Towards Automatic Marking of Pepeha: a Formulaic Māori Language Speech

Catherine I. Watson[1], Piata Allen[2], Peter J. Keegan[2], Keoni Mahelona[3], Peter-Lucas Jones[3]

[1]Faculty of Engineering, [2]Faculty of Education and Social Work, University of Auckland Waipapa Taumata Rau, Aotearoa New Zealand
[3]Te Hiku Media, Aotearoa New Zealand

c.watson, p.allen, p.keegan@auckland.ac.nz, keoni,peterlucas@tehiku.co.nz

Abstract

This study looks at the assessment of the pronunciation of te reo Māori (Māori language) within a short formulaic speech, known as a pepeha. It is a comparison between marks awarded to the pehepa by trained markers, and scores awarded to the pepeha by Arero, a speech recognition platform which is purpose-build to assess te reo Māori. Pepeha recordings of 304 people were analysed. The study found that there were many similarities between the two assessment methods and they were correlated, albeit weakly. It is argued that the results suggest automatic marking of pepeha is feasible. The next step is to understand acceptable phonetic variation in the pepeha pronunciation via phonetic analysis of a large number of the pepeha recordings.

Index Terms: speech recognition, automatic marking, speech production, te reo Māori (Māori language)

1. Introduction

Te reo Māori (Māori language) is the indigenous language of Aotearoa New Zealand. It is also one of the official languages of the country, although it is only spoken "fairly well" by 7.9 % of the current population, which is 5.2 million people [1,2]. After many years of colonial neglect and repression, the importance of the language beyond the Māori community, is finally being recognized countrywide. In 2019 the central government launched an initiative to have 1 million speakers of te reo Māori (henceforth referred to as Māori) by 2040 [3].

The phonology of Māori is straightforward [4]. It has 10 consonants /p t k m n ŋ f r w h/. There are no consonant clusters, and no voicing contrasts within the stop consonants, and in addition stops were originally unaspirated [4]. The vowel system is usually analysed in terms of five short vowels, /i, e, a, o, u/. These may occur alone or in sequences, vowel length is phonemic. Sequences of mid or low vowels, followed by a high vowel can form diphthongs. Sequences of long vowels and diphthongs may be formed across word and morpheme boundaries. All Māori syllables are open, and 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.

Māori is written in a Roman script. It has 13 graphs, for 8 of the consonants, and all 5 vowels. There are two digraphs, "ng" for the velar nasal, and "wh" for the labio-dental fricative. Vowel length is typical represented by a macron over the vowel, but this is not absolute. For a more in depth discussion please refer to [4]

Over 60 % of the New Zealanders now believe Māori should be taught in schools [1], and the Teaching Council of Aotearoa New Zealand requires all teachers to develop the use of Māori to meet the standards of the teaching profession and to fulfil the Code of Professional Responsibility [5]. At the University of Auckland (Waipapa Taumata Rau), EDUCM199 is a compulsory online course for all initial education students (trainee teachers) to learn (or validate as some students can already pronounce Māori correctly) the pronunciation of Māori. Since 2020 over 1000 students have completed the semester long course teaching the pronunciation of the language. The course includes a pepeha which is a highly formulaic tribal Māori language proverb that is commonly used in modern times as a short introductory speech. A pepeha situates a speaker, identifying local geographical features, and communities (traditionally a particular tribal group (iwi)) to which they belong or in whose region they are domiciled. The pepeha learnt by the students is specifically for the Faculty of Education and Social Work Campus, located at Epsom (Maungawhau), Auckland, i.e., the faculty to which they belong and where they attend classes. We will refer to this pepeha as the Epsom pepeha and it is listed in Figure 1.

