INTERSPEECH 2017
August 20–24, 2017, Stockholm, Sweden

The motivation and development of *MPAi*, a Māori Pronunciation Aid

C. I. Watson [1], *P.J. Keegan*[1], *M.A. Maclagan*[2], *R. Harlow*[3], *J. King*[2]

[1]University of Auckland, New Zealand
[2]University of Canterbury, New Zealand
[3]The University of Waikato, New Zealand

c.watson@auckland.ac.nz, p.keegan@auckland.ac.nz, margaret.maclagan@canterbury.ac.nz, rharlow@waikato.ac.nz, j.king@canterbury.ac.nz

Abstract

This paper outlines the motivation and development of a pronunciation aid (*MPAi*) for the Māori language, the language of the indigenous people of New Zealand. Māori is threatened and after a break in transmission the language is currently undergoing revitalization. The data for the aid has come from a corpus of 60 speakers (men and women). The language aid allows users to model their speech against exemplars from young speakers or older speakers of Māori. This is important, because of the status of the elders in the Māori speaking community, but it also recognizes that Māori is undergoing substantial vowel change. The pronunciation aid gives feedback on vowel production via formant analysis, and selected words via speech recognition. The evaluation of the aid by 22 language teachers is presented and the resulting changes are discussed.

Index Terms: Māori language, human-computer interaction, Pronunciation Aid, formants

1. Introduction

Māori, the indigenous language of New Zealand has become severely compromised in the last 100 years due to the impact of English. Since the 1980s it has been undergoing extensive revitalization (see [1-3]). This paper presents MPAi, a Māori Pronunciation Aid, developed as part of our contribution to the revitalization process. Because it is important to understand the factors that led to the development of MPAi, we will first provide more details about the status of Māori in New Zealand, and about our project on sound change in Māori, which was crucial to the development of MPAi.

1.1. Māori Language

According to the 2013 census [4] the total Māori population in New Zealand (NZ), is around 600,000, and about 11% of these claim to speak Māori well or very well [5]. In terms of the total population of NZ, about 3.5 % speak the language to some extent. This is around 150,000 people and includes non–Māori.

Māori is an official language of NZ but English is the dominant language of the country. Because NZ is so isolated, it was not settled until the 1200s. Māori, the first people to populate NZ, had no contact with people speaking other languages until about 180 years ago, when the colonization of NZ, particularly by English speaking people, began in earnest. By about the 1860s the European English speaking population became the majority, and from then on outnumbered the Māori population. Orthography for the Māori language was produced in the early 1800s, and during the latter part of that century, there was a large number of works printed in Māori.

However by the end of the 1800s most Māori were receiving schooling in English, and speaking Māori in schools was often actively discouraged. Between the 1950s and 1980s, there was a sharp decline in the number of fluent Māori speakers and the language became seriously threatened ([3, chapter 7]). Since the mid-1980s there has been a strong revitalization effort, leading to an increasing number of first language (L1) speakers, who are also fluent in English. However because there was a break in the intergenerational transmission of Māori [6], L1 speakers often learn Māori from second language (L2) Māori speakers, either their parents and/or teachers. All Māori speakers now are bilingual. Despite the break in transmission, traditional Māori ceremonies retain their high prestige. In particular Māori oratory skills are highly revered. Thus elders in the community are regarded as the "gold standard" for Māori pronunciation.

1.2. Sound Change in Māori

The contact between Māori and English over the last 180 years has affected the Māori language. The MAONZE project [1] has been investigating sound change in Māori from a corpus of 60 bilingual speakers (men and women), whose birthdates span over 100 years. The corpus comprises three groups. The first group of speakers, historical elders, were born in the late nineteenth century and recorded mostly in 1946-48. The second group are present day elders born between 1920-40, and the third group are two sets of young speakers (L1 and L2) born between 1970-90.

The phonology of Māori is fairly straightforward. The Māori vowel system is usually described in terms of the five short vowels /i e a o u/ [3]. There are 10 consonants /p t k m n ŋ f r w h/. There are no voicing contrasts within the stop consonants, and in addition they were originally unaspirated [3]. Māori syllables take the form (C)V(V(V)), they have an optional onset consisting of a single consonant, and a peak of up to three morae in length. Phonetically there is some regional variation in Māori, but only in the consonants, not the vowels.

