



Derrick, Donald & Mayer, Connor & Gick, Bryan. 2024. Uniformity in speech: The economy of reuse and adaptation across contexts. *Glossa: a journal of general linguistics* 9(1). pp. 1–35. DOI: <https://doi.org/10.16995/glossa.11681>



Uniformity in speech: The economy of reuse and adaptation across contexts

Donald Derrick, University of Canterbury, NZ, donald.derrick@canterbury.ac.nz

Connor Mayer, University of California, Irvine, US, cjmayer@uci.edu

Bryan Gick, University of British Columbia, CAN and Haskins Laboratories, US, gick@mail.ubc.ca

North American English (NAE) flaps/taps and rhotic vowels have been shown to exhibit extreme variability that can be categorized into subphonemic variants. This variability provides known mechanical benefits in NAE speech production. However, we also know languages reuse gestures for maximum efficiency during speech production; this uniformity of behavior reduces gestural variability. Here we test two conflicting hypotheses: Under a *uniformity* hypothesis in which extreme variability is inherent to rhotic vowels only, that variability can still transfer to flaps/taps and non-rhotic vowels due to adaptation across similar speech contexts. But because of the underlying reliance on extreme variability from rhotic vowels, this uniformity hypothesis does not predict extreme variability in flaps/taps within non-rhotic English dialects. Under a *mechanical* hypothesis in which extreme variability is inherent to all segments where it would provide mechanical advantage, including flaps/taps, such variability would appear across all English dialects with flaps/taps, affecting adjacent non-rhotic vowels through coarticulation whenever doing so would provide mechanical advantage. We test these two hypotheses by comparing speech-rate-varying NAE sequences with and without rhotic vowels to sequences from New Zealand English (NZE), which has flaps/taps, but no rhotic vowels at all. We find that NZE speakers all use similar tongue-tip motion patterns for flaps/taps across both slow and fast speech, unlike NAE speakers who sometimes use two different stable patterns, one for slow and another fast speech. Results show extreme variability is not inherent to flaps/taps across English dialects, supporting the *uniformity* hypothesis.

Glossa: a journal of general linguistics is a peer-reviewed open access journal published by the Open Library of Humanities. © 2024 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

OPEN ACCESS



1 Introduction

Two speech sounds have been prominently identified in the languages of the world as exhibiting extreme articulatory variability, such that articulatory pathways can involve completely different directions and patterns of motion – the English rhotic approximant (Delattre & Freeman 1968; Hagiwara 1995; Westbury et al. 1998; Guenther et al. 1999; Mielke et al. 2010; 2016) and the North American English (NAE) flap/tap (Derrick & Gick 2011; 2014; Derrick et al. 2015b). Curiously, both of these speech sounds occur in at least some dialects of a single language. In this study, we test whether such extreme articulatory variability is an independent property of both these sounds, or whether it has been transferred from rhotics to flaps/taps.

It is well known that English /ɹ/ exhibits a great deal of articulatory variation within and between speakers (e.g. Delattre & Freeman 1968; Mielke et al. 2010; 2016). While some speakers use one strategy, others display idiosyncratic but consistent patterns of allophony where different variants are employed in different contexts. It has been claimed that this idiosyncratic variability is permitted because it is largely imperceptible (Guenther et al. 1999). The most salient feature of English /ɹ/, a low F3, can be achieved by several distinct articulatory strategies, allowing speakers to use choose production strategies that minimize articulatory difficulty in different contexts without sacrificing acoustic goals (Guenther et al. 1999).

Similar results have been found for NAE taps and flaps, which show extreme variability via the motion direction of tongue-tip contact (or approximation) towards and away from the alveolar region (Derrick & Gick 2011; 2014; Derrick et al. 2015b). Although flap/tap variability is often conditioned by proximity to /ɹ/, it appears even in contexts where there is no /ɹ/ nearby (Derrick & Gick 2011; 2014). This variability is also apparent in differences in the articulation of rhotic-flap/tap sequences between slow and fast speech in NAE (Derrick & Gick 2021), resulting in sometimes very different patterns of motion for /VrVrV/ sequences (where V can be a rhotic or non-rhotic vowel) in slow vs. fast speech. Previous work has also independently shown that there is a mechanical benefit to using extreme variation in taps and flaps, even in non-rhotic contexts (Derrick & Gick 2014). Similarly to rhotics, the acoustic differences between different tap/flap strategies are dominantly found in F2, and likely difficult to perceive (Derrick & Schultz 2013).

Although mechanical ease accounts for much of the variability observed in these sounds, a competing pressure also appears to be at play. There are well-known speech constraints that favor reuse and sharing of structural components in speech, be they phonological features (Clements 2003; Archangeli et al. 2011), gestures, or other components of speech production. Chodroff and Wilson (2017), and Faytak (2018) extend Keating (2003)'s notion of a *uniformity constraint* to targets of acoustic and articulatory phonetic realization. In this analysis, reuse of patterns of behavior themselves provide a kind of ease of articulation that can and does compete with mechanical ease of articulation (Keating 2003). The conflict between mechanical ease and uniformity is evident in both /ɹ/ and tap/flap productions: while some speakers show a high

degree of contextual variability that facilitates mechanical ease, others use the same strategy across most or all contexts.

The question we aim to address in this paper is why these sounds (in particular, flaps/taps) exhibit the level of variability they do. The first hypothesis we will consider is the *mechanical hypothesis*. Under this hypothesis, we predict both /ɹ/ and taps/flaps to vary *independently*. This variance arises from the fact that both sounds can be produced using categorically distinct articulatory strategies that achieve essentially the same perceptual outcome, and speakers exploit these strategies for mechanical ease (modulo pressures towards uniformity).

However, there is a confound in the currently available data: while it is clear that English /ɹ/ production varies even in dialects without taps/flaps, variability in tap/flap production has not been studied in a non-rhotic dialect of English. Because so much of the variability in tap/flap production is conditioned by the presence of rhotics (particularly syllabic rhotics, which can occur adjacent to taps/flaps), this raises the question of whether variability in rhotic production is a *precondition* for variability in tap/flap production. We will call this the *uniformity hypothesis*.

Under this hypothesis, the pressure for uniformity in articulatory realization will dominate unless there is sufficient pressure from contextual factors to allow for the development of multiple strategies. However, once a new strategy is available, the principle of uniformity allows it to be reused in contexts where it was not originally developed. To generalize: Suppose a speaker frequently produces a segment /S/ in two contexts, C_1 and C_2 . For mechanical reasons, C_1 strongly favors articulation $[S_1]$ and C_2 strongly favors articulation $[S_2]$, and the speaker comes to use these strategies in each context. In some new context C_3 that weakly favors articulation $[S_2]$, this speaker is free to deploy $[S_2]$, because it is part of her repertoire of solutions. Now suppose that a different speaker only produces /S/ in context C_1 , and hence only develops articulation $[S_1]$. In this case, uniformity will require that the speaker use $[S_1]$ in C_3 , even though a more mechanically optimal strategy $[S_2]$ is possible. In other words, there is a strong pressure to reuse entrenched strategies rather than developing new ones, unless the mechanical cost is too high.

In our case, segment S would be the NAE flap/tap, context C_1 might be preceding a non-rhotic vowel, context C_2 might be preceding a rhotic vowel, and context C_3 might be part of a sequence of intervocal taps/flaps with no rhotics. Uniformity allows variability conditioned by one context (an adjacent rhotic) to be used in other contexts where no rhotics are present.

The *mechanical hypothesis* and the *uniformity hypothesis* make different predictions about the behavior of taps/flaps in a non-rhotic dialect of English. Under the mechanical hypothesis, non-rhotic dialects should also display extreme variability in tap/flap production in a way that facilitates mechanical ease. Under the uniformity hypothesis, non-rhotic dialects are not predicted to exhibit this variability, because of a lack of strong contextual conditioning (namely rhotic vowels) that allows for the development of variable articulation strategies.

To test the predictions of the *mechanical hypothesis* vs. the *uniformity hypothesis*, we look at sequences involving only non-rhotic vowels and flaps in NAE, a rhotic English dialect, and New Zealand English (NZE), a non-rhotic dialect. NZE has both flaps/taps and highly variable rhotic prevocalic consonants (Heyne et al. 2018), but no rhotic vowels or other extreme variability-bearing vocalic segments adjacent to flaps/taps, which only occur intervocalically (see Zue & Laferriere 1979 for an NAE analogy). Previous work has already shown that many English speakers will switch from one tap/flap production strategy to another as speech rate increases (Derrick and Gick 2021). In the current paper, we test whether NZE speakers produce different stable sequences of tongue motion at slow and fast speech rates in the same way NAE does. We then measure tongue position at the midpoint of /VrVrV/ sequences in the absence of rhotic vowels under different speech rate conditions to identify if NZE shows the degree of variability we see in NAE.

1.1 Background

English rhotics have long been described as exhibiting extreme articulatory variability across speakers, even extending to the categorical (e.g., Delattre & Freeman 1968; Tiede et al. 2004). Westbury et al. (1998) referred to NAE /ɹ/ as “infamously variable”, and Guenther et al. (1999) says “The American English phoneme /ɹ/ has long been associated with relatively large amounts of articulatory variability”, noting that /ɹ/ seems to be subject to substantial within-speaker articulatory variation, while maintaining relatively stable acoustic (F3) targets (see also Delattre & Freeman 1968; Espy-Wilson & Boyce 1994; A Hagiwara 1995; Alwan et al. 1997; Ong and Stone 1998; Westbury et al. 1998).

Extreme variability in /ɹ/ variation is constrained in that it seems to be stochastically distributed; at least some /ɹ/ variation has been associated with distribution patterns influenced by adjacent segments (Mielke et al. 2010; 2016; Heyne et al. 2018). Among speakers who can produce bunched (tongue tip-down) and retroflex (tongue tip-up) rhotic consonants, prevocalic rhotic consonants are more likely to be tip-up than postvocalic rhotic consonants. Prevocalic rhotic consonants are more likely to be produced tip-up before low and back vowels, while rhotic consonants are less likely to be produced tip-up following coronal consonants and following fricatives (see Figure 4 in Mielke et al. 2016).

Some of this contextual variation has been attributed to biomechanical factors. A biomechanical simulation study (Stavness et al. 2012) found that movements are less costly between certain rhotic shapes and certain vowels, showing that movements from tongue tip-up rhotic to /a/, and from tip-down rhotic to /i/, produce comparatively less muscle stress, strain, and volume displacement. Heyne et al. (2018) demonstrated similar patterns of behavior in the production of rhotic consonants in New Zealand English (NZE), despite the lack of rhotic vowels, indicating

that the variation patterns are likely to be based on biomechanical constraints on tongue motion interacting with the ability of at least some speakers to produce rhotics with multiple different tongue shapes. That is, extreme variability in English rhotics appears to be driven by *mechanical* constraints.

But rhotics are not the only example of extreme variability in NAE, and it is their interaction with other segments – in this case NAE taps and flaps – that we propose drives the sharable *uniformity* nature of extreme variability in a system: NAE taps and flaps can be articulated one of 4 ways: the tongue can impact the alveolar ridge via a tap from below (alveolar tap), above and back (post-alveolar tap), or by a tangential impact from below (up-flap) or above (down-flap) (Derrick & Gick 2011). The likelihood of any particular flap pattern is based on several factors:

The vowels that precede and follow flaps influence flap motion direction depending on whether the vowels are rhotic or non-rhotic (Derrick & Gick 2011); individual variation across speakers shows that, while some speakers do not show categorical variation, others will vary productions across repetitions of the same word, even in the same phonetic context. That is, there are many subphonemic tongue motion patterns for flaps. The full reason for such variability was not identified, but it must be noted that this previous study did not account for speech rate.

