Te Reo
the Journal of the Linguistic
Society of New Zealand

Volume 66
Issue 2 (*Special Issue*)

Research Article

2024
Pages 105-131

January, 2024

Ngā āhuatanga o te /r/ o te reo Māori: preliminary investigations into the acoustics of Māori /r/

Isabella Shields
Waipapa Taumata Rau || University of Auckland

Catherine I. Watson
Waipapa Taumata Rau || University of Auckland

Peter J. Keegan
Waipapa Taumata Rau || University of Auckland

This paper is a peer-reviewed contribution from https://nzlingsoc.makeitso.nz/journal/current-issue

Guest Editors: Julie Barbour and Peter Keegan

Ngā āhuatanga o te /r/ o te reo Māori: preliminary investigations into the acoustics of Māori /r/

ISABELLA SHIELDS[1], CATHERINE I. WATSON[1]
AND PETER J. KEEGAN[2]
[1]*Te Kura Pūhanga Hiko, Rorohiko me te Pūmanawa || Department of Electrical, Computer and Software Engineering*
Waipapa Taumata Rau || University of Auckland

[2] *Te Puna Wānanga || School of Māori & Indigenous Education*
Waipapa Taumata Rau || University of Auckland

Abstract

In this article we present three preliminary investigations into the acoustics of the /r/ sound in te reo Māori, the Māori language. Investigation A considers the spectrographic characteristics and duration of /r/ as produced by male *kaumātua* (elders) of the MAONZE corpus. These speakers produce /r/ as a tap/flap in most contexts, but with variations in their production as evidenced by spectrographic inspection. An approximant /r/ also appeared in conditioned locations where it was predicted to do so. Investigation B considers the Ngā Mahi corpus, honing in on the potential influence of lexical stress on /r/ duration. Duration was found to increase when /r/ was in closer proximity to the stressed syllable. Investigation C hones in on the fourth formant (F4) in the Ngā Mahi speech. We identify notable F4 lowering timed to /r/ in a wide range of segmental contexts. These investigations provide a starting point for further analysis of Māori /r/.

Keywords

Te reo Māori, Māori language, rhotics, acoustic phonetics.

1 Introduction

Te reo Māori is the Indigenous language of Aotearoa New Zealand. Following the European colonisation of Aotearoa, Māori has been increasingly influenced by English, which became and remains the dominant spoken language. According to 2013 Census data, approximately 125,000 Māori indicated they could hold a conversation in Māori about everyday things (Stats NZ 2013). This proportion of speakers decreased almost 5% from that reported in 2006. Analysis of a survey of people of Māori ethnicity/descent following the 2018 Census estimated that around 70% of Māori believed there was at least some importance in using the language in daily life (Stats NZ 2018). Revitalisation efforts have improved the outlook for the Māori language, but there is still much work to be done, including in understanding the acoustics of Māori /r/.

Rhotic sounds, or *r-like* sounds, are notoriously heterogeneous, with the class encompassing various sounds with different articulatory and acoustic characteristics. A prominent cross-linguistic characteristic of rhotics is their predisposition to variation. In several languages there are various acceptable allophones of a single rhotic phoneme. In Cibaeño Spanish, the trill phoneme can appear as a glottal fricative, a tap, a trill, a tap combined with breathiness, or a trill combined with breathiness (Willis 2007). Similar situations are attested in various other languages; see, for example, Rafat (2010) for Persian, Son (2008) for Korean (in which a liquid phoneme varies between rhotic and lateral), and Cohen et al. (2019) for Modern Hebrew. Diachronic variation of rhotics is also common, as illustrated in Scottish English (Nance 2014), German dialects (Wiese 2003), and Brazilian Portuguese (Rennicke 2015).

While the Māori language has only one rhotic sound, we suspect it is not without its intricacies. A simple Google search querying "pronunciation of Māori r" returns varied and sometimes contradictory results. These search results are not linguistically thorough descriptions but they do serve to demonstrate the confusion that surrounds the pronunciation of the sound. The acoustics of Māori /r/ have not yet been sufficiently investigated; this paper outlines preliminary investigations which extend existing research (Shields 2021a), and have framed subsequent research since their completion (Shields 2021b; 2022). While we by no means attempt to provide a comprehensive picture of the acoustics of this sound in this article, we present findings from three preliminary investigations which provide a foundation for further investigation and better characterisation of Māori /r/. These preliminary investigations consider some acoustic and spectrographic characteristics of the sound as it appears in different word, vowel and lexical stress environments. In particular, we aim to address the following questions:

> What are the spectrographic characteristics of Māori /r/ produced by present-day male *kaumātua* (elders)? Are there differences in the realisation of /r/ across speakers?
> Does proximity to word stress influence the duration of Māori /r/?
> Is a lowered fourth formant (F4) a correlate of Māori /r/?

1.1 The phonology of Māori

The Māori vowel system has five monophthongs that have long and short qualities: /i e a o u/. These monophthongs can be combined to form diphthongs. Analysis undertaken as part of the Māori and New Zealand English (MAONZE) project has identified changes over time in vowel quality and quantity in Māori (Harlow et al. 2009: 135–140). Vowel quantity is phonemic, and the distinction between some long and short vowel pairs has diminished over time. There are ten consonants in Māori: /p t k m n ŋ f h r w/. Changes in the pronunciation of some of these consonants has also been attested in acoustic studies; voice onset times (VOT) of stops have increased along with increasing aspiration for certain sounds (Maclagan & King 2007; Maclagan et al. 2009).

Descriptions of stress in Māori include both lexical and phrasal stress. Here we limit our scope to primary lexical stress and its potential interactions with /r/. Morae[1] are the basis for the application of lexical and phrase stress as proposed by Biggs (summarised in Biggs 1998: 172–173). Syllables are also relevant in the placement of lexical stress. Syllables in Māori are always open, and onsets are optional, and if filled, contain only a single consonant. The syllable structure is described by (C)V(V)(V), where 'V' is a short monophthong. A syllable could have a diphthong composed of a long monophthong followed by a short monophthong as its nucleus. Bauer (1993: 546) outlines a hierarchy of syllables; syllables with a long vowel (or a diphthong with a long vowel) rank above syllables with a diphthong (composed of short vowels) which in turn rank above a syllable with a single short vowel. The lexical stress of a word should fall on the highest ranking syllable within four morae of the end of the word, assuming the word is monomorphemic. In the present study, we do not consider words consisting of more than four morae, so syllable rank is sufficient to determine word stress placement.

Correlates of stress in Māori have not been the focus of wide investigation. Perceived prominence has been found to align with expected phrase and word stress placement (Thompson et al. 2010; 2011a). Pitch peaks have been found to align more often than not with locations of phrase stress, although this is not always the case (Thompson et al. 2011b). Phonetic correlates of Māori stress are discussed by Bauer (1993: 545), who points to pitch fall and vowel duration as the most reliable markers. These correlates can be accompanied by an *emphatic onset* (of the mora/syllable) and increased loudness. For /r/, this is described as a longer contact and for the approximant /w/, a closer approximation of articulators. In this article we present an investigation considering the interaction between lexical stress and Māori /r/.

1.1.1 The /r/ sound in Māori

Generally speaking, linguistic descriptions of Māori /r/ are largely in agreement. The sound is usually designated as some variety of alveolar tap or flap. It is unclear in many descriptions whether taps and flaps are differentiated. The place of articulation of this sound is usually described as alveolar, and sometimes as post-alveolar (Bauer 1993: 523–524; Harlow 2007: 77). Bauer describes /r/ as a voiced lamino-alveolar or apico-alveolar tap, and indicates that articulation of the sound involves considerable contact between the sides of the tongue and the gums. Biggs (1961: 9) defines /r/ as an alveolar flap which is produced with the tip of the tongue "curved upwards" to contact the alveolar ridge.

Hohepa (1965: 10) states that the sound may also appear as an apical alveolar trill; however, Bauer (1993: 523–524) notes no evidence of trilling in her observations. Another potential variant is suggested by Harlow (2007: 77), who indicates an approximant /r/ may appear in fast speech, especially in situations where the sound appears repeatedly with intervening unstressed vowels (in the word *kōrero*, for example). Outside of these contexts, Harlow (2007: 77) affirms that an approximant /r/ in Māori is a direct result of interference from New Zealand English (NZE), a language in which the dominant realisation of /r/ is a post-alveolar approximant. There is also reference to historical lateral realisations of /r/. Bauer (1993: 524) observes central and lateral releases of /r/ in speakers from across the country, while Biggs (1961: 9) observes the phenomenon only in some Eastern dialect speakers. Conversely, Harlow (2007: 45) considers the lateral variant more likely to appear in South Island Māori dialects and less frequently in North Island dialects, except for Northland speech.

[1] Māori is described by Bauer (1981) as a mora-timed language, with a mora defined as any short vowel preceded by an optional consonant (that is, $\mu = (C)V$, where C is a consonant and V is a short vowel).

The evidence for this historical lateral variant is found in historical texts and place names (see, for example, the southern New Zealand town of *Waihola*).

In addition to the possible increase in /r/ duration due to stress, Bauer (1993: 524) points towards changes in /r/ duration which are dependent on its placement in a word: intervocalic /r/ is very brief, while word-initial /r/ is "sometimes a little longer". These variations in /r/ duration are not quantified.

1.1.2 Previous studies of Māori /r/

Previous studies investigating Māori /r/ have largely been perceptual. Two studies have investigated /r/ in the MAONZE corpus; the first investigated the speech of a single speaker, the second several groups of speakers. The first investigation into /r/ considered the speech of a historical (Mobile Unit) speaker from the MAONZE corpus (Maclagan & King 2004). This perceptual study considered all /r/ tokens of this speaker in the word *Māori*, both in Māori and in English. The speaker generally used the canonical /r/ for each language (that is, a tap for Māori and an approximant for English). When Māori words were used in English speech, there were observations of approximant /r/ where a tap would be expected. No lateral observations were found in the speaker's Māori speech. For Māori words that appeared in predominantly English speech, consideration was given to segmental context and stress. It was concluded that these factors did not seem to impact whether or not the /r/ appeared as flapped, and that instead it was relatively random. The word *Māori* pronounced in predominantly English speech was found to have a greater proportion of non-flapped /r/ tokens than other words. This was attributed to its relative frequency in NZE speech compared to other words. A further study focused specifically on the pronunciation of the word *Māori* expanded into other parts of the MAONZE corpus (Maclagan et al. 2019). This work, which also considered changes in vowel pronunciation, involved a perceptual-auditory assessment of the /r/ in the word *Māori*. This occurred both for Māori and English speech. The expectation was that there would be more flapping of the /r/ in Māori speech, which was the case for most speaker groups, save the women who often also flapped in predominantly English-language speech. Young first-language male speakers appeared to differ from other male groups in the corpus, producing more flapped /r/ tokens than other /r/ types.

