

Articulation of vowel length contrasts in Australian English

Louise Ratko 

Department of Linguistics, Macquarie University, Australia
louise.ratko@mq.edu.au

Michael Proctor

Department of Linguistics, Macquarie University, Australia
michael.proctor@mq.edu.au

Felicity Cox

Department of Linguistics, Macquarie University, Australia
felicity.cox@mq.edu.au

Acoustic studies have shown that in Australian English (AusE), vowel length contrasts are realised through temporal, spectral and dynamic characteristics. However, relatively little is known about the articulatory differences between long and short vowels in this variety. This study investigates the articulatory properties of three long–short vowel pairs in AusE: /i:–ɪ/ *beat* – *bit*, /e:–ɛ/ *cart* – *cut* and /o:–ɔ/ *port* – *pot*, using electromagnetic articulography. Our findings show that short vowel gestures had shorter durations and more centralised articulatory targets than their long equivalents. Short vowel gestures also had proportionately shorter periods of articulatory stability and proportionately longer articulatory transitions to following consonants than long vowels. Long–short vowel pairs varied in the relationship between their acoustic duration and the similarity of their articulatory targets: /i:–ɪ/ had more similar acoustic durations and less similar articulatory targets, while /e:–ɛ/ were distinguished by greater differences in acoustic duration and more similar articulatory targets. These data suggest that the articulation of vowel length contrasts in AusE may be realised through a complex interaction of temporal, spatial and dynamic kinematic cues.

1 Introduction

The acoustic characterisation of vowel length contrasts in Australian English (AusE) has been clearly documented. Vowel length contrasts in this variety are realised through temporal (Bernard 1967, Cochrane 1970, Fletcher & McVeigh 1993, Cox 2006, Cox & Palethorpe 2011, Cox, Palethorpe & Miles 2015), spectral (Bernard 1970, Cox 2006, Elvin, Williams & Escudero 2016), and dynamic characteristics (Harrington & Cassidy 1994, Watson & Harrington 1999, Cox 2006, Elvin et al. 2016). Less is known about the articulatory characteristics of vowel length contrasts in AusE. Fletcher, Harrington & Hajek (1994) compared jaw displacement in the long–short vowel pair /e:–ɛ/ (*barb* – *bub*) in /bVb/ syllables for three speakers, and found that /e:/ was consistently characterised by a lower target jaw position

than /e/. Blackwood Ximenes, Shaw & Carignan (2017) examined the articulation of a subset of AusE vowels produced in /sVd/ context by four speakers, and found that the average tongue dorsum position of /ɪ/ was lower and more retracted than its long equivalent /i:/. The present study expands on previous articulographic work by examining the lingual articulation of multiple long–short vowel pairs in AusE, allowing us to characterise the kinematic properties that underlie the realisation of this contrast.

1.1 Phonetic correlates of vowel length contrast

The primary cue to vowel length contrasts in languages such as AusE, is vowel duration. Long vowels are prototypically produced with a greater acoustic duration than short vowels (Lehiste 1970, Lindau 1978). The acoustic duration of vowels is commonly measured from the onset to offset of vowel voicing (House 1961, Lehiste & Peterson 1961, Bell-Berti & Harris 1981, Hertrich & Ackermann 1997). This measure is dependent upon the duration of laryngeal activity associated with vowel articulation (Bell-Berti & Harris 1981, Hertrich & Ackermann 1997). However, the durations of the supralaryngeal articulatory movements of the lips, jaw and tongue have been relatively understudied. Hertrich & Ackermann (1997) examined the duration of lip-opening gestures associated with German vowels, finding that, on average, the lip-opening movement of short vowels was approximately 80% the duration of those of long vowels, while the acoustic duration of short vowels was 60% that of long vowels. These results demonstrate that while vowel length-related durational contrast is specified across multiple articulators (e.g. lips, larynx), this durational contrast appears to be specified differently across these different articulators. However, it remains an open question whether differences between acoustic and articulatory characteristics of vowel duration occur in other languages.