line 1	Ko Maungawhau, ko Maungakiekie ngā Maunga
	The mountains are Maungakiekie and Maungawhau..
line 2	Ko Waitematā, ko Manuka ngā Whanga
	The harbours are Waitematā and Manuka
line 3	Ko Tūtahi tonu te whare
	The meeting house is Tūtahi Tonu
line 4	Ko te Aka Matua o te Pou Hawaiki te marae
	The courtyard is Te Aka Matua o Te Pou Hawaiki
line 5	Ko Niwaru te waka
	Niwaru is the waka (canoe)
line 6	Ko Tuputupu Whenua te tangata
	Tuputupu Whenua is the ancestor

Figure 1: The Epsom pepeha, with the English translation in italics

In EDUCM199 the assessment of the pepeha is very time consuming and requires many people. If the pepeha has not met the required standard students are invited to resubmit. Students are provided with written feedback about where they are making mistakes, and what they need to improve. The students who are really struggling are provided with individual face to face tuition. The students are able to make multiple submissions. Consequently there a lot to be marked, in a very short amount of time. This has lead the teaching team to investigate whether an automatic marking solution would be feasible. If an automatic marking system could mark the well pronounced pepeha, the teaching team could focus on the students that were struggling.

There is a substantial body of work on the automatic marking of well-resourced languages such as English, for example [6-11]. This scoring is for second language assessment, which is comparable to the pepeha exercise described above. The scoring in [6-11] is based around pronunciation, fluency, grammar, and intelligibility. The scoring system proposed in [11] has become the bench mark in many of these studies. The correlation between these automatic marking systems, and human markers has found to be quite high, ranging from 0.7 [7,9,10] to well over 0.9 [6]. Training these systems involves a lot of data, which for some Englishes, such as American English, there is much available. There is also a large body of study around pronunciation assessment (e.g. [12-16]). Here machine learning is used to provide feedback on what phones were mispronounced, and what were the speech errors. Again it is the well resource languages for which these systems are typically built.

Due to the small numbers of people with expertise in speech technology, and an awareness of spoken Māori, only a few people have worked on using speech technology to help with the learning of the language. The first known development was in 1992 [17] where Kingi built an automatic speech recognition system that provided feedback on the pronunciation of words. Ten years later Laws [18] developed the first text to speech (TTS) system, which was intended to be part of a talking dictionary. Fifteen years on from that was MPAi [19], a system that provides real-time feedback on Māori vowel pronunciation. Then just under 20 years after Law's diphone TTS system, an HMM based TTS system was developed [20,21] which is intended to be part of the MPAi platform. Despite the intention that all the above platforms would be used to aid teaching Māori, only MPAi was been formally evaluated by te reo Māori teachers.

In preliminary work [22] towards automatic marking of pepeha, a system was built on the framework proposed in [13]. It is based a combination of deep learning networks to create an end-to-end speech recognition system that does mispronunciation detection and diagnosis. The training dataset comprised the Ngā Mahi corpus [20,21] which is 2.3 hours of a single speaker, and 1 ½ hours of read speech from 28 speakers of the MAONZE corpus [23-25]. When this network was tested with pepeha recordings from the EDUCM199 corpus we got an accuracy of 60%. Whilst this work was promising, we needed the speech recognition platform to be trained on much more data. But we did not have anything suitable, therefore we had to look at alternative approaches.

A notable exception to lean data in Māori speech technology developments is the work done by Te Hiku Media with their multi-lingual language platform Papa Reo (https://papareo.nz/) [26]. This broadcasting and technology iwi-based company has been leading Māori speech technology development since 2018 when they received a large government grant. Te Hiku Media have had the support from their large community of te reo Māori speakers and have created a large spoken language corpora of 400 hours of recordings from 2,200 speakers of Māori, saying 5000 unique sentences [26]. From this large corpora they have created multiple speech technology tools including Arero which is a Māori speech recognition engine. Arero is based on recurrent neural networks and calculates confidence scores for phones. The output from Arero is used by Papa reo in a number of ways including Rongo, which is an application released in 2022 which teaches and provides feedback on Māori pronunciation.

Arero, therefore, seemed a very suitable platform to investigate the feasibility of automatic marking of pepeha. This study outlines our comparison of the results of the hand marked pepeha, with the output from Arero, the Māori speech recognition engine. The question we seek to answer is how well speech recognition scores from Arero align with the marks awarded to the pepeha from the Marker.