1.2 Rhotic sounds

Rhotics are a subset of the liquid class of sounds. The rhotic class encompasses a wide array of sounds: trills, taps/flaps, approximants, and fricatives. Class-membership is contested and complicated, and is outside of the scope of this discussion. The acoustic characteristics of rhotic sounds are varied given the range of different sounds in the class. As mentioned above, there are references to Māori /r/ appearing as a tap/flap, a trill, and an approximant. These sounds involve very different articulatory movements and can have different acoustic characteristics.

A tap/flap involves muscle movement where an active articulator (usually the tip of the tongue) contacts a passive articulator (usually somewhere on the roof of the mouth). This contact is short (usually reported as being between 20 and 30 ms) (Zue & Laferriere 1979; Recasens 1991; Willis 2007; Warner et al. 2009; Bradley & Willis 2012). It is reported by some that no release burst should occur upon release of contact of the articulators as the duration of contact is short enough so as not to permit pressure build-up behind the constriction (Zue & Laferriere 1979; Derrick & Schultz 2013). Contrary to this, several studies have identified evidence of release bursts while inspecting audio signals and spectrograms. In a study of the

American English[2] (AE) flap variant of /t/, bursts were found to occur in 30% of observations (Warner et al. 2009). The same study found formant energy continued (either weakly or strongly) through the articulation of a majority of productions, indicating a constriction insufficient to fully block airflow during articulation. Other formant behaviours provide insight into the articulatory approaches to tap/flap sounds: asymmetrical formant trajectories into and out of the sound in AE indicate an anticipatory articulatory movement, while symmetrical formant trajectories about the consonant in Spanish are interpreted as less anticipatory movement with articulation limited to the tongue tip (Ladefoged & Maddieson 1996: 232).

There are two (very) generalised approaches to producing an approximant /r/: retroflexed and bunched. A retroflex sound is produced when the tongue is curled backwards. The articulation of a retroflex approximant /r/ involves the raising of the tongue tip and dorsum, creating a cavity below the tongue. The articulation of the bunched /r/ involves the lowering of the tongue tip and raising of the dorsum. This movement is slower relative to that employed in tap/flap production (see Gurugubelli et al. 2020: 15, for example, where the reported mean /r/ duration was 91 ms when in an intervocalic position, and 66 ms when word-initial). The NZE approximant /r/ is described as alveolar, and an ultrasound study investigating /r/ articulation of NZE speakers found that both of these two articulatory approaches were present (Heyne et al. 2018). Acoustic correlates of approximant /r/, specifically AE variants, often centre around the shape of and lowering of the third formant (F3). F3 trajectory shape is said to resemble an inverted parabola (Boyce & Espy-Wilson 1997: 3741). In some instances, the F3 minimum has been found to occur outside of the /r/ segment, a phenomenon attributed to anticipatory movement of the tongue into the target articulatory position (Espy-Wilson 1992: 756). Formants higher than F3, that is, the fourth and fifth formants (F4 and F5), were found to differ depending on whether or not there was a bunched or retroflex articulatory approach (Zhou et al. 2008).

The articulatory process of trilling is more complex than taps/flaps or approximants. Generally speaking, a trill sound involves the vibration of one articulator against another. This vibration is not solely controlled by muscle contractions but rather is subject to aerodynamic conditions. Conditions for the initiation of apical trilling include the placement of the tongue with acceptable shape and elasticity, and an adequate pressure differential across the constriction point (Solé 2002). The tongue-tip is usually relaxed. After the trilling has begun, the vibration is self-sustaining as long as these conditions are still met. A trill cycle is made up of one closed phase and one open phase. The closed phases of trilling correspond to periods of reduced energy and amplitude in the speech waveform. This results in spectrograms with alternating segments of low energy (closed phases) and high energy (open phases). Sometimes a short burst will occur at the point where the constriction is forced open. In spectrograms, open phase sections usually appear similar to a vowel with regions of higher energy corresponding to vocal tract resonances in that position.

2 Available corpora

Two sets of recordings were used in the analysis presented here. They differ significantly in their nature; one is a sociolinguistic corpus composed of several speakers, while the other consists of a single speaker's read speech.

[2] We refer to several studies which analyse speakers of "American English." In the absence of specifics on the dialect of United States English spoken, we follow the authors' designation and refer to these as American English (AE).

2.1 The MAONZE corpus

The MAONZE corpus is a collection of Māori and English speech recordings (King et al. 2010; 2011a). It was developed with the goal of investigating sound change in Māori, both over time and under the influence of English. Speech of three separate generations are captured in the MAONZE corpus; historical elders (aged 55–77 at time of recording), present-day elders (aged 63–87 at time of recording), and present-day younger speakers (aged 17–35 at time of recording). Across these three generations, the birth dates of the recorded speakers span from 1871 to 1992. The recording equipment and recording environments of the different speaker groups are not consistent. The historical recordings were sourced from the Radio New Zealand Sound Archive, Sound Archives/Ngā Taonga Kōrero (SANTK), and the Television New Zealand Archive. The more contemporary recordings – the present-day elders and younger speakers – consist of the speaker in conversation with an interviewer. These interviews took place in various locations, including in the interviewee's home or workplace, with the speaker sitting. Audio recordings were made using Sony TCD-D8 digital audio tape or Marantz PMD recorders, and Sony ECM-T145 lapel microphones. The interviews focused on various topics as the interviewers aimed to guide the conversation towards topics of interest to the interviewee. Transcription of the MAONZE corpus speech was made using Transcriber[3] and Praat (Boersma & Weenik 2010). The overall corpus consists of 109 hours of speech, 57 of which are Māori (King et al. 2011b). Only a subset of the MAONZE corpus was considered in the study presented here: present-day male kaumātua. Overall, this speaker group comprises 10 speakers.

2.2 The Ngā Mahi recordings

The Ngā Mahi recordings consist of a collection of read-speech recordings of a single middle-aged male speaker of Māori. They were developed as part of the creation of a Māori text-to-speech synthesiser (Shields et al. 2019, 2020; James et al. 2020). The speaker recorded is of Waikato-Maniapoto and Ngāti Porou descent and was in his early 50s at the time of recording. He is a fluent, second-language speaker of Māori who learned the language post-adolescence.

The content of these recordings was taken from *Ngā Mahi a Ngā Tūpuna* (Grey & Williams 1971), a text collection of Māori myths and legends, and was recorded in segments. A segment was usually a sentence, although some sentences were segmented when too long to be comfortably spoken aloud. This collection of recordings includes over two hours of speech. The recordings were made in a WhisperRoom[4] Sound Isolation chamber, a recording space designed to attenuate undesirable noise. The speaker, while seated in this room, was recorded using a Rode Lavalier Lapel microphone, placed approximately 15 cm from his mouth. A Roland OCTA-CAPTURE was used for analogue-to-digital conversion, with a sampling rate of 44.1 kHz. Signal capture was achieved using Audacity (Audacity Team 2021). Sentences were shown one at a time to the speaker on a computer monitor inside the recording space. Time was allotted to practise producing each sentence before recording. While each recording was made, the first author listened to the produced speech in order to identify if there were any errors or other acoustic interference (e.g. movement noise). Sentences were re-recorded if the first author noticed such issues.

[3] http://trans.sourceforge.net/en/presentation.php
[4] https://whisperroom.com

3 Data pre-processing

The pre-processing for the MAONZE corpus and Ngā Mahi recordings was the same; orthographic and phonetic transcription data were aligned with audio, and formant values were estimated. Both alignment and formant estimation were corrected as part of the work presented here. Analysis was undertaken using R (version 3.6.3) (R Core Team 2020) and RStudio (2021).

Automated alignment generation was undertaken for both corpora using Montreal Forced Aligner (MFA) prior to these studies (McAuliffe et al. 2017). Alignment using MFA requires audio recordings in WAV format accompanied by text transcriptions of the speech, as well as a dictionary with all words present along with their broad phonetic transcription. The dictionary used for these automated alignment tasks is an existing dictionary maintained by the second and third authors. This automated alignment was completed for parts of the MAONZE corpus and for the Ngā Mahi recordings (Shields et al. 2019; Stoakes et al. 2019). The output of MFA is a TextGrid associated with each input WAV audio recording. The TextGrid and WAV file pairs were used to create an emuDB in emuR using the R package *emuR* (version 2.3.0) (Winkelmann et al. 2021). This package allows for the visualisation of audio and formant estimation along with the associated spectrograms, as well as the correction of boundary placement and acoustic features. All hand correction of MFA boundary estimations was undertaken by the first author using emuR's EMU-webApp. Phone start and end boundaries for /r/ were placed based on reduced amplitude of the waveform envelope, reduction in root mean square waveform amplitude, and reduction in formant energies in the spectrogram and were placed at the nearest zero-crossing.

Formant trajectories were estimated for both corpora's recordings using the *forest()* function with default settings in the R package *wrassp* (version 1.0.1) (Bombien et al. 2021). These were subsequently hand corrected by the first author using emuR's EMU-webApp. Corrections were made only for tokens used in the analyses presented here. In most cases, the errors in formant estimation that required large adjustment occurred in the higher frequency ranges, with the third and fourth formants (F3 and F4) especially impacted. Occasionally, the first and second formant (F1 and F2) trajectories were clearly incorrect. For example, F1 estimations would track close to zero, and F2 trajectories would become confused with F1 or F3. All formant trajectories were plotted using the *ggplot2* package in R (Wickham 2016).