In Dutch, English, German, and Swedish, long/tense and short/lax vowels¹ often also differ with regard to their position in the vowel space, with the acoustic and articulatory targets of short vowels produced closer to the centre of the vowel space compared to their long equivalents (Lindblom 1963, Hadding-Koch & Abramson 1964, Lindau 1978, Nooteboom & Doodeman 1980, Jessen 1993, Hoole & Mooshammer 2002, Cox 2006, Harrington, Hoole & Reubold 2012, Elvin et al. 2016). Early accounts of vowel quality differences between long and short vowels proposed a physiological explanation, whereby the centralisation of short vowel targets was said to be due to biomechanical limitations on achieving the same phonological target as their long equivalents in a shorter time span (Lindblom 1963). In this undershoot account, the primary determinant of centralisation is vowel duration: the shorter the vowel the more centralised its target (Lindblom 1963). However, vowel quality may be manipulated independently of vowel duration in the realisation of vowel length contrasts. In German unstressed syllables, short (lax) vowels are centralised but not shorter in duration than long vowels (Mooshammer & Fuchs 2002, Mooshammer & Geng 2008). Furthermore, listeners appear to use both vowel quality differences and durational differences as cues to vowel length contrasts (Delattre 1962, Hadding-Koch & Abramson 1964, Mooshammer & Fuchs 2002, Gussenhoven 2007, Mády & Reichel 2007, Mooshammer & Geng 2008, Lehnert-LeHouillier 2010, Meister, Werner & Meister 2011, Tomaschek, Truckenbrodt & Hertrich 2015). Vowel quality differences and durational differences appear to be in a trading relationship in some languages; listeners rely less on durational cues to vowel length when presented with stimuli in which long–short vowel quality differences are exaggerated, and rely more on durational cues when quality differences are minimised (Delattre 1962, Hadding-Koch & Abramson 1964, Lehnert-LeHouillier 2010).

Vowel length contrasts are also characterised by differences in formant dynamics. The proportionate duration of three acoustic components: the acoustic onglide, acoustic steady-state (target) and the acoustic offglide have been shown to differ between long and

¹ In this paper the term ‘vowel length contrast’ is used to refer to what is variably described as a vowel length contrast or a tense-lax contrast in different languages.

short vowels. In American English (Lehiste & Peterson 1961), Canadian English (Nearey & Assmann 1986), German (Strange & Bohn 1998), and AusE (Bernard 1967, Watson & Harrington 1999, Cox 2006), short vowels have proportionately shorter acoustic steady-states and proportionately longer acoustic offglides than their long counterparts. These differences have also been observed in articulation, with short vowels in German and Slovak exhibiting proportionately shorter articulatory steady states than their long equivalents (Kroos et al. 1997, Hoole & Mooshammer 2002, Beňuš 2011). In German, short vowels also exhibit proportionately longer release intervals (the articulatory transition to following tautosyllabic consonants) than long vowels (Kroos et al. 1997, Hoole & Mooshammer 2002). Little is known about the dynamic articulatory properties of AusE vowels, or whether AusE exhibits similar vowel-length dependent patterns of articulation as those found in German.

1.2 Australian English

The AusE vowel inventory consists of 18 stressable vowels (Cox 2006, Cox & Fletcher 2017),² including six long /i: e: o: u: ɜ:/ and six short /ɪ e ə ʊ ɔ u/ monophthongs (Figure 1; Cox 2006, Cox & Fletcher 2017).

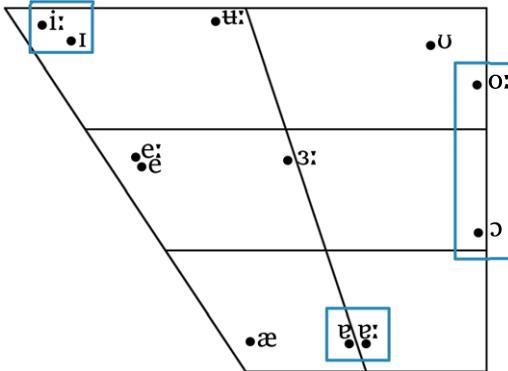


Figure 1 (Colour online) Schematic illustrating the distribution of AusE monophthongs in the acoustic vowel space. Overlaid blue boxes indicate vowel pairs examined in this study. Based on Cox & Palethorpe (2007).