2. Methodology

2.1 Speakers and Material

Ethics approval was given to analyse the pepeha recordings of the students. The EDUCM199 2020 corpus consists of 318 recordings of students saying pepeha. The recordings were all uploaded by the students to CANVAS (https://canvas.auckland.ac.nz/), a web-based course delivery platform. The students recorded themselves, typically using the recording functionality on their smart phones. The recordings were saved in various media (audio or video) on CANVAS, as m4p or m4a files. The markers would listen to recordings on CANVAS and mark the pepeha. Based on the feedback from the markers the students could decide to resubmit a different recording of the pepeha to improve their mark. Each time a student resubmitted, CANVAS wrote over the previous recording. CANVAS only allowed the most recent recording to be saved.

For this study, all the recordings were downloaded from CANVAS and were converted to wav files for analysis, typically these had a 44.1 kHz sampling rate, and were 24 bit. Not all of the recordings were useable. We had to discard all those that were not the Epsom pepeha (listed in Figure1). In addition whilst some of the recordings were present, the associated hand marks were missing, so these recordings too were discarded. Further some of the recordings had extra material such as general Māori greetings, these were removed so the recording was only of the pepeha. This left us with a corpus of 304 recordings of different students saying the Epsom pepeha. It is these that are studied in this paper.

2.2 Marking

The pepeha marking team fluctuated depending on availability, but was usually a team of two to four. One person marked the whole pepeha but different markers could mark different submissions by the same student. Each marker is given training on exemplar scripts and their marking was checked and they were given feedback. The marking team met regularly to moderate each other's work. Borderline submissions are moderated by a second marker.

Each pepeha was given a mark out of 100 and the marks were given on a syllable basis, with three exceptions: "kie", "tua" and "nua", where the two syllables were treated as one. This was done to simplify the marking, i.e., keep total marks to 100. The syllables that were anticipated as being harder to pronounce were marked out of 2, the other were marked out of 1. Figure 2 outlines the marking schedule. When a syllable was marked out of two the possible marks were 2,1, or 0, when the mark is out of 1, the possible marks are either 1 or 0. It can be seen from Figure 2 that the first two lines of the pepeha were anticipated as being difficult to pronounce, as all the syllables are marked out of 2.

Unlike the recordings, it is possible to access the markers feedback for all submissions made by the students. However we only looked at the marks from the last submission, which corresponded to the recording we could access. Throughout this study the data from the marks given by people will be called marks.

Line 1	Ko	Mau	nga	whau,	ko	Mau	nga	kie	kie	ngā	mau	nga				Mark Total
Marks	2	2	2	2	2	2	2	2	2	2	2	2				24/24
Line 2	Ko	Wai	te	ma	tā,	ko	Ma/Mā	nu	ka/kau	ngā	Wha	nga				
Marks	2	2	2	2	2	2	2	2	2	2	2	2				24/24
Line 3	Ko	Tū	ta	hi	To	nu	te	wha	re							
Marks	1	1	1	1	1	1	1	1	1							9/9
Line 4	Ko	Te	A	ka	Ma	tua	o/ki	Te	Pou	Ha	wai	ki	te	ma	rae	
Marks	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	16/16
Line 5	Ko	Ni	wa	ru	te	wa	ka									
Marks	1	1	1	1	1	1	1									7/7
Line 6	Ko	Tu	pu	tu	pu	whe	nua	te	ta	nga	ta					
Marks	2	1	2	1	2	2	2	2	2	2	2					20/20

Figure 2: The Marking schedule for the speech, giving the marks per syllable, and per line

