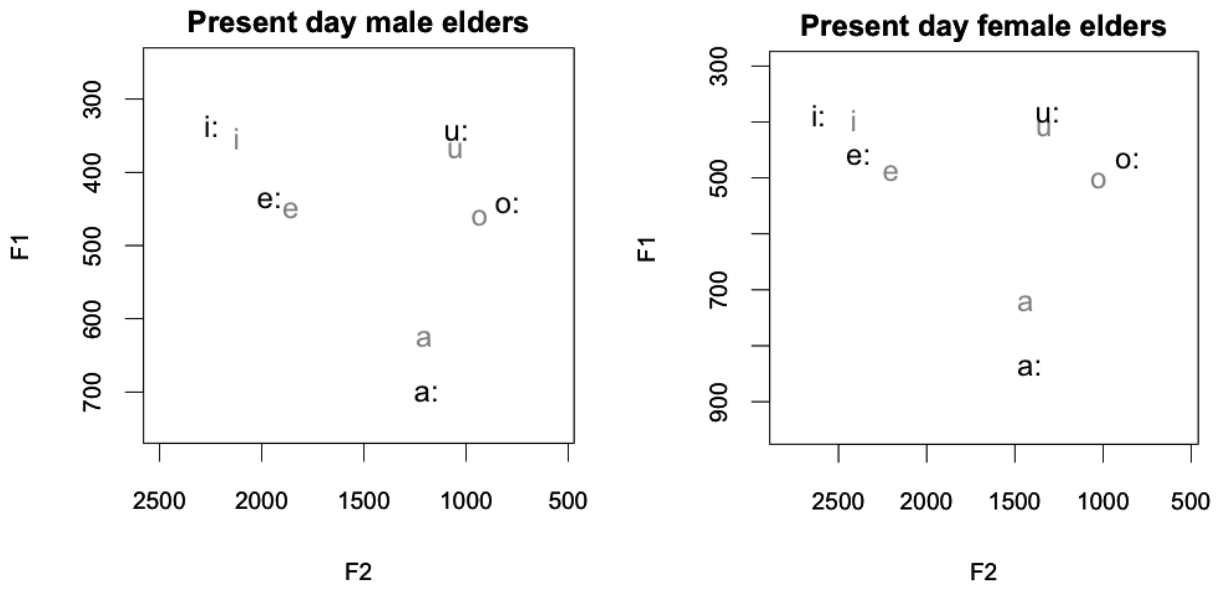


Based on King, Watson, Maclagan, Harlow & Keegan, 2010, p. 197, Figure 10-1.

Figure 4: Long and short vowel F1 and F2 means of the Māori speech of historical male elders, present day male elders and young males. Formant values are shown in Hz.

The vowel space for female speakers exhibits the same changes as noted for the men, except that Māori women have been a generation ahead with some of the changes. This parallels the situation of other examples of sound change where women 'set the standards' for sound change: leading when a change is below the level of consciousness, and holding back on changes which become salient and also stigmatised (Holmes, 1997). Examples of both these processes in Māori can be seen in Figure 5 which shows the vowel space of present day male elders alongside that of present day female elders. Because women's vocal tracts are smaller than men's, women's formant frequencies are higher than men's. Rather than use a formal normalisation procedure (e.g., Lobanov, 1971; for more

information see http://ncslaap.lib.ncsu.edu/tools/norm/norm_methods.php) which may introduce misleading artefacts (Disner, 1980), we have changed the scales slightly so that the men's and women's vowel plots are approximately the same size and can be more easily compared.



Based on King, Watson et al., 2010, p. 201, Figure 10-5.

Figure 5: Long and short vowel F1 and F2 means of the Māori speech of present day male elders and present day female elders. Formant values are shown in Hz.

In Figure 5 women can be seen to be leading the raising of the mid-vowels KĒ/KETE and MŌ/MOKO. By studying the vowel plots of all the speaker groups we note that the raising of these vowels progressed unchecked over all the speaker groups. We can therefore conclude that speakers never became aware of or concerned about this sound change. Figure 5 also shows that women are slightly leading the men with the fronting of TŪ/TUKU. However, the women are more conservative than the men in maintaining a clear qualitative distinction between long and short vowel pairs. This indicates that at some point speakers have become aware of, and resistant to, the merging of long and short vowel qualities and that women therefore became more conservative with this feature. Thus present day female elders demonstrate their role as setting the pronunciation standards for the following generation in both advancing sound change which is below the level of consciousness and holding back with change that becomes salient and stigmatised (see King, Watson et al., 2010).