The MAONZE study focused mainly on the vowels, investigating sound change in the long and short monophthongs and the five most common diphthongs /ai, ae, au, ou, ao/. The acoustic analysis was mainly based on the first and second formants, and also on vowel duration. The results showed changes in vowel quality and vowel duration (for both genders). In particular the /e: e/ pair has risen considerably, and the /u: u/ pair has fronted. In addition there was evidence of diphthong mergers, such as /ai/ and /ae/, /au/ and /ou/, especially amongst the younger speakers (e.g. see [7]). Some of these changes are clearly due to internal influences, such as the merger between /ai/ and /ae/ but others are due to external influences of NZ

http://dx.doi.org/10.21437/Interspeech.2017-215

English. The /e e:/ and /u u:/ movements are linked with DRESS rising, and GOOSE fronting respectively [7]

We have not only disseminated the findings to research audiences, but also to the larger Māori speaking community including language teachers [8]. In general the academic audiences were predominantly interested in the mechanisms and motivations of sound change. However the Māori community were excited about the systematic presentations of the vowel spaces of the elders compared to the young speakers, and the fact there were measurable differences. This coincided with what the community was hearing, but had been unable to quantify. Within the Māori speaking community sound change was not seen as a natural consequence of a living language, but rather as a consequence of a break in inter-generational transmission, and therefore to be regretted. The Māori language of the current elders is highly valued, and many younger speakers consider it a compliment if they are told they sound like an elder. The MAONZE project received requests to create a platform to enable Māori speakers to compare their speech with the current day elders in the MAONZE corpus. This led to the development of *MPAi*.

2. *MPAi* Background

There are many Computer-Assisted pronunciation training packages and more are in development (e.g. [9-16]). However with the exception of *MPAi* there are no tools which provide real-time feedback to assist learners to pronounce Māori correctly. Most of the available resources provide models of Māori exemplars via audio recordings, and the descriptions of Māori phonetics are simplistic and often inaccurate, referring to their supposed equivalents in English (see, e.g., [17]). There is in fact very little computer assisted learning for any language undergoing revitalization. Ojibwemodaa [18], ACORNS [19], and Hika [20] are some of the more robust examples available. They provide a series of audio and video resources, but none of focus on automated pronunciation feedback.

The *MPAi* project has been under development for 10 years [21, 22]. *MPAi* has two main modules. The first allows users to record and analyze their own vowels, displaying the user's vowels on a formant plot. The reason for using the formant plots was the well-known relationship of the first formant (plotted on the y axis) being inversely proportional to jaw opening, and the second formant (plotted on the x axis) being inversely proportional to tongue backing. An example of the formant module is given in Figure 1. The black dots are calculated from the user's input, the red vowels and the ellipses are the reference or target vowels (here the female present-day elders). The x and y axis are labelled with Jaw and Tongue position, rather than the actual formant values, which are not of interest to the target user. Initially we used only the vowel spaces of the present-day elders, but preliminary trials revealed that some of the learners wanted to match their productions against the young L1 Māori speakers. The historical recordings were not used as exemplars because the recordings are of poor quality and the materials are limited. Users can therefore choose between the present day elders or the young Māori, and also between male or female speakers for their models.

The words module is based on a purpose built list of words. It provides feedback on the production of the monophthongs and diphthongs. The latter was important because users wanted the chance to practise words which could potentially be impacted by diphthong mergers. Users get feedback telling them what word and phones have been identified, which phones are correct, and which are wrong. The data for the speech recognition system was trained from recordings of word lists read by the present-day elders, male and female.

As well as the analysis features, *MPAi* also enables users to access audio recordings from the MAONZE corpus and additional video material and to hear their own recordings looping with recordings from the exemplar database. The main portion of *MPAi* is written in C Sharp, Python was used to do formant analysis using the tkSnack package [23]. Automatic Speech recognition is performed using HTK (C executable) [24], Perl is also required in the training phase. The default values for the parameters were used in both tkSnack and HTK [23, 24]. *MPAi* runs on Windows 7 or higher.

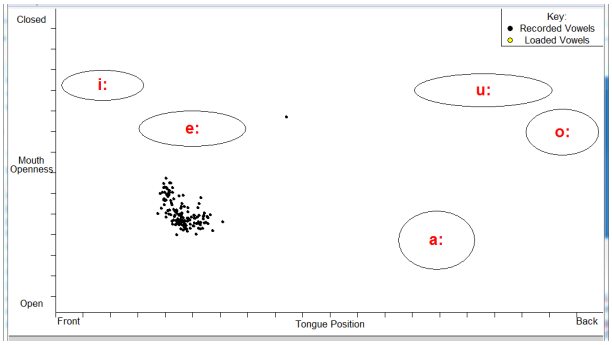

Figure 1: *Formant plot for a speaker practicing /e:/*

3. *MPAi* Evaluation

Whilst we performed a number of demonstrations of *MPAi* in its early years, the lack of robustness in the software meant that a standalone trial was impossible. By mid 2016 the software was stable and two user trials were completed. The first trial, with eleven participants, ensured there were no unexpected glitches. The main trial took place in November 2016 and we now report the results. Ethics approval for both trials was given by the University of Auckland.

