Te Reo Māori Voice for TTS

Isabella Shields[1] (Ngāti Porou), Catherine I. Watson[1], Peter J. Keegan[2] (Waikato-Maniapoto, Ngāti Porou), Rebekah Berriman[1] (Ngāi Tahu), Jesin James[1]

Engineering[1], Te Puna Wānanga[2], University of Auckland

THE UNIVERSITY OF AUCKLAND Te Whare Wānanga o Tāmaki Makaurau NEW ZEALAND | ENGINEERING

THE UNIVERSITY OF AUCKLAND Te Whare Wānanga o Tāmaki Makaurau NEW ZEALAND | EDUCATION AND SOCIAL WORK TE PUNA WĀNANGA

Introduction

Te Reo Māori -The Speakers

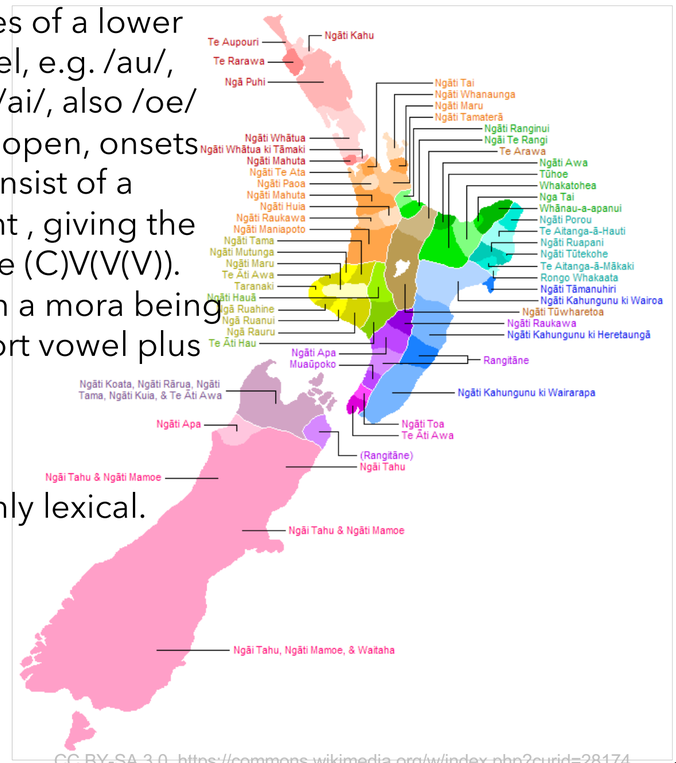
- Indigenous language of Aotearoa/New Zealand
- Aotearoa/New Zealand last major land mass to be discovered by humans, and was settled about 1200 AD.
- Until late 1700s no contact with other languages
- Over the last 200 years Māori has had increasing contact with New Zealand English.
  - late 18th C: earliest contacts with English speakers
  - 1st half 19th C: beginning of European settlement and missionary activity
  - late 19th C: schooling in English widespread
    - older Māori still monolingual
  - 1st half 20th C: Māori population bilingual
- In 2019 Māori = 15% of NZ pop of 4. 9 million
- 14% of Māori can speak Māori 'well' or 'very well'
- 25% can conduct a basic conversation

Te Reo Māori - Revitalisation

- Shift away from Māori occurred post 1920s, leading to a break in intergenerational transmission.
- Initiatives began in 1980s
  - Immersion schooling
  - Broadcasting
  - Largely urban, largely second language speaking adults
- Treaty settlements from late 1990s given rise to tribal authorities which have become important commercial and societal entities with the ability to devote resources to aspects such as revitalisation initiatives.

Te Reo Māori – The Language

- 10 monophthongs, 5 long short pairs:
  /i: i e: e a: a o: o u: u/
- 10 consonants:
  /p, t, k, m, n, ŋ, f, h, r, w/
- Diphthongs: at least within syllables are all mainly lexical sequences of a lower and higher vowel, e.g. /au/, /ou/, /ao/, /ae/, /ai/, also /oe/
- All syllables are open, onsets are empty or consist of a single consonant , giving the syllable structure (C)V(V(V)).
- Mora timed-with a mora being defined as a short vowel plus any preceding consonant.
- Regional differences mainly lexical.



Building the Voice

Kua hangaia e mātou he reo rorohiko tāne. Ka noho tēnei reo rorohiko ki roto i te pūmanawa MaryTTS. Ka whakaurua e te tangata he tuhinga, ā, ka hurihia he reo-ā tangi. I ahu mai ngā kōrero i te pukapuka Ngā Mahi a Ngā Tupuna. Ka āta whiriwhiria kia mau te katoa o ngā tangi oro o te reo Māori. Kia hangai atu ngā oro tuhi ki ngā oro tangi, ka whakamahia te Montreal Forced Aligner, nā te tangata anō i atā tirotiro te tika o ngā hononga. Ka tukua ēnei raraunga me tētahi papakupu whakaahua ki a MaryTTS.