In summary, the Māori monophthongs have undergone many changes as a result of contact with the English language. The extent of these changes can be seen in Figure 6 which shows the vowel plot of the historical male speakers alongside that of the present day young female speakers.

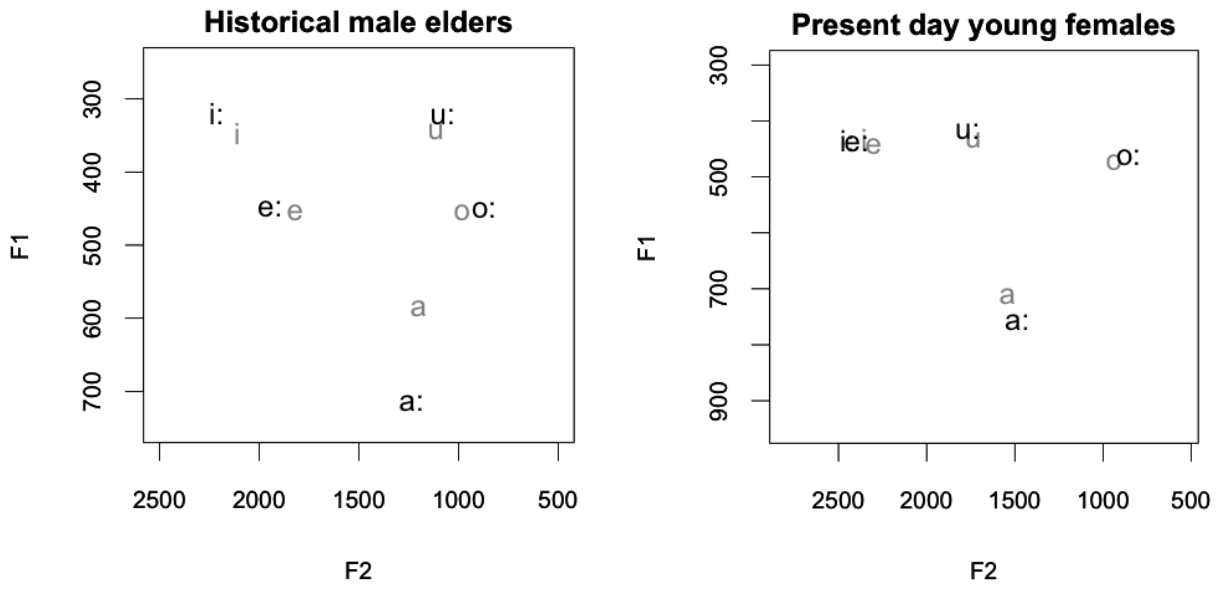


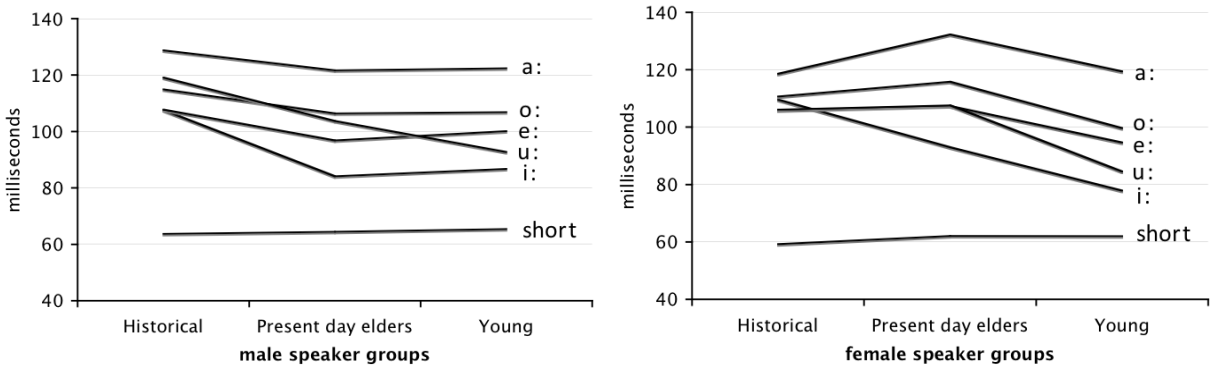
Figure 6: Long and short vowel F1 and F2 means of the Māori speech of historical male elders and present day young females. Formant values are shown in Hz.

