INTERSPEECH 2019
September 15–19, 2019, Graz, Austria

Articulation of vowel length contrasts in Australian English

Louise Ratko, Michael Proctor, Felicity Cox

[1]Centre for Language Sciences, Department of Linguistics, Macquarie University
`louise.ratko@hdr.mq.edu.au`

Abstract

The articulatory realisation of phonemic vowel length contrasts is still imperfectly understood. Australian English (AusE) /ɐː/ and /ɐ/ differ primarily in duration and therefore provide an ideal case for examining the articulatory properties of long vs. short vowels. Patterns of compression, acceleration ratios and VC coordination were examined using electromagnetic articulography (EMA) in /pVːp/ and /pVp/ syllables produced by three speakers of AusE at two speech rates. Short vowels were less compressible and had higher acceleration ratios than long vowels. VC rimes had proportionately earlier coda onsets than VːC rimes. These findings suggest that long and short vowels are characterised by different patterns of both intra- and intergestural organisation in AusE.

Index Terms: articulation, duration, speech rate, vowel length

1. Introduction

Studies of vowel length contrast[1] in German and Slovak have shown that there are systematic articulatory differences between long and short vowels that indicate length-specific patterns of intra- and intergestural organisation [1, 2, 3, 4].

First, long and short vowels differ in their compressibility under changes to speech rate: as speech rate increases, short vowels decrease in duration to a lesser extent than long vowels [1, 3, 4, 5]. Temporal compression of long vowels is primarily manifest through reduction of the duration of the articulatory steady-state (gestural nucleus); for short vowels, on the other hand, the duration of the gestural nucleus remains relatively unaffected by speech rate [1, 3, 4]. Unlike gestural nucleus durations, transitions into the vowel (opening intervals) and transitions out of the vowel (closing intervals) have been shown to be equally compressed across long and short vowels by increases to speech rate in German [3, 4].

Second, long and short vowels differ in their gestural stiffness [1, 6]. In task dynamic models e.g. [7, 8] the global durations of gestures are determined by intrinsic stiffness, with stiffer gestures exhibiting shorter durations [9, 10]. A metric which has been used to characterise stiffness of vowel gestures is the Acceleration Ratio [2, 3, 4, 9, 10], defined as:

$$\text{acceleration ratio (AR)} = \frac{\text{acceleration interval}}{\text{vowel onset to vowel target}} \tag{1}$$

The acceleration interval is defined as the interval between vowel onset and initial peak velocity (Figure 1). In longer, less stiff vowels, peak velocity is reached earlier in the transition between the vowel onset and vowel target ($\text{AR} \leqslant 0.5$), while shorter, stiffer vowels have higher AR values ($\text{AR} > 0.5$) [9, 10]. Greater AR have been found in short vowels in Slovak [1] and German [3, 4]. English speech rate studies, have shown that AR increases along with increases to speech rate [9, 10]. However, to date, AR has not been examined for vowel-length contrasts in English. Moreover, past studies of Englisy have not examined whether speech rate impacts AR of long and short vowels to the same extent.

Finally, long and short vowels differ in their intergestural coordination with coda consonants. Data from German [2, 3, 4, 6] and Slovak [1] show that short vowel gestures are more overlapped with following coda consonants than long vowel gestures. Fast rate vowels have also been shown to have greater overlap with surrounding gestures than normal rate vowels [3, 4, 11]. However, to date, the relationship between phonological vowel length and VC coordination in English has not been examined. Past studies have also not examined whether speech rate affects coordination of long and short vowels with a following coda to the same extent. However, if short vowels are already more overlapped with coda consonants than long vowels when produced at a normal speech rate, it may be that at fast speech rates short vowels will resist further increases in VC overlap.

In AusE, German and Slovak, long and short vowels share similar dynamic and kinematic properties: short vowels have proportionately shorter articulatory and acoustic steady-states and proportionately longer transitions to following consonants [1, 2, 3, 12, 13, 14]. However, unlike German and Slovak, vowel length contrasts are not systemic in AusE, but rather are restricted to a subset of long-short vowel pairs, /iː-ɪ/, /eː-e/ and /ɐː- ɐ/ [12, 15]. Furthermore, in AusE, the low vowels /ɐː/ 'cart' and ɐ/ 'cut' are distinguishable only in vowel duration and not spectral quality [12, 13, 14, 15], unlike in some other varieties of English. Australian English therefore provides an ideal starting point for examining the articulatory properties of vowel length contrast in languages without a systemic vowel length contrast. This preliminary study will examine the effect of speech rate on patterns of compression, acceleration ratios, and VC coordination of the AusE long-short vowel pair /ɐː- ɐ/.