2.3 Speech recognition Scoring

The speech recognition scoring was done using Arero developed by Papa Reo (Te Hiku Media), it gives a log likelihood probability score on each phone. The input into Arero was a wav file and text of the Epsom pepeha, and the output json file with probability scores for each phone. Due to the fact the orthography of Māori is almost phonetic, Arero uses the text to identify the phones. A visual representation of the output of Arero is given in Figure 3. The characters of each word in the pepeha are given a specific colour, depending on the log likelihood score of the associated phone. For example any phone with a log likelihood of greater than 0.9 was associated with a character coloured purple. Arero gives the two letters associated with the diagraphs the same log likelihood score, this is why the characters in “ng” are always the same colour, as are the characters in “wh”, these two diagraphs can be seen in the word “maungawhau” in the top line in Figure 3. Arero treats diphthongs are vowel sequences, this is why the diphthong /au/ which occurs twice in “maungawhau:” is represented as two characters with different colours.

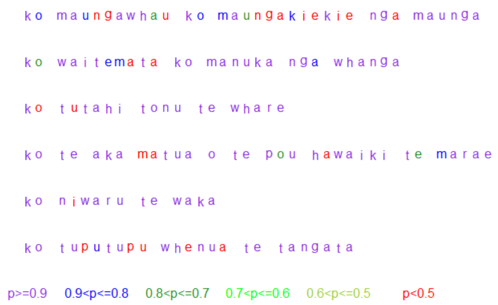

Figure 3: A visual depiction of the output of Arero from a recorded pepeha, with colour presenting the log likelihood range.

To be able to compare the scoring from Arero with that from the hand marked data we needed the Arero scores to be on a syllable basis, rather than phone basis. To get the log likelihood value for each syllable we looked at the log likelihoods of the phones within the syllable and took the minimum. This was a deliberately conservative approach, which is appropriate in pronunciation assessment. The phones within the syllable are given in Figure 2. Through this study the data derived from the output of Arero will be called scores.

The processing of the mark and score data was done using functionality from both R and Excel.

3. Results

In the comparison of the marks awarded by people and the speech recognition scores from Arero, we considered the overall marks and scores for the pepeha. The overall mark for each pepeha was out of 100. An overall score was calculated for the pepeha by treating the log-likelihoods for each syllable as a number, and adding these together. Thus the overall possible score for the pepeha was 66.

.

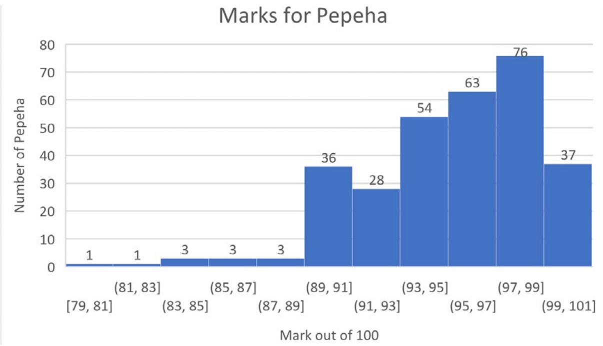

Figure 4 The marks awarded for the pepeha

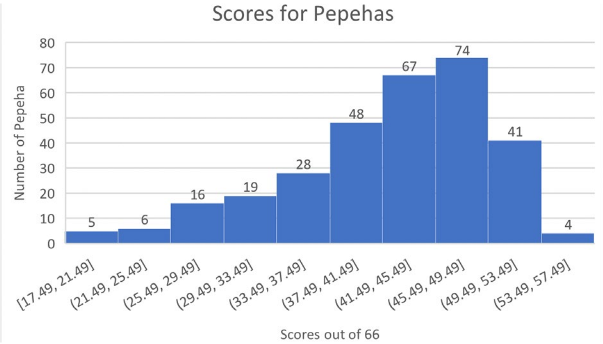

Figure 5 The scores from Arero for the pepeha

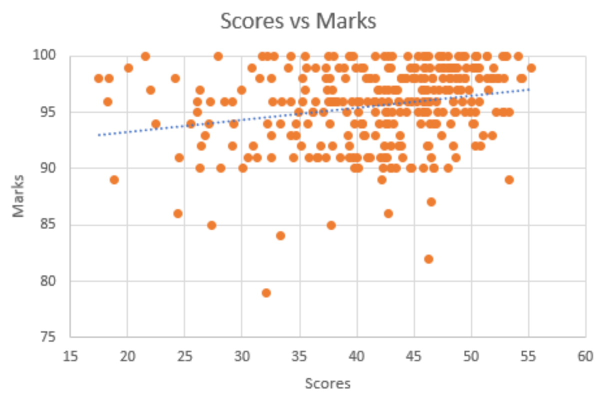

Figure 6: The overall summed scores per pepeha (x axis) compared to the marks per pepeha (y axis), a trend line in blue has been plotted.