Australian English uses a vowel length contrast (Cox & Palethorpe 2007), rather than a tense–lax contrast, characteristic of other English varieties such as American English (Peterson & Lehiste 1960, House 1961, Lehiste & Peterson 1961). This is due to the contrastive status of duration in signalling the distinction between some vowel pairs; in particular, /ɛ:–ɐ/, /e:–e/ and (for some speakers) /i:–ɪ/ (Watson & Harrington 1999, Cox 2006, Cox & Palethorpe 2007).

1.2.1 Duration

On average, AusE short vowels are 60% the duration of their long equivalents in voiced coda contexts (Cox 2006, Elvin et al. 2016). This is more distinct than the tense/lax contrast of General American English, in which lax vowels are approximately 75% the duration of their tense equivalents (Peterson & Lehiste 1960, House 1961). Relative durational differences are consistent across various phonetic contexts (Elvin et al. 2016). However, the absolute

² This study uses the Harrington, Cox & Evans (HCE) system for phonemic transcription of AusE vowels. See Cox & Fletcher (2017) for details.

duration of AusE vowels is affected by vowel height similar to other English dialects (House 1961, Chen 1970, Cochrane 1970, Klatt 1976, Cox 2006, Elvin et al. 2016). In citation form /hVd/ context, /ɪ/ has a shorter absolute duration (~140 ms) than both the short open vowel /e/ (~160 ms) and the short mid-open vowel /ɔ/ (~170 ms) (Cox 2006). No previous studies of AusE have examined the duration of lingual activity associated with vowels, so it is not known how these durational contrasts manifest in the articulatory domain.

1.2.2 Target quality

Although duration is the primary cue to vowel length in AusE (Harrington & Cassidy 1994, Watson & Harrington 1999, Cox 2006), long and short vowel pairs also differ in target quality, with some vowel pairs intrinsically more spectrally and spatially differentiated than others.

/e:/ and /e/ have largely overlapping vowel targets, and thus differ primarily in duration (Bernard 1970, Cochrane 1970, Harrington & Cassidy 1994, Watson & Harrington 1999, Cox 2006, Elvin et al. 2016). Cox (2006) examined 960 /hVd/ tokens from 120 female adolescent speakers of AusE. The mean F1 and F2 values of /e:/ (F1 = 856 Hz, F2 = 1451 Hz) did not differ significantly from those of /e/ (F1 = 842 Hz, F2 = 1469 Hz). Likewise, Bernard (1970), in his cineoradiographic study of AusE vowels, found a high degree of similarity between the lingual articulatory targets of /e:-e/ in /hVd/ syllables. Conversely, Fletcher et al. (1994) found small but significant differences in jaw displacement between /e:/ and /e/ in /bVb/ syllables, with /e/ showing a more centralised jaw trajectory.

/i:-i/ also share similar acoustic vowel targets. Cox (2006) found no significant difference in the mean F1 and F2 values of /i:/ (F1 = 391 Hz, F2 = 2729 Hz) and /i/ (F1 = 402 Hz, F2 = 2697 Hz) produced by adolescent females in the 1990s. However, studies based on more recent acoustic data have suggested that in young AusE speakers /i/ is marginally lower and more retracted than /i:/ (Cox, Palethorpe & Bentink 2014). These acoustic results are supported by recent articulatory studies where /i/ is produced with a significantly more retracted and lowered tongue dorsum than /i:/ (Blackwood Ximenes et al. 2017).

Unlike /e:-e/ and /i:-i/, /o:/ and /ɔ/ can be differentiated through their target formant values alone, independent of durational information (Bernard 1970, Watson & Harrington 1999, Cox 2006, Elvin et al. 2016, Cox & Fletcher 2017). The primary difference between /o:/ and /ɔ/ is in target F1 (/o:/ = 494 Hz, /ɔ/ = 708 Hz) although the pair also differs in F2 (/o:/ = 954 Hz, /ɔ/ = 1182 Hz; Cox 2006). Early articulatory analysis of /o:/ and /ɔ/ shows a clear differentiation of target tongue position for this pair (Bernard 1970), however, recent articulatory analyses highlight that the tongue dorsum positions at the target of /o:/ and /ɔ/ have much larger degree of articulatory overlap than reflected in the target F1 and F2 values of these vowels: /o:/ is articulated with a similar tongue dorsum height and a slightly more retracted posture than /ɔ/ (Blackwood Ximenes, Shaw & Carignan 2016, 2017; Ratko et al. 2016). Instead, differences in lip rounding may also contribute to the F1 and F2 differences between /o:/ and /ɔ/. Blackwood Ximenes et al. (2017) observed that the long /o:/ had a greater degree of lip protrusion than the short /ɔ/ in three out of four recorded participants. More research is needed to determine whether differences in lip rounding are also present in other samples of AusE speakers.