Gravity and elasticity also influence rhotic and flap motion (Derrick et al. 2015b), such that some word/phrase sequences, such as /VrɚrV/ (e.g. “Saturday”) exhibit a remarkably stable tongue tip path for most speakers. In our previous study “fully 180 of 213, or 84.5%, (of the) sequences in “Saturday” were produced as...sequences involv[ing] a single up/down arc of motion...As expected, the up-down flap sequence is thus dramatically overrepresented in our production results” (Derrick et al. 2015b, page 1499). This tongue-tip up-down motion allows for successful production of all three vowels in a /VrɚrV/ sequence, and biomechanical simulations showed that gravity and muscle elasticity allow the entire sequence of tongue-tip motion be executed with a single initial burst of muscle activity for the up-flap, followed by relaxation into the down-flap (Derrick et al. 2015b). That is, it is possible to use one tongue motion activation to span many segments. This shows evidence of a mechanical advantage for extreme variability in both rhotic and flap production.

NAE flaps also influence surrounding vowels by accommodating end-state comfort (Derrick & Gick 2014). End-state comfort is a measure of motion planning, evidenced by a willingness to start complex motion in an uncomfortable state to end in a comfortable state (Rosenbaum et al. 1996). A classic example is rotating the wrist upside down if you know you need to pick up a glass and flip it over so that your wrist is comfortable at the end of the motion (see similar examples in Rosenbaum et al. 1992). That is, when a speaker plans a /VrVrɚ/ sequence like “editor” or “auditor”, the tongue tip motion often begins with a tip-down vowel, followed by an up-flap into a second tip-up vowel, followed by a postalveolar tap to a final tip-up rhotic vowel.

In this tongue tip motion pattern, the middle vowel's production is altered to allow for end-state comfort. End-state comfort itself is used as evidence of sequence planning in motor control (Rosenblum et al. 1992; 1996). The result is that in these sequences, speakers can produce non-rhotic vowels with a tip-up tongue shape. That is, segments that should be produced with tongue tip-down can be produced in a different way to accommodate the needs of the motion system. There was even evidence of support for extreme variability in non-rhotic vowels in contexts with no nearby rhotic vowels: Many speakers have more up-flaps in the first flap of “edit/audit a” compared to alveolar taps for the flap in “edit/audit the” (Derrick & Gick 2014). This result shows that extreme variability in flap/tap provides mechanical advantage making end-state comfort in complex sequences easier to achieve.

In addition to the above several factors, speech rate has been found to play an important role in within-individual variation of flaps and taps. Derrick & Gick (2021) found that many NAE speakers employ very different tongue tip movement patterns for slow vs. fast speech. This result may be thought of as “gait change” in tongue motion, analogous to the well-known gait change that occurs between walking and running in humans. Further, just as the gait change between walking and running expands the locomotive speed range of humans, the difference in tongue movement patterns gives those speakers who shift patterns access to a wider range of speech-rates than speakers who do not shift. The rate-related shifts in movement patterns observed by Derrick & Gick (2021) reveal extreme variability in both rhotic vowels and flaps.

So, our central question is whether flap/taps show cross-dialectal extreme variability the same way rhotic consonants do, supporting a mechanical hypothesis, or whether flaps/taps fail to show extreme variability in non-rhotic English dialects, supporting a uniformity hypothesis.

We address this question by comparing two language variants, both having the relevant characteristics in common (including flaps/taps and highly variable rhotic consonants) except for the lack of rhotic vowels in NZE. We compare open access data of North American English sequences with and without rhotic vowels in the environment to sequences from our New Zealand English which has no rhotic vowels at all, and we do so under different speech rates.

We perform this analysis in part by looking at the middle vowel of /VrVrV/ sequences, which, while being both unstressed and produced between two flaps, is also simply the midpoint of the complex tongue tip motion sequence under analysis. The position and unstressed state allow the vowel to be reduced more easily, particularly in fast speech, and makes the vowel potentially subject to end-state-comfort effects.

1.2 Hypotheses

These hypotheses are built on two general hypotheses in linguistic and speech research, and are here applied to data originally collected to test for speech-rate induced gait-change in NAE and

NZE. These include the 1) *Mechanical Hypothesis* in which extreme, categorizable articulatory variability is a property of English flap/tap, resulting from mechanical constraints, as previously documented for English rhotics, and 2) *Uniformity Hypothesis*, in which extreme variability in NAE flap/tap exists because it has transferred from adjacent rhotic vowels and has generalized to non-adjacent contexts.

In order to distinguish between these two hypotheses, we evaluate speech sequences in both a rhotic and non-rhotic variant of English under conditions most likely to elicit the greatest mechanical pressure for variation, by 1) observing tightly interdependent movement sequences, and 2) varying speech rate (Gay 1981).

1.3 Predictions

These two hypotheses generate several predictions. To begin, baseline prediction sets 1 and 2 are included below to provide another contextual view of already-known NAE speech behavior (Derrick & Gick 2021), for comparison with, and to facilitate, analyses of the predictions from test sets 3 and 4.

Baseline Set 1: *Adjacency*: The tongue tip position of rhotic vowels in NAE is highly variable and may influence the tongue tip position of surrounding non-rhotic vowels.

- Prediction 1a) For V2 in a NAE /VrVrV/ sequences, the tongue tip will be higher and farther back for rhotic vowels than non-rhotic vowels.
- Prediction 1b) For V2 in NAE /VrVrV/ sequences, rhotic vowel adjacency will influence tongue tip position such that rhotic vowel (NAE VRV) tongue tip height and backness > non-rhotics bounded by rhotics (NAE RVR) > non-rhotic vowels followed by a rhotic vowel (NAE VVR) > non-rhotic vowels surrounded by other non-rhotic vowels (NAE VVV).

Baseline Set 2: *Speech rate*: NAE vowels (both rhotic and non-rhotic) constrained by flaps on either side will be strongly affected by speech rate:

- Prediction 2) For NAE, for V2, all three vowel groups will have tongue tip V2 higher and farther back for faster speech rates.

Test Set 3 – *Uniformity model*: NZE vowels, not having access to the variability of rhotic vowels elsewhere having been transferred to flaps/taps, will show a significantly smaller shift between slow and fast speech rates than NAE vowels:

- Prediction 3a) For NZE, for V2, we predict significantly less of a change in tongue tip position based on speech rate compared to NAE.

Test Set 3 – *Mechanical model*: NZE flaps and taps, for mechanical reasons, already have extreme variability, which will influence adjacent NZE vowels.

- Prediction 3b) For NZE, for V2, we predict a similar change in tongue tip position based on speech rate compared to NAE.

Test Set 3 – *All models*:

- Prediction 3c) The shift for NAE vowels will take place at a similar speech-rate threshold for rhotic and non-rhotic vowels.

Test Set 4 – *Uniformity model*: NZE, not having a rhotic vowel, does not show extreme variability in /VrVrV/ sequences:

- Prediction 4a) NZE will not demonstrate stable categorical differences between slow and fast speech the way NAE does: It will not show tongue-tip gait change.

Test Set 4 – *Mechanical Model*: NZE, for mechanical reasons, shows extreme variability in /VrVrV/ sequences:

- Prediction 4b) NZE will demonstrate stable categorical differences between slow and fast speech the way NAE does: It will show tongue-tip gait change.

2 Methods

We designed our study to align with the methods described in Derrick & Gick (2021). Both the North American English and New Zealand English data in the present paper were collected using similar methods and procedures. We differ here only in the stimuli used, as the stimuli that result in flaps in New Zealand English differ from those in American English. Following the best-fit solutions described in the supplemental materials of Derrick & Gick (2021), the specific measures used in our best-fit model for tongue motion displacement range (used to produce **Figure 4a** and **Table 4**) differ from those used in Derrick & Gick (2021) as they include only angular displacement and not distance. For all of our NAE data, we use the data collected and discussed in Derrick and Gick (2021). For all of our NZE data, the methods are described here.

2.1 Ethics and consent

The University of Canterbury’s Human Research Ethics Committee (HREC) approved ethics for this study (HEC 2012/19). The experiments were performed in accordance with the procedures listed in the HEC 2012/19 document. Each participant provided informed consent before

participating in the experiment. Participants were compensated with \$40 New Zealand Dollars worth of local Westfield mall vouchers.

2.2 Participants

We recorded 12 participants (6 female and 6 male). All participants were native New Zealand English (NZE) speakers. Participants reported normal hearing following the Noble paradigm, where participants are asked about any difficulty hearing, any difficulty following television programs at a socially acceptable volume, and their ability to converse in large groups or noisy environments. These questions form William Noble's 3-question summary of Gatehouse & Noble's (2004) "Speech, Spatial, and Qualities of Hearing Scale", and were intended for non-clinical hearing screening (Noble 2011).

2.3 Materials

Setup included an NDI Wave EMA machine with 100 Hz temporal resolution and 16 five degrees-of-freedom (5D) sensor ports. Setup also included a General Electric Logiq E 2012 ultrasound machine with an 8C-RS wide-band micro-convex array 12 × 22 mm, 4–10 megahertz imaging frequency transducer. Audio was collected using a USB Pre 2 pre-amplifier (Sound Devices, LLC) connected to a Sennheiser MKH-416 short shotgun microphone mounted to a Manfrotto "magic-arm" for directional control. Ultrasound data were captured using an Epiphan VGA2USB Pro frame grabber connected to a MacBook Pro (late-2013) with a solid-state drive. The USB-Pre 2 audio output and NDI wave machine were connected to a Windows 7 desktop computer with NDI's Wavefront control and capture software installed. This setup allows simultaneous ultrasound, EMA, and audio recording of participants. In this study, the ultrasound measurements were used for visual confirmation of tongue movements only.

2.4 Stimuli

We selected five two-word utterances, or token types, with double-flap sequences ('added a'), and embedded them in carrier phrases that have no directly adjacent tongue motion-generating consonants (e.g. 'We have added a book'). All of these token types are structured in a /V(1)rV(2)rV(3)/ frame. The stimuli are all listed in **Table 1**. Stimuli were chosen to allow for a variety of surrounding vowel contexts, while simultaneously keeping the experiment short enough to allow the equipment to work effectively.

The phrase structures ensure speakers place primary stress on the syllable before the first flap, a context in which speakers are most likely to produce flap sequences (Zue & Laferriere 1979).

	Token type in carrier phrase	Token type
1	We have added a book	added a
2	We have bordered a book	bordered a
3	We have murdered a book	murdered a
4	We have ordered a book	ordered a
5	We have worded a book	worded a

Table 1: Experiment stimuli list.

2.5 Setup and procedure

After completing initial screening, each participant was seated in a comfortable chair and heard a detailed description of the experimental procedure. An ultrasound transducer was held in place beneath the chin using a soft, non-metallic stabilizer (Derrick et al. 2015a), allowing participants' tongue movements to be recorded using ultrasound. The ultrasound measurements were used for visual confirmation of tongue movements, but were otherwise not included in the analysis. Five-dimensional (5D) electromagnetic articulometry (EMA) sensors were taped to the skin over the mastoid processes behind the ears and the nasion. Sensors were then taped and glued along the midsagittal line to the upper and lower lips on the skin next to the vermilion border of the lip using Epiglu. One sensor was then glued to the lower incisor, and three to the tongue: One approximately 1 centimeter away from the tongue tip, one at the back — just avoiding the gag reflex — and a middle sensor half-way between the front and back sensor. Tongue sensors were then coated in Ketac, a two-part epoxy cement normally used in dental implants. Both the Epiglu and Ketac are slowly broken down by saliva, allowing about 45–50 minutes of experiment time.

Once sensors were connected, an MKH-416 short shotgun microphone attached to a Manfrotto magic arm was placed on the opposite side of the head from the NDI wave electric field generator. The microphone was far enough away to avoid electro-magnetic interference with the NDI sensors, but close enough to reduce the acoustic interference from the many machine fans used to cool equipment during the recordings. The NDI wave recordings were captured at 100 cycles per second (Hz), and the audio recordings were synchronously captured at 22,050 Hz using 16 bit pulse-code-modulation (a standard .wav file format). Once the setup was complete, participants read 10 blocks each containing the 5 sentences in **Table 1**, at 5 different speech rates, presented on a computer using Psychopy (Pierce 2007).