There were 22 participants in the main trial. They were mainly students from the Māori medium teacher training programme at the University of Auckland. In addition to trialing *MPAi*, each participant filled out a questionnaire and undertook a ratings exercise. Seventeen women and five men with a mean age of 36.4 years (sd 10.7 years) were involved in the trial. Twelve of the participants learnt Māori as a child. When participants self-rated their Māori proficiency, five said they spoke it well or very well, 8 said they spoke it fairly well, 8 said they spoke it not very well, and one said they had only words and phrases. The participants therefore reflected the target users of *MPAi*, i.e. speakers who can speak Māori, but wish to improve their pronunciation. More participants would have been desirable, but the number of Māori to draw from is small. During the trial, participants were guided throughout by one of the authors (PK) or a trained research assistant. Participants spent between 20-40 minutes on *MPAi* (mean time 25.5 minutes). They used the two main modules, as well as playing the various videos and audio recordings. The results we present here focus on the userability of *MPAi*, and its perceived effectiveness. We have yet to measure the real effectiveness.

Feedback was solicited from the participants using items focusing on effectiveness and usability via a 5-point Likert scale. With regards of effectiveness: 18 participants (82%) felt the visual feedback from *MPAi* was effective and 16 participants (73%) felt it was effective in making them aware

of their articulators when speaking Māori. When assessing *MPAi* in terms of usability 18 of the participants felt confident using it. In addition 15 found *MPAi* easy to use, and that was also the number that felt other people would find it easy to use too. Despite the perceived effectiveness, only 10 participants agreed they would use *MPAi* frequently, although a further 9 were neutral on their perceived usage. There was a very strong perception that the structure of *MPAi* could be improved as only 9 participants felt the various functions were well integrated. The full distribution of responses about the effectiveness and usability issues from the questionnaire are given in the Likert plots in Figure 2.

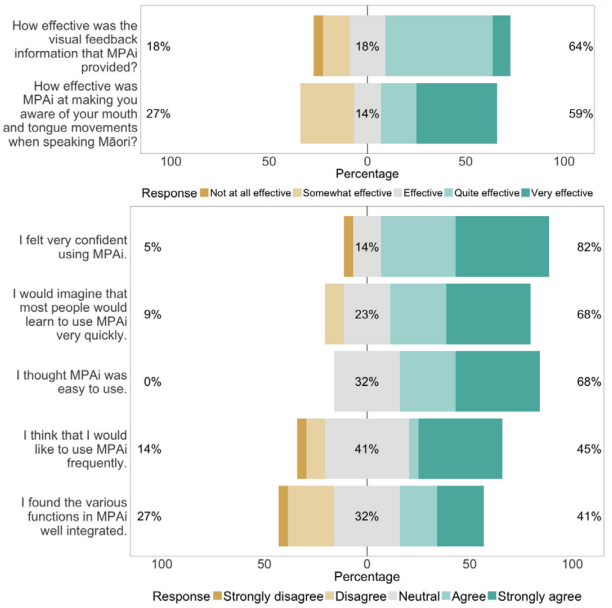

Figure 2: *Likert plots of the participants' responses, with the questions on the left. The top two plots are for the effectiveness questions, the remaining five are the responses for the usability questions. The right most percentages are the combined scores for the upper two ratings, the left most percentages are the combined scores for the lower two ratings, and the middle percentage is the score for the middle rating.*

From the open ended questions and comment sections we got insights into the possible reasons for the Likert responses. It was clear the participants had concerns about the words module which used speech recognition, 5 participants questioned its accuracy. The main motivation for this module was to help participants practise words that were minimal pairs, with the distinguishing phonemes being the diphthongs undergoing mergers (e.g. "pai" vs "pae"). The task, therefore, is a difficult one, the differences are very subtle, and potentially beyond the capacity of our recognition module. To further confound the issue our sound change results showed that this merger was also happening for some of the present-day elder models as well as the young speakers [7]. As the recognition module was trained on the present day elders, and the acoustic models were trained from phonemic labelled segments, the acoustic models may not be good enough to distinguish some of the minimal pairs.