Figure 4 shows a histogram of the marks awarded for the pepeha. The data distribution is highly negatively skewed, with a median mark is 96/100, and the vast majority of the marks are above 90/100, These are expected as the students were able to make multiple submissions. Further they were required to achieved 90 % across two equally weighted assignments to pass the course, and the pepeha production was one of the assignment. Figure 5 shows the scores awarded to the pepeha by Arero. The data distribution is also skewed to the right, but to a lesser degree than for the marks. The median score is 43/66, and additionally there is a much larger spread of the scores for the pepeha in comparison to the marks; the score range is 40/66 compared to 21/100 mark range.

Figure 6 shows the scatter plot of the overall Arero scores versus the hand marked total per pepeha. Whilst there is a trend that pepeha with high marks also achieve high scores, the correlation between the marks and scores is only weak, at 0.23. It is quite clear from Figure 6 that there are a notable number of pepeha that got high marks but did not get high scores from Arero.

4 Discussion and Future Work

The correlation between the automatic speech recognitions system and the base line (hand marked data) was considerably weaker than the systems presented earlier [6,7,9,10]. The weak correlation between the total marks and final score for the pepeha may be for a number of reasons. Firstly the quality of the recordings was very variable, and whilst the markers may have been able to ignore the noise, its presence may negatively impact the Arero scores. In addition the markers were only marking on pronunciation, not fluency, but Arero, trained on continuous speech, would be marking on a combination of both. Finally there is also some suggestion of marker variability. A phonetic analysis of 40 of the pepeha recordings suggests that the markers did not pick up on all the speech errors, for instance 19/40 of the velar nasals in the particle “ngā” in the first line were pronounced as alveolar nasals, but this was not picked up by the markers.

The correlation between the marks and scores may have been confounded by the fact that by the majority of marks were above 90%, so perhaps there was a ceiling effect. To test this, we looked at a subset of the pepeha where the marks were more variable. The six lines of the pepeha varied in difficulty, with the first line having the largest number of difficult sounds (see Figure 2 for the syllables involved). As mentioned in Section 2.2 this was why each syllable in the line was marked out of two. Line 1 had the lowest average mark and score, of all six lines. When we investigated the correlation between the marks and scores for this line we found it was 0.25. This is an increase from the overall correlation but it still remained weak. Thus we could not confirm a ceiling effect.

Looking at what the two grading approaches implied about syllable pronunciation there were many similarities. Notably both the hand marking and the pepeha scores suggested that syllables with the diphthong /au/ (i.e. “mau”, and “whau” in line 1 of the pepeha) were difficult to pronounce. Further, both approaches also suggested that the syllable “hi:” and any syllable with the short vowel /a/ (e.g. “ka”) were easy to pronounce. There were a few differences between the approaches. Arero identified that the syllable “ngā”, which has a long /a/ was not well pronounced, scoring an average log likelihood of 0.09. However, “ngā” was marked highly in the hand marked data. As suggested above this may be because the markers were not picking up that the velar nasal was not being pronounced correctly. But it might also be related to fluency, which Arero was better at detecting. Another difference noted was the relatively low average mark for “ru” (0.74 out of 1) from the hand marked data, however the averaged log likelihood value was 0.68.