1.2.3 Dynamic formant structure

Finally, long and short vowels differ in their dynamic formant structure in AusE (Harrington & Cassidy 1994, Watson & Harrington 1999, Cox 2006, Cox et al. 2014). On average, short vowels are produced with proportionately shorter acoustic steady-states and proportionately longer acoustic offglides than phonologically long vowels (Cox 2006). However, /i:/ and /i/ differ further in dynamic formant structure, with /i:/ characterised by a prolonged acoustic onglide for some AusE speakers, giving it a semi-diphthongal quality [ʔi] (Harrington &

Cassidy 1994, Harrington, Cox & Evans 1997, Watson & Harrington 1999, Cox 2006, Cox et al. 2014, Cox et al. 2015).

Collectively, this work suggests that different long–short vowel pairs may vary in the extent to which vowel length contrast is expressed by temporal (duration) or spectral/spatial (target formant values or target tongue position) information (Watson & Harrington 1999, Cox 2006).

This study will focus on the articulation of three long–short vowel pairs /i:–ɪ/, /e:–ɛ/ and /o:–ɔ/.³ These pairs are distributed across three peripheral areas of the AusE vowel space (Figure 1). /i:–ɪ/ *beat* – *bit* are considered to contrast primarily in vowel length, although this pair also has an additional onglide contrast present in /i:/ (Cox 2006, Cox & Palethorpe 2007). /e:–ɛ/ *cart* – *cut* contrast primarily in length. The third pair, /o:–ɔ/ *port* – *pot* are distinguishable by acoustic height in addition to length (Cox 2006), but have a high degree of lingual articulatory similarity (Blackwood Ximenes et al. 2016, 2017; Ratko et al. 2016).

1.3 Aims and predictions

The aim of this paper is to provide an empirical investigation of the lingual articulation of vowel length contrasts in AusE. The present study builds upon a largely acoustic description of AusE vowels. The few prior articulatory studies of AusE vowels have not focused on length contrasts (Bernard 1967, Blackwood Ximenes et al. 2017) or have included examination of only a single long–short vowel pair (Fletcher et al. 1994). We make the following predictions:

1. Durational differences in the lingual gestures (gesture onset to gesture offset) of long and short vowels should follow similar patterns as acoustic duration differences, with short vowel gestures having a shorter duration than long vowel gestures (Cox 2006, Elvin et al. 2016), but the magnitude of durational differences between long and short vowels should be reduced in the articulatory domain, as has been found in German (Hertrich & Ackermann 1997).
2. Although all long–short vowel pairs should exhibit similar articulatory targets, the degree of similarity is predicted to differ by vowel pair. The low vowel pair /e:–ɛ/ will have the most similar articulatory targets, whereas /i:–ɪ/ and /o:–ɔ/ will have less similar pairwise articulatory targets (Bernard 1970, Cox 2006, Elvin et al. 2016).
3. There will be a trading relationship between acoustic duration and spatial and kinematic differences in the realisation of vowel length contrast (Delattre 1962, Hadding-Koch & Abramson 1964, Lehnert-LeHouillier 2010). That is, the long–short pair with the most similar articulatory targets will exhibit the largest pairwise difference in acoustic duration, whereas the long–short pair with the least similar articulatory targets will exhibit the smallest difference in acoustic duration. This is in opposition to Lindblom's (1963) target undershoot account, which predicts that the vowel pair with the least similar articulatory target would exhibit the largest durational differences.
4. /o:–ɔ/ will be produced with more lip rounding than /ɔ/.
5. In line with acoustic studies (Watson & Harrington 1999, Cox 2006, Elvin et al. 2016), long and short vowel gestures will be characterised by different dynamic articulatory patterns. Short vowels should exhibit a proportionately shorter period of articulatory stability around their mid-points and proportionately longer articulatory transitions to following consonants. However, the long vowel /i:/ should be characterised by a lengthy phonological onglide as is characteristic of AusE (Cox 2006).

