THE DYNAMICS OF CLOSING DIPHTHONG FORMANT TRAJECTORIES IN TE REO MĀORI

Hywel M. Stoakes[1] , Catherine I. Watson[1], Peter J. Keegan[1], Margaret A. Maclagan[2], Jeanette King[2] and Ray Harlow[3]

[1]University of Auckland, [2]University of Canterbury, [3]University of Waikato, Aotearoa/New Zealand
{h.stoakes;c.watson;p.keegan}@auckland.ac.nz,{margaret.maclagan;j.king}@canterbury.ac.nz,ray.harlow@waikato.ac.nz

ABSTRACT

Te reo Māori, an Eastern Polynesian language, is one of the official languages of Aotearoa/New Zealand and is spoken by approximately 148,400 people, 84.5% of whom identify as Māori. Two pairs of closing diphthongs are usually described as merging phonetically in the speech of modern language speakers. Some diphthong pairs can be confused for second language learners, although they occur in grammatically and semantically distinct words. The current study examines recordings of 18 elder speakers (9M, 9F) of te reo Māori, collected as part of the MAONZE project, an ongoing investigation of sound change within the language. The aim is to describe these diphthongs in terms of trajectory, duration and overall prominence, using measurements of averaged trajectories of formants (F1 and F2) and duration. Results show that there are differences in duration as well as the formant trajectories within diphthong pairs /ai/, /ae/ and /au/, /ou/, suggesting contrasts beyond formant dynamics.

Keywords: diphthongs, dynamics, te reo Māori language, Polynesian languages, Aotearoa/New Zealand

1. INTRODUCTION

Te reo Māori, one of the national languages of Aotearoa/New Zealand, is a member of the Eastern Polynesian language branch of the large Austronesian language family [6]. The segmental phonology is regular with a small consonant inventory typical for the branch (see [1] and [6]). Present day elder Speakers of te reo are pivotal in the revitalisation movement [10, 17]. These elder speakers are very fluent adult speakers of the language, although due to the extended period of colonial contact and historical break in language transmission [9, 2], there have been significant influences from New Zealand English evident in the phonetics of these speakers [7, 8, 18]. In terms of vowels there has been gradual change over time and this was already in progress from the early part of the last century. This has been explored in great depth in recent years by the MAONZE research team [18, 11]. However, there are still some unanswered questions, particularly concerning the dynamic properties of diphthongs.

Vowels in te reo Māori have a lexical length alternation and generally carry a large phonological load due to the relatively small consonant inventory. Every combination of two vowels can be adjacent and some scholars consider long vowels to be sequences of like segments or geminated vowels (see [3]). Consequently there is the opportunity for extreme coarticulation, although there are many environments, particularly when spanning word boundaries, when vowels are observed as articulated separately possibly to maintain intelligibility. Insertion of glides, in close vowel followed by open vowel sequences, are prevalent in contemporary speech. Devoicing by glottalisation is found at the ends of short utterances although glottal stops at utterance boundaries are rare. Initial and medial stops are voiceless with a long duration but very short or coincident voice onset times (VOT). Voice onset time has been getting progressively longer diachronically, however [14], and greater levels of aspiration are evident in word initial environments, which is thought to be directly due to contact and interaction with New Zealand English. These changes in interarticulator timing have likely affected the rhythmic and metrical organisation of the language and will be investigated more fully in future work.

1.1. Diphthongs in te reo Māori

Custodians of the language as well as te reo Māori teachers usually describe falling diphthongs (close to open) as articulated with each vowel pronounced separately. The raising diphthongs (open to close) are instead blended (or diphthongal) [1]. This study builds on two earlier papers examining diphthong articulation in te reo Māori. Our focus is on the articulation of the two pairs of raising diphthongs in the language, drawing on previous work to inform our study. Watson *et al.* [18] investigated first and second formant values extracted at the first and second

target of /ai ae au ou/. They found when comparing the speech of present day elders and young te reo speakers that /ai/ and /ae/ were distinguished on the basis of F1 for the first target. In addition, the present day elders and young men also contrasted the two diphthongs at the second target, however the young women speakers did not. With regard to the second pair /au/ and /ou/, these were only distinguished at F1 for the first target by the both the present day elder and young speaker groups. This merger was suggested to be quite advanced, and influenced by the fronting of /u/ and /u:/ in the monophthongs [18]. Diphthong trajectories of these four diphthongs were investigated in [12] for the male speakers of the MAONZE corpus. Visual inspection of the trajectories suggested the major changes over time were in F2 for all the diphthongs. The observations made were consistent with the statistically significant findings from the static target analysis in [18]. Whilst these findings were illuminating, the analyses were incomplete as the investigation of the diphthong trajectories only included male speakers, but the results from [18] suggested there may be additional gender differences. Further limitations were, that [18] and [12] did not control for the phonetic environment of the initial consonant which may be a factor influencing the realisation of the contours.