We hypothesise that, compared to long vowels:

- short vowel gestural durations will be less compressed by increases to speech rate [1, 2, 3, 4]
- short vowel gestural nucleus durations should be less compressed by increases to speech rate [1, 3, 4]
- short vowel closing intervals should be equally compressed by increased to speech rate [3, 4]
- short vowels will have greater acceleration ratios [3, 4]
- codas following short vowels should begin proportionately earlier in the VC rime [1, 2, 3, 4, 6]
- codas following short vowels should resist increases in VC overlap as speech rate increases

---

[1] Because the focus of this paper is the temporal dimension of vowel-length/tensity contrasts, we will exclusively refer to this contrast as a length distinction.

http://dx.doi.org/10.21437/Interspeech.2019-2995

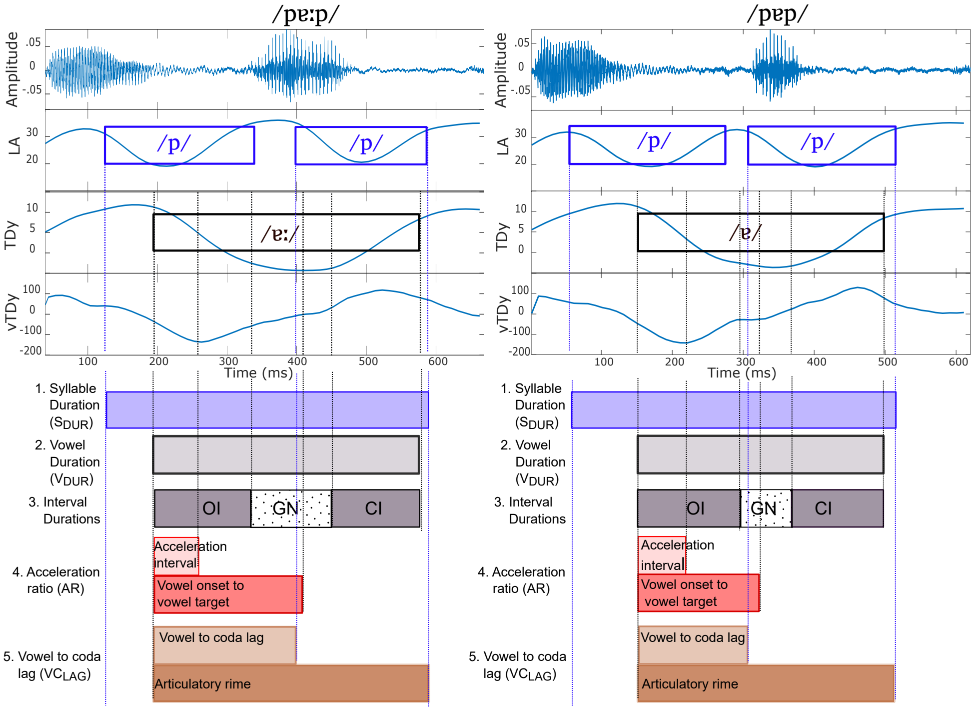

Figure 1: ***Articulatory measurements of syllables contrasting long vs. short vowels**. Top row: acoustic waveforms of 'parp' (left) and 'pup' (right). Tokens produced by participant W1. LA: Lip Aperture (mm), TDy: vertical displacement of tongue dorsum sensor (mm) and vTDy: velocity of tongue dorsum sensor (mm/s). Vertical lines denote key articulatory landmarks: Syllable onset (blue), vowel onset, initial peak velocity, gestural nucleus onset, coda onset (blue), vowel target, gestural nucleus offset, vowel offset, syllable offset (blue). Horizontal bars indicate gestural and intergestural intervals used in the analysis: 1) syllable duration (SDUR), 2) vowel gestural duration (VDUR), 3) vowel gesture intervals: opening interval (OI), gestural nucleus (GN), closing interval (CI), 4) acceleration ratio (AR), 5) vowel to coda lag (VCLAG).*

2. Methods

2.1. Participants

Three monolingual speakers of Australian English (two males; age: 25-26, mean: 25.7 years), participated in the study. All reported normal hearing, and no history of speech or language problems. All participants had completed primary, secondary and tertiary education in New South Wales.