- We have built a synthetic male te reo Māori voice, which runs on MaryTTS [1]. It enables the creation of synthetic speech to any input text in te reo Māori (see Fig1).
- The recordings took over 8 hours to complete, but condensed down to two hours of continuous speech.
- The recordings were done in a soundproof booth, the lapel microphone was a Rodes Lavellier, and a Roland Octo-capture was used for digitisation and amplification.
- The recordings excerpts from *Ngā Mahi a Ngā tūpuna*.
- Text was broken up into phrases, phrasing was informed by punctuation. 1030 sentences were recorded, the selection criteria ensured full diphone coverage.
- The phonetic labels were first determined automatically using a model obtained from Montreal Forced aligner [2]. The phonetic boundaries were hand checked.
- This labelled data is then passed to MaryTTS, along with a te reo Māori Phonetic dictionary, and using this information a Synthetic voice is created.
- The te reo Māori Phonetic dictionary we created contains over 10,000 words along with phonetic transcription of word, syllable boundaries and stress mark up. The latter two are determined automatically by a set of rules based on Bigg's stress rules and mora rules [3] (see Fig. 2).

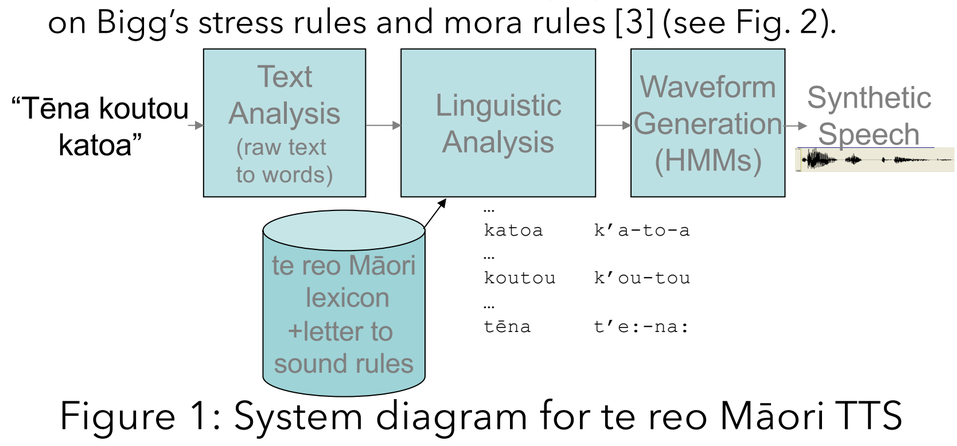


Figure 1: System diagram for te reo Māori TTS

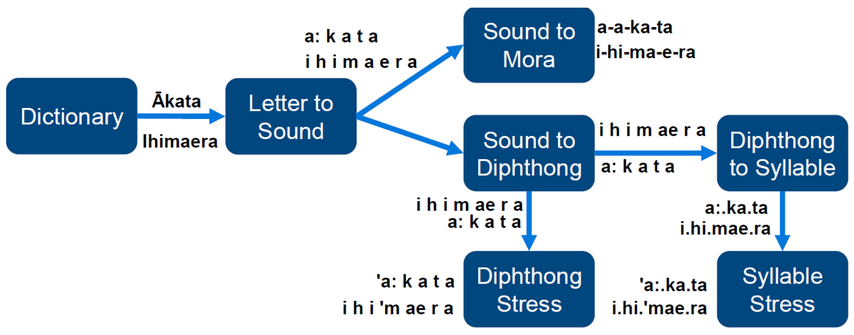


Figure 2: Application of phonetic rules for Words not in lexicon

Acknowledgements

We thank Te Hiku Media (Te Rarawa) for their generous support in funding this project. (https://tehiku.nz/). We also thank the New Zealand SFTI funded Te Tātari Raraunga project for its support.

Bibliography

[1] Schröder, M., Charfuelan, M., Pammi, S., and Steiner, I.. Open source voice creation toolkit for the MARY TTS platform. In Piero Cosi, Renato De Mori, Giuseppe Di Fabbrizio, and Roberto Pieraccini, editors, Interspeech, pages 3253--3256, Florence, Italy, August 2011

[2] 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.

[3] Biggs, B. Let's learn Māori: a guide to the study of the Māori language. Wellington: A.H. & A.W. Reed, 1969.