We induced different speech rates by having participants hear reiterant speech (spoken 'ma ma **ma** ma ma ma', with the stress on the third syllable.) produced at one of five different speech rates (3, 4, 5, 6, or 7 syllables per second) before being asked to read the relevant phrase at the preceding reiterant speech rate. Within each block, sentences and speech rates were randomly

presented. Participants read sentences at the reiterant speech rate as instructed and to the best of their ability. Each example was randomly presented as 25 phrases per block, with 10 blocks in total, such that the entire task took 35 min to complete.

In the event of sensor detachment, the area around the sensor was quickly dried with a paper towel, and the sensor was reattached with Epiglu only, within 1 mm of the original attachment point. No sensor was reattached a second time.

Once the experiment was complete, the participant was asked to hold a protractor between their teeth with the flat end against the corners of the mouth, and three (3) 10-second recordings of the occlusal (bite) plane were recorded. Setup took between 30 and 45 minutes; recording took about 45 minutes; recording of the occlusal plane, palate, and head rotation took no more than 10 min; and removal of sensors took 5 minutes. The entire process was typically completed within under 2 hours.

2.6 Data processing

EMA data were loaded from NDI-wave data files, and smoothed with a discrete cosine transform technique that effectively low-pass-filters the data and restores missing samples using an all-in-one process (Garcia, 2010; 2011). This process was implemented through MVIEW (Tiede 2010). Data were then rotated to an idealized flat (transverse-cut) occlusal plane with the tongue tip facing forward. This was accomplished using the recorded occlusal plane and the recorded planar triangle between the nasion and two mastoid processes, allowing all of the participants' data to be rotated and translated to a common analysis space. Tongue palate traces were generated using the highest tongue sensor positions along the midsagittal plane, after removing extreme outliers. Acoustic recordings were transcribed, isolating the phrases in one transcription tier, the vowel-flap-vowel-flap-vowel sequences under analysis in a second tier, and the two flap contacts in a third tier.

Flap contacts were identified by the acoustic amplitude dip (Zue & Laferriere 1979), or by ear if the flap was approximated enough to not have an amplitude dip (such approximants were rare, accounting for less than 10% of the data). In order to compare different speech rates, the acoustic and vocal tract movement information was subdivided into 31 time slices: Eleven (11) from the onset of the first vowel to the point of lowest acoustic intensity of the first flap, 10 more from that point to the point of the lowest acoustic intensity of the second flap, and from there, 10 more to the end of the following vowel. The entire time span constitutes the duration of each token type. These Procrustean fits allowed comparison of tongue motion and acoustic information at the same relative timing regardless of speech rate. Acoustic cues were chosen because our previous research showed that flaps in English can be categorized in at least four patterns. Two of them, alveolar-taps and post-alveolar taps, involve tongue tip and blade motion

towards the teeth or hard palate, making light contact, and moving away again. Two others, up-flaps and down-flaps, involve the tongue making tangential contact with the teeth or hard palate (Derrick & Gick 2011).

These subphonemic differences mean that it is impossible to identify flap contact through articulatory gesture identification tools such as `FindGest` (Tiede 2010). However, there is almost always a direct and simultaneous relationship between the point of lowest amplitude in the acoustic signal and the timing of tongue to palate/teeth contact during flap production (Zue & Laferriere 1979). This makes acoustic cues the most suitable method of isolating the underlying articulatory motion patterns for this dataset.

2.7 Visualization

Movement data from these Procrustean fits were visualized on millimeter-grid graphs. The graphs show the palate and position traces of the tongue tip, tongue mid, tongue back, lower incisor, upper lip, and lower lip throughout token production for each reiterant speech rate from 3 to 7 syllables/s. These graphs were produced for each participant and token type, with movement traces averaged over all the blocks. Versions of this graph tracing each block separately were used to identify cases where EMA sensors became unglued from participants' tongues, or sensor wires had tiny breakages. These tokens were excluded from analysis. Lastly, visual comparison of the different speech-rate traces revealed a wide variety of tongue motion pattern differences between participants, token types, and speech rates. Testing for sub-hypothesis 1 requires analysis of Angular displacement and speech-rate range, as well as critical fluctuations over the time-course of token production.

2.8 Analysis: Angular displacement and speech-rate range

Tongue motion patterns from either the NAE or NZE dataset cannot be properly compared using ordinary statistical methods. Most statistical tests involve comparisons of lines or curves corresponding to each speaker, finding an average line or shape for each group and comparing them based on how much lines in each group vary from that average. However, each speaker, token type, and speech rate could and sometimes did have wildly varying patterns that do not conform to any of the typically describable statistical distribution patterns. As a result, the basic mathematical assumptions underlying most methods of statistical analysis were not met.

Instead, what we did is build a comparison of actual speech rates, measured by the auditory duration of the recorded token. We grouped those durations by speaker, token, and reiterant speech rate, giving us point data for each of these groups. These points correspond to the filled-in circles in **Figure 4**, and were placed on the y-axis. We also computed the angle of the tongue tip position based on changes in tongue tip position throughout the production of each token. We averaged the sum for each speaker and word produced following the slowest reiterant speech rate (3 syllables/second), and subtracted the average sum for each speaker and word produced

following the fastest reiterant speech rate (7 syllables/second). This gave us a cumulative angular displacement range for each speaker and word. We place each of the speech rate dots along the x-axis based on this measure. In this way, each speaker and word has 5 different y-axis positions, one for each reiterant speech rate, but only 1 x-axis position. This measure provided a uniform way of comparing tongue motion paths that did not otherwise conform to normally expected patterns of statistical distribution. Detailed explanations of how to implement these comparison algorithms can be found in Derrick & Gick (2021).

Note that we ran initial tests comparing versions with this angular displacement z-scored and summed with a z-score of the motion distance of the tongue tip sensor. We also included tongue mid, tongue back, and lip sensors, and ran comparisons of models carefully removing sensors and distance measurements until we found a version that produced a statistical comparison model that accounted for the most data. This process of comparison is called a backwards iterative model fit (back-fit), and the best version is slightly different for North American English, which included tongue mid and tongue tip cumulative displacement (angular displacement plus distance), whereas for New Zealand English, the best model only included tongue tip data, and only included cumulative angular displacement. The process of back-fitting is a standard and well-known statistical method for finding optimal models. The output of the final best-fit NZE model is shown in Formula 1. This analysis allows comparison of recorded speech rate range and tongue tip angular displacement range so that individual tongue traces could be visually compared for those speakers and words with the least angular displacement difference and those with the most. This is a way of visually identifying presence or absence of different tongue gaits for slow and fast speech, and the results can be seen in **Figure 5**.

2.9 Analysis: Critical fluctuations

As noted in Derrick & Gick (2021), different patterns of tongue motion do not conclusively demonstrate different tongue gaits. To do that, evidence that the two patterns are both stable and commonly used is required. There are many ways of obtaining that information, but only one of the standard ways was available from this research paradigm, that is, to identify how much effort, as measured through critical fluctuation data, was produced throughout the time course of each speech utterance. Higher effort at the beginning of a sequence of complex motion compared to the end is a measure of “end-state comfort”, recognized as evidence of motion planning (see Rosenbaum et al. 1992; 1996). In contrast, more effort at the end, or “beginning-state-comfort” indicates less preparation, requiring more effort towards the end of a complex sequence.

We measured this effort using the formula in Schiepek & Strunk (2010), originally formulated to calculate the likelihood of mental breakdowns amongst psychiatric patients; it is a measure of effort that uses information from velocity, acceleration and jerk from short sequences of measured data. The results of this formula were placed into a generalized additive mixed-effects model that

produced a three-dimensional surface that identifies regions of greater and lesser effort over the time course of utterances based on the tongue-tip angular displacement range. The results of this comparison can be seen in **Figure 6**. Details of how to replicate this formula and analysis process can be seen in Derrick & Gick (2021).

2.10 Analysis: Tongue tip position comparisons for V2

Tongue tip comparisons for testing sub-hypotheses 2–4 require comparing tongue-tip height and tongue-tip frontness for the middle vowel between:

- 1) NAE rhotic vowels (hereafter *NAE VRV*): (“We have Saturday books”, “We have bettered a book”).
- 2) NAE non-rhotic vowels bounded by rhotic vowels (hereafter *NAE RVR*): (“We have herded her books”, “We have worded her books”)
- 3) NAE non-rhotic vowels followed a rhotic vowel (hereafter *NAE VVR*): (“We have editor books”, “We have auditor books”)
- 4) NAE non-rhotic vowels with no nearby rhotic vowels (hereafter *NAE VVV*): (“We may edit a book”, “We may audit a book”)
- 5) NZE non-rhotic vowels (hereafter *NZE VVV*): (“added a”, “bordered a”, “murdered a”, “ordered a”, “worded a”)

These comparisons were completed using generalized mixed-effects linear models comparing the vowel groups, reiterant speech rate, and the interaction between both. All P-values in the models are based on Wald z-scores. Optimal model fit was based on the `buildmer` (Voeten 2023) function in R (R Core Team 2023). following a forward and then backward iterative model fitting, with model fits fitted using the `bobyqa` optimizer (Bartoń 2023).

2.11 Open access

The EMA data for our NZE and NAE English data, as well as statistical tests and code for producing images for this paper can be found at the Open Science Foundation at https://osf.io/n65t8/?view_only=e5712e0862994b9b81f6e7b2505bb77e. These include the statistical tests used to decide the best-fit model that produced **Table 4** and **Figure 4a**.

3 Results

Figure 1 shows the average tongue tip positions for the second vowel in $/VrVrV/$, with each vowel group at each speech rate, and so descriptively addressed predictions 1a and 1b. The NAE VRV group’s tongue tip is higher and further back than is seen from any of the other vowel groups. The NAE VVV group has the lowest tongue tip positions. NAE VVR group and the NAE

RVR group are both similar, having higher and further back tongue-tip positions as compared to NAE VVV group. NZE VVV are all further front than any of the NAE vowels, and mid height between the low NAE VVV and the high NAE VRV group.

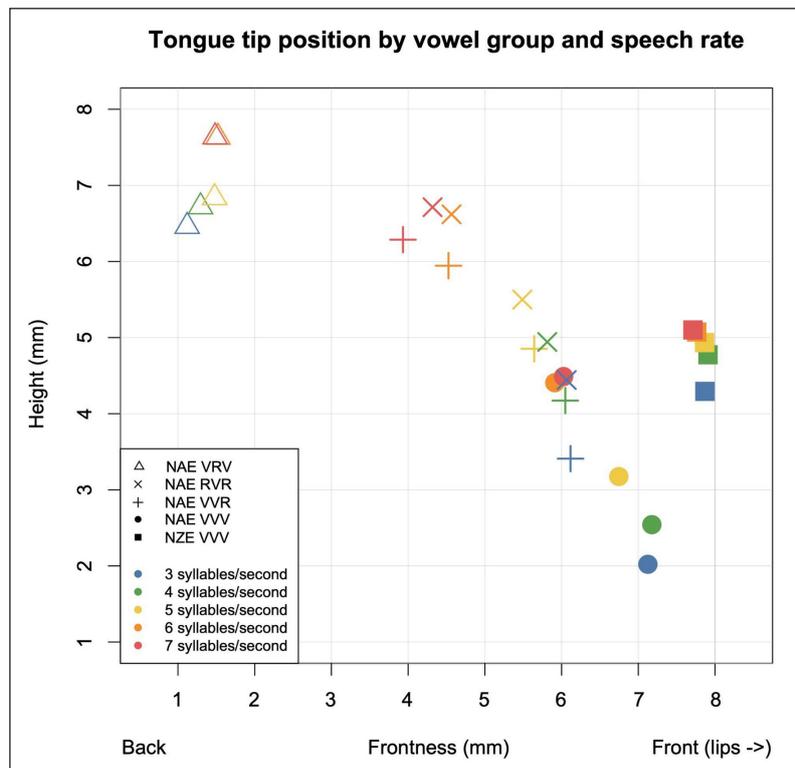


Figure 1: Tongue tip position by vowel group. Graph presented to show the average (mean) physical scale of the events under analysis.