2.2. Experiment Materials

Long /ɐː/ and short /ɐ/ were contrasted in /pVp/ syllables: 'parp' and 'pup'. Target syllables were elicited in the carrier phrase *'See pVp heat'* to control for tongue position prior to and following target items. Stimulus materials were paced to elicit two distinct speech rates from participants: each elicitation sentence was presented for 1000 ms at the normal speech rate and 500 ms at the fast rate. 15 repetitions of each target word were elicited at each rate, resulting in a total of 60 tokens per participant.

2.3. Data acquisition and analysis

Speech movements were tracked at a sampling rate of 100 Hz using a Northern Digital Inc. Wave Electromagnetic Articulography (EMA) system. Sensors were placed on the mid-sagittal plane on the lips, tongue and jaw. Labial and lingual sensor signals were corrected for head movement and rotated into a common coordinate system defined with respect to three reference sensors located on participants' left and right mastoids and nasion. Vertical sensor displacement was expressed relative to the occlusal plane; horizontal displacement with respect to the rear of the upper incisors. Analyses presented in this study used data from two articulatory signals: LA (lip aperture) and TD (tongue dorsum). LA is the euclidean distance between the upper and lower lip sensors. The TD sensor is the rearmost lingual sensor. The TD sensor was chosen as it provided the richest information about vowel articulation. Sensor traces were low-pass filtered and conditioned using a DCT-based discretised smoothing spline [16] and synchronised with companion audio recorded using a shotgun microphone at a sampling rate of 22,050 Hz.

Gestural landmarks were semi-automatically located with the *findgest* algorithm in Mview [17]. Syllable durations (SDUR) were measured from the onset of the initial /p/ gesture to the offset of the final /p/ gesture (Figure 1). Vowel gesture durations (VDUR) spanned from vowel onset to vowel offset (Figure 1). Three intervals were demarcated in each vowel gesture (Figure 1). Opening interval = Vowel onset to gestural nucleus onset, Gestural nucleus = Gestural nucleus onset to gestural nucleus offset, Closing interval = Gestural nucleus offset to vowel offset. Durations of gestural nucleus (GNDUR) and closing interval (CIDUR), will be analysed here. GNDUR and CIDUR are expressed as a proportion of the total gestural duration of the vowel (VDUR). Gestural stiffness was measured through acceleration ratio (AR) calculated as shown in Equation 1 and Figure 1. VC coordination was measured by vowel to coda lag (VCLAG): the lag between the onset of the vowel gesture and the onset of the coda consonant as a proportion of total rime (VC) duration. VCLAG were expressed as proportions to control for the intrinsically different durations of long and short vowel gestures. Of 180 total utterances elicited (15 items x 2 target words x 2 speech rates x 3 speakers), 148 were analysed. 32 items were excluded due to mispronunciation and tracking errors.

3. Results

We examined acoustic durations of normal rate vowels to confirm that our participants produced /ɐː/ and /ɐ/ with similar durations as previous studies of AusE. The grand mean acoustic duration of /ɐ/ produced by these three speakers at the normal speech rate was 0.58 that of /ɐː/, consistent with Cox's finding that /ɐ/ was 0.57 of the acoustic duration of /ɐː/. Grand mean gestural duration (VDUR) of /ɐ/ produced at normal speech rate was 0.81 of the duration of /ɐː/. This is consistent with a previous study of Australian English, which found grand mean gestural duration of /ɐ/ to be 0.82 the duration of /ɐː/ [14].

We constructed linear mixed effects models using the lme4 package in R [18] with the equation: Dependent Variable $\sim$ vowel length (LONG = 0; SHORT = 1) $\times$ speech rate (NORMAL = 0, FAST = 1) + (1 | speaker). The dependent variables examined were: SDUR, VDUR, GNDUR, CIDUR, AR, VCLAG. *p*-values for main effects were obtained through maximum likelihood tests with Satterthwaite approximations to degrees of freedom [19].

First we examined whether our elicitation method successfully elicited speaking rate differences for the two rate conditions. SDUR was 404 ms (s.d. 46) in the fast rate, and 474 ms (s.d. 35) in the normal rate. We modelled SDUR as a linear function of speech rate with random effects of speaker. The effect of speech rate on SDUR was significant (β = -85.2 ms, t = -11.7, $p < .001$; Figure 2).