In contrast the participants were very positive about the videos and the module based on formant analysis. 8 participants indicated they found the formant module useful. They became aware of tongue movement, and how it influences the vowel production. When we asked "What, if anything, did the aid teach you about pronunciation of Māori?" and "What did you like about the aid" the comments centred mainly around heightened awareness of the impact of tongue movement on vowel production. This can be seen in the word cloud representation of the responses in Figure 3, where words with a frequency of greater than 3 are shown.

Five of the participants found the formant plot too complicated, and wanted it simplified. One wrote "A chart or graph of some sort to show where the tongue should be when pronouncing vowels". 9 participants requested changes to *MPAi*, wanting it to be more colourful, appealing, with a less complex layout. A further 9 requested improvements to the supporting resources. They wanted videos of the lower face of a man and a woman saying the isolated vowels, and they also wanted to have videos of the tongue movements.

In spite of an earlier request to include younger L1 speakers' exemplars, none of the participants wanted to model their speech exclusively on this cohort. Two participants indicated age was irrelevant, four wanted to model on their own age group, but the majority (14) wanted to model on the present day elder group, and a further two selected both their own age group or the present day elder group.

Figure 3: *Word cloud generated from the questions "What did you like about the aid" and "What, if anything, did the aid teach you about pronunciation of Māori?" This plot was generated using TagCloud [25].*

4. Improvements to *MPAi*

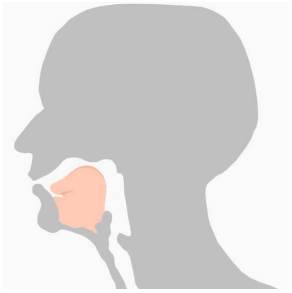

Figure 4: *Still from the /u:/ tongue movement animation.*

Between December 2016 and February 2017 we worked on improvements to *MPAi* based on the above feedback. To reduce the complexity of the aid, *MPAi* has been split into two standalone modules, one for practising words and one for practising vowels. In the vowel module, we created more reference resources. There are now vowel videos showing lip movements for two speakers (a male and a female). We have also generated tongue movement animations. These animations were generated from static midsagittal MRI images of the vocal

tract showing tongue positions of the five vowels, plus the vocal tract at rest. GIMP was used to create stylized images, removing the physiological details such as the brain stem, and spine, and it was also used to create the animation frames. These animations were then timed to coincide with the appropriate audio. Figure 4 shows a still from the /u:/ animation.

Vowel spaces on formant plots are meaningful to phoneticians, and speech scientists, but for most people they are overwhelming. From the feedback we have created a new target vowel plot, based on formant analysis. Users are able to focus on one vowel at a time, and their task is to get the trace as close to the target as possible. The target values are determined by the mean values from the MAONZE corpus for the speaker groups already mentioned (young male and female, present day elder male and female). The user identifies which group they wish to match their own speech to. Figure 5 shows the result for a user practising an /a:/ vowel.

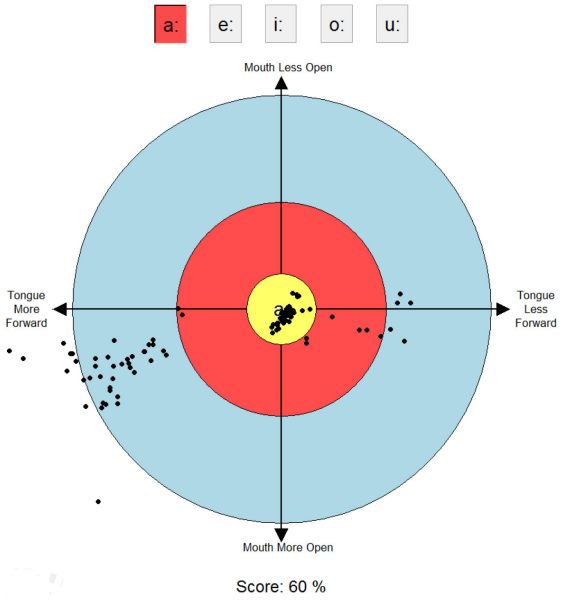

Figure 5: *Screen for the vowel target module.*

The outer edge of the blue circle is 3 standard deviations from the mean, the outer edge of the red circle is 1.5 standard deviations from the mean, and the outer edge of the yellow circle is 0.5 standard deviations from the mean. The scoring is set so that any trace in the yellow circle is 100% correct, and from the outer edge of the yellow circle to the outer edge of the blue circle is a linear decrement from 100% to 0% correct. In this particular example the user obtained a mean score of 60%. The interpretive axis gives the user feedback on how to improve their pronunciation. Initially this user had their /a:/ too front, then they hyper-corrected, before finishing on the target sound. The score for each attempt of the user is kept, and the user can view their progress report in a browser.