4 Investigation A: Realisations of /r/ as produced by MAONZE kaumātua in three high-frequency content words

The first investigation undertaken considered the speech of the MAONZE present-day male kaumātua (elders). Our focus was on the spectrographic realisation of Māori /r/, with the aim of exploring the acoustic characteristics of the sound as it appeared in three high-frequency words.

4.1 Method

Categories of /r/ were developed based on the range of spectrographic realisations seen in the MAONZE kaumātua speaker group. Each /r/ token in the analysed MAONZE kaumātua data and Ngā Mahi speaker data was categorised according to these criteria. The development of these categories began with a first-pass visual inspection of /r/ in the database and an aggregation of salient features. These features included the presence of a short burst, presence of formant energy in the segment, degree of formant strength (judged visually), and number of cycles (in the case where a trill had been produced).[5] The categories outlined below cover the range of /r/ realisations present in the kaumātua recordings, and cover three different classes of /r/ sound: taps/flaps (Categories 1–3), approximants (Category 4), and trills (Category 5). The characteristics of each of these categories are outlined in Table 1. Phone start and end boundaries and formant estimations were corrected as described in Section 3.

Table 1. Summary of /r/ categories based on spectrographic and acoustic realisation

/r/ spectrographic category	**Description**
(1) clear tap/flap [ɾ]	No formant presence or little formant presence in the closure period of articulation, followed by a clear burst. See Figure 1(a).
(2) tap/flap with small burst [ɾ]	Some formant presence accompanied by a short burst. This burst may be less clear than those seen in a *clear tap/flap*, having only higher frequency components.
(3) tap/flap with no burst [ɾ]	No burst of any kind but does exhibit a reduced intensity of formants. This intensity reduction can vary in degree. See Figure 1(b).
(4) approximant [ɹ]	May include some lowering of the third formant (F3). Usually some reduction in formant intensities, and a relatively longer duration than Categories 1–3.
(5) trill [r]	Alternating sections of low and high intensity of formants that correspond to open and closed phases of the trill articulation. Release bursts may be present at the beginning of open phases.

[5] An alternative approach to characterising speech data using spectrographic information is the landmark feature-cue-based approach (Stevens 2002; Yun et al. 2020). This approach involves the application of a set of binary features to represent a segment. In the case of a flap occurring as an allophone of /t/ in AE, the binary features include the beginning of the closure, the beginning of the release (i.e. the burst). If these do not occur, they are marked as not present. Such an approach is not easily applicable here given the occurrence of different sounds (such as an approximant and a tap/flap). The binary features applicable to an approximant do not directly align with those applicable to a tap/flap.

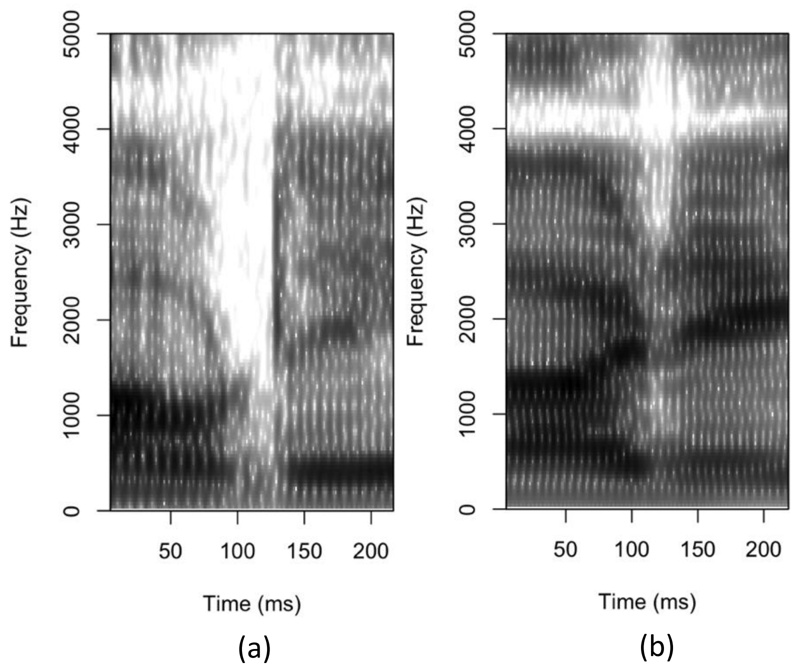

Figure 1. Examples of (a) a clear tap/flap (category 1) in the sequence /ori/ and (b) a tap/flap with no burst (category 3) in the sequence /uri/, both taken from the Ngā Mahi speaker recordings

It is possible for trills with only one trill cycle to occur (that is, one contact is made between the tongue and the passive articulator). This complicates the separation of taps/flaps and trills. The underlying assumption made here is that a token is a tap/flap unless there is clear evidence it is a trill (that is, there is clear evidence of multiple contacts of the tongue in the spectrogram/audio signal).

4.1.1 Speaker group and speaker selection

The speaker group selected for this study was the MAONZE present-day male kaumātua. These speakers were born between 1925 and 1938 and were recorded between 2001 and 2009 (King et al. 2010). These speakers were chosen for this exploratory study as they were likely to have Māori pronunciation less influenced by English than that of the younger speakers. While the historical speaker group in the MAONZE corpus is known to have Māori pronunciation less influenced by English, their recordings are bandlimited to 5 kHz (King et al. 2010), making analysis of higher formants difficult. Visual inspection of spectrograms of the historical speaker recordings also indicated that F3 was not reliably visible. While the lowering of F3 has been disproven as an acoustic feature of all rhotic sounds (see Chabot 2019 for a summary of this and wider discussion of the rhotic class), the absence of F3 in a preliminary acoustic investigation into Māori /r/ was considered disadvantageous. While present-day male kaumātua recordings are of higher quality than the historical speaker group, formants higher than F3 are not reliably visible throughout the production of the target word. The present investigation considers F1–F3, which were all visible more often than not for the tokens considered.

The subset of the speaker group selected for analysis was included based on the frequency of the target words in their recordings. Of the 10 speakers in the present-day male kaumātua group, six are analysed here.

4.1.2 Target word selection

We considered three words produced by the kaumātua speaker group: *haere* (‘to go; journey or trip’, a verb or noun), *reo* (‘language, dialect, or speech’, a noun), and *kōrero* (‘to speak; speech or story’, a verb or noun). These words were selected on the basis of where /r/ appeared in the word and the relative frequency of the word in the corpus. The location of /r/ in these words is word-medial, word-initial, and word-medial in a repeated sequence, respectively. Often *reo* is directly preceded by a word, meaning the word-initial /r/ is usually intervocalic. We consider it here as a separate category to /r/ in *haere* and *kōrero* (which are also intervocalic) as it is not yet clear how word position and word stress interact with /r/ realisation in Māori. Instances of these words (along with their associated recording start and end times) were identified in the MAONZE corpus using the emuR database querying functionality and extracted for analysis.

4.2 Results

Exclusion of tokens resulted in a total of 210 *haere* tokens, 171 *reo* tokens, and 201 *kōrero* tokens being available for analysis. The majority of exclusions were a result of inadequate formant data, predominantly for F3 in the surrounding vowels. This was not unexpected given some of the issues with the acoustic quality of the recordings. There were also exclusions due to environmental noise.

Table 2 summarises the number of tokens for each word and speaker analysed.

Table 2. Tokens analysed for each speaker in the kaumātua speaker group

Speaker	# *haere*	# *reo*	# *kōrero*
K0001	31	54	38
K0002	40	19	30
K0003	35	6	30
K0005	46	49	41
K0006	20	30	21
K0009	38	13	41
Total	210	171	201

4.2.1 Rates of occurrence of different /r/ categories

The most common /r/ category encountered in these tokens was some variant of a tap/flap ([ɾ], Categories 1 through 3), which made up 79.8% of all /r/ observations. Among the tap/flap /r/ variants, the most common was a tap/flap with no burst ([ɾ], Category 3; 38.4% of observations), followed by a tap/flap with small burst and clear tap/flap ([ɾ], Category 2 and 1, respectively; 21.7% and 19.7% of observations, respectively). It follows that just over half of the tap/flap realisations had some sort of release burst. This is a rate relatively higher than that observed in the AE flap allophone (30% with a burst) (Warner et al. 2009: 3318), and in the Korean flap allophone (42% with a burst) (Son 2008: 694). Approximant /r/ ([ɹ], Category 4) was the next most common after tap/flap variants, making up 20.1% of all observations. A single trilled /r/ ([r], Category 5) was observed.

On a word by word basis, tap/flap variants were by far the most common for instances of /r/ in *haere* and *reo*, appearing in over 98% of observations in both cases. For these two words, all other observations were approximants. The two /r/s in *kōrero* are considered separately. In the case of the first /r/, tap/flap variants ([ɾ], Categories 1 through 3), were also the most common, appearing in 95% of observations. Save the single trilled /r/ ([r], Category 5), all other observations in this position were approximants ([ɹ], Category 4). The second /r/ of *kōrero* appeared as an approximant in 71.3% of observations, and as a tap/flap variant otherwise.

The appearance of approximant /r/ in the second position in *kōrero* was expected given Harlow's (2007: 77) description of Māori /r/ in repeated sequences with intervening unstressed vowels. Also in agreement with Harlow (2007: 77), we do not find many examples of approximant /r/ outside of this environment. We suspect the prevalence of approximant /r/ in this repeated context may be a result of the articulatory difficulty in producing two flaps in quick succession. Rácz et al. (2016) argue that obligatory contour principle effects can be seen in Māori; that is, there is a tendency for sequences of consonants that share a place of articulation to be avoided. They note that this is particularly so when the consonants are identical (in our case, with two flaps and intervening vowels). It is possible then that we are seeing some sort of analogue of this in phonetic realisation; rather than producing those two flaps, the speakers are producing a flap followed by an approximant. However, we are of course observing a change in manner of articulation rather than place of articulation.