3.1. Vowel gestural durations

To examine whether short vowel gestures were less compressible than long vowel gestures, we constructed a linear model of VDUR as a function of vowel length $\times$ speech rate with random effects of speaker. VDUR was shorter for short vowels (β = -79.7 ms, t = -6.8, $p < .001$, Figure 2). VDUR was shorter for fast rate vowels (β = -72.6 ms, t = -6.1, $p < .002$). There was an interaction of vowel length $\times$ speech rate (β = 40.9 ms, t = 2.3, p = .016) indicating that VDUR of short vowels were less impacted by increases to speech rate than VDUR of long vowels.

3.2. Interval durations

To examine whether the gestural nucleus was less compressible in short vowels compared to long vowels, we constructed a linear model of GNDUR as a function of vowel length $\times$ speech rate with random effects of speaker. GNDUR was shorter in short vowels (β = -5.4%, t = -4.7, $p < .001$, Figure 3). GNDUR was shorter in fast vowels (β = -3.4%, t = -3.1, p = .002). There was an interaction of vowel length $\times$ speech rate (β = 3.8%, t = 2.4, p = .017) indicating that gestural nucleus duration was less impacted by changes to speech rate in short, compared to long vowels.

To examine whether the gestural closure interval was equally compressible in short and long vowels, we constructed a linear model of CIDUR as a function of vowel length $\times$ speech rate with random effects of speaker. CIDUR did not differ significantly between long and short vowels (p = .486, Figure 3).

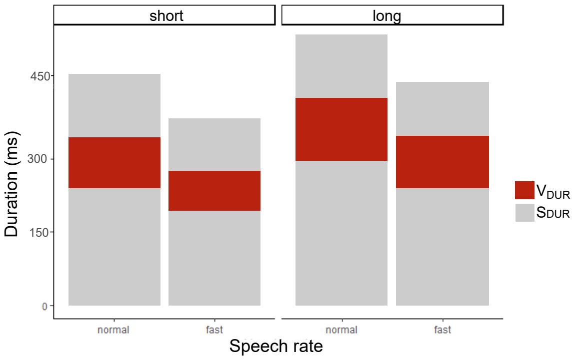

Figure 2: ***Articulatory syllable durations (SDUR, entire column) and vowel gestural durations (VDUR)*** *by vowel length and speech rate. Duration in ms. Intervals determined as shown in Figure 1.*

CIDUR was shorter in fast vowels (β = 4.9%, t = 2.9, p = .005). There was no significant interaction between vowel length * speech rate (p = .361).

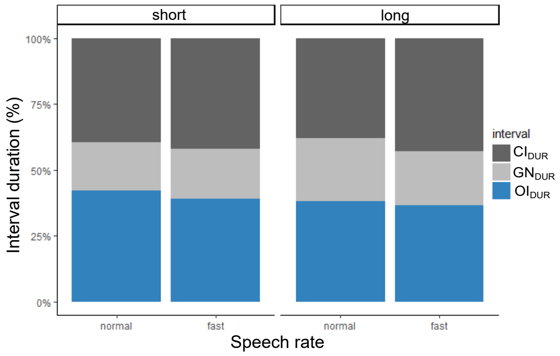

Figure 3: ***Opening interval (OI), gestural nucleus (GN), and closing interval (CI)*** *(expressed as proportions of entire vowel gestural duration) by vowel length and speech rate. Intervals determined as shown in Figure 1.*

3.3. Acceleration ratio (AR)

To examine whether acceleration ratio (AR) was greater for short vowels compared to long vowels, we constructed a linear model of AR as a function of vowel length $\times$ speech rate with random effects of speaker. AR was greater in short vowels (β = 0.07, t = 5.4, $p < .001$, Figure 4). There was no effect of speech rate on AR (p = .067). There was also no significant interaction of vowel length $\times$ speech rate (p = .251).

3.4. VC coordination

To examine whether short vowel coordination with a following coda consonant differed from that of long vowels, we constructed a linear model of VCLAG as a function of vowel length $\times$ speech rate with random effects of speaker. Short vowels had shorter VCLAG than long vowels (β = -6.8%, t = -9.7, $p <$ .001; Figure 5). Fast vowels had shorter VCLAG than normal rate vowels (β = -4.8%, t = -6.8, $p < .001$). There was an interaction of vowel length $\times$ speech rate (β = 2.4%, t = 2.4, p = .016), indicating that short vowel VCLAG was less impacted by increases to speech rate than long vowel VCLAG.