Whilst not comprehensively conclusive, the scores from Arero do align with the marks of the markers, and therefore automatic marking of pepeha is feasible. The comparison suggests that Arero is much more sensitive picking up variations in speech than the markers. This sensitivity can be seen Figure 3 which the visual depiction of the log-likelihood scores from the exemplar recording of the pepeha – the one provide to the students to listen to when learning how to pronounce the pepeha. It is therefore reasonable to expect that it is well pronounced. However it can be seen that whilst many phone did have a log likelihood greater that 0.9 (purple text), there were notable number of text coloured red, which means those phones had a log likelihood of less than 0.5. To be useful in automatic marking it will be necessary to identify which is the log likelihood threshold for good pronunciation in Arero.

However comparing the Arero scores to the marks may not be the best way to identify what is the log likelihood threshold for a good pronunciation. There is a strong suggestion that the markers became selective in the errors they were assessing. In order to truly assess how the output from Arero can help with automatic marking of pepeha we need to have a much better sense of the types of speech errors that are occurring, and how sensitive Arero is to them. We are about to embark on an in depth phonetic analysis of the EDUCM199 2020 corpus. We need the analysis to both understand the types of errors that occur, but also we need to understand the limits of acceptable variation. Unlike well-resourced languages such as English, there has been no systematic analysis of speech errors for learners of the language, nor on what is acceptable variation in pronunciation.

In 2026 the University of Auckland aims to offer a compulsory course teaching the pronunciation of te reo Māori, all students will be required to complete the course. pepeha will also be used in this course as a means of assessing pronunciation, albeit a different pepeha (which is based on the main city campus). As the University of Auckland has over 10,000 new undergraduate students per annum, the need to develop an automatic marking platform for pepeha is absolute.

5 Conclusions

In this study we have looked at the assessment of the pronunciation of te reo Māori (Māori language) within a short formulaic proverb, known as a pepeha. We have compared the marks awarded to the pepeha by trained markers, to the scores awarded to the pepeha by the speech recognition platform Arero, which is purpose-build to assess te reo Māori. We were assessing whether the marks and scores were aligned. Whilst there was only a weak correlation between the overall marks and scores, there was a lot of similarity in the overall conclusions that could be drawn about the pronunciation of the pepeha at both the line level and the syllable level. This was both in terms of what was pronounced well, and what was not. The next step is to do an in-depth phonetic analysis of the pepeha recordings to understand what the speech errors are, and to understand the acceptable variability in spoken Māori.

6 Acknowledgements

We thank Tūreiti Keith, and Jesin James for their insightful comments and thank Rochai Taiaroa, Carrie Ng, Ayush Jani, Caleb Baptista and Phillipp Governale for their contributions to this project.

References

[1] StatsNZ "Te reo Māori proficiency and support continues to grow" https://www.stats.govt.nz/news/te-reo-maori-proficiency-and-support-continues-to-grow/,2022, site accessed 26/7/2023

[2] StatsNZ "Population" https://www.stats.govt.nz/topics/population, 2023, site accessed 26/7/2023

[3] Radio New Zealand "Plan to have 1 million people speaking te reo Māori by 2040" https://www.rnz.co.nz/news/te-manu-korihi/383063/plan-to-have-1-million-people-speaking-te-reo-maori-by-2040, 2019, site accessed 26/7/2023

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

[5] Education Council New Zealand *Our code our standards . Ngā tikanga matatika ngā paerewa.* https://teachingcouncil.nz/assets/Files/Code-and-Standards/Our-Code-Our-Standards-Nga-Tikanga-Matatika-Nga-Paerewa.pdf, n.d.

[6] Cheng, J., D'Antilio, Y. Z., Chen, X., & Bernstein, J. Automatic assessment of the speech of young English learners. In *Proceedings of the ninth workshop on innovative use of NLP for building educational applications*, June, 2014, pp. 12-21.

[7] Yu, Z., Ramanarayanan, V., Suendermann-Oeft, D., Wang, X., Zechner, K., Chen, L., ... & Qian, Y. Using bidirectional LSTM recurrent neural networks to learn high-level abstractions of sequential features for automated scoring of non-native spontaneous speech. In *2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*, December, 2015, pp. 338-345.