Evidence was found of inter-speaker variation in /r/ category observations. For example, in observations of /r/ in the second position in *kōrero*, one speaker (K0001) almost exclusively produced approximant /r/ (Category 4), while speaker K0003 produced a similar number of approximants and tap/flap variants (Categories 1–3).

4.2.2 Duration of /r/

The mean durations of the Category 1 ([ɾ], clear tap/flap) and Category 2 ([ɾ], tap/flap with a small burst) /r/ tokens were 27.6 ms (s.d. 8.3 ms) and 24.6 ms (s.d. 6.3 ms), respectively. The mean duration for Category 3 ([ɾ], tap/flap with no burst) was marginally higher at 29.7 ms (s.d. 9.2 ms). The measured durations for these three tap/flap variants are within the expected range observed for similar sounds in other languages (we refer the reader back to our description of taps/flaps in Section 1.2). Finally, the mean duration of Category 4 /r/ ([ɹ], approximants) was 64.7 ms (s.d. 15.9 ms).

Figure 2 shows the durations of /r/ tokens by category and word context. For Category 1 and 2 /r/ tokens, we see very little difference in duration across word contexts. Conversely, there appears to be some variation in duration of Category 3 and 4 /r/ tokens across word contexts. This is evident for Category 3 tokens in the final syllable of *kōrero*, where /r/ duration is increased, particularly compared to /r/ in *reo* or the first occurrence in *kōrero*. We believe it is possible that we are observing something of a continuum from the 'clear' tap to the approximant, which we find, above, in Section 4.2.1, is common in this context.

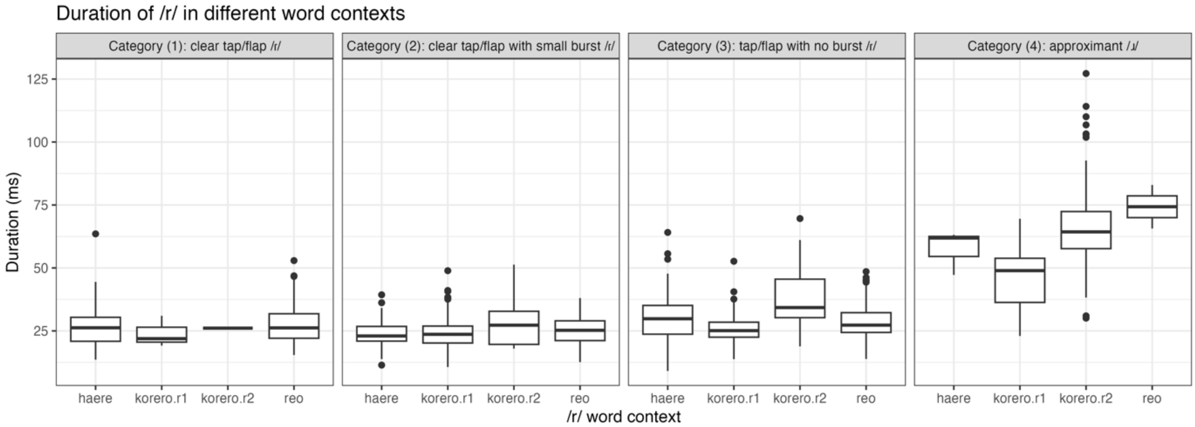

Figure 2. Duration (ms) of different /r/ categories: (1) clear tap/flap, (2) tap/flap with small burst, (3) tap/flap with no burst, and (4) approximants (as there were very few examples of (5) (trilled /r/), these have been excluded)

4.2.3 Formant behaviours

We initially considered the formant trajectories of the different /r/ categories in isolation. It quickly became apparent that it was necessary to consider the formant behaviours in a wider context (that is, including the vowels surrounding the /r/). In *haere* (/ˈhae.re/), the formant trajectories were largely very flat, regardless of /r/ category. In terms of articulation, the tongue position required to produce /e/ is not far from that of an alveolar tap/flap, and the tongue is already moving upward in the articulation of the diphthong /ae/, which may explain the lack of notable formant movement timed to the /r/. The formants of /r/ in *reo* (/ˈre.o/) exhibited more movement, with approximant /r/ (Category 4) tokens exhibiting lower F3 than other /r/ types. This is also evident for *kōrero* (shown in Figure 3), where the F3 trajectory of Category 4 /r/ tokens resembles an inverted parabola (as is also reported for AE approximants (Boyce & Espy-Wilson 1997)). Figure 3(a) and Figure 3(b) show LOESS (local polynomial regression) fits of smoothed formant trajectories (F1–F3) of speaker K0005 and K0009's productions of *kōrero* (n = 41 for each speaker). The initial /k/ of /ˈkoː.re.ro/ was excluded. A 95% confidence interval for each fit is shown in grey. Fits and confidence intervals were implemented in *geom_smooth()* function in the *ggplot2* package (Wickham 2016). In the first position in *kōrero* (/ˈkoː.**r**e.ro/) all /r/ variants had notably lower F2 and F3 values than in *haere* or *reo*, with F2 and F3 bunched closer together. In the second position in *kōrero* (/ˈkoː.re.**r**o/) this was generally not the case.

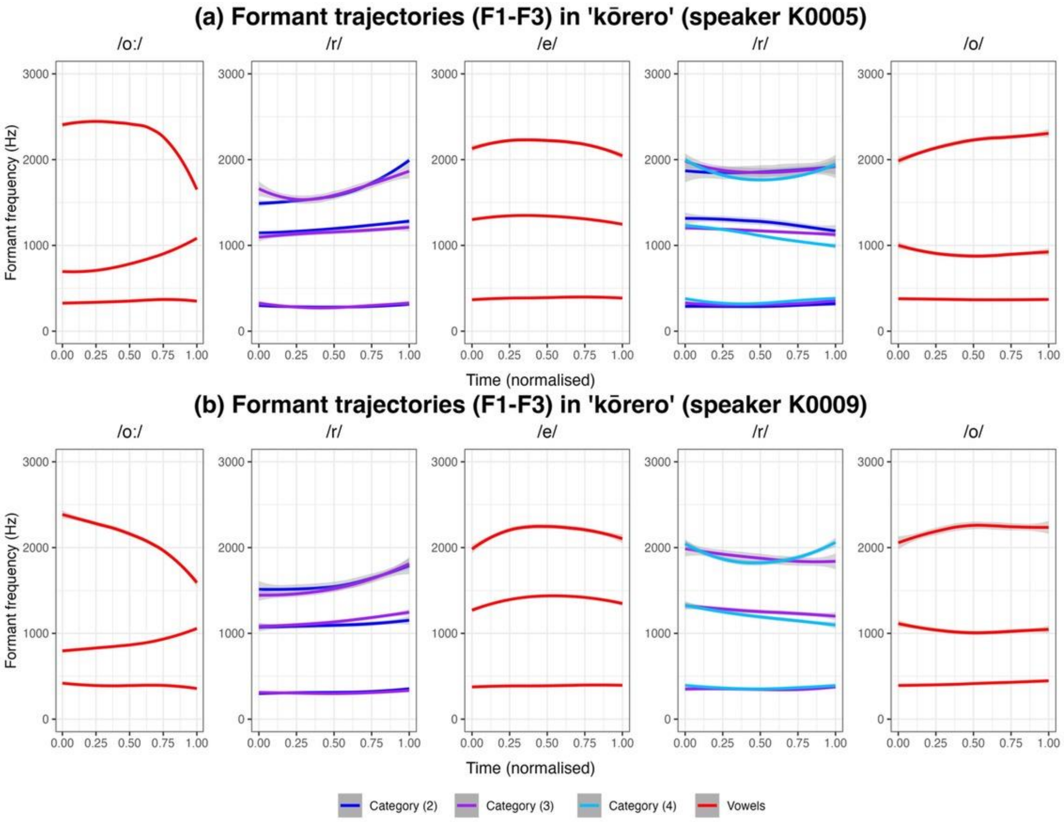

Figure 3. LOESS-smoothed formant trajectories (F1–F3) for (a) speaker K0005's kōrero productions and (b) speaker K0009's kōrero productions (Category 2 shows [ɾ] with a small burst, Category 3 shows [ɾ] with no burst, Category 4 shows [ɹ]. Grey sections about the coloured smooth lines indicate 95% confidence intervals. Note that in cases where formant energy was not present throughout the /r/ production, the smooths are fitting across empty space.)

We note raising of F2 and lowering of F3 in /oː/ preceding /r/ occurring as early as or before halfway through the vowel. Notable reduction of /e/ (between the two /r/s) also occurs. We did not identify this taking place in for /r/ in *haere* or *reo*, indicating the repeated-/r/ environment in *kōrero* may be the cause of the reduction of /e/. For one speaker analysed in Maclagan et al. (2004), the mean frequency of F3 was 2985 Hz, ~700 Hz higher than the highest mean value reached for both speakers shown in Figure 3.

4.3 Summary

From this investigation, the most common spectrographic variant was some sort of flap. Of these, around half included a release burst. Approximant /r/ was indeed largely restricted to the second position in *kōrero*, where it appears as part of a sequence of /r/s with intervening unstressed vowels. An approximant only occurred a few times outside of this position. We consider this rate of occurrence low enough to be explained as mispronunciations and not as evidence of interference from NZE. As previously indicated, consideration of the formants of /r/ in their wider context was necessary, and allowed an insight into the influence of /r/ on its surrounding vowels. We note examples of vowel reduction apparently caused by /r/, as well as the early impact that /r/ has on the vowel formant trajectories. Changes can be seen as early as halfway through the production of the vowel preceding /r/. We also find evidence of both intra-speaker and inter-speaker variation. As only three target words were considered in this analysis, we emphasise that conclusions from these findings are limited.

5 Investigation B: Lexical stress interaction with /r/ in the Ngā Mahi recordings

The investigation into the MAONZE kaumātua revealed a preference shown by the speakers for using different types of /r/ in different word positions. There are several different factors that could be influencing this variation. We investigate one such factor, lexical stress, in the Ngā Mahi speech and seek to verify whether or not a feature of Bauer's (1993: 545) *emphatic onset* appears for this speaker. We investigate the measure of /r/ duration as influenced by lexical stress.