It is likely that the vowel space of the historical male elders is representative of the Polynesian vowel system which is noted as being stable over a long period of time (Krupa, 1982). In comparison, the vowel space of the present day young females shows the results of the raising of the mid-vowels and fronting of TŪ/TUKU. With the fronting of TŪ/TUKU the back point vowels are now MŌ/MOKO. With regard to the front vowels, the raising of KĒ/KETE amongst present day young females has proceeded to the point where the quality of KĒ/KETE and PĪ/PIKI are indistinguishable with all the tokens from the four vowels occupying the same acoustic space. This does not seem to have affected intelligibility because context usually facilitates disambiguation.

Vowel quantity

As shown in the vowel plots in Figures 4-6, with the exception of WĀ/WAKA, there has been a reduction in the qualitative difference between short and long vowel pairs over time. It is therefore not surprising that analysis showed that there have been corresponding changes to the quantity (duration) of the vowels. Figure 7 shows the changes amongst the three groups of male and female speakers. It can be seen that the short vowel lengths in both the male and female groups have remained quite consistent over time. Amongst both sets of historical speakers long vowel length was approximately twice that of the short vowels, consistent with the long vowels' phonemic attribute of comprising two of the same short vowels. Amongst present day male and female speakers there has been a dramatic reduction in the length of most of the long vowels, most particularly the two high vowels PĪ and TŪ. The only long vowel which retains its relative length is WĀ. This is consistent with the vowel space results presented above and partly reflects its

greater functional load, in that more word pairs are distinguished by WĀ/WAKA than by other long/short vowel pairs.



Based on King, Watson et al., 2010, p. 205, Figure 10-9.

Figure 7: Long and short vowel lengths for male and female speaker groups.

Diphthongs

As a consequence of the changes in the monophthong system, there have also been changes in the diphthong system of Māori. Figure 8 shows the diphthong plots for the historical male elders and the present day young males indicating that over time two diphthong pairs are merging. The raising of the mid-front vowels KĒ/KETE towards the space occupied by PĪ/PIKI has implications for the pronunciation of the second target of the diphthongs MAE and WAI resulting in a merging of the pair. In addition the fronting of TŪ/TUKU has affected the F2 values of the second targets of the diphthongs RAU and HOU and this, together with changes in their first targets has also resulted in a merger.

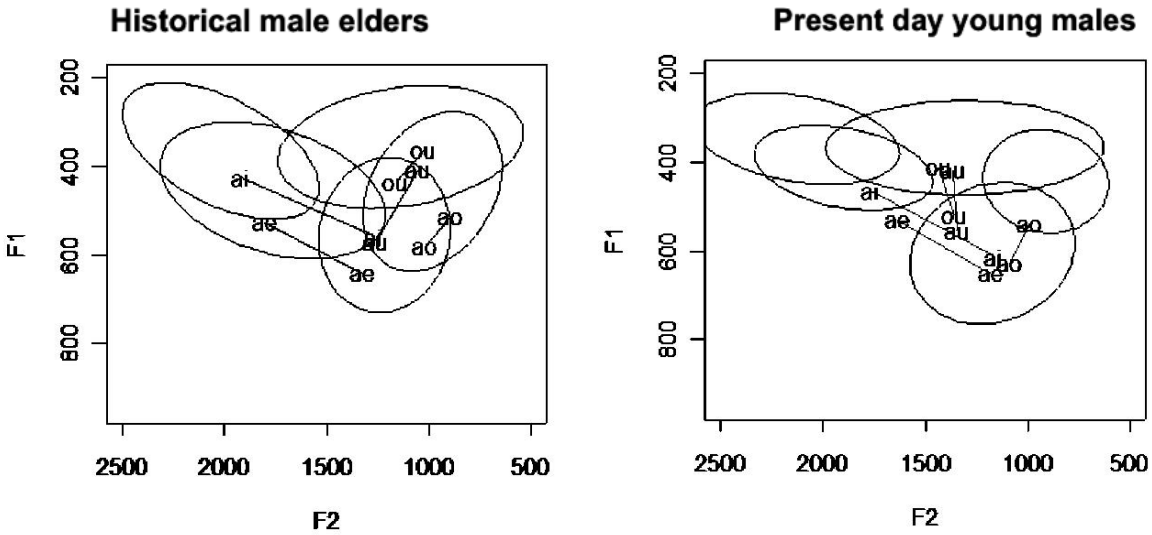
More details on the diphthong analysis for male speakers are available in Harlow, Keegan, King, Maclagan and Watson (2009) and for the female speakers in King, Watson et al. (2010) and the influence of the English language on these changes are discussed in Maclagan et al. (2004).