2. AIMS AND METHOD

The aim of this study is to examine the phonetic differences between two sets of diphthong pairs using a dynamic acoustic analysis. This allows us to explore whether there is evidence of acoustic mergers between the two diphthong pairs in terms of formant trajectory or measured duration. The speech and language data used in this study have been drawn from the MAONZE corpus, a large-scale project looking at sound change in the language [11]. This corpus contains many hours of phonetic and higher linguistic annotation which have been constructed over a period of more than 15 years. With recent rapid advances in forced alignment techniques and machine-based speech-to-text annotation systems, it is now possible to use the utterance level transcriptions to annotate larger portions of the corpus in further phonetic detail. For the current study we have chosen a small subset of the main corpus which contains speech from elder speakers of te reo Māori born between 1920 and 1944. This research would not have been possible without the close links between researchers and language community members and the continued support of *iwi* (tribe, extended kinship group).

2.1. Speakers and corpus

The study uses real words drawn from approximately 20 hours of recordings which predominantly consists of connected speech, both conversation and monologue narrative. This in turn derives from a subset of the entire MAONZE corpus. Eighteen fluent speakers, nine men and nine women, of te reo Māori were recorded and their speech automatically segmented and labelled (see below). Conversations, narrative and controlled reading lists and also short reading passages (< 3mins) were included. This analysis selects a limited set of monosyllabic words uttered within the connected speech. One set of monosyllables that have glottal fricatives in initial position and one set with a voiceless unaspirated bilabial stop initial were extracted. All words were spoken in various positions within the utterance and the only prosodic control was that they were identified as containing full vowels, not undergoing reduction. Due to the spontaneous nature of the speech data, the counts of each word are not equal, with *pai* overrepresented within the corpus (see table 1). Only durations less than 350ms are included in the analysis to remove outliers with extreme lengthening when a word was uttered in isolation within its own phrase (N = 798).

Table 1: Words included in the study, with counts

Word	Diph.	Gloss	*N*
pai	/ai/	(verb) like, approve	483
pae	/ae/	(noun) horizon, perch	45
pau	/au/	(verb) exhausted	57
pou	/ou/	(verb/noun) to erect/post	15
hai	/ai/	(particle) at, in, with	28
hae	/ae/	(verb) (-a) to scratch	12
hau	/au/	(noun) wind, vital essence	74
hou	/ou/	(modifier) new, recent	88

2.2. Analysis and statistics

Speech data were first prepared for forced alignment using Kaldi [16] and a set of scripts developed as part of the ELPIS project [5]. This facilitated the production of a phonetic dictionary used in the subsequent forced aligning of the data within the Montreal Forced Aligner (v. 1.0) [15]. The forced aligned data builds on the existing analyses by allowing consonant environment to be queried. After alignment, all boundaries in the target words were manually checked for accuracy. The speech data are field recordings of spontaneous speech and conversation that captured the resonant characteristics of the room which was not ideal for analyses of voiceless segments. In general, the right edges of segments were well placed by the forced align-

ment algorithm, with consonant to vowel boundaries consistently and accurately segmented. The left edge of consonants were consistently temporally misaligned, however with post-closure activity in the frequency spectrum associated with the vowel rather than the preceding consonant.

The forced aligned data were outputed as Praat TextGrids which were then compiled into an Emu Speech Database [20]. Formants trajectories were calculated in `R` using the `forest` function (gender specified, all else with default values), within the `wrassp` package [4]. Second formant (F2) values were corrected when obviously mistracked. These were usually cases when there a formant was misidentified as either F3 or F1. The first formant (F1) was corrected and interpolated only when there were zero values present, otherwise these formants were left uncorrected.