The contents of predictions 1–3 were also tested using statistical models that compare the second vowel in /VrVrV/ sequences. To test predictions 1a and 1b, the model needed to compare vowel types: 1) NAE VRV, 2) NAE RVR, 3) NAE VVR, 4) NAE VVV, and 5) NZE VVV. To test prediction 2, the model needed to test speech rate, as measured in the syllables-per-second used in the reiterant speech rate cue (3, 4, 5, 6 and 7 syllables/second). To test prediction 3a, the model needed to compare NZE and NAE. Prediction 3b was tested as a by-product of the test for prediction 2 as the specific speech rates are isolated in the same way for both languages and all 4 vowel groups.

The optimal formula for tongue tip frontness, as uncovered from *buildmer* (Voeten 2023), is shown in Formula 1:

Formula 1: Tongue tip frontness $\sim 1 + \text{vowel group} + \text{syllables-per-second} + \text{vowel group} \cdot \text{syllables-per-second} + (1 \mid \text{subject})$

Where *vowel group* includes 1) NAE VRV, 2) NAE RVR, 3) NAE VVR, 4) NAE VVV, and 5) NZE VVV.

The results are shown in **Table 2**. The results show a significant main effect difference between NZE vowels and NAE rhotics, as well as an interaction between vowel type and reiterant speech rate (SPS) for every vowel group compared to New Zealand English.

	Estimate	Std. Error	t-value	p-value
(Intercept)	0.389	0.268	1.451	0.147
NAE VRV	-1.324	0.388	-3.409	*** <0.001
NAE RVR	-0.353	0.388	-0.910	0.363
NAE VVR	-0.344	0.388	-0.885	0.376
NAE VVV	-0.148	0.388	-0.382	0.703
SPS 4	0.008	0.022	0.356	0.722
SPS 5	-0.001	0.022	-0.036	0.971
SPS 6	-0.021	0.022	-0.940	0.347
SPS 7	-0.031	0.022	-1.368	0.171
NAE VVV : SPS 4	0.005	0.043	0.110	0.913
NAE VVV : SPS 5	-0.071	0.043	-1.634	0.102
NAE VVV : SPS 6	-0.217	0.043	-4.999	*** <0.001
NAE VVV : SPS 7	-0.185	0.043	-4.255	*** <0.001
NAE RVR : SPS 4	-0.061	0.043	-1.398	0.162
NAE RVR : SPS 5	-0.113	0.043	-2.597	** 0.009
NAE RVR : SPS 6	-0.275	0.043	-6.332	*** <0.001
NAE RVR : SPS 7	-0.310	0.043	-7.138	*** <0.001
NAE VVR : SPS 4	-0.024	0.043	-0.561	0.575
NAE VVR : SPS 5	-0.092	0.043	-2.124	* 0.034
NAE VVR : SPS 6	-0.295	0.043	-6.797	*** <0.001
NAE VVR : SPS 7	-0.396	0.043	-9.137	*** <0.001
NAE VRV : SPS 4	0.026	0.043	0.609	0.542
NAE VRV : SPS 5	0.071	0.043	1.64	0.100
NAE VRV : SPS 6	0.100	0.043	2.304	* 0.021
NAE VRV : SPS 7	0.102	0.043	2.356	* 0.018

Table 2: Results for GLMM model fit for Formula 1. * $\alpha = 0.05$, ** $\alpha = 0.01$, *** $\alpha = 0.001$.

The results show for tongue frontness that the tongue is significantly farther back at faster speech rates for all the NAE non-rhotic vowel groups, regardless of rhotic vowel context. These results are highlighted based on alpha level in **Figure 2**. **Figure 2** zeros all of the tongue backness results based on the slowest reiterant speech rate for each vowel group. This makes it easier to see how far the tongue backness diverges between slow and fast speech, and visualizes how much divergence is required for the difference to be statistically significant.

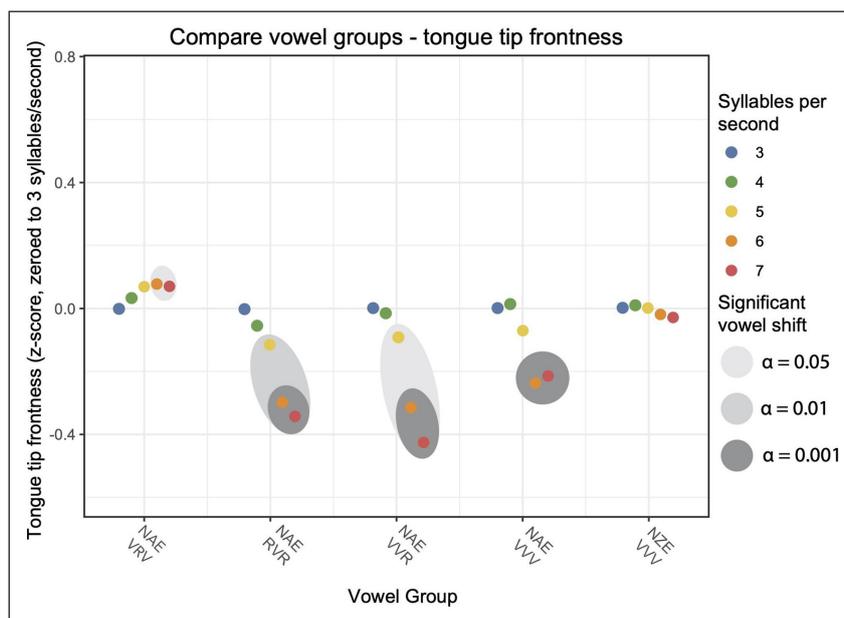


Figure 2: Tongue tip frontness (z-score) by vowel group and elicited speech rate. Data zeroed to 3 syllables/second for illustrative purposes of making regions of significance easier to understand. Significant differences in tongue tip frontness between the NAE vowel groups and NZE vowels are highlighted in gray, following the significance levels highlighted in the listed alpha groups.

The optimal formula for tongue tip height as uncovered from `buildmer` (Voeten 2023) is shown in Formula 2:

Formula 2: Tongue tip height $\sim 1 + \text{vowel group} + \text{syllables-per-second} + \text{vowel group:syllables-per-second} + (1 + \text{vowel group} \mid \text{subject})$

Formula 2 differs from Formula 1 in that the random effect is more complex, allowing the model to accurately factor out more differences between participants. The results are shown in **Table 3**. The results show a significant main effect difference between NZE non-rhotics and NAE rhotics, as well as an interaction between vowel type and reiterant speech rate (SPS) for every vowel group compared to New Zealand English.

	Estimate	Std. error	t-value	p-value
(Intercept)	-0.206	0.226	-0.909	0.363
NAE VRV	0.592	0.342	1.73	0.083
NAE RVR	0.037	0.334	0.112	0.911
NAE VVR	-0.239	0.366	-0.655	0.512
NAE VVV	-0.63	0.363	-1.737	0.082
SPS 4	0.132	0.027	4.86	*** < 0.001
SPS 5	0.175	0.027	6.477	*** < 0.001
SPS 6	0.213	0.027	7.878	*** < 0.001
SPS 7	0.221	0.027	8.156	*** < 0.001
NAE VVV : SPS 4	0.012	0.053	0.224	0.823
NAE VVV : SPS 5	0.141	0.053	2.682	** 0.007
NAE VVV : SPS 6	0.439	0.053	8.356	*** < 0.001
NAE VVV : SPS 7	0.453	0.053	8.601	*** < 0.001
NAE RVR : SPS 4	0.004	0.053	0.079	0.937
NAE RVR : SPS 5	0.114	0.052	2.166	* 0.03
NAE RVR : SPS 6	0.382	0.053	7.267	*** < 0.001
NAE RVR : SPS 7	0.404	0.053	7.682	*** < 0.001
NAE VVR : SPS 4	0.076	0.053	1.447	0.148
NAE VVR : SPS 5	0.219	0.052	4.181	*** < 0.001
NAE VVR : SPS 6	0.477	0.053	9.083	*** < 0.001
NAE VVR : SPS 7	0.568	0.053	10.804	*** < 0.001
NAE RVR : SPS 4	-0.061	0.052	-1.168	0.243
NAE RVR : SPS 5	-0.072	0.052	-1.374	0.169
NAE RVR : SPS 6	0.107	0.052	2.038	* 0.042
NAE RVR : SPS 7	0.101	0.052	1.917	. 0.055

Table 3: Results for GLMM model fit for Formula 2. * $\alpha = 0.05$, ** $\alpha = 0.01$, *** $\alpha = 0.001$.

The results show that the tongue is significantly higher at faster speech rates for all NAE vowels, regardless of rhotic vowel context. These results are highlighted based on degree of significance in **Figure 3**. Like **Figure 2**, **Figure 3** zeros all of the tongue backness results based on the slowest reiterant speech rate for each vowel group. This makes it easier to see how far the tongue height diverges between slow and fast speech, and visualize how much divergence is required for the difference to be statistically significant.

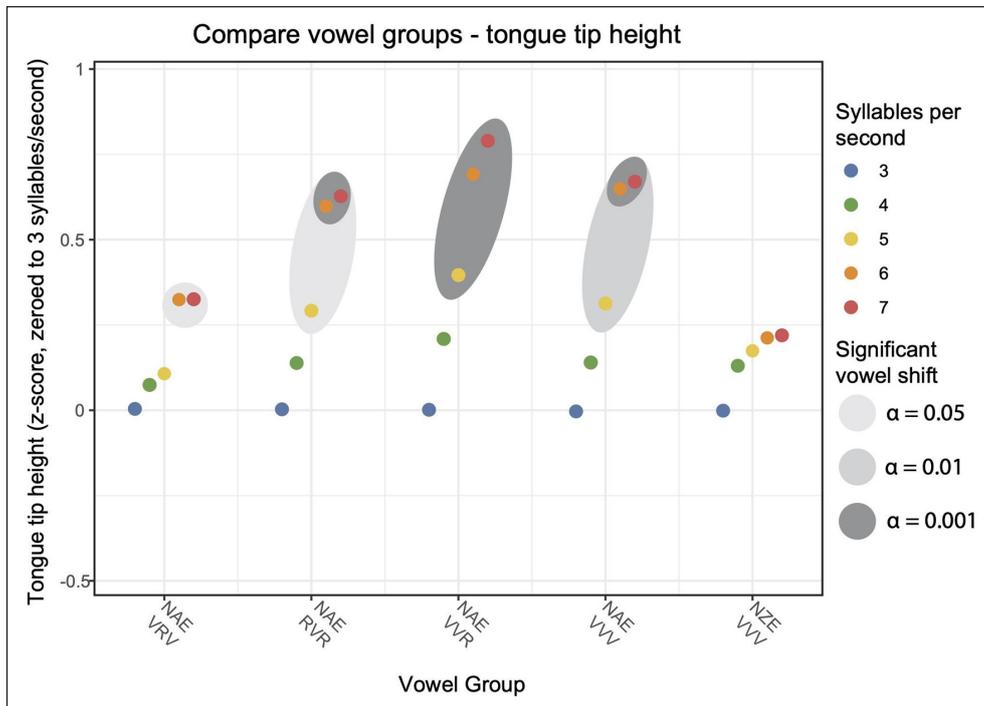


Figure 3: Tongue tip height (z-score) by vowel group and elicited speech rate. Data zeroed to 3 syllables/second for illustrative purposes of making regions of significance easier to understand. Significant differences in tongue tip frontness between the NAE vowel groups and NZE vowels are highlighted in gray, following the significance levels highlighted in the listed alpha groups.

3.1 Prediction 4: Gait change

Our comparison of token duration by subject and token type (y-axis) and cumulative angular displacement range (x-axis) are shown in **Figure 4**. The lines on the graph were generated from the results of running a generalized linear mixed-effects model seen in Formula 3:

$$\text{Formula 3: Token duration} \sim \text{cumulative angular displacement} * \text{syllables per second} + (1 + \text{cumulative angular displacement} \mid \text{subject})$$

This was the best-fit model that converged, and the results of the model fit are reported in **Table 4**.