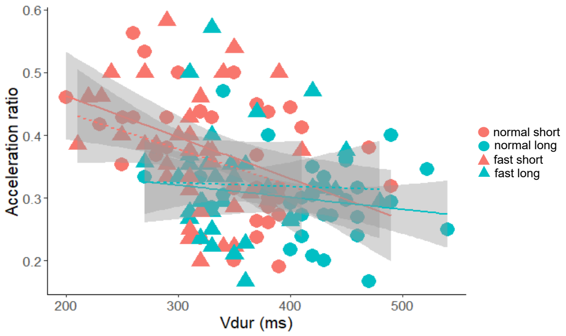

Figure 4: ***Acceleration ratios*** *(AR) by vowel length and speech rate. Solid regression line = normal rate, dashed regression line = fast rate. AR = acceleration interval as ratio of vowel onset to vowel target as shown in Figure 1. Higher AR = greater gestural stiffness.*

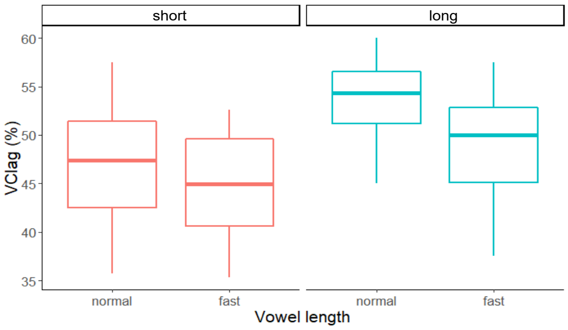

Figure 5: ***Lag between vowel onset and coda onset*** *(VCLAG) by vowel length and speech rate. VCLAG = Coda onset − Vowel onset, expressed as proportion of total VC rime duration, as shown in Figure 1. Lower values indicate proportionately earlier onset of coda consonant.*

3.5. Summary of results

Compared to long vowels:

- short vowel VDUR was less affected by increases to speech rate
- short vowel GNDUR was less affected by increases to speech rate
- short vowel CIDUR was equally affected by increases to speech rate
- short vowel AR was higher
- AR was unaffected by speech rate
- VCLAG was proportionately shorter for rimes containing short vowels
- VCLAG for rimes containing short vowels was less affected by increases to speech rate

4. Discussion

In this study, we investigated articulatory differences between /pV:p/ and /pVp/ AusE syllables at normal and fast speech rates.

First, consistent with previous research in Slovak and German, we found that short vowels had shorter vowel gestural durations (VDUR) and shorter gestural nucleus durations (GNDUR) at both normal and fast speech rates. Short VDUR and short vowel GNDUR were less compressible in the fast speech rate compared to long vowels (Figures 2, 3) [1, 2, 3, 4, 6]. We also compared long and short vowel closing interval durations (CIDUR), and found that unlike vowel GNDUR, CIDUR compressed equally for both long and short vowels.

Second, we found acceleration ratios (AR) differed by vowel length, with short vowel gestures having greater AR than long vowel gestures (Figure 4). This is consistent with the hypothesis that short vowels have intrinsically greater gestural stiffness than long vowels [1, 6]. However, contrary to speech rate studies of English we did not find significantly greater AR for fast vs. normal rate vowels [9, 10]. While this may not be surprising for short vowels, which are generally less impacted by increases to speech rate [1, 3, 4, 5], it is unexpected for long vowels. Past studies have shown that the articulatory strategies used to increase speech rate can vary considerably across speakers, with not all speakers increasing stiffness along with speech rate [10, 11]. Therefore our results may be sensitive to our small participant number (N = 3). More data are needed from more speakers before further conclusions can be drawn.

Finally, we examined the lag between vowel onset and coda onset (VCLAG), and found that coda consonants began proportionately earlier in VC vs V:C rimes. Codas also began proportionately earlier in fast vs. normal rate vowels (Figure 5). We also found that increases to speech rate affected VCLAG more for long than short vowels. These findings are consistent with the suggestion that vowel length is linked to changes in intergestural VC coordination, with short vowels sharing a closer relationship with following coda consonants than long vowels [1, 2, 3, 4, 6]. However, it is important to note that VCLAG does not provide a comprehensive account of VC overlap. Short vowels have intrinsically shorter durations than long vowel gestures [6, 14], therefore while codas begin proportionately earlier in VC than V:C rimes, it may also be the case that short vowel gestures end proportionately earlier in the rime, resulting in comparable degrees of VC overlap between VC and V:C rimes. More research is needed to clarify whether earlier coda onset results in more VC overlap for tautosyllabic short compared to long vowels.