5.1 Method

To establish if lexical stress impacts the realisation of /r/, different stress environments were selected. These are: (1) ˈrV(V) words, where /r/ is in the onset position of the stressed syllable, (2) ˈrVCV words, where /r/ is in the onset position of the stressed syllable, (3) ˈ(C)VrV words, where /r/ is in the onset of an unstressed syllable directly following the stressed syllable, and (4) ˈ(C)V(V)C(V)rV and ˈ(C)V(V)C(V)rVCV words, where /r/ is in the onset of an unstressed syllable placed further from the stressed syllable.

After these stress environments were designated, we selected words for analysis that fit these criteria and appeared more than five times in the corpus. The words selected for analysis are summarised in Table 3. The processes for formant correction and phone start/end boundary correction were the same as for Investigation A, and are outlined in Section 3. It must be noted that no categories consider the situation where /r/ is word-medial and in the stressed syllable onset (as in *porou*), or indeed, where /r/ is word-initial and not in the stressed syllable onset (as in *rerekē*). This means, at least in the present study, we cannot ignore the possibility that word position is playing a role. Phone duration was measured by computing the difference between the start and end boundaries of each /r/ observation.

Table 3. Target words analysed in stress investigation

Stress category	Target words analysed
(1) /r/ in stressed S, 1S	*rā*
(2) /r/ in stressed S, 2S	*rangi, rapu, rawa, ringa, rite, rongo, roto, runga, rupe*
(3) /r/ not in stressed S, 2S	*ara, whare, kāore, kore, māori, muri, ngaro, tārai*
(4) /r/ not in stressed S, 3+S	*ehara, engari, tāwhiri, tahuri, tamariki, kāhore*

5.2 Results

The token counts for each stress category are shown in Table 4. The token counts by word and stress category are shown in Table 5. In total, 731 /r/ tokens were analysed across the stress categories and target words.

The durations of each /r/ analysed are shown in a box plot in Figure 4. There is a tendency for /r/ duration to be longer in situations where it is in the onset of the stressed syllable. This is coherent with the observation made by Bauer. Of the four stress categories considered, /r/ in the stressed onset of a two syllable word (Stress 2) exhibited the longest durations for /r/, followed by Stress 1. Stress 3 and Stress 4 had similar durations. A number of the observations in Stress 2 were longer in duration than expected. These are almost all approximant realisations of /r/. This context does not fit that outlined by Harlow (2007: 77): that is, approximant /r/ should only appear in fast speech, or where the /r/ appears repeatedly with intervening unstressed vowels. We considered whether there was a difference in articulatory difficulty in producing an approximant versus a tap/flap following a period of non-phonation, as a small number of the ˈrVCV words appeared in utterance-initial position. However, some other factor must be considered to explain the presence of approximant /r/ where the ˈrVCV word was not utterance-initial. Given that the Ngā Mahi speaker is a fluent but post-adolescent speaker of Māori, it is plausible that the NZE approximant /r/ is indeed influencing his production of Māori /r/, at least in this stress environment.

Table 4. Summary of token counts for stress investigation

Stress category	**Count**
(1) /r/ in stressed S, 1S	117
(2) /r/ in stressed S, 2S	342
(3) /r/ not in stressed S, 2S	170
(4) /r/ not in stressed S, 3+S	102
Total	731

Table 5. Details of token counts analysed in stress investigation

Stress category	**Word**	**Count**	**Stress category**	**Word**	**Count**
(1) /r/ in stressed S, 1S	*rā*	117	(3) /r/ not in stressed S, 2S	*kāore*	14
(2) /r/ in stressed S, 2S	*rātou*	75	cont.	*kore*	24
	rangi	42		*māori*	8
	rapu	17		*muri*	31
	rawa	61		*ngaro*	31
	ringa	5		*tārai*	16
	rite	6	(4) /r/ not in stressed S, 3+S	*engari*	14
	rongo	32		*ehara*	33
	roto	42		*kāhore*	12
	runga	41		*tāwhiri*	18
	rupe	21		*tahuri*	10
(3) /r/ not in stressed S, 2S	*ara*	14		*tamariki*	15
	whare	32			

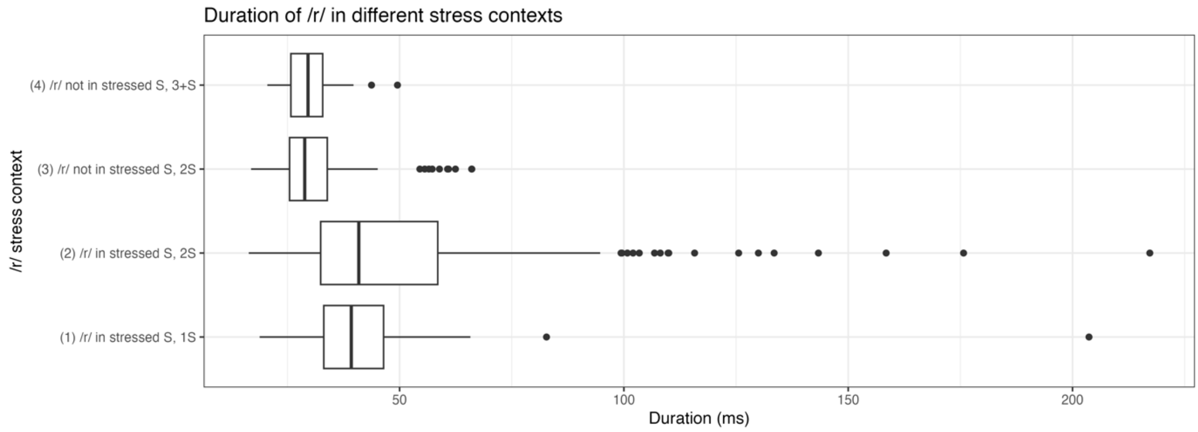

Figure 4. Duration of /r/ in different word stress environments, as produced by the Ngā Mahi speaker
(Stress categories are as follows: (1) /r/ in stressed syllable in monosyllabic word, (2) /r/ in stressed syllable in disyllabic words, (3) /r/ in unstressed syllable in disyllabic word, (4) /r/ in unstressed syllable in multisyllabic words.)

5.3 Summary

In this investigation, we consider one possible acoustic correlate of stress for /r/ as suggested in Bauer (1993: 545): increased duration. We observe some interaction between lexical stress and /r/ duration for the Ngā Mahi speaker, whose /r/ productions increase in duration with greater proximity to the stressed syllable. This is coherent with Bauer's description of duration increase when /r/ is stressed. In our data, we observe a difference in mean /r/ duration between stressed and unstressed /r/. We note a much larger range of durations for words where /r/ is in the stressed onset, and identify that a number of these are approximants. We note again, however, that the categories used in this analysis conflate word position and stress. This means the observed increased duration may be a result of /r/ occurring word-initially rather than as a direct result of word stress placement. This is something that should be teased out in a further study. We also reiterate that we consider only one speaker's /r/ productions and it is entirely possible that the observed feature behaviours are idiolectal.

6 Investigation C: /r/ fourth formant (F4) lowering in the Ngā Mahi corpus speech

While completing the boundary and formant estimation correction for the Ngā Mahi speaker's recordings, we identified what appeared to be a notable drop in (F4) that was timed to /r/. F4 lowering was observed and analysed in the AE flap allophone (Warner & Tucker 2017; Faytak et al. 2019), but has not been reported or investigated in relation to Māori /r/. It is not yet clear why lowered F4 occurs in AE flaps. Retroflexion has been connected with F4 lowering in stops (Stevens & Blumstein 1975) and the retroflex gesture in the AE retroflex approximant variant (Zhou et al. 2007). We have found no descriptions of retroflexion in descriptions of Māori phonology.

6.1 Method

To confirm if F4 lowering appears with /r/ in all environments in the Ngā Mahi data, analysis of a range of VrV sequences was performed. The VrV sequences selected cover a range of different vowel qualities and combinations thereof. While other sequences were included in a broader investigation, we limit discussion here to sequences where V_1 and V_2 are the same, providing as this provides an environment where the surrounding sounds are acoustically similar. The VrV sequences analysed are /iri/, /ere/, /ara/, /oro/, and /uru/. The initial goal was to amass 30–40 tokens of each sequence where possible. For each VrV sequence, the corpus database was queried using emuR for instances of the short vowel followed by /r/ followed by the same short vowel. No exclusions were made for examples spanning word boundaries, except for cases where a pause preceded /r/. Exclusions were also made where the speaker obviously mispronounced one of the vowels surrounding /r/. Finally, while results of Investigation B point towards a stress/word position interaction with /r/, we do not select tokens based on stress context here.

The formant tracks are plotted from the midpoint of the preceding vowel (V_1) to the midpoint of the following vowel (V_2). This was informed by the results of the MAONZE kaumātua investigation covered in Section 4, where formant movement related to /r/ began as early as halfway through the preceding vowel. Consideration of the formant behaviours in this frame allows us to investigate the influence of /r/ on its immediate segmental neighbours.

In an investigation of F4 lowering in AE flaps, the change in F4 was quantified separately for the transition from the preceding vowel into the flap (V to C) and from the flap into the following vowel (C to V) (Warner & Tucker 2017). The values are taken from the maximum of the vowel's F4 estimation, and the minimum of the tap F4 estimation (where it was visible). For comparison with this study, we also compute the percentage of tokens with F4 lowering exceeding 1 kHz. We acknowledge that AE is a language that is unrelated to Māori. As such, the relevance of a direct comparison between the AE flap and the Māori tap/flap is limited. However, in the absence of a similar study in a more closely related language (say, a Polynesian language, or even NZE), we compare the behaviours of the two *similar* sounds in Māori and AE.

6.2 Results

The token counts for each VrV sequence are given in Table 6, along with counts of /r/ type as discussed in Investigation A (see Table 1 for a summary of /r/ categories). The relative frequency of the various VrV sequences differs significantly. The first sequence to be analysed was /ara/, and many tokens (n = 203) were corrected (that is, formants and start/end boundaries were corrected) before other sequences were considered. As in the results of Investigation A, the most commonly observed /r/ was some sort of tap/flap (Categories 1–3), with 91.4% of observations falling into these categories. Approximants, (Category 4) make up 8.6% of observations, although this may well be an overrepresentation given the relatively high proportion of /ara/ tokens.