Consonants

While the main focus of the initial work of the MAONZE project has been on the changing vowel space in Māori there has been some work on consonants including changes in the aspiration of the plosives (Maclagan & King, 2007), the pronunciation of /r/ (Maclagan & King, 2005) and the pronunciation of /f/ (Maclagan & King, 2002).

Changes in the aspiration of Māori plosives over time were examined by analysing the Māori and English plosive consonants of three male speakers, one from each of the three speaker groups. The analysis shows that both the number of aspirated plosives and the degree of aspiration (measured by VOT) have increased from the oldest speaker (born in 1885) to the youngest speaker (born in 1972) in both

languages. Table 2 shows that the mean VOT for plosives in Māori has been increasing over time to parallel the VOT times for English. There may be some language internal factors at work, but influence from English is a likely cause for this change.



Based on Harlow et al., 2009, p. 141, Figure 5.

Figure 8: Beginning and end targets for diphthongs of historical male and present day young male speaker groups plotted against ellipses for short vowels.

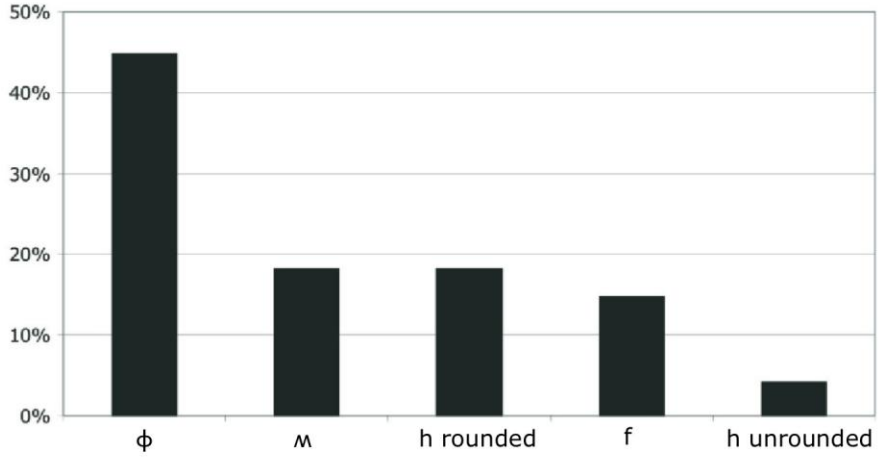
A further analysis investigated the relationship between the increasing aspiration of the plosives over time and the fronting of TŪ/TUKU. This revealed that the fronting of TŪ/TUKU has not simply followed GOOSE fronting in English. Rather the motivator has been the introduction of aspiration into the previously unaspirated Māori /t/. It is this increased aspiration, together with the frequently occurring contextual environment whereby TŪ/TUKU vowels regularly follow aspirated /t/, that has facilitated their fronting. For many modern speakers, the VOT has increased to such an extent that /t/ is heard as affricated rather than merely aspirated. The relationship between the increasing aspiration of /t/ and the fronting of TŪ/TUKU is discussed in Maclagan, Watson, Harlow, King and Keenan (2009).

Table 2: Mean voice onset time (VOT) in ms for plosives /p/, /t/ and /k/ in Māori and English for three male speakers.

	Historical male speaker		Present day male elder		Present day young male	
	Māori	English	Māori	English	Māori	English
mean	25	43	41	66	57	68
sd	10	16	18	15	22	21
n	135	98	246	98	114	101

Adapted from Maclagan and King 2007, p. 2, Table 5.