³ The corresponding lexical sets for these vowels are /i:/ FLEECE, /ɪ/ KIT, /e:/ START, /ɛ/ STRUT, /o:/ THOUGHT, /ɔ/ LOT (Wells 1982: xviii–xix). Note that AusE is non-rhotic.

2 Method

2.1 Participants

Participants were seven monolingual speakers of AusE (four females). Average age was 20.4 years (*s.d.* = 2.82). All participants were born in Australia and had at least one Australian-born parent. All reported no history of speech or hearing disorders. All received primary and secondary education within New South Wales, and were residents of the Greater Sydney region at time of recording.

2.2 Experiment materials

Vowel pairs /i:–ɪ/, /e:–e/, /o:–ɔ/ were elicited in two symmetrical consonant contexts: /pVp/ and /tVt/ (Table 1). Consonant context conditions the duration, quality and formant dynamics of vowels (Stevens & House 1963, Klatt 1976, Jenkins, Strange & Edman 1983, Strange, Jenkins & Johnson 1983, Sussman et al. 1997, Strange & Bohn 1998, Hillenbrand, Clark & Nearey 2001, Strange et al. 2007, Pycha & Dahan 2016). As such, we included two consonant contexts to better determine which intrinsic differences between long and short vowels are maintained across consonant contexts, and which are contingent upon surrounding consonant identity.

Table 1 Orthographic and phonemic representations of target words.

	Vowel	Labial		Coronal	
/i:–ɪ/	/i:/	/pi:p/	peep	/ti:t/	teat
	/ɪ/	/pɪp/	pip	/tɪt/	tit
/e:–e/	/e:/	/pɛ:p/	parp	/tɛ:t/	tart
	/e/	/pɛp/	pup	/tɛt/	tut
/o:–ɔ/	/o:/	/pɔ:p/	porp	/to:t/	tort
	/ɔ/	/pɔp/	pop	/tɔt/	tot

The experiment was designed to carefully control for the effects of phonetic context on vowel articulation, which necessitated the use of a combination of both words and non-words. Studies have shown that participants may hyperarticulate novel or unfamiliar words (Umeda 1975, Klatt 1976, Fowler & Housom 1987). To minimise the potential influences on articulation due to lexical status and familiarity, all participants in this task undertook two practice sessions prior to recording.

A carrier phrase was used to create an antagonistic tongue dorsum position prior to and following the target item. /i:–ɪ/ were presented within the carrier phrase *Star CVC heart* /stɛ: CVC hɛ:t/. /e:–e/ and /o:–ɔ/ were presented within the carrier phrase *See CVC heat* /si: CVC hi:t/, with focus on the target word. Prior to recording, all participants were familiarised with elicitation materials and instructed to read them aloud. If a participant pronounced the target word incorrectly or was unsure how to pronounce it, they were shown a written word that rhymed with the desired pronunciation. Participants then read each phrase from orthographic presentation on a computer screen in a sound attenuated room. Presentation was self-paced.

The 12 target words (Table 1) were divided into two blocks: block one consisted of target words containing /i:/ and /ɪ/, and block two containing /e: e o: ɔ/. Target words were randomised within blocks. Ten repetitions of each word within its carrier phrase (120 items) were elicited from each participant. Participant W1 terminated the experiment early, resulting in only eight repetitions for that participant.

2.3 Data acquisition

Articulatory data were recorded using a Northern Digital Inc. Wave Electromagnetic Articulography (EMA) system (Northern Digital Inc. 2016) at a sampling rate of 100 Hz. The placement of sensors is shown in Figure 2. Three lingual sensors were placed at the (1) tongue tip (~6 mm from anatomical tongue tip), (2) tongue body (~22 mm from tongue tip) and (3) tongue dorsum (~40 mm from tongue tip). Sensors were also placed on the (4) upper lip, (5) lower lip and (6) lower gum line, to track jaw height. Reference sensors were placed on the (7) nasion and the protrusion of the (8) left mastoid and (9) right mastoid processes. Speech audio was recorded using a Røde NT1-A shotgun microphone at a sampling rate of 22050 Hz.