Table 6. Summary of tokens analysed in present study (asterisk indicates that not all available tokens were included because there were more than 30 possible tokens)

Sequence	(1) Clear tap/flap	(2) Tap/flap with small burst	(3) Tap/flap with no burst	(4) Approximant	Total tokens
iri	6	7	40	0	53
ere	9	10	20	1	*40
ara	4	9	169	21	*203
oro	1	3	26	10	40
uru	4	2	29	0	35
	(6.5%) 24	(8.4%) 31	(76.5%) 284	(8.6%) 32	371

The averaged formant tracks for sequences where V_1 and V_2 are the same are shown
Figure **5**. Given the relatively small number of Category 1, 2, and 4 observations, we only plot F1 through F4 for Category 3 /r/ (tap/flap with no burst). Time-normalisation of formant estimations was undertaken using the *normalize_length()* function in the emuR package, interpolating formant estimations to 21 points for each phone (Winkelmann et al. 2021). These sequences were then concatenated. The normalised time period between 0 and 0.25 corresponds to the second half of the vowel preceding /r/ in the sequence. The normalised period between 0.75 and 1 corresponds to the first half of the vowel following /r/ in the sequence. The period between these two (0.25–0.75) corresponds to the /r/. Smoothed formant trajectories (F1–F4) are plotted using the *geom_smooth()* function in the *ggplot2* package (method = 'gam' and formula 'y ~ s(x, bs = "cs")') (Wickham 2016). A 95% confidence interval for each fit is shown in grey, although this is very narrow, particularly as we plot all formants on a single plot ranging from 0 Hz to 4 kHz.

6.2.1 Lower formants (F1–F2)

The movement of (F1) is as expected and largely similar across all VrV sequences; F1 has a largely symmetrical movement across the sequence, lowering slightly into the /r/ beginning approximately in the last quarter of the preceding vowel, and rising into the following vowel. The degree of lowering of F1 varies across the different vowel environments. This lowering of F1 timed to /r/ is coherent with the articulatory process of tongue raising into an alveolar tap/flap articulation. The relative degree of tongue movement is congruently reflected in the degree of F1 movement; of the five vowel sequences pictured in
Figure **5**, the greatest change is seen in /ara/ where the tongue begins in the lowest position possible and therefore F1 is the highest of the five vowels. For both /iri/ and /ere/, (F2) lowers slightly into the /r/ and rises similarly out of it. In the sequence /oro/ the inverse happens. In the sequence /uru/ F2 rises into the /r/, but does not lower as the second /u/ is produced. This indicates the /u/ articulation following /r/ is potentially more fronted as is consistent with a preceding alveolar consonant (Maclagan et al. 2009).

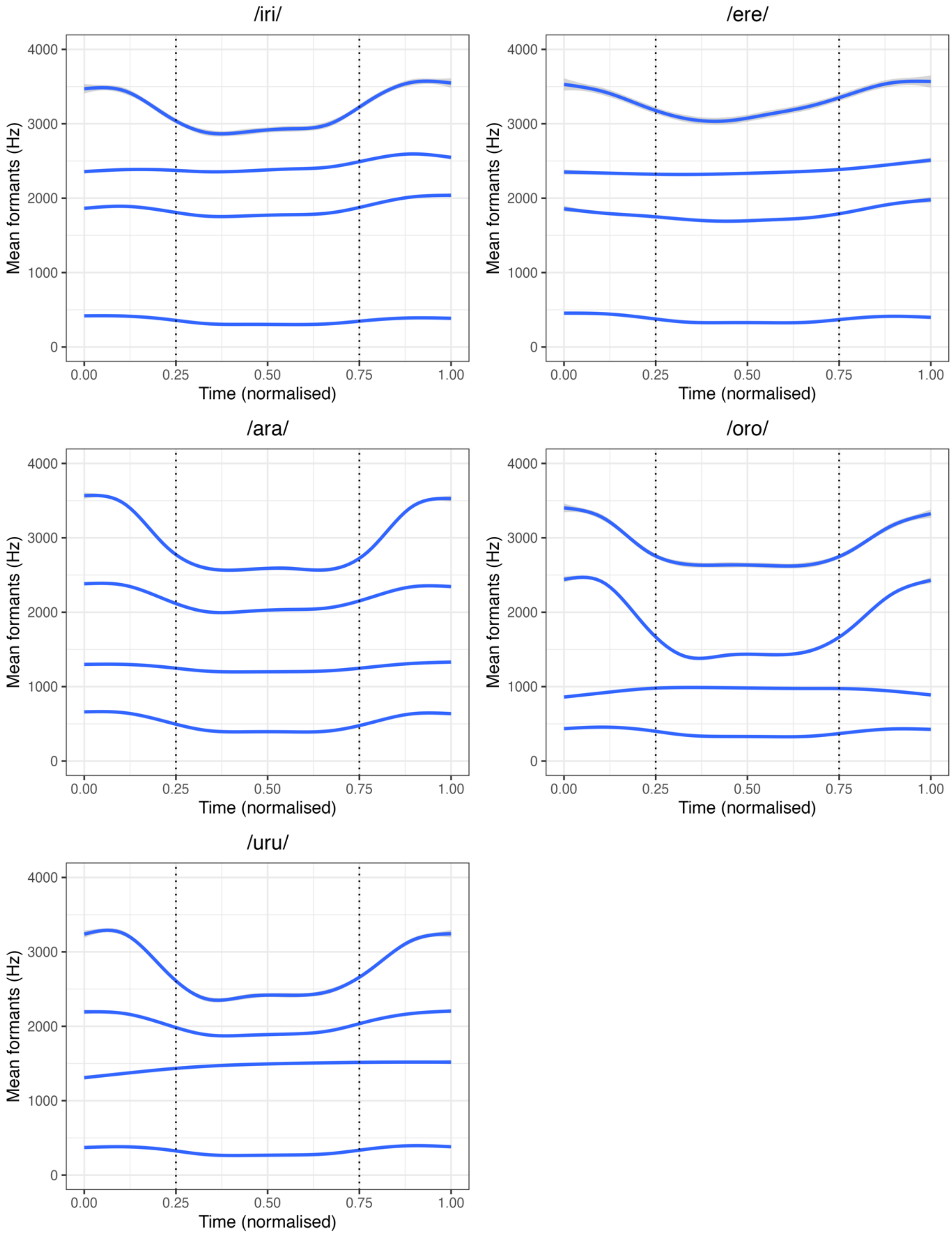

Figure 5. Formant trajectories of Category 3 /r/ tokens (tap/flap with no burst) extracted from the Ngā Mahi recordings (F1–F4) for /iri/, /ere/, /ara/, /oro/, and /uru/ sequences (Note that in cases where formant energy was not present throughout the /r/ production, the smooths are fitting across empty space.)

6.2.2 Higher formants (F3–F4)

The third formant (F3) behaves relatively similarly across the different VrV environments, except in the /oro/ sequence. In the /iri/, /ere/, /ara/, and /uru/ sequences (see Figure **5**) there was relatively small lowering of F3, although this occurred to different degrees and with different degrees of symmetry. Lowering of F3 in /oro/ is conspicuous, with the mean value decreasing by more than 1000 Hz. We suspect there is a particular aspect of the Ngā Mahi speaker's approach to articulating the /oro/ sequence, which results in the changes in resonance necessary to reduced F3 so notably. Without articulatory data, we cannot discern the exact source of this lowering. In approximant /r/ variants, F3 lowering has been attributed as a resonance of a sub-lingual front cavity in the vocal tract produced when the tongue is raised (Stevens 1999: 535–543). Furthermore, the lowering of F3 in the sequence /oro/ is likely intensified by lowering due to lip-rounding.

Generally speaking, fourth formant (F4) lowering clearly occurs to some degree in all /VrV/ sequences considered. F4 movement appears to be symmetrical in some sequences (e.g., /ara/, /oro/), but mostly asymmetrical elsewhere. The shape of the lowering roughly follows that of an inverted parabola which can be skewed in cases where there is asymmetrical movement (these generally occur, unsurprisingly, in sequences where the preceding and following vowels are different). The aforementioned AE flap study reports that, for the VC transition, 4.6% of observations have F4 lowering exceeding 1 kHz (Warner & Tucker 2017: 14). In the present study of Māori, we find 51.4% of tokens have lowering exceeding 1 kHz. For the CV transition, 2.0% of AE tokens had lowering exceeding 1 kHz, compared to 39.1% of Māori /r/ observations in this analysis. Overall, we observe a higher rate of F4 lowering across a wide range of segmental contexts than was reported in the AE investigation.

6.3 Summary

In Investigation C, we observe the behaviours of lower and higher formants (F1–F4) in a range of VrV sequences produced by the Ngā Mahi speaker. The observed F1 and F2 behaviours are coherent with the expected articulatory movements required to produce an alveolar tap/flap. The movements of F4 were of particular interest, and we find that notable F4 lowering occurred with /r/ in all considered vowel contexts in the Ngā Mahi speech. This was even the case for contexts where the vowel was mid-front or high-front, which we consider to be an articulatory target in close proximity to that of an alveolar tap/flap. It therefore seems unlikely that the observed F4 behaviours are the result of some sort of anticipatory articulation when the degree of lowering does not appear to be greatly reduced in these contexts, as is suggested in the AE flap investigation (Warner & Tucker 2017). It is thus still very much unclear what the articulatory source of the lowered F4 is, and without an articulatory study, we cannot reach any firm conclusions on its source. The lowering could be a characteristic of the Ngā Mahi speaker's approach to producing Māori /r/, or it could point towards a more widespread articulatory approach to the sound. This is an avenue for future research.

7 General discussion

In this study we consider different acoustic measures in relation to Māori /r/. These basic measures of duration, intensity, and formant behaviour are fundamental in the acoustic characterisation of a sound. We find that the suprasegmental environment of /r/ may factor in the realisation of these features.