All plots were made using `ggplot2` [19] and error bars are calculated and plotted using a Generalized Additive Model (GAM) analysis using the formula $y \sim s(x, bs = `cs')$ within the '`bam`' function, a part of the `mgcv` package [21]. Statistics were calculated using the `lmer` function in the `lmerTest` package and using the `step` function [13].

3. RESULTS

Within this sample of present day elder speakers of te reo Māori, there are clear differences between the diphthong pairs. Although, sequence differentiation is not made in the formant space alone. In the /ae/ and /ai/ pairs there are observed durational differences between the diphthong pairs, with the /ai/ and /au/ shorter in duration than /ae/ and /ou/ (see figures 2 and fig 4).

3.1. Formant trajectories over time

The results from the formant trajectory analysis show that when normalised over time the trajectory shapes for both men and women in the /ae/ and /ai/ pair are similar (see fig. 1 and fig. 3), although there is greater variation within the sample for the /ae/ diphthongs for both shown by the grey ribbon on the plot. Within the /au/ /ou/ pair, the F2 is relatively static across the entire trajectory for the female speakers, there is greater formant movement for the male speakers.

The duration measurements show that for each diphthong pair there is a durationally short member of the set and a durationally long member (see fig. 2 and fig 4). All diphthongs show a statistical difference when speaker is included as a random intercept in the model, however. Statistical analysis of the durations show that when speaker is included as a random intercept in a linear mixed effects

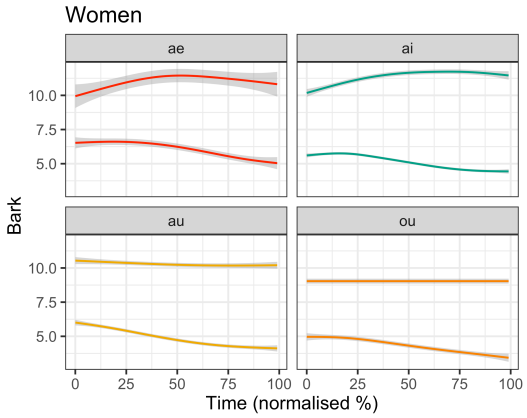

Figure 1: F1 and F2 trajectories for diphthongs /ae/, /ai/, /au/, /ou/: Female Elders

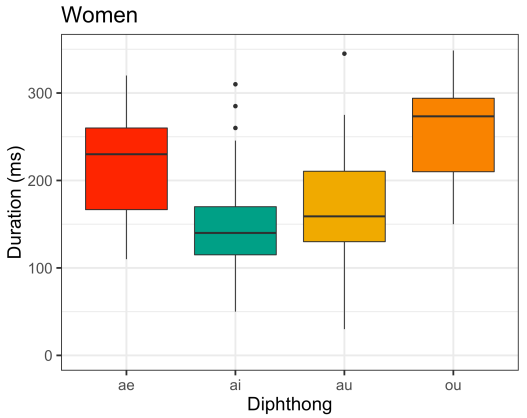

Figure 2: Durations for diphthongs /ae/, /ai/, /au/, /ou/: Female Elders

regression model, a main effect of diphthong duration is statistically significant across all diphthongs ($\chi^2(3, N = 798) = 148.83, p < .001$), calculated using a *post hoc* Satterthwaite approximation.

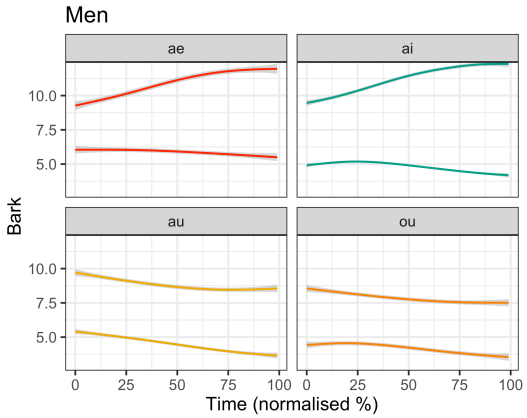

Figure 3: F1 and F2 trajectories for diphthongs /ae/, /ai/, /au/, /ou/: Male Elders