5. Discussion and Conclusion

It is tempting, in a pronunciation aid, to include lots of speech processing techniques. However these highly technical displays are of little use to the average language learner [15]. They require too much specialist knowledge to interpret. We used real-time formant analysis because we were able to construct vowel space exemplars from four groups of 10 speakers (young male and female, present day elder Male, and Female). The exemplars came from nearly 20,000 tokens. Whilst formant plots are used in Computer Assisted Pronunciation Training (CAPT) tools, the tendency is to recreate the formant plots in research papers in x and y scales given in Hertz (e.g. [9-11]). However this is not meaningful to most lay users. In *MPAi* the vowel plots were labelled according to jaw and tongue position, rather than frequency values. The feedback was clear: people found the linking between vocal tract shape and vowel type insightful. As one participant put it "Mouth openness and tongue position are pivotal to pronunciation". However the feedback was that the plots with the entire vowel system were still were too complex. The single vowel target displays (Figure 5) are our attempt to hone in on a single feature (a single vowel), and are yet to be tested with users.

Automatic Speech recognition (ASR) is an increasingly common technology to use in CAPT tools (e.g. [11-13]). However as Walker et al. [13] noted CAPT tools depend on ASR technology working well. Most of the CAPTs using ASR are targeted at L2 learners learning a foreign language, and this can have an impact on the selection of training materials (e.g. [13-15]). *MPAi* is designed currently to assist language revitalization. Currently the acoustic models are based on the present day elders group. One of the reasons for the perceived poor performance of the ASR module is potentially the lexicon the aid is currently trained on. As mentioned in Section 3 it contains many minimal pairs, where the distinguishing phonemes are potentially merging. We are currently compiling a new lexicon of common Māori words and phrases, and anticipate the perception of the ASR system will improve once this is incorporated into *MPAi*. We also intend to provide two sets of acoustic models, one from the present day elders group, and the other from the young group, enabling users to select which group they wish to match their pronunciation to. Like other studies (e.g. [12]) we believe there must be many exemplars to help the user create pronunciation models.

In the latest release of *MPAi* we have also increased the number of supporting resources, adding more videos and animations of vowel productions. The feature where users were able to listen to different speakers producing the same word was very popular with all our trial participants, who could already speak Māori. However any users who had little or no experience of Māori found this feature confusing.

Hansen [16] prescribes four essential characteristics of good feedback in CAPT: (1) it should be easily understandable, (2) it should be qualitative (correct phoneme?), (3) it should be quantitative (correct phoneme length?), and (4) it should suggest corrections. This has been our philosophy too. As noted by a participant in the trial "A lot of trust will be given to the application, so it needs to work well before [it's] released." Whilst this statement is true for any system that provides automated feedback, it is particularly crucial for a language which is seriously under resourced and is undergoing revitalization. Future enhancements of *MPAi* will be driven by further trials along with more guidance and input from the Māori-speaking community, cognizant of the high prestige attributed to the Māori elder's speech.

6. Acknowledgements

We thank the speakers in the MAONZE corpus, their recordings have made *MPAi* possible. Thanks also to the anonymous trial participants, the many *MPAi* project research assistants, Rashina Hoda for her help in compiling the survey, and the University of Auckland Strategic Research Initiative Fund grant for funding the major development of *MPAi* .

7. References

[1] J. King, M. Maclagan, R. Harlow, P. Keegan and C. Watson. "The MAONZE project: Changing uses of an indigenous language database." *Corpus Linguistics and Linguistic Theory*, 7, no. 1, pp. 37-57, 2011.

[2] R. Harlow and J. Barbour. "Māori in the 21st century: climate change for a minority language?" in *Language Ecology for the 21st Century: Linguistic Conflicts and Social Environments.* W. Vandenbussche, E. Håkon Jahr and P. Trudgill, Eds. Oslo: Novus Press, 2013, pp. 241-266.

[3] R. Harlow. *Māori: A Linguistic Introduction*. Cambridge: Cambridge University Press, 2007.

[4] Statistics New Zealand. "2013 Census QuickStats about Māori." Internet: http://www.stats.govt.nz/Census/2013-census/profile-and-summary-reports/quickstats-about-māori-english.aspx [Mar. 12, 2017].