[8] Chen, L., Tao, J., Ghaffarzadegan, S., & Qian, Y. End-to-end neural network based automated speech scoring. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* April, 2018, pp. 6234-6238.

[9] Qian, Y., Lange, P., Evanini, K., Pugh, R., Ubale, R., Mulholland, M., & Wang, X. Neural approaches to automated speech scoring of monologue and dialogue responses. In *ICASSP 2019 IEEE international conference on acoustics, speech and signal processing*. May, 2019, pp. 8112-8116.

[10] Tao, J., Ghaffarzadegan, S., Chen, L., & Zechner, K. Exploring deep learning architectures for automatically grading non-native spontaneous speech. In *2016 IEEE International conference on acoustics, speech and signal processing (ICASSP).* March, 2016, pp. 6140-61441.

[11] Zechner, K., Higgins, D., Xi, X., & Williamson, D. M. Automatic scoring of non-native spontaneous speech in tests of spoken English. *Speech Communication*, *51*(10), 883-895. , 2009

[12] Arora, V., Lahiri, A., & Reetz, H. Phonological feature-based speech recognition system for pronunciation training in non-native language learning. *The Journal of the Acoustical Society of America*, *143*(1), 98-108. 2018.

[13] Leung, W. K., Liu, X., & Meng, H. CNN-RNN-CTC based end-to-end mispronunciation detection and diagnosis. In *ICASSP 2019 IEEE International Conference on Acoustics, Speech and Signal Processing* May, 2019, pp. 8132-8136.

[14] Li, K., Qian, X., & Meng, H. Mispronunciation detection and diagnosis in l2 english speech using multidistribution deep neural networks. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *25*(1), 193-207. 2017.

[15] Zhang, L., Zhao, Z., Ma, C., Shan, L., Sun, H., Jiang, L., ... & Gao, C. End-to-end automatic pronunciation error detection based on improved hybrid ctc/attention architecture. *Sensors*, *20*(7), 1809. 2020.

[16] Yan, B. C., & Chen, B. End-to-end mispronunciation detection and diagnosis from raw waveforms. In *2021 29th European Signal Processing Conference (EUSIPCO)*). August, 2021, pp. 61-65.

[17] P. Kingi, "*Te whakahuatia reo Māori/A Māori speech training system.*" Unpublished BSc Dissertation Thesis, University of Waikato, New Zealand. 1992.

[18] Laws, Mark R. Speech Data Analysis for Diphone Construction of a Māori Online Text-to-speech Synthesizer. In: *IASTED Int. Conference,* USA. 2003, pp. 103–108.

[19] Watson, Catherine I, Peter Keegan, Margaret Maclagan, Ray Harlow & J King. The Motivation and Development of MPAi, a Māori Pronunciation Aid. In *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 2017., 2063–2067.

[20] Shields, Isabella, Catherine I. Watson, Peter J. Keegan, Rebekah Berriman & Jesin James. Te Reo Māori Voice for TTS. Language *Technology for All Conference.* Paris, 2019., https://lt4all.elra.info/media/papers/P1/136.pdf.

[21] James, Jesin, Isabella Shields, Rebekah Berriman, Peter J. Keegan & Catherine I. Watson. Developing resources for te reo Māori text to speech synthesis system. In *International Conference on Text, Speech, and Dialogue*, 2020, 294–302. Cham: Springer. DOI: 10.1007/978-3-030-58323-1_32.

[22] Governale, P. "*Mispronunciation detection and diagnosis for English and te reo Māori using deep learning*" A Report written for the partial requirement of a Bachelors of Engineering Honours (Electrical and Electronic), University of Auckland. 2019.

[23] MOANZE, MAONZE project website https://maonze.blogs.auckland.ac.nz/, n.d., site accessed 26/7/2023

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

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

[26] Moses, C., Thompson, M., Mahelona, K., & Jones, P. L. Scoring pronunciation accuracy via close introspection of a speech recognition recurrent neural network. *Conference on Neural Information Systems, NeurIPS 2020*. 2020.