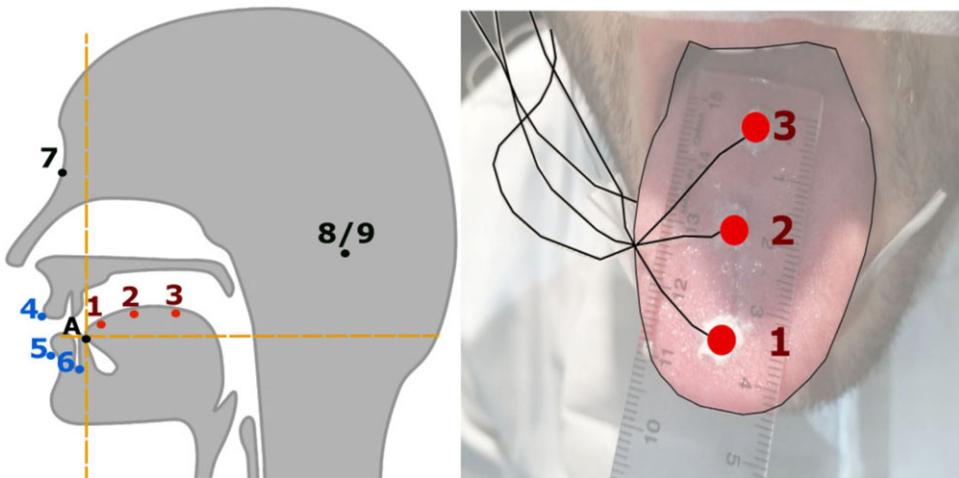


Figure 2 (Colour online) Configuration of EMA sensors. Left: Midsagittal view of sensor locations. Horizontal dashed line = occlusal plane; vertical dashed line = maxillary occlusal plane. Right: Location of the lingual sensors.

2.4 Data processing

Articulatory sensor signals were corrected for head movement and rotated to a common coordinate system defined with respect to the rear of the upper incisors using the three reference sensors. For the analysis presented in this study we used data from the tongue dorsum (TD) sensor (Sensor 3 in Figure 2). The TD sensor was chosen as it exhibited the greatest displacement during vowel gesture production for all participants and vowel pairs (see Appendix Table A1). Articulatory signals were low-pass filtered and conditioned using a DCT-based discretised smoothing spline (Garcia 2010) and synchronised with the audio data.

2.5 Acoustic segmentation

Two acoustic landmarks were identified for each vowel: acoustic onset (Figure 3: (A)) and acoustic offset (Figure 3: (B)). In each recording, RMS energy was calculated in 20 ms 75% overlapped Hamming-windowed intervals over the length of a 1.5-second interval centred on the target vowel. Working outwards from the peak RMS energy, the first and last points in time were located at which signal energy fell below 0.5% of maximum RMS energy (Tiede 2005). These acoustic estimates were superimposed on time-aligned waveforms and short-time spectrograms plotted up to 10000 Hz, and inspected and manually adjusted by a trained phonetician when necessary (approximately 5% of tokens). Vowel acoustic duration (AcDur) was calculated as the difference between acoustic limits (B–A).