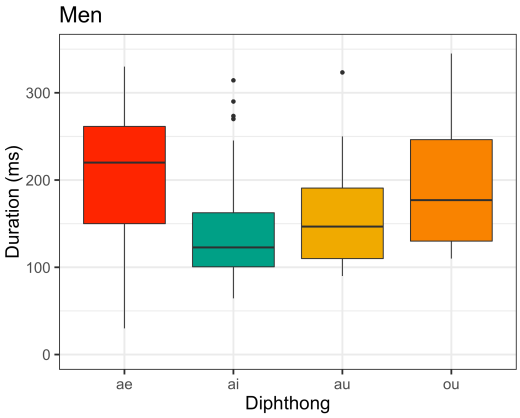

Figure 4: Vowel Duration for diphthongs /ae/, /ai/, /au/, /ou/: Male Elders

3.2. Formants in articulatory space

When plotted within F1 and F2 articulatory space we can see that there is a clear separation between /ae/ and /ai/ and to a lesser extent /au/ and /ou/ (see Fig. 5). Over time there has been considerable fronting of the /u/ phoneme which is phonetically realised as a close, central, rounded vowel [ʉ]. The /a/ vowel is open and central [ɐ] and the /i/ vowel is close, front and unrounded, [i].

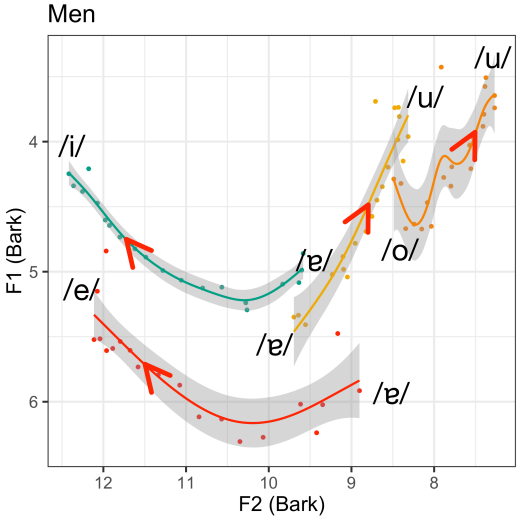

Figure 5: F1 and F2 Trajectories for Male Elders

4. DISCUSSION

The formant findings in this study are consistent with previous observations reported in [12] and [18]. The study finds greater variability of the F2 trajectory of /ae/ when compared to /ai/. This may contribute to the perception that /ae/ and /ai/ have merged for modern listeners of the language, although the second target of /ae/ and the overall F1 are considerably lower than /ai/ for the men, as shown in Figure 5. The durational findings in this analysis are also interesting, as it suggests that within the pairs there are articulatory differences. It seems unusual that /ai/ is shorter than /ae/ despite the fact it traverses a greater proportion of the vowel space. Similar observations can be made for /ou/ and /au/. To fully interpret the variability within the sample will require further investigation, however. The diphthong trajectories show the region of the vowel space through which the diphthongs travel although this gives no explicit information regarding proportion of time a vowel is held at the extrema of the diphthong or the absolute transition speed between first and second target. This is one disadvantage of statistical techniques such as GAMMS that time-normalise across the entire sequence of interest. A registration landmark such as first target, indicated by a steady F1 and F2 values, or the turning point values of F2 may fully capture the phonetic contrasts observed between the phone sequences contrasted in this study. Importantly, the duration differences may be primarily lexical, as each of words in the pairs belong to different grammatical categories. This also means that as this sample is drawn from uncontrolled conversational speech, each word will occur syntactically at different positions within the phrase (see Table 1). Further investigation into the other diphthongs in te reo Māori and a wider variety of lexical items, as well as control for phrase position is needed in order to form a full representation of diphthong articulation within the language.

4.1. Conclusion

The current study has contributed to our understanding of the dynamic articulation in raising diphthongs and when considered with findings of the MAONZE project [12, 18], gives insight into language community reports regarding diphthong mergers. Furthermore, it shows that formant trajectory analysis alone is not sufficient to fully describe the difference between these pairs. The rate at which the formants move across the vowel space may also provides a phonetic cue to the difference in each pair. Further analysis will involve increasing the number of vowel tokens including a greater variety of phonetic contexts andcontrolling for position of the target word within the intonational phrase.

5. ACKNOWLEDGEMENTS

Thanks to the speakers that participated in this study. Thanks to three anonymous reviewers for their comments, all errors remain our own. This research was funded by a Faculty Development Grant, the University of Auckland (3714653).