Figure 4a shows that some NZE speakers have a very narrow realized speech-rate range. The speakers with the narrowest tongue tip cumulative angular displacement ranges always spoke fast, at between 0.3 and 0.4 tokens per second. At the other extreme, speakers with the widest tongue tip angular displacement ranges spoke from 0.3 to 0.8 tokens per second depending on the reiterant speech with which they were prompted.

	Estimate	Std. err.	df	t value	p-value
(Intercept)	0.647	0.0114	17.0	57.0	*** <0.001
Cumulative angular displacement	0.0642	0.00779	23.1	8.24	*** <0.001
Syllables per second (4)	-0.117	0.00698	289	-16.7	*** <0.001
Syllables per second (5)	-0.182	0.00698	289	-26.1	*** <0.001
Syllables per second (6)	-0.263	0.00698	289	-37.7	*** <0.001
Syllables per second (7)	-0.284	0.00698	289	-40.7	*** <0.001
Angular displacement: Syllables per second (4)	-0.0338	0.00699	289	-4.83	*** <0.001
Angular displacement: Syllables per second (5)	-0.0457	0.00699	289	-6.54	*** <0.001
Angular displacement: Syllables per second (6)	-0.0740	0.00699	289	-10.6	*** <0.001
Angular displacement: Syllables per second (7)	-0.0785	0.00699	289	-11.2	*** <0.001

Table 4: Results of generalized linear mixed-effects model (GLMM) model fit in formula 1. Std. err. = standard error, df = degrees of freedom. * $\alpha = 0.05$, ** $\alpha = 0.01$, *** $\alpha = 0.001$.

In comparison, **Figure 4b** is a reproduction of **Figure 5** from Derrick & Gick (2021). This figure shows NAE has a similarly wide speech-rate range to NZE. There is, however, less clustering of speech-rate ranges around the more narrow (left-hand side) of **Figure 4b**. That is, more NAE speakers have wider displacement and speech-rate ranges. While this comparison is not intended to show a statistically significant comparison between **Figure 4a** and **4b**, it does make the comparison of tongue tip movements in **Figure 5a/b** with those in **Figure 5c/d** easier to understand visually.

Figure 5 compares tongue-tip motions in response to the slowest (3 syllables/second) and fastest reiterant speech rate (7 syllables/second) for NZE (5a and 5b) and NAE (5c and 5d) speakers. For each language, 10 speakers/words with the narrowest displacement ranges are shown on the left, and the 10 speakers/words with the widest displacement ranges are on the right.

For NZE, there is very little difference between the two groups, a result that contrasts starkly to the wide range and variety of differences one can easily see in **Figure 5c** vs. **Figure 5d** (see Derrick & Gick 2021).

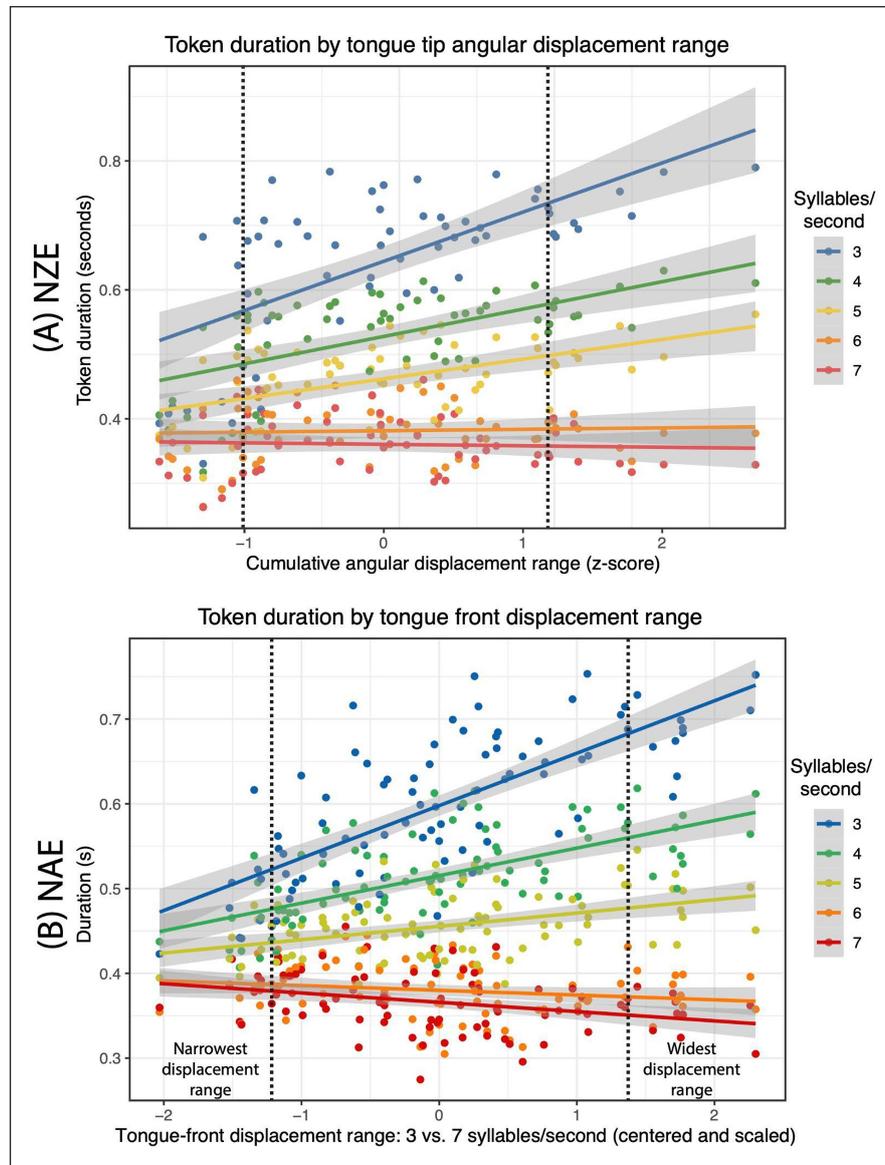


Figure 4a: Comparison of tongue-tip angular displacement range and token duration (comparing the responses to reiterant speech at 3 syllables/s and 7 syllables/s for all participants/token types) and token duration for NZE. Values for the 10 participants/token types with the narrowest tongue-tip angular displacement ranges are on the left side of the leftmost dashed-black line, and their tongue-tip motions are highlighted in Figure 5a. Values for the 10 participants/token types with the widest tongue-tip displacement ranges are on the right side of the rightmost dashed-black line, and their tongue-tip motions are highlighted in Figure 5b. **Figure 4b:** Similar to 4a but using tongue-front (tip and blade) displacement range. Tongue tip motion from the narrowest 10 displacement ranges are shown in Figure 5c, and the widest 10 are shown in Figure 5d (reproduced from Figure 5, Derrick & Gick 2021).

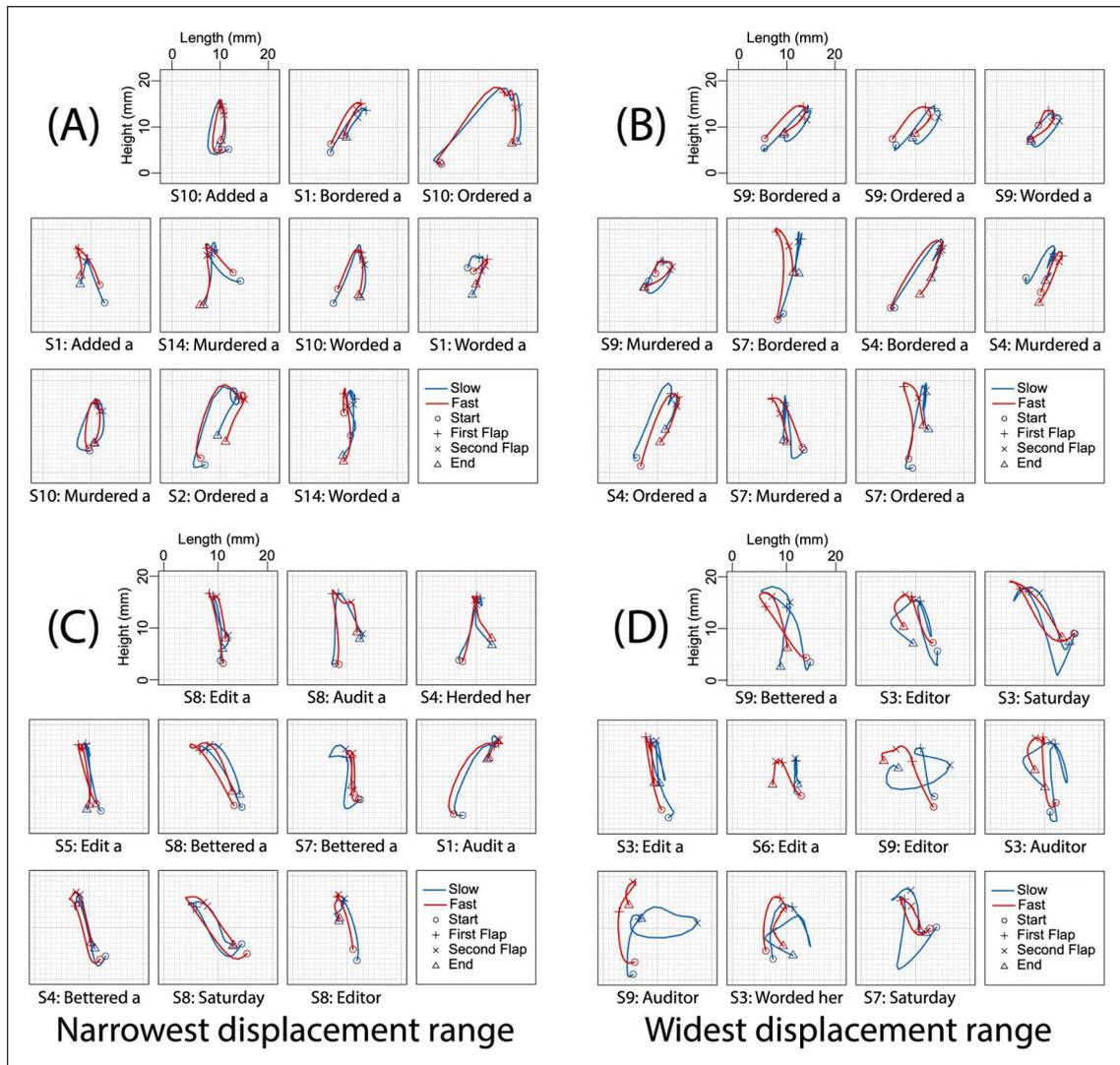


Figure 5: Averaged tongue tip motion patterns comparing the 10 participants/token types with the narrowest tongue-tip displacement ranges for NZE (a – top left) and NAE (c – bottom left) with the 10 participants/token types with the widest tongue-tip displacement ranges for NZE (b – top right), and NAE (d – bottom right), as identified in Figure 4a for NZE and Figure 4b for NAE. Each grid-box shows tongue motion in response to reiterant speech at 3 syllables/s in blue (labeled slow in the legend), and 7 syllables/s in red (labeled fast in the legend). (Figure 5c and 5d reproduced from Figures 6a and 6b, Derrick & Gick 2021).

To confirm whether there was no difference in gait between slow and fast speech for any of the participants for NZE, we ran a GAMMS comparing critical fluctuations during the time course of token production (**Figure 6** x-axis) against cumulative angular displacement ranges (**Figure 6** y-axis). The GAMMS model used is shown in Formula 2.