Our findings are coherent with descriptions of Māori /r/ given by Bauer (1993) and Harlow (2007), with a few exceptions. The vast majority of /r/ productions analysed were variants of a tap/flap. Through analysis of spectrographic data, we identified a range of realisations of this tap/flap which combined salient landmarks: taps/flaps with or without formant energies, and taps/flaps with or without burst energies. Analysis of these landmarks in taps/flaps in other languages has confirmed similar acoustic variation (Son 2008; Warner, Fountain, & Tucker 2009; Warner & Tucker 2011; Bradley & Willis 2012). In the MAONZE kaumātua group, approximant /r/ appears in the position anticipated by Harlow, with few exceptions. The Ngā Mahi speaker also tends to follow this trend, but with more exceptions to the rule. The observations of approximant /r/ outside the expected position could indicate the emergence of approximant /r/ as an increasingly prevalent allophone of Māori /r/. For the MAONZE kaumātua, approximant /r/ appears outside of this expected position at a rate small enough to be explained as mispronunciation error. As was expected, we found very limited evidence of trilled /r/. While there is reference to trilled /r/ in at least one description of Māori, it is uncommon and markedly absent from contemporary descriptions, save for the occasional mention in more colloquial descriptions (e.g. in some Google Search results). In accordance with Harlow's (2007) descriptions, and the previous /r/ studies, we do not find lateral realisations in our analyses, neither in the MAONZE kaumātua data nor the Ngā Mahi speaker data. We do note, however, that we have heard anecdotal reports from fluent speakers of lateral productions of /r/ in their area (in particular, speakers from Te Tai Tokerau (Northland)). These findings address half of our first research question: there are indeed various realisations of Māori /r/ (taps/flaps, approximants, and the occasional trill) and these demonstrate notable variation in their acoustics, as evidenced by spectrographic analysis.

Investigation B addressed our second research question concerning word stress and its interaction with /r/ duration, one possible correlate of stress in Māori. The results of this investigation confirm Bauer's (1993: 545) description of stressed /r/ in Māori. In the Ngā Mahi speaker data we find a tendency for duration to increase when /r/ appears in the onset of the stressed syllable. There is every possibility, however, that the inverse could be found in a different speaker's /r/ productions. A shortcoming of the stress analyses is the lack of consideration of phrasal stress and intonation. While we do not understand fully how these factors manifest acoustically in Māori speech, we suspect that they may interact with /r/. Bauer also indicates /r/ in the stressed onset may be accompanied by increased intensity, but we do not report on this here. We recommend each of these factors be considered in further studies. Finally, the present investigation into stress conflates the position of /r/ in a word and whether it is stressed/unstressed. In order to tease out any interactions between /r/ and stress, this needs to be considered in subsequent work.

Investigation C honed in on our final research question regarding the fourth formant, where we identified considerable lowering of F4 in the Ngā Mahi speaker data. This lowering is confirmed to occur across a range of vowel contexts, implying that it may be a consistent feature of Māori /r/, at least for the single speaker considered. As discussed, F4 lowering has been associated with retroflexion (such as in Zhou et al. 2008), but the F4 lowering observed in our data could well be the result of some other articulatory approach preferred by the Ngā Mahi speaker. Without an articulatory study, the cause of the observed F4 lowering will be difficult to ascertain. An acoustic investigation of a wider speaker group would better reflect the population and confirm whether this feature is indeed a feature of Māori /r/. The investigation into F4 lowering supports the argument for analysing high-quality recordings of speech. The discovery of F4 lowering in the Ngā Mahi recordings was accidental and would not have been possible if we analysed only the MAONZE data. This is due to issues with audio quality. While the MAONZE corpus offers a potential trove of information pertaining to Māori /r/ production, it will not be sufficient to paint a complete picture of the acoustics of Māori /r/, particularly where analyses are concerned with the higher formants (F3 and F4). Neither, obviously, will the Ngā Mahi recordings in their current state, given the corpus comprises data from a single speaker. Further analysis requires quality audio recordings of more speakers to determine if F4 lowering is indeed a feature of Māori /r/ and not an idiosyncrasy of the analysed speaker.

Much like rhotic sounds in other languages, we find evidence of inter-speaker variation in production as well as intra-speaker variation, which we have already discussed. In our investigation of the MAONZE kaumātua, we identified different preferences for /r/ variants across speakers and target words. A broader systematic investigation with a wider set of target words could shed light on these speaker preferences, and identify if they are possibly conditioned by factors such as segmental and suprasegmental environment. These conclusions were the major motivation for the development of a Māori /r/ production study (see Shields 2022, for preliminary findings of this work). This research focuses on the possible interactions between /r/ and its segmental and suprasegmental environment. Recordings for this production study have been undertaken in the same WhisperRoom Sound Isolation chamber used in the development of the Ngā Mahi corpus. Analysis of these recordings will make clear how widespread F4 lowering is in a broader speaker group. Presently, the corpus developed consists of recordings from 17 speakers of Māori, including both first- and second-language speakers.

8 Conclusions

The primary objective of these investigations was to begin the groundwork of understanding the acoustics of Māori /r/. In line with contemporary descriptions of the language, we have confirmed that /r/ appears predominantly as a flap, and occasionally as an approximant. We found a range of realisations of the sound through spectrographic analysis. With regard to potential acoustic features, we found that duration appears to be influenced by lexical stress (or indeed word position) for the Ngā Mahi speaker, with duration found to increase in greater proximity to the stressed syllable. In the Ngā Mahi corpus speech, we identified notable lowering of F4 in all vowel environments. We suggest this feature be investigated in a wider speaker group to ascertain if it is a widespread correlate of /r/.

Acknowledgements

We extend our thanks to Margaret Maclagan, as well as the wider MAONZE research group for their support and guidance. We express our gratitude to the speakers recorded in the MAONZE corpus, and to the Ngā Mahi speaker. Finally, we thank the reviewers for their insightful comments. This research has been funded by the University of Auckland Doctoral Scholarship.

References

Audacity Team. 2021. Audacity®.

Bauer, Winifred. 1981. Hae. re vs. ha. e. re: A note. *Te Reo* 24. 31–36.

Bauer, Winifred. 1993. *Maori*. London: Routledge.

Biggs, Bruce. 1961. The structure of New Zealand Maaori. *Anthropological Linguistics* 3(3). 1–54.

Biggs, Bruce. 1998. *Let's learn Maori: A guide to the study of the Maori language*. 3rd edn. Auckland: Auckland University Press.

Boersma, Paul & Weenik, David. 2010. *Praat: Doing phonetics by computer*. https://doi.org/10.1097/aud.0b013e31821473f7

Bombien, Lasse & Winkelmann, Raphael & Scheffers, Michel. 2021. *Wrassp: An R wrapper to the ASSP library*. *https://rdrr.io/cran/wrassp/*

Boyce, Suzanne & Espy-Wilson, Carol Y. 1997. Coarticulatory stability in American English /R/. *The Journal of the Acoustical Society of America* 101(6). 3741–53. https://doi.org/10.1121/1.418333

Bradley, Travis G. & Willis, Erik W. 2012. Rhotic variation and contrast in Veracruz Mexican Spanish. *Estudios de Fonética Experimental* 21. 43–74.

Chabot, Alex. 2019. What's wrong with being a rhotic? *Glossa: A Journal of General Linguistics* 4(1). 1–24.

Cohen, Evan-Gary & Laks, Lior & Savu, Carmen. 2019. The phonetics of Modern Hebrew rhotics. *Brill's Annual of Afroasiatic Languages and Linguistics* 11(1). 28–48. https://doi.org/10.1163/18776930-01101003

Derrick, Donald & Schultz, Benjamin. 2013. Acoustic correlates of flaps in North American English. In *Proceedings of Meetings on Acoustics ICA2013*. Vol. 19. Acoustical Society of America. https://doi.org/10.1121/1.4798779

Espy-Wilson, Carol Y. 1992. Acoustic measures for linguistic features distinguishing the semivowels /w j r l/ in American English. *The Journal of the Acoustical Society of America* 92(2). 736–57. https://doi.org/10.1121/1.403998

Faytak, Matthew & Aziz, Jacob & Barnett, Phillip & Jo, Jinyoung & Kuo, Jennifer & Teixeira, G. & Wu, Joy, & Zhou, Z. L. & Keating, Patricia. 2019. Flap articulation and lowered fourth formant. In *178th Meeting of the Acoustical Society of America*. https://github.com/mfaytak/flaps-f4-asa/blob/master/5aSC12.pdf

Grey, George & Williams, Herbert W. 1971. *Nga mahi a nga tupuna*. 4th ed. Auckland: Reed.

Gurugubelli, Krishna & Vuppala, Anil K. & Narendra, N. P. & Alku, Paavo. 2020. Duration of the rhotic approximant /ɹ/ in spastic dysarthria of different severity levels. *Speech Communication* 125. 61–68. https://doi.org/10.1016/j.specom.2020.09.006

Harlow, Ray. 2007. *Maori: A linguistic introduction*. Cambridge: Cambridge University Press.

Harlow, Ray & Keegan, Peter & King, Jeanette & Maclagan, Margaret & Watson, Catherine. 2009. The changing sound of the Māori language. In Stanford, J. N. & Preston, D. R. (eds.), *Variation in Indigenous minority languages*, 129–52. Amsterdam/Philadelphia: John Benjamins Amsterdam.

Heyne, Matthias & Wang, Xuan & Derrick, Donald & Dorreen, Kieran & Kevin Watson. 2018. The articulation of /ɹ/ in New Zealand English. *Journal of the International Phonetic Association* 50(3). 366–88. https://doi.org/10.1017/S0025100318000324

Hohepa, Patrick Wahanga. 1965. *A profile-generative grammar of Maori*. Bloomington: Indiana University. (Doctoral dissertation.)

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

King, Jeanette & Maclagan, Margaret & Harlow, Ray & Keegan, Peter J. & Watson, Catherine I. 2010. The MAONZE corpus: Establishing a corpus of Maori speech. *New Zealand Studies in Applied Linguistics* 16(2). 1–16.