[5] Statistics New Zealand. "Te Kupenga." Internet: http://www.stats.govt.nz/tekupenga, May 6, 2016 [Mar. 12, 2017].

[6] J.A. Fishman. *Reversing language shift: Theoretical and empirical foundations of assistance to threatened languages.* Clevedon: Multilingual matters, 1991.

[7] C.I. Watson, M.A. Maclagan, J. King, R. Harlow and P.J. Keegan. "Sound change in Māori and the influence of New Zealand English." *Journal of the International Phonetic Association*, vol. 46, no. 2, pp. 185-218, Aug. 2016.

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

[9] C. Lie-Lahuerta. "Fix your vowels: Computer-assisted training by Dutch learners of Spanish." *Tijdschrift voor Skandinavistiek*, vol. 32, nos. 1 & 2, Jan. 2011.

[10] A. Dowd, J. Smith and J. Wolfe. "Learning to pronounce vowel sounds in a foreign language using acoustic measurements of the vocal tract as feedback in real time." *Language and Speech*, vol. 41, no. 1, pp. 1-20, Jan. 1998.

[11] D. Brett. "Computer generated feedback on vowel production by learners of English as a second language." *ReCALL*, vol. 16, no. 1, pp. 103-13, May 2004.

[12] A. Neri, C. Cucchiarini, H. Strik and L. Boves. "The Pedagogy-Technology Interface in Computer Assisted Pronunciation Training." *Computer Assisted Language Learning*, vol. 15, no. 5, pp. 441-467, 2002.

[13] N.R. Walker, P. Trofimovich, H. Cedergren and E. Gatbonton. "Using ASR Technology in Language Training for Specific Purposes: A Perspective from Quebec, Canada." *CALICO Journal*, vol. 28, no. 3, pp. 721-743, 2011.

[14] Y. Ohkawa, M. Suzuki, H. Ogasawara, A. Ito and S. Makino. "A speaker adaptation method for non-native speech using learners' native utterances for computer-assisted language learning systems." *Speech Communication,* vol. 51, pp. 875-882, 2009.

[15] C. Cucchiarini, A. Neri and H. Strik. "Oral proficiency training in Dutch L2: The contribution of ASR-based corrective feedback." *Speech Communication*, vol. 51, pp. 853–863, 2009.

[16] T.K. Hansen. "Computer assisted pronunciation training: the four 'K's of feedback," in *4th Internat. Conf. on Multimedia and Information and Communication Technologies in Education*, 2006, pp. 342–346.

[17] H. Niwa. *Pronounce Māori with Confidence.* Auckland: Reed, 2003.

[18] Grassroots Indigenous Multimedia. "Speak Local." Internet: http://gim-ojibwe.org/ [Mar. 12, 2017].

[19] D. Harvey. "ACORNS overview." Internet: http://cs.sou.edu/~harveyd/acorns/index.php [Mar. 12, 2017].

[20] Hika Group. "Hika, Languages Connecting People." Internet: http://hikagroup.com/ [Mar. 12, 2017].

[21] J. King, R. Harlow, C. Watson, P. Keegan and M. Maclagan. "Changing Pronunciation of the Māori Language: Implications for Revitalization" in *Indigenous Language Revitalization Encouragement, Guidance & Lessons Learned.* J. Reyhner and L. Lockhead, Eds. Flagstaff: AZ: Northern Arizona University, 2009, pp. 75-86.

[22] P.J. Keegan, C.I. Watson, J. King, M. Maclagan and R. Harlow. "The Role of Technology in Measuring Changes in the Pronunciation of Māori over Generations," in *Language Endangerment in the 21st Century: Globalisation, Technology and New Media, Proceedings of Conference FEL XVI.* T. Ka'ai, M. ÓLaoire, N. Ostler, R. Ka'ai-Mahuta, D. Mahuta and T. Smith, Eds. Auckland: Te Ipukarea - The National Māori Language Institute, AUT University/Foundation for Endangered Languages, 2012, pp. 65-71.

[23] K. Sjölander "Snack Manual." Internet: http://www.speech.kth.se/snack/man/snack2.2/tcl-man.html [Mar.12, 2017].

[24] Cambridge University Engineering Department. "HTK Speech Recognition Toolkit." Internet: http://htk.eng.cam.ac.uk/ [Mar. 12, 2017].

[25] D. Steinbock. "Tag Crowd." Internet: http://tagcrowd.com/ [Mar. 12, 2017].