Formula 4: Critical fluctuation \sim te(time slice, angular displacement) + s(time slice, angular displacement, subject, bs = “fs”, m = 1) + s(syllables per second, subject, bs = “re”) + s(token type, subject, bs = “re”)

Where “te(time slice, angular displacement)” stands for the tensor, which is a 3-d surface showing the time slice (time course of speech production) on the x-axis, the cumulative angular displacement on the y-axis, and the degree of critical fluctuation in the orange-blue diverging gradient, with dark orange having the highest critical fluctuation, and dark blue having the least critical fluctuation. The “s(time slice, angular displacement, subject, bs = “fs”, m = 1)” component contains the random effects surfaces based on each participant, “s(syllables per second, subject, bs = “re”)” contains the random effects for each reiterant speech rate, and “s(token type, subject, bs = “re”)” contains the random effects for each token type. The results of the GAMMS model from Formula 2 are shown in **Table 5**.

	edf	ref.df	F	p-value
(intercept)	0.0635	0.00434	14.6	*** <0.001
te(time slice, angular displacement)	15.5	17.5	3.45	*** <0.001
s(time slice, angular displacement, subject)	194	357	11.1	*** <0.001
s(syllables per second, subject)	48.8	59.0	11.9	*** <0.001
s(token type, subject)	36.4	58.0	2.62	*** <0.001

Table 5: Results for Generalized additive mixed-effects (GAMM) model fit for Formula 2. R-sq. (adj) = 0.708, Deviance explained = 72%, fREML = -17842, Scale est. = 0.0006277, n = 6900. * $\alpha = 0.05$, ** $\alpha = 0.01$, *** $\alpha = 0.001$.

The results show each component is strongly statistically significant. The results are graphed in **Figure 6**. The figure shows end-state-comfort effects for all cumulative angular displacement ranges, with beginning-state effort significant for the narrowest angular displacement ranges.

In comparison, in **Figure 7** (reproduced from Figure 7, Derrick & Gick 2021), we see evidence of two gaits in fast NAE speech. In this case, the narrowest and widest displacement ranges both have the highest degrees of critical fluctuation as sequence onset, indicating most of the productions involved well-planned motion sequences, whereas the middle section had the highest degree of critical fluctuation at the second flap near the end of the sequence, indicating less well-planned motion sequences.

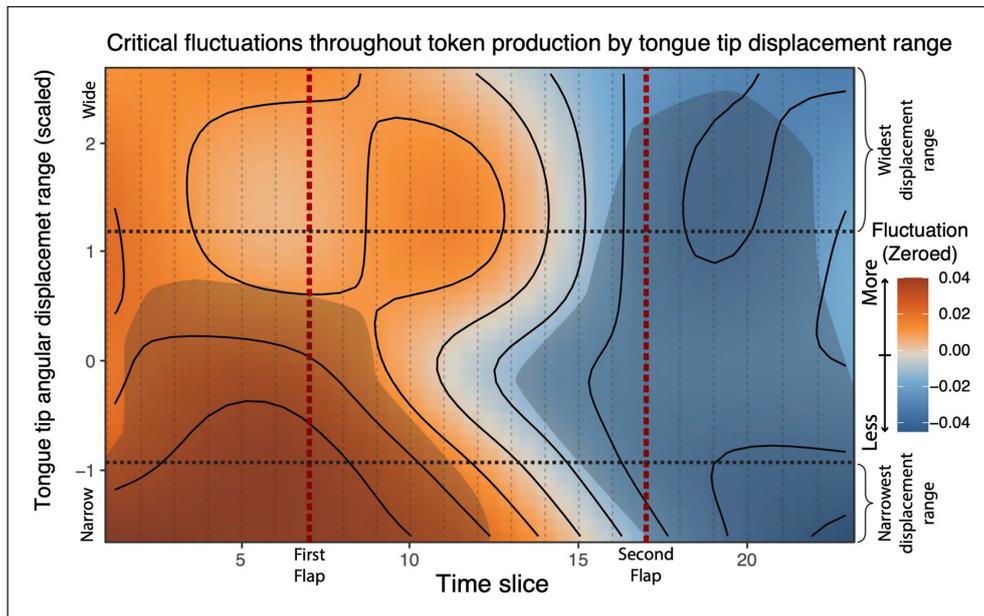


Figure 6: Comparison of critical fluctuations throughout the time course of token production in our NZE data, comparing speakers producing token types based on tongue-front displacement range. The x-axis is divided into fluctuation (F) time slices, which represent the time-course of token production over the contents of 7 Procrustean time slices. The y-axis is the scaled tongue-front displacement range for each speaker and token type. The z-dimension, shown as orange-blue diverging gradient, includes fluctuation data for each fluctuation time slice. The two red dashed lines show the time slices centered around the first and second flap, referencing Figure 4. The lower dividing line shows speakers producing token types with the narrowest tongue-tip angular displacement range, corresponding to the motion patterns shown in Figure 5a. The upper dividing line shows speakers producing token types with the widest tongue-tip angular displacement range, corresponding to the motion patterns shown in Figure 5b.

Taken together, these results from NZE show consistent end-state comfort, indicating one stable pattern of motion across speech rates. This contrasts with the NAE data that shows consistent end-state comfort only in the fastest and slowest speech, indicating one stable pattern of motion in slow speech, and two in fast speech.

4 Discussion

The results support prediction 1a: For V2 in an NAE /VrVrV/ sequences, rhotic vowels have a less fronted and somewhat higher tongue-tip position compared to NAE non-rhotic vowels. The results also support prediction 1b: For V2 in NAE /VrVrV/ sequences, non-rhotic vowels have less fronted and higher tongue-tip position when they are followed by or are bounded by rhotic vowels, as can be seen in **Figure 1**. Note also that NZE VVV were further front and higher than NAE VVV.

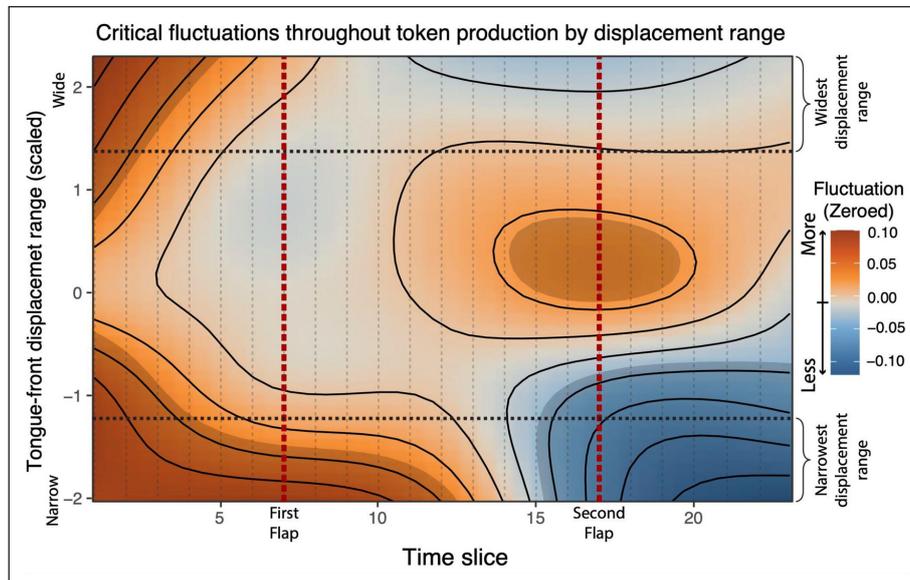


Figure 7: Comparison of critical fluctuations throughout the time course of token production in our NAE data, comparing speakers producing token types based on tongue-front displacement range. The x-axis is divided into fluctuation (F) time slices, which represent the time-course of token production over the contents of 7 Procrustean time slices. The y-axis is the scaled tongue-front displacement range for each speaker and token type. The z-dimension, shown as orange-blue diverging gradient, includes Fluctuation data for each fluctuation time slice. The two red dashed lines show the time slices centered around the first and second flap, referencing Figure 6. The lower dividing line shows speakers producing token types with the narrowest tongue-front displacement range, corresponding to the motion patterns shown in Figure 5c. The upper dividing line shows speakers producing token types with the widest tongue-front displacement range, corresponding to the motion patterns shown in Figure 5d. (reproduced from Figure 7, Derrick & Gick 2021).

The results strongly support prediction 2: **Figures 1 and 2** and **Tables 2 and 3** show that NAE VVV are all higher and further back at faster speech rates. Note that the differences for all the NAE non-rhotic vowel groups were greater than they were for rhotic vowels, as shown in **Figures 1–3**. This is likely true because the two token types used for rhotic vowels (“Saturday” and “bettered a”) are both $/Vr\partial rV/$ sequences. Derrick & Gick (2015b) showed that these sequences exhibited greater stability of flap production than typically observed in Derrick & Gick (2011) because the interaction of flap motion direction, tongue elasticity, and gravity allow the production of this sequence with but one motor action of tongue tip motion spanning over the entire $/r\partial r/$ sequence. This tongue front gesture produces an up-flap—tip-up $/\partial/$ —down-flap sequence. Even so, the data clearly show that this sequence is still produced differentially based on speech rate – the distinction in tongue tip height patterns, in a smaller but statistically significant way, with the other NAE vowels. That is, the tongue tip is higher for the faster speech rates. Taken together,

the supports for predictions 1a, 1b, and 2 form the backdrop of our reanalysis of NAE results needed to assess predictions 3 and 4.

The results of this research also strongly support prediction 3a (Uniformity), but not 3b (Mechanical): **Figures 2 and 3** and **Tables 2 and 3** show that speech rate had no significant effect on the tongue tip position for NZE vowels, whereas there was a significant difference in tongue tip position for all of the NAE non-rhotic vowel groups. The support for prediction 3a and not 3b support the uniformity hypothesis over the mechanical hypothesis.

Prediction 3c is visually supported in **Figures 2 and 3**. However, we note that the two gaits split at slightly different speech rates for different vowel groups: significant differences in tongue tip height/backness occurred between speech rates of 3–4 and 5–7 syllables/second for some groups, and between 3–5 and 6–7 syllables/second for others. This suggests that there are likely additional variables not measured in this study that contribute to the relationship between sequence, speech rate, and mechanical advantage.

Prediction 4a (Uniformity) was also supported, whereas prediction 4b (Mechanical) was not: Derrick & Gick (2021) demonstrated that NAE /VrVrV/ sequences may exhibit a gait-change-like differences between the tongue front motion patterns for slow and fast speech; this gait change even occurred in /VrVrV/ sequences with no rhotic vowels (specifically S3 and S6's "edit a", as shown here in **Figure 5d**). These results are reflected in differences in tongue-tip frontness and height and different speech rates seen for all NAE non-rhotic groups in the V2 (/V1rV2rV3/) position. However, such gait-change was not observed for NZE, as shown in **Figures 4–6** and **Tables 4–5**. This result strongly supports the Uniformity hypothesis over the Mechanical hypothesis.

The results show that the extreme variability previously observed for NAE rhotics (Mielke et al. 2010; 2016; Derrick & Gick 2011; 2014; Derrick et al. 2015b) transfers to flaps, as evidenced by the appearance of extreme variability in non-rhotic vowels adjacent to NAE flaps. While we have previously observed such variability in NAE non-rhotic vowels (Derrick & Gick 2014; 2021), here we see it more clearly in relationship to speech rate: Higher/farther back tongue tip position for fast speech, and lower/farther front tongue tip position for slow speech. This pattern is not an exact tongue-tip positioning overlap with NAE rhotic vowels. The NAE non-rhotic groups still have on average much more fronted tongue tip positions than NAE VRV, as seen in **Figure 1**.

In contrast, NZE, which has rhotic consonants with extreme variability (Heyne et al. 2018) similar to that seen in NAE rhotic consonants (Mielke et al., 2010; 2016), does not transfer that variability to flaps, or in turn to flap-adjacent vowels. Instead, the NZE speakers have a significantly smaller shift in tongue tip position based on speech rate compared to the NAE speakers, suggesting that the extreme variability may be context-dependent based on syllable position and possibly other higher-order phonetic structures.