An auditory analysis was undertaken of the varying pronunciations of *wh* in one of the historical male speakers (Maclagan & King, 2002). Figure 9 shows that this Ngāti Maniapoto speaker produced a number of variants for *wh* in his speech, and that the most common variant today, excepting recognised tribal variants, (/f/), was not the most frequent. A comparison with a present day elder and a young speaker showed that this variety of pronunciations had reduced to one variant by the time of the modern day elder.



Adapted from Maclagan and King 2002: 49, figure 1.

Figure 9: Relative frequencies for varying pronunciations of *wh* in one historical male speaker.

English

The project has also been interested in the pronunciation of English by the speakers in the MAONZE corpus. Analysis of the vowel space of the three groups of male speakers shows that the English pronunciation of these speakers is largely similar to that of their non-Māori contemporaries (Watson, Maclagan, King, & Harlow, 2008). However, the speakers who are first language speakers of Māori (in particular the historical elders and the present day elders) show influence of the Māori vowel space on their English in that they are relatively conservative in their pronunciations with regard to some of the recognised changes and produce relatively back versions of GOOSE and START. This is likely to be the reason that earlier commentators noted a "purity of vowels" in the English of Māori speakers (Richards, 1970, p. 131).

Rhythm

In contrast to English which has a relatively stress timed rhythm Māori is regarded as having a mora timed rhythm (Bauer, 1981). The attested changes in the duration of Māori vowels (outlined above), where most long vowels are approaching their shorter counterparts in length, has implications for the rhythm of the Māori language. Current work by the MAONZE project is examining changes in rhythm.

The PVI (Pairwise Variability Index), which compares the variability of the length of adjacent vowels separated by consonants, is a method often used to compare the rhythm of various languages (Grabe & Low, 2002). However, the presence of long sequences of vowels in Māori (for example, the phrase *ki a ia*) makes this method unsuitable for making effective assessments of the rhythm of Māori (Maclagan, Watson, King et al., 2009).

Nevertheless, a series of perception experiments determined that both Māori and non-Māori listeners were able to distinguish between Māori and English speech from prosodic information which includes rhythm. They were presented with short excerpts of low pass filtered speech. The filtering removes segmental information but still retains prosodic information on pitch, intensity and timing (Maclagan, Watson, King et al., 2009). Further investigation is focussing on the perception of rhythmic prominences in Māori (see Thompson et al., 2010).

During the vowel analysis we noted that amongst the historical speakers even unstressed vowels seemed to be fully articulated and that there was no unstressed vowel variant such as there is in English with the schwa vowel. As described above, the MAONZE vowel analysis included only stressed vowels as tokens. A current analysis is investigating all short and long vowels not adjacent to other vowels in a six minute stretch of speech for one to two male speakers from each of the three speaker groups. The results indicate that increasingly centralised variants are being produced for all vowels over time, suggesting the possibility of a centralised vowel emerging in Māori in the future (Kaefer et al., 2010).

The verbal particle *ka*

Most modern younger speakers of Māori are either second-language speakers or have been raised in a context where second-language speakers predominate. This has led to changes in the intergenerational language transmission process, and means that some phonetic distinctions have not been acquired by younger speakers. An example of this process is the loss of the distinction between short and long versions of the verbal particle *ka*. Historically, the Māori tense/aspect marker *ka* has two allomorphs, one, /ka:/, which is used when the rest of the verb phrase consists of only two moræ (*ka noho* ‘sits, lives’ with [ka:]), and the other, /ka/, for longer phrases (*ka tū ake* ‘stands up’ with [ka]) (Biggs, 1969, p. 28). The MAONZE miner software has facilitated an analysis of the distribution of these two variants in the speech of the three speaker groups (Harlow & Bauer et al., 2011). Results show that the historical elders do indeed observe the traditional rule but that the rule has been lost amongst present day young adult speakers who invariably produce the short variant in all contexts. Modern day elders observe the traditional rule approximately half the time. This shift is attributable both to a proportional increase in the use of longer phrases over the same period and to the decreasing use of Māori generally, so that opportunities to acquire the inherited rule have diminished considerably.