In German and Slovak long and short vowels show systematic articulatory differences that suggest different patterns of intra- and inter-gestural organisation. Short vowels are less compressible than their long equivalents, have higher gestural stiffness and have different patterns of VC coordination [1, 2, 3, 4]. The patterns observed in these data are consistent with these findings, and suggest that long and short AusE vowels also share these systematic articulatory contrasts.

5. Future directions

More data is needed to determine whether these patterns hold in larger populations of speakers, and in other long-short vowel pairs in AusE which differ in both vowel duration and spectral quality, such as /iː-ɪ/ (‘sheep’-‘ship’). It is also an open question as to whether these articulatory differences hold in a wider range of languages with systemic and non-systemic vowel length contrasts. Finally, the incorporation of a slow speech rate may reveal additional insight into differences in intra- and inter-gestural organisation between long and short vowels

6. References

[1] S. Beňuš, "Control of phonemic length contrast and speech rate in vocalic and consonantal syllable nuclei," *The Journal of the Acoustical Society of America*, vol. 130, pp. 2116–2127, 2011.

[2] P. Hoole, C. Mooshammer, and H. G. Tillmann, "Kinematic analysis of vowel production in German," *Proceedings of ICSLP 94*, vol. 1, pp. 53–56, 1994.

[3] P. Hoole and C. Mooshammer, *Silbenschnitt and Tonakzente*. Tübingen: Niemeyer, 2002, ch. Articulatory analysis of the German vowel system, pp. 129–152.

[4] C. Kroos, P. Hoole, B. Kühnert, and H. G. Tillmann, "Phonetic evidence for the phonological status of the tense-lax distinction in German," *Forschungsberichte des Instituts für Phonetik and Spachliche Kommunikation der Universität München*, vol. 35, pp. 17–25, 1997.

[5] R. Port, "Linguistic timing factors in combination," *The Journal of the Acoustical Society of America*, vol. 69, pp. 262–274, 1981.

[6] I. Hertrich and H. Ackermann, "Articulatory control of phonological vowel length contrasts: Kinematic analysis of labial gestures," *The Journal of the Acoustical Society of America*, vol. 102, pp. 523–536, 1997.

[7] C. Browman and L. Goldstein, "Towards an articulatory phonology*," *Phonology Yearbook*, vol. 3, pp. 219–152, 1986.

[8] C. P. Browman and L. Goldstein, "Articulatory phonology: An overview," *Phonetica*, vol. 49, pp. 155–180, 1992.

[9] D. J. Ostry and K. G. Munhall, "Control of rate and duration of speech movements." *The Journal of the Acoustical Society of America*, vol. 77, pp. 640–648, 1985.

[10] S. G. Adams, G. Weismer, and R. D. Kent, "Speaking rate and speech movement velocity profiles," *Journal of Speech and Hearing Research*, vol. 36, pp. 41–54, 1993.

[11] S. Shaiman, "Kinematics of compensatory vowel shortening: Intra and interarticulatory timing," *The Journal of the Acoustical Society of America*, vol. 106, pp. 89–107, 2001.

[12] F. Cox, "The Acoustic Characteristics of /hVd/ Vowels in the Speech of some Australian Teenagers," *Australian Journal of Linguistics*, vol. 26, pp. 147–179, 2006.

[13] J. Elvin, D. Williams, and P. Escudero, "Dynamic acoustic properties of monophthongs and diphthongs in Western Sydney Australian English," *The Journal of the Acoustical Society of America*, vol. 140, pp. 576–581, 2016.

[14] L. Ratko, "Articulatory characterisation of length contrasts in Australian English vowels," Master's thesis, Department of Linguistics, Macquarie University, 2017.

[15] I. Yuen, F. Cox, and K. Demuth, "Three year olds' production of Australian English vowel length contrasts as a function of prosodic context," *Journal of the Acoustical Society of America*, vol. 135, no. 3, pp. 1469–1479, 2014.

[16] D. Garcia, "Robust smoothing of gridded data in one and higher dimensions with missing values," *Computational statistics and data analysis*, vol. 54, pp. 1167–1178, 2010.

[17] M. Tiede, "Mview: software for visualization and analysis of concurrently recorded movement data." 2005.

[18] D. Bates, "lme4: Mixed effects modelling with R." Springer, 2010.

[19] A. Kuznetsova, P. Brockhoff, and R. Christensen, "Lmertest package: Tests in linear-mixed effects models," *Journal of Statistical Software*, vol. 82, pp. 1–26, 2017.