4.1 Speaker and situation-specific exceptions

Researchers have found speakers may use multiple articulatory solutions to solve the same speech problem. This variability is, to be sure, quite constrained: Derrick et al. (2015b) found gravity can stabilize production patterns for what would otherwise be more variable sequences such as those found in the English word “Saturday”. Derrick & Gick (2014) also found similar stabilization effects likely stemming from end-state-comfort in sequences such as “editor” and “edit a”. Nevertheless, these papers show that rhotic and flap production variability occurs not just between speakers, but there remains several instances of stochastic variability within speakers. In speech without as many such constraints, Derrick & Gick (2011) showed that subphonemic categorical variation can occur in repetitions of the same word at roughly the same speech rate simply because of unknown changes to the speaker over the period of an experiment, which might include fatigue, recent speech errors, or other unknown influences.

In addition, Tiede et al. (2010) also found that with rhotics, perturbations of the tongue tended to make /ɹ/ more retroflexed, and Harandi et al. (2017) noted that when modeling even simpler speech that involves moving the bulk of the tongue forward (/ə-gis/) as compared to backward (/ə-suk/), individual speaker vocal tract morphology should be taken into account. Taken together, these types of studies show that while speech is regularly constrained by specific speech production conditions and needs that can foster the generation of uniformities, the speech production system also encounters other speaker-specific and situation-specific conditions that lead to speakers using different sequences of motion to resolve otherwise similar speech production problems. So, in the real-world of speech production, as opposed to just in laboratory experiments, while we would expect that speakers conform to uniformities as we see in the data from this experiment, such outcomes will not be universally true.

4.2 The acoustics of flap/tap

While this paper does not focus on the acoustics of flaps and taps in either variety of English because of the distortions that are caused from EMA recordings, it is worth noting that this is a suitable subject for future research. We know from listening to our data that flaps were often produced as either partially or fully devoiced flaps or sometimes even stops in slower speech, and as partial or full approximants in fast speech such that the only way to detect the flap center was through careful listening rather than through spectrographic or waveform analysis. Our anecdotal observations match with the rigorous results of Warner & Tucker (2011). Warner et al. (2009) also showed that listeners are quite good at identifying reduced approximant and even vowel-like flaps as flaps. In addition, Warner & Tucker (2017) found a relationship between F4 drop and flap-rhotic adjacency, which produced a greater drop than flap-non-rhotic adjacency. While beyond the scope of this paper, it would be possible to run comparisons of our NAE and

NZE data to see if there are differences in acoustic speech-rate and F4 effects between the two languages, and to identify whether perceivers are as good at identifying reduced NZE flaps as they are at identifying reduced NAE flaps.

4.3 Timing and potential L2 effects

Other avenues of future research include the possibility of alternative explanations for our results: NZE has been described as a dialect of English that has been tending away from stress-to syllable-timed as it has contacted Māori (Warren 1999). Nokes & Hay's research also shows NZE has diachronically transitioned into a variety of English that had an overall faster speech rate, from 4.5 syl/sec for NZE speakers born in 1860, to about 5.3 syl/sec for NZE speakers born from the 1960s onward (Nokes & Hay 2012). At the same time, NZE had a reduction of vocalic normalized Pairwise Variability Index (PVI), dropping from 68 to 64 over the same century, indicating a shift away from a stress towards a syllable-timed dialect of English (Nokes & Hay 2012). This result shows that "stressed and unstressed vowels are less differentiated by duration in modern NZE" (Nokes & Hay 2012). In addition, NZE speakers produce more peripheral vowels in unstressed position than speakers of British English (Warren 1999; Hay et al. 2008). This tendency toward peripheral unstressed vowels actually increases at higher speech rates for NZE (Warren 1999). This change from stress to syllable timing reduces the opportunity for speakers to vary production around stress for slow as compared to fast speech. This reduction in potential variation may itself explain the lack of gait-change in our NZE data, as seen in **Figure 7**.

This transition to syllable-timing and a lack of reduction of unstressed vowels has been attributed to contact with Māori, but is also directly correlated with the length of time since NZE has been its own non-rhotic variety of English. However, Nokes and Hay (2012) has contradicted a Māori influence-based analysis, with researchers arguing that a merger between KIT and schwa vowels has reduced the distinguishability of stressed and unstressed vowels (Nokes & Hay 2012). Langstrof (2006) also found that during the intermediate period of NZE, distinction in allophonic KIT duration based on following consonant voicing disappeared. In addition, Maclagan and Hay (2007) found that the DRESS vowel shortened as it raised, removing distinction between stressed and unstressed duration. So, disambiguating between a possibly Māori-adjacency influence reducing unstressed vowel variability, our own hypothesis of the lack of rhotic vowel reducing unstressed vowel variability, and a reduction in stressed vowel variability giving the illusion of reduction in unstressed vowel variability would be useful. We propose a few methods:

NZE Southern accent famously has a rhotic NURSE vowel, but does not have the other rhotic vowel variants. The Southland dialect also developed far away from most Māori language influence. Therefore, Southland NZE may also have different speech rates and vowel reduction in unstressed position than the rest of NZE. It might be possible that the Southland variant

of NZE would allow speakers to have gait-change between slow and fast /VrVrV/ sequences. However, the limited cases of such rhotic vowels might also involve a limited variability in production, which might then fail to allow speakers to produce different gaits for fast and slow /VrVrV/ sequences. There are also many British English (BE) non-rhotic dialects, and it may be possible to find gait-change in those dialects, and that may be the case because they are not as thoroughly influenced by a fully mora/syllable timed language the way NZE is by Māori. Future study and comparison of Southland NZE and BE's PVI, speech rate, and /VrVrV/ sequences may disambiguate between these three possible influences on the lack of gait-change in NZE between slow and fast speech. Lastly, following Nokes and Kay (2012), vowel measurements must be carefully designed, and following Liu and Takeda (2021), speech rate must be taken into account in any such study. This care is required to accurately distinguish mora, syllable, and stress-timing.

While we note that L2-induced changes that are moving some English dialects away from stress timing (Liu & Takeda 2021) could be a factor in reducing extreme variability (Nokes & Hay 2012), there are many reasons to believe that L2 interactions would often instead increase, rather than decrease, extreme variability. One example is the ongoing emergence of rhotic vowels in Canadian French rhotics (Mielke 2015; Lipari 2023). While Mielke (2015) shows that none of these rhotics is as extreme in F3 lowering as NAE rhotics, there is an ongoing change resulting in both bunched and retroflex varieties clearly shown in Figure 7 from that paper. This change might be coming from bilingual speakers of NAE rhotic vowels (Mielke 2015) or from English loanwords without any impact on speaker status (Lipari 2023), that is, as a “change from below.” In either case, there is a definite influence from the extreme variability seen in NAE rhotics.

4.4 From the subphonemic to phonemic – implications for language change

Kirparsky (1965) argues that phonological changes all stem from sound changes or imperfect learning. Highly variable speech acts therefore all provide rich opportunities for such sound changes. Our data especially so since they show how extreme variability in one segment (NAE rhotics) can induce changes in the structure and size of whole “chunks” (Schmidt & Lee 2011, Segawa, et al. 2019) of speech: for example, a rhotic vowel in the center of a /VrVrV/ sequence has an influence on the production of every sound in the sequence, which might lead to idiosyncratic tap/flap realization for words like ‘Saturday’ relative to other words with similar sequences.

Also, while our research focuses on the effects of extreme variability on individual speakers, all speakers live in communities, and as a result, these cases of extreme variability provide a mechanism for phonetic changes to become word- and phrase-level effects, as described by Bermúdez-Otero (2015). The potential of transfer from speaker-internal subphonemic segmental variability to later phonemic word or phrase-level changes in populations is intriguing. Taking the literature as it exists, an argument can be made that such an event may be in-progress:

Researchers have identified extreme variability in NAE rhotics (i.e. Delattre & Freeman 1968; Mielke et al. 2016) and flap/taps (Derrick & Gick 2011; 2014). This extreme variability has spread through second-language contact and between speakers in a community to alter the Canadian French mid rounded vowel into a sometimes rhotic vowel (Mielke 2015; Lipari 2023) such that the vowel's production forms an in-progress free-variation cline across communities of speakers (Mielke 2015). Since 2015, this cline has been shown to be influenced by age, gender, and dialect (Lipari 2023).

This Canadian French example is already a documented transfer of subphonemic observation to sociolinguistic effect over time and across languages. All that remains to complete this possibility-space would be to observe phonologization of the distinction between rhotic and non-rhotic variants of the vowel in question on a word-by-word basis. At the moment, the Canadian French change appears to be a synchronic phonological process similar to the “free variation” of the [dʒ ~ j] alternation in Emirati Arabic (Szreder & Derrick, 2024), but not (yet) like the the [k ~ tʃ] alternation in Emirati Arabic, which is now a completed phonemic change (Szreder & Derrick 2024).

Since we also know the extreme variability of NAE flaps and rhotic vowels extends into /VrVrV/ sequences (Derrick & Gick 2021) as well, we can imagine even broader connections to language change spanning from subphonemic segment variability into sociolinguistic and phonemic changes to entire word and phrase production patterns. While these would be even more difficult to study and document than the case of rhotacization of the Canadian French mid-rounded vowel, the possibility space is very much worth exploring.

Data accessibility statement

Data and materials for all experiments can be found here: https://osf.io/n65t8/?view_only=e5712e0862994b9b81f6e7b2505bb77e.

Supplementary files

The EMA data for our NZE and NAE English data, as well as statistical tests and code for producing images for this paper can be found at the Open Science Foundation at https://osf.io/n65t8/?view_only=e5712e0862994b9b81f6e7b2505bb77e. These include the statistical tests used to decide the best-fit model that produced **Table 4** and **Figure 4a**.

Ethics and consent

The University of Canterbury’s Human Research Ethics Committee (HREC) approved ethics for this study (HEC 2012/19). The experiments were performed in accordance with the procedures listed in the HEC 2012/19 document.

Funding information

This research was funded by a New Zealand MARSDEN fast-start grant (12-UOC-081) “Saving energy vs. making yourself understood during speech production” to Donald Derrick.

Acknowledgements

Thanks to the people of the University of British Columbia’s Integrated Speech Research Laboratory for helpful discussions. Thanks to the people at New Zealand Institute of Language, Brain, and Behaviour, Simon Todd and Jacqui Nokes for their insights into the applied math used in this article. Special thanks to Wei-Rong Chen for writing the palate estimation program and Mark Tiede and Michael Proctor for writing the NDI wave data visualization software used in this research. Dedicated to the memory of Romain Fiasson, who performed most of the acoustic labeling and segmenting for this research.

Competing interests

The authors have no competing interests to declare.