King, Jeanette & Maclagan, Margaret, Harlow, Ray & Keegan, Peter J. & Watson, Catherine I. 2011a. The MAONZE corpus: Transcribing and analysing Maori speech. *New Zealand Studies in Applied Linguistics* 17(1). 32–48.

King, Jeanette & Maclagan, Margaret & Harlow, Ray & Keegan, Peter J. & Watson, Catherine I. 2011b. The MAONZE project: Changing uses of an Indigenous language database. *Corpus Linguistics and Linguistic Theory*. 7(1). 37–57. https://doi.org/10.1515/cllt.2011.003

Ladefoged, Peter & Maddieson, Ian. 1996. *The sounds of the world's languages*. Oxford/Cambridge, MA: Blackwell.

Maclagan, Margaret & Harlow, Ray & King, Jeanette & Keegan, Peter J. & Watson, Catherine I. 2004. Acoustic analysis of Maori: Historical data. In *Proceedings of the 2004 Conference of the Australian Linguistic Society*, Sydney, July 13–15.

Maclagan, Margaret & Harlow, Ray & King, Jeanette & Watson, Catherine I. & Keegan, Peter J. 2019. The pronunciation of 'Māori.' In *NZ Linguistics Society Conference*. Christchurch: University of Canterbury, 28–29 November.

Maclagan, Margaret & King, Jeanette. 2004. A note on the realisation of /r/ in the word Maori. *New Zealand English Journal* 18. 35–39.

Maclagan, Margaret & King, Jeanette. 2007. Aspiration of plosives in Māori: Change over time. *Australian Journal of Linguistics* 27(1). 81–96.

Maclagan, Margaret & Watson, Catherine I. & Harlow, Ray & King, Jeanette & Keegan, Peter J. 2009. /u/ fronting and /t/ aspiration in Māori and New Zealand English. *Language Variation and Change* 21(2). 175–92. https://doi.org/10.1017/S095439450999007X

McAuliffe, Michael & Socolof, Michaela & Mihuc, Sarah & Wagner, Michael & Sonderegger, Morgan. 2017. Montreal Forced Aligner: Trainable text-speech alignment using Kaldi. In *INTERSPEECH*, 2017: 498–502. https://www.isca-speech.org/archive/Interspeech_2017/pdfs/1386.PDF

Nance, Claire. 2014. Phonetic variation in Scottish Gaelic laterals. *Journal of Phonetics* 47(1). 1–17. https://doi.org/10.1016/j.wocn.2014.07.005

R Core Team. 2020. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

Rácz, Pè & Hay, Jennifer & Needle, Jeremy & King, Jeanette & Pierrehumbert, Janet B. 2016. Gradient Māori phonotactics. *Te Reo* 59. 3-21.

Rafat, Yasaman. 2010. A socio-phonetic investigation of rhotics in Persian. *Iranian Studies* 43 (5). 667–82.

Recasens, Daniel. 1991. On the production characteristics of apicoalveolar taps and trills. *Journal of Phonetics* 19(3–4). 267–80. https://doi.org/10.1016/S0095-4470(19)30344-4

Rennicke, Iiris E. 2015. *Variation and change in the rhotics of Brazilian Portuguese*. Belo Horizonte: Federal University of Minas Gerais. (Doctoral dissertation.)

RStudio Team. 2021. *RStudio: Integrated development environment for R.* Boston. http://www.rstudio.com/

Shields, Isabella & James, Jesin & Watson, Catherine I. & Keegan, Peter J. 2020. *Māori TTS report.* Auckland: University of Auckland.

Shields, Isabella & Watson, Catherine I. & Keegan, Peter J. & Berriman, Rebekah & James, Jesin. 2019. Te reo Māori voice for TTS. In *Language Technology for All Conference*, Paris, 4–6 December. https://lt4all.elra.info/media/papers/P1/136.pdf

Shields, Isabella & Watson, Catherine I. & Keegan, Peter J. & Maclagan, Margaret. 2021a. A preliminary investigation of the acoustics of Māori /r/. In *Language and Society Conference*, Hamilton: University of Waikato, 10-12 February.

Shields, Isabella & Watson, Catherine I, & Keegan, Peter J. 2021b. Acoustics of te reo Māori /r/: First impressions and corpus building for fourth formant and stress analysis. In *R-atics 7 Conference*, Lausanne: University of Lausanne, 18-19 November.

Shields, Isabella & Watson, Catherine I. & Keegan, Peter J. 2022. Preliminary analysis of /r/ acoustics and features in three Māori speakers. In *Proceedings of the Eighteenth*

Australasian International Conference on Speech Science and Technology, Canberra, December 13–16, 96–100.

Solé, Maria-Josep. 2002. Aerodynamic characteristics of trills and phonological patterning. *Journal of Phonetics* 30(4). 655–88. https://doi.org/10.1006/jpho.2002.0179

Son, Minjung. 2008. Pitfalls of spectrogram readings of flaps. *Journal of the Acoustical Society of America* 123(5). 3079.

Stats NZ. 2013. 2013 Census QuickStats about Māori. Wellington: Stats NZ. https://www.stats.govt.nz/assets/Uploads/Retirement-of-archive-website-project-files/Reports/2013-Census-QuickStats-about-Maori/qs-maori.pdf

Stats NZ. 2018. Te Kupenga 2018. https://www.stats.govt.nz/information-releases/te-kupenga-2018-final-english

Stevens, Kenneth N. 1999. *Acoustic phonetics*. Cambridge: MIT Press.

Stevens, Kenneth N. 2002. Toward a model for lexical access based on acoustic landmarks and distinctive features. *The Journal of the Acoustical Society of America* 111(4). 1872–91. https://doi.org/10.1121/1.1458026

Stevens, Kenneth N. & Blumstein, Sheila E. 1975. Quantal aspects of consonant production and perception: A study of retroflex stop consonants. *Journal of Phonetics* 3(4), 215–233.

Stoakes, Hywel M. & Watson, Catherine I. & Keegan, Peter J. & Maclagan, Margaret A. & King, Jeanette & Harlow, Ray. 2019. The dynamics of closing diphthong formant trajectories in te reo Māori. In Calhoun, S. & Escudero, P. & Tabain, M. & Warren, P. (eds.), *Proceedings of the 19th International Congress of Phonetic Sciences*, 989–93. Canberra, Australia: Australasian Speech Science and Technology Association Inc. https://assta.org/proceedings/ICPhS2019/papers/ICPhS_1038.pdf

Thompson, Laura & Watson, Catherine I. & Charters, Helen & Harlow, Ray & Keegan, Peter J. & King, Jeanette & Maclagan, Margaret. 2010. An experiment in *mita*-reading: investigating perception of rhythmic prominence in the Māori language. In *Proceedings of the Thirteenth Australasian International Conference on Speech Science and Technology*, Melbourne, December 14–16, 150–153.

Thompson, Laura & Watson, Catherine I. & Harlow, Ray & Maclagan, Margaret & Charters, Helen & King, Jeanette & Keegan, Peter J. 2011a. Adventures in *mita*-reading: examining stress rules and perception of prosodic prominence in the Māori language. In *XVIIth International Congress of Phonetic Sciences*. Hong Kong, 17-21 August, 1982–1985.

Thompson, Laura & Watson, Catherine I. & Harlow, Ray & King, Jeanette & Maclagan, Margaret & Charters, Helen & Keegan, Peter J. 2011b. Phrases, Pitch and Perceived Prominence in Māori. In *Twelfth Annual Conference of the International Speech Communication Association*. Florence, 27–31 August, 1365–1368.

Warner, Natasha & Fountain, Amy & Tucker, Benjamin V. 2009. Cues to perception of reduced flaps. *The Journal of the Acoustical Society of America* 125(5). 3317–27. https://doi.org/10.1121/1.3097773

Warner, Natasha & Tucker, Benjamin V. 2011. Phonetic variability of stops and flaps in spontaneous and careful speech. *The Journal of the Acoustical Society of America* 130(3), 1606–1617. https://doi.org/10.1121/1.3621306

Warner, Natasha & Tucker, Benjamin V. 2017. An effect of flaps on the fourth formant in English. *Journal of the International Phonetic Association* 47(1). 1–15. https://doi.org/10.1017/S0025100316000219

Wickham, H. 2016. *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Wiese, Richard. 2003. The unity and variation of (German) /R/. *Zeitschrift Für Dialektologie Und Linguistik* 70(1). 25–43.

Willis, Erik W. 2007. An acoustic study of the 'pre-aspirated trill' in narrative Cibaeño Dominican Spanish. *Journal of the International Phonetic Association* 37(1). 33–49. https://doi.org/10.1017/S0025100306002799

Winkelmann, Raphael & Jaensch, Klaus & Cassidy, Steve & Harrington, Jonathan. 2021. *EmuR: Main package of the EMU speech database management system*. https://rdrr.io/cran/emuR/

Yun, Suyeon & Choi, Jeung-Yoon & Shattuck-Hufnagel, Stefanie. 2020. A landmark-cue-based approach to analyzing the acoustic realizations of American English intervocalic flaps. *The Journal of the Acoustical Society of America* 147(6). EL471–477. https://doi.org/10.1121/10.0001345

Zhou, Xinhui & Espy-Wilson, Carol Y. & Boyce, Suzanne & Tiede, Mark & Holland, Christy & Choe, Ann. 2008. A magnetic resonance imaging-based articulatory and acoustic study of "retroflex" and "bunched" American English /R/. *The Journal of the Acoustical Society of America* 123(6). 4466–4481. https://doi.org/10.1121/1.2902168

Zhou, Xinhui & Espy-Wilson, Carol Y. & Tiede, Mark & Boyce, Suzanne. 2007. Acoustic cues of "retroflex" and "bunched" American English rhotic sound. *The Journal of the Acoustical Society of America* 121(5). 3168. https://doi.org/10.1121/1.4782272

Zue, Victor W. & Laferriere, Martha. 1979. Acoustic study of medial /t,d/ in American English. *The Journal of the Acoustical Society of America* 66(4). 1039–1050. https://doi.org/10.1121/1.383323