Conclusion

Much of the sound change in Māori documented by the MAONZE project has been influenced by English. This is perhaps to be expected, as although English and Māori are both official languages in New Zealand, the use of the Māori language has declined considerably since the mid-1900s and has been subject to massive revitalisation efforts since the mid-1980s.

The results of the analyses to date have implications for the revitalisation efforts of other indigenous languages in that we predict that vowel systems will lose contrasts which are not present in the colonising language, and that sound change in the colonising language will affect the indigenous language (King, Harlow, Watson, Keegan & Maclagan, 2009). The results of the vowel analysis have also been published in Māori in order that the implications reach as wide a local audience as possible (Harlow, Keegan, King, Maclagan, & Watson, 2005; Keegan, King, Harlow, Maclagan, & Watson, 2008). Other publications have addressed the issue of the implications of the sound changes on the teaching of Māori (Keegan, King, Maclagan, Watson, & Harlow, 2009). A practical outcome of the MAONZE project is work on designing a computer-based pronunciation aid for Māori (Watson, Smith, et al., 2009) which allows learners to hear model words being spoken and test their own pronunciation.

The MAONZE project demonstrates a methodology for analysing Māori/English speech which could readily be applied to other indigenous languages being influenced by majority languages. We have only provided examples of some of types of analyses that can be undertaken. We expect the MAONZE corpus to continue to be useful for us and other researchers into the future and that many new and different analyses will be carried out on the present data.

In assembling this corpus (see the previously published companion paper) and in achieving the detail of analysis made possible by it, the MAONZE project has broken new ground in the study of change in a minority indigenous language whose ecology crucially involves fragile transmission and very heavy influence from a major language within the same community.

References

Bauer, W. (1981). Hae.re vs. ha.e.re: A note. *Te Reo, 24*, 31-36.

Baayen, R. H., Piepenbrock, R. & van Rijn, H. (1993). The CELEX lexical database (CD-ROM). Philadelphia: Linguistic Data Consortium.

Biggs, B. (1969). *Let's learn Māori*. Wellington: A. H. and A. W. Reed.

Boersma, P. & Weenink, D. (2010). Praat: Doing phonetics by computer [Computer program]. Version 5.2.06, retrieved from http://www.praat.org

Disner, S. F. (1980). Evaluation of vowel normalization procedures. *Journal of the Acoustical Society of America, 67*(1), 253-261.

Fromont, R., & Hay, J. (2008). ONZE Miner: Development of a browser-based research tool. *Corpora, 3*(2), 173-193.

Gordon, E., Campbell, L., Hay, J., Maclagan, M., Sudbury, A., & Trudgill, P. (2004). *New Zealand English: Its origins and evolution*. Cambridge: Cambridge University Press.

Grabe, E. &. Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In C. Gussenhoven & N. Warner (Eds), *Laboratory phonology 7 (*pp. 515-46). Berlin: Mouton de Gruyter.

Harlow, R., Bauer, W., Maclagan, M., Watson, C., Keegan, P. & King, J. (2011). Interrupted transmission and rule loss in Māori: The case of *ka*. *Oceanic Linguistics, 50*(1), 51-65.

Harlow, R., Keegan, P., King, J., Maclagan, M., & Watson, C. (2005). Te whakahuatanga i te reo Māori: Kua ahatia e tatou i roto i nga tau 100 kua hipa nei? *He Puna Korero - Journal of Māori & Pacific Development*, *6*(1), 45-57.

Harlow, R., Keegan, P., King, J., Maclagan, M., & Watson, C. (2009). The changing sound of the Māori language. In J. N. Stanford & D. R. Preston (Eds.), *Variation in indigenous minority languages* (pp. 129-152). Amsterdam: John Benjamins.

Holmes, J. (1997). Setting new standards: Sound changes and gender in New Zealand English. *English World Wide*, *18*(1), 107-142.

Kaefer, S., King, J., Watson, C., Maclagan, M., Harlow, R. & Keegan, P. (2010, 22-23 November). Evidence for the development of a centralised vowel in Māori and its effect on rhythm. 12th New Zealand Language & Society Conference, Auckand.

Keegan, P., King, J., Harlow, R., Maclagan, M., & Watson, C. (2008). Ngā nekehanga o te whakahua i te reo Māori i roto i te rautau kua hipa nei. *AlterNative*, *4*(2), 180-197.