References

Alwan, Abeer A. & Narayanan, Shrikanth S. & Haker, Katherine. 1997. Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part II. The rhotics. *The Journal of the Acoustical Society of America* 101(2). 1078–1089. DOI: <https://doi.org/10.1121/1.417972>

- Archangeli, Diana & Baker, Adam & Mielke, Jeff. 2011. Categorization and features: Evidence from American English /ɪ/. Where do phonological features come from? In Clements, Nick G. & Ridouane, Rachid (eds.), *Where do phonological features come from?: Cognitive, physical and developmental bases of distinctive speech categories. [Language faculty and beyond 6]*, 173–196. Amsterdam: John Benjamins. DOI: <https://doi.org/10.1075/lfab.6.07arc>
- Bartoń, Kamil. 2023. *MuMin: Multi-model inference*. R package version 1.47.5, URL: <https://CRAN.R-project.org/package=MuMin>
- Bermúdez-Otero, Ricardo. 2015. Amphichronic explanation and the life cycle of phonological processes. In Honeybone, Patrick & Salmons, Joseph (eds.), *The Oxford handbook of historical phonology*, 374–399. DOI: <https://doi.org/10.1093/oxfordhb/9780199232819.013.014>
- Chodroff, Eleanor & Wilson, Colin. 2017. Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English. *Journal of Phonetics* 61. 30–47. DOI: <https://doi.org/10.1016/j.wocn.2017.01.001>
- Clements, George N. 2003. Feature economy in sound systems. *Phonology* 20(3). 287–333. DOI: <https://doi.org/10.1017/S095267570400003X>
- Delattre, Pierre & Freeman, Donald C. 1968. A dialect study of American Rs by x-ray motion picture. *Linguistics* 44. 29–68. <https://api.semanticscholar.org/CorpusID:144205881>. DOI: <https://doi.org/10.1515/ling.1968.6.44.29>
- Derrick, Donald & Best, Catherine T. & Fiasson, Romain. 2015a. Non-metallic ultrasound probe holder for co-collection and co-registration with EMA. In *Proceedings of 18th International Congress of Phonetic Sciences (ICPhS)*, 1–5.
- Derrick, Donald & Gick, Bryan. 2011. Individual variation in English flaps and taps: A case of categorical phonetics. *Canadian Journal of Linguistics* 56(3). 307–319. DOI: <https://doi.org/10.1017/S0008413100002024>
- Derrick, Donald & Gick, Bryan. 2014. Accommodation of end-state comfort reveals subphonemic planning in speech. *Phonetica* 71(3). 183–200. DOI: <https://doi.org/10.1159/000369630>
- Derrick, Donald & Gick, Bryan. 2021. Gait change in tongue movement. *Scientific Reports* 11(16565). 1–14. DOI: <https://doi.org/10.1038/s41598-021-96139-4>
- Derrick, Donald & Schultz, Benjamin. 2013. Acoustic correlates of flaps and taps in North American English. In *Proceedings of Meetings in Acoustics* 19(1), AIP Publishing. DOI: <https://doi.org/10.1121/1.4798779>
- Derrick, Donald & Stavness, Ian & Gick, Bryan. 2015b. Three speech sounds, one motor action: Evidence for speech-motor disparity from English flap production. *Journal of the Acoustical Society of America* 137(3). 1493–1502. DOI: <https://doi.org/10.1121/1.4906831>
- Espy-Wilson, Carol & Boyce, Suzanne. 1994. Acoustic differences between “bunched” and “retroflex” variants of American English /ɪ/. *Journal of the Acoustical Society of America* 95(5). 2823. DOI: <https://doi.org/10.1121/1.409691>
- Faytak, Matthew D. 2018. *Articulatory uniformity through articulatory reuse: insights from an ultrasound study of Sūzhōu Chinese*. PhD Dissertation, Berkeley: University of California. DOI: <https://doi.org/10.5070/P7141042486>

- Garcia, Damien. 2010. Robust smoothing of gridded data in one and higher dimensions with missing values. *Computational Statistics & Data Analysis* 54. 1167–1178. DOI: <https://doi.org/10.1016/j.csda.2009.09.020>
- Garcia, Damien. 2011. A fast all-in-one method for automated post-processing of piv data. *Experiments in Fluids* 50. 1247–1259. DOI: <https://doi.org/10.1007/s00348-010-0985-y>
- Gatehouse, Stuart & Noble, William. 2004. The speech, spatial and qualities of hearing scale (SSQ). *International Journal of Audiology* 43(2). 85–99. DOI: <https://doi.org/10.1080/14992020400050014>
- Gay, T. 1981. Mechanisms in the control of speech rate. *Phonetica* 38. 148–158. DOI: <https://doi.org/10.1159/000260020>
- Guenther, Frank H. & Espy-Wilson, Carol Y. & Boyce, Suzanne E. & Matthies, Melanie L. & Zandipour, Majid & Perkell, Joseph S. 1999. Articulatory tradeoffs reduce acoustic variability during American English /r/ production. *Journal of the Acoustical Society of America* 105(5). 2854–2865. DOI: <https://doi.org/10.1121/1.426900>
- Hagiwara, Robert. 1995. Acoustic realizations of American /r/ as produced by women and men. PhD Thesis, UCLA.
- Harandi, Negar N. & Woo, Jonghye & Stone, Maureen & Abugharbieh, Rafeef & Fels, Sidney. 2017. Variability in muscle activation of simple speech motions: A biomechanical modeling approach. *Journal of the Acoustical Society of America* 141(4). 2579–2590. DOI: <https://doi.org/10.1121/1.4978420>
- Hay, Jennifer & Maclagan, Margaret & Gordon, Elizabeth. 2008. *New Zealand English*. Scotland: Edinburgh University Press. DOI: <https://doi.org/10.1515/9780748630882>
- Heyne, Matthias & Wang, Xuan & Derrick, Donald & Dorreen, Kieran & Watson, Kevin. 2018. The articulation of /ɪ/ in New Zealand English. *Journal of the International Phonetic Association*, 1–23. DOI: <https://doi.org/10.1017/S0025100318000324>
- Keating, Patricia A. 2003. Phonetic and other influences on voicing contrasts. In Solé, Maria-Josep & Recasens, Daniel & Romero, Joaquin. (eds.), *Proceedings of the 15th international congress of the phonetic sciences*, 20–23. Barcelona: Spain.
- Kirparsky, Paul. 1965. Phonological change. PhD dissertation, Massachusetts Institute of Technology.
- Langstrof, Christian. 2006. Acoustic evidence for a push-chain shift in the intermediate period of New Zealand English. *Language Variation and Change* 18. 141–164. DOI: <https://doi.org/10.1017/S0954394506060078>
- Lipari, Massimo. 2023. The emergence of rhotic vowels in Quebec French: a change from below? In *Proceedings of international congress of the phonetic sciences (ICPhS 2023)*.
- Liu, Sha & Takeda, Kaye. 2021. Mora-timed, stress-timed, and syllable-timed rhythm classes: Clues in English speech production by bilingual speakers. *Acta Linguistica Academica* 68(3). 350–369 DOI: <https://doi.org/10.1556/2062.2021.00469>
- Maclagan, Margaret & Hay, Jennifer. 2007. Getting fed up with our feet: Contrast maintenance and the New Zealand English ‘short’ front vowel shift. *Language Variation and Change* 19. 1–25. DOI: <https://doi.org/10.1017/S0954394507070020>

- Mielke, Jeff. 2015. An ultrasound study of Canadian French rhotic vowels with polar smoothing spline comparisons. *Journal of the Acoustical Society of America* 137(5). 2858–2869. DOI: <https://doi.org/10.1121/1.4919346>
- Mielke, Jeff & Baker, Adam & Archangeli, Diana. 2010. Variability and homogeneity in American English /ɹ/ allophony and /s/ retraction. In Fougeron, Cécile & Kuehnert, Barbara & D’Imperio, Mariapaola & Valée, Nathalie (eds.), *Laboratory Phonology* 10, 699–730. Berlin: Mouton de Gruyter. DOI: <https://doi.org/10.1515/9783110224917.5.699>
- Mielke, Jeff & Baker, Adam & Archangeli, Diana. 2016. Individual-level contact limits phonological complexity: Evidence from bunched and retroflex /ɹ/. *Language* 92(1). 101–141. <https://www.jstor.org/stable/24672200>. DOI: <https://doi.org/10.1353/lan.2016.0019>
- Nokes, Jacqui & Hay, Jennifer. 2012. Acoustic correlates of rhythm in New Zealand English: A diachronic study. *Language Variation and Change* 24. 1–31. DOI: <https://doi.org/10.1017/S0954394512000051>
- Noble, William. 2011. Identifying normal and non-normal hearing: Methods and paradoxes. *WARC talk, MARCS Auditory Laboratory*.
- Ong, Darryl & Stone, Maureen. 1998. Three-dimensional vocal tract shapes in [r] and [l]: A study of MRI, ultrasound, electropalatography, and acoustics. *Phonoscope* 1. 1–14.
- Pierce, Jonathan W. 2007. PsychoPy: Psychophysics software in Python. *Journal of Neuroscience Methods* 162. 8–13. DOI: <https://doi.org/10.1016/j.jneumeth.2006.11.017>
- R Core Team. 2023. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna: Austria. URL: <https://www.R-project.org/>
- Rosenbaum, David A. & Van Heugten, Caroline M. & Caldwell, Graham E. 1996. From cognition to biomechanics and back: the end-state comfort effect and the middle-is-faster effect. *Acta Psychologica (Amsterdam)* 94. 59–85. DOI: [https://doi.org/10.1016/0001-6918\(95\)00062-3](https://doi.org/10.1016/0001-6918(95)00062-3)
- Rosenbaum, David A. & Vaughan, Jonathan & Barnes, Heather J. & Jorgensen, Matthew J. 1992. Time course of movement planning: selection of handgrips for object manipulation. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 18. 1058–1073. DOI: <https://doi.org/10.1037/0278-7393.18.5.1058>
- Schiepek, Günter & Strunk, Guido. 2010. The identification of critical fluctuations and phase transitions in short term and coarse-grained time series—a method for real-time monitoring of human change processes. *Biological Cybernetics* 102. 197–207. DOI: <https://doi.org/10.1007/s00422-009-0362-1>
- Schmidt, Richard A. & Lee, Timothy D. 2011. *Motor control and learning: A behavioral emphasis* (5th ed.) Human Kinetics.
- Segawa, Jennifer & Masapollo, Matthew & Tong, Mona & Smith, Dante J. & Guenther, Frank. H. 2019. Chunking of phonological units in speech sequencing. *Brain and Language* 195. 104636. DOI: <https://doi.org/10.1016/j.bandl.2019.05.001>
- Stavness, Ian & Gick, Bryan & Derrick, Donald & Fels, Sidney. 2012. Biomechanical modeling of English /r/ variants. *Journal of the Acoustical Society of America – Express letters* 131(5). EL355–EL360. DOI: <https://doi.org/10.1121/1.3695407>

- Szreder, Marta & Derrick, Donald. 2024. Phonological conditioning of affricate variability in Emirati Arabic. *Journal of the International Phonetic Association* 54(1). 146–164 DOI: <https://doi.org/10.1017/S0025100323000166>
- Tiede, Mark K. 2010. MVIEW: Multi-channel visualization application for displaying dynamic sensor movements.
- Tiede, Mark K. & Boyce, Suzanne E. & Espy-Wilson, Carol Y. & Gracco, Vincent L. 2010. Variability of North American English /r/ production in response to palatal perturbation. In Maassen, Ben & van Lieshout, Pascal (eds.), *Speech Motor Control: New developments in basic and applied research*, 53–68. DOI: <https://doi.org/10.1093/acprof:oso/9780199235797.003.0004>
- Tiede, Mark K. & Boyce, Suzanne E. & Holland, Carol K. & Choe, K. Ann. 2004. A new taxonomy of American English /r/ using MRI and ultrasound. *Journal of the Acoustical Society of America* 115. 2633–2634. DOI: <https://doi.org/10.1121/1.4784878>
- Voeten, Cesko C. 2023. *buildmer: Stepwise elimination and term reordering for mixed-effects regression*. R package version 2.8. URL: <https://CRAN.R-project.org/package=buildmer>
- Warren, Paul. 1999. Timing properties of New Zealand English. *International Congress of the Phonetic Sciences (ICPhS99)*, 1843–1846.
- Warner, Natasha & Fountain, Amy & Tucker, Benjamin V. 2009. Cues to perception of reduced flaps. *Journal of the Acoustical Society of America* 125(5). 3317–3327. DOI: <https://doi.org/10.1121/1.3097773>
- Warner, Natasha & Tucker, Benjamin V. 2011. Phonetic variability of stops and flaps in spontaneous and careful speech. *Journal of the Acoustical Society of America* 130(3). 1606–1617. DOI: <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. DOI: <https://doi.org/10.1017/S0025100316000219>
- Westbury, John R. & Hashi, Michiko & Lindstrom, Mary J. 1998. Differences among speakers in lingual articulation for American English /ɹ/. *Speech Communication* 26. 203–226. DOI: [https://doi.org/10.1016/S0167-6393\(98\)00058-2](https://doi.org/10.1016/S0167-6393(98)00058-2)
- Zue, Victor W. & Laferriere, Martha. 1979. Acoustic study of medial /t, d/ in American English. *Journal of the Acoustical Society of America* 66, 1039–1050. DOI: <https://doi.org/10.1121/1.383323>