6. REFERENCES

[1] Bauer, W. 2003. *Maori*. London ; New York: Routledge.

[2] Benton, R. 1997. *The Maori language : dying or reviving? : a working paper prepared for the East-West Center Alumni-in-Residence Working Paper Series*. Working paper (International Association of East-West Center Alumni). Wellington, N.Z.: New Zealand Council for Educational Research.

[3] Biggs, B. 1961. The structure of New Zealand Maaori. *Anthropological Linguistics* 1–54.

[4] Bombien, L., Winkelmann, R., Scheffers, M. 2018. *wrassp: an R wrapper to the ASSP Library*.

[5] Foley, B., Arnold, J., Coto-Solano, R., Durantin, G., Ellison, M., van Esch Daan, , Heath, S., Kratochvíl, F., Maxwell-Smith, Z., David, N., Olsson, O., Richards, M., Nay, S., Stoakes, H., Thieberger, N., Wiles, J. 2018. Building speech recognition systems for language documentation: The CoEDL endangered language pipeline and inference system (ELPIS). *The Proceeding of 6th Intl. Workshop on Spoken Language Technologies for Under-Resourced Languages 29-31 August 2018* Gurugram, India.

[6] Harlow, R. 2007. *Māori: A linguistic introduction*. Cambridge University Press.

[7] Harlow, R., Keegan, P., King, J., Maclagan, M., Watson, C., others, 2005. Te whakahuatanga i te reo Māori: Kua ahatia e tatou i roto i nga tau 100 kua hipa nei? (the pronunciation of Māori: What have we done to it in the last 100 years?). *He Puna Korero: Journal of Maori and Pacific Development* 6(1), 45.

[8] Keegan, P., Watson, C., Maclagan, M., King, J. 2014. Sound change in Māori and the formation of the MAONZE project project. In: *He Hiringa, He Pūmanawa: Studies on the Māori language*. Wellington: Huia Publishers 33–54.

[9] King, J. 2018. Māori: revitalization of an endangered language. In: Rehg, K., Campbell, L., (eds), *The Oxford Handbook of Endangered Languages*. Oxford: Oxford University Press 592–612.

[10] King, J., Cunningham, U. 2017. Tamariki and fanau: Child speakers of Māori and Samoan in Aotearoa/New Zealand. *Te Reo* 60(1), 29–46.

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

[12] King, J., Watson, C., Maclagan, M., Keegan, P., Ray, H. 2014. Diphthong trajectories in Māori. *The proceedings of 15th Australasian International Conference on Speech Science and Technology SST 2014* Christchurch, New Zealand. ASSTA 243.

[13] Kuznetsova, A., Brockhoff, P., Christense, R. 2017. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software* 82(13), 1–26.

[14] Maclagan, M., King, J. 2007. Aspiration of plosives in Māori: Change over time. *Australian Journal of Linguistics* 27(1), 81–96.

[15] McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., Sonderegger, M. 2017. Montreal Forced Aligner: trainable text-speech alignment using Kaldi. Lacerda, F., others, , (eds), *Proceedings of Interspeech 2017 Stockholm, Sweden*. ISCA 498–502.

[16] Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., Hannemann, M., Motlicek, P., Qian, Y., Schwarz, P., Silovsky, J., Stemmer, G., Vesely, K. 2011. The Kaldi speech recognition toolkit. *IEEE 2011 Workshop on Automatic Speech Recognition and Understanding*. IEEE Signal Processing Society. IEEE Catalog No.: CFP11SRW-USB.

[17] Watson, C. I., Keegan, P. J., Maclagan, M. A., Harlow, R., King, J. 2017. The motivation and development of MPAi, a Māori pronunciation aid. *Proc. Interspeech 2017* 2063–2067.

[18] Watson, C. I., Maclagan, M. A., King, J., Harlow, R., Keegan, P. J. 2016. Sound change in Māori and the influence of New Zealand English. *Journal of the International Phonetic Association* 46(2), 185–218.

[19] Wickham, H. 2016. *ggplot2: elegant graphics for data analysis*. Springer.

[20] Winkelmann, R., Harrington, J., Jansch, K. 2017. EMU-SDMS: Advanced speech database management and analysis in R. *Computer Speech & Language* 45, 392–410.

[21] Wood, S. N. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73(1), 3–36.