Keegan, P. J., King, J., Maclagan, M., Watson, C., & Harlow, R. (2009). Changes in the pronunciation of Māori and implications for teachers and learners of Māori. In S. May (Ed.), *Refereed Conference Proceedings of LED 2007, 2nd International Conference on Language, Education and Diversity*. Hamilton, N.Z.: Wilf Malcolm Institute of Educational Research.

King, J., Harlow, R., Watson, C., Keegan, P., & Maclagan, M. (2009). Changing pronunciation of the Māori language: Implications for revitalization. In J. Reyhner & L. Lockard (Eds.), *Indigenous language revitalization: Encouragement, guidance and lessons learned* (pp. 85-96). Flagstaff: Northern Arizona University.

King, J., Maclagan, M., Harlow, R., Keegan, P. & Watson, C.I. (2010). The MAONZE corpus: Establishing a corpus of Māori speech. *New Zealand Studies in Applied Linguistics, 16*(2), 1-16.

King, J., Maclagan, M., Harlow, R., Keegan, P. & Watson, C.I. (2011). The MAONZE project: Changing uses of an indigenous language database. *Corpus Linguistics and Linguistic Theory. 7*(1), 37-57.

King, J., Watson, C., Maclagan, M., Harlow, R., & Keegan, P. (2010). Māori women's role in sound change. In J. Holmes & M. Marra (Eds.), *Femininity, feminism and gendered discourse* (pp. 191-211). Newcastle upon Tyne: Cambridge Scholars Publishing.

Krupa, V. (1982). *The Polynesian languages: A guide*. Languages of Asia and Africa, Vol. 4. London: Routledge and Kegan Paul.

Lobanov, B. M. (1971). Classification of Russian vowels spoken by different listeners. *Journal of the Acoustical Society of America, 49,* 606-08.,

Maclagan, M., Harlow, R., King, J., Keegan P., & Watson, C. (2004). New Zealand English influence on Māori pronunciation over time, *Te Reo, 47,* 7-27.

Maclagan, M. & Hay, J. (2011) Transcription. In M. Di Paolo & M. Yaeger-Dror (Eds.), *Sociophonetics: A student's guide* (pp. 36-45). New York: Routledge.

Maclagan, M., & King, J. (2002). The pronunciation of wh in Māori: A case study from the late nineteenth century. *Te Reo, 45,* 45-63.

Maclagan, M., & King, J. (2005). A note on the realisation of /r/ in the word Māori. *New Zealand English Journal, 18,* 35-39.

Maclagan, M. A., & King, J. (2007). Aspiration of plosives in Māori: Change over time. *Australian Journal of Linguistics, 27*(1), 81-96.

Maclagan, M., Watson, C. I., Harlow, R., King, J., & Keegan, P. (2009). /u/ fronting and /t/ aspiration in Māori and New Zealand English. *Language Variation and Change*, *21*(2), 175-192.

Maclagan, M., Watson, C., King, J., Harlow, R., Thompson, L., & Keegan, P. (2009). Investigating changes in the rhythm of Māori over time. Poster presented at the 10th Annual Conference of the International Speech Communication Association, Brighton, England.

Richards, J.C. (1970). The language factor in Māori schooling. In J. Ewing & J. Shallcrass (Eds), *Introduction to Māori education* (p122-132). Wellington: New Zealand Universities Press.

Thompson, L., Watson, C., Charters, H., Harlow, R., Peter Keegan, P., King, J. & Maclagan, M. (2010). An experiment in mita-reading: Investigating perception of rhythmic prominence in the Māori language. In *Proceedings of the 13th Australasian International Conference on Speech Science and Technology* (pp. 150-154).

Watson, C., Maclagan, M., King, J., & Harlow, R. (2008). The English pronunciation of successive groups of Māori speakers. In *Proceedings of the 9th Annual Conference of the International Speech Communication Association* (pp. 338-341).

Watson, C., Smith, J., King, J., Maclagan, M., Keegan, P., & Harlow, R. (2009). Vowel mergers in modern Māori and a speech aid designed to promote vowel distinction. In *HCSNet Summerfest'09 Conference Handbook* (p. 158).

Wells, J. C. (1982). *Accents of English*. Cambridge: Cambridge University Press.