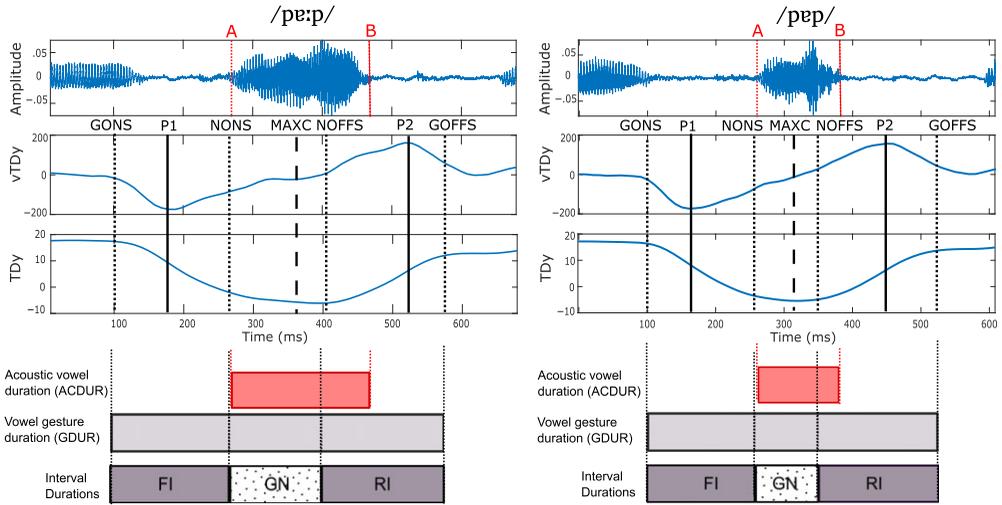


Figure 3 (Colour online) Articulatory measurements of syllables contrasting *parp* and *pup*. Items produced by participant W4. Top row: acoustic waveform of *parp* (left) and *pup* (right). vTDy: vertical velocity of tongue dorsum sensor (mm/s) and TDy: vertical displacement of tongue dorsum sensor (mm). GONS = gesture onset, P1 = velocity peak of movement towards vowel target, NONS = nucleus onset, MAXC = vowel target, point of maximum TD displacement, NOFFS = nucleus offset, P2 = peak velocity of movement away from target, GOFFS = gesture offset. Horizontal bars indicate acoustic and gesture intervals used in analysis: (i) acoustic vowel duration (AcDur), (ii) vowel gesture duration (GDur), and (iii) vowel gesture intervals: formation interval (FI), gesture nucleus (GN), release interval (RI).

2.6 Articulatory analysis

Acoustic and articulatory landmarks are illustrated for tokens of *parp* and *pup* in Figure 3. The topmost panel is the acoustic waveform, the middle panel is the velocity of the tongue dorsum (TD) sensor and the lower panel is the TD trajectory. For simplicity, velocity and displacement are shown only in the vertical dimension, however, measurements were based on the tangential velocity of the TD sensor in both horizontal (TD_x) and vertical (TD_y) dimensions. A trained phonetician located a lingual vowel gesture in each target word using the *findgest* algorithm in the MATLAB-based software package MVIEW (Tiede 2005). The *findgest* algorithm uses the tangential velocity of a given sensor to locate several gesture landmarks (Figure 3).

Gestural onset (GONS) was the point before P1 where velocity dropped to 20% of P1 velocity, nucleus onset (NONS) was the point after P1 where velocity dropped to 20% of P1 velocity, nucleus offset (NOFFS) was the point before P2 where velocity dropped to 20% of P2 velocity, gestural offset (GOFFS) was the point after P2 where velocity dropped to 15% of P2 velocity. Vowel gesture durations (GDur) spanned from vowel gesture onset (GONS) to vowel gesture offset (GOFFS).

Three intervals were demarcated in each vowel gesture (Chitoran, Goldstein & Byrd 2002, Gafos 2002): (i) Formation interval (FI) = vowel gesture onset (GONS) to gesture nucleus onset (NONS), (ii) Gesture nucleus (GN) = gesture nucleus onset (NONS) to gesture nucleus offset (NOFFS), (iii) Release interval (RI) = gesture nucleus offset (NOFFS) to vowel gesture offset (GOFFS). The duration of these three intervals were represented as proportionate durations of the entire vowel gesture duration (FI%, GN% and RI%).

The choice of the current tripartite division of vowel gestures is informed by theories of gestural grammar (Browman & Goldstein 1990, Chitoran et al. 2002, Gafos 2002, Davidson 2004). Several studies have shown that linguistic grammars have access to and utilise the internal temporal structures of vowel and consonant gestures (see Gafos 2002 for review).

Acoustic studies of vowels also utilise a tripartite division, particularly in reference to vowel length contrasts, where the durations of these three sub-vocalic intervals are important to differentiating vowel length in many languages, including AusE (Cochrane 1970, Watson & Harrington 1999, Cox 2006).

We also report Euclidean distances between the articulatory targets (TargDiff) of the three long–short vowel pairs. Euclidean distances were calculated for each participant between the centroid of the articulatory targets of each of the three long vowels (MAXC in Figure 3; [(CentroidTD_{xl}, CentroidTD_{yl})] and the individual tokens of their short equivalents [(TD_{xs_i}, TD_{ys_i})]:

$$\text{TargDiff} = \sqrt{(\text{CentroidTD}_{xl} - \text{TD}_{xs_i})^2 + (\text{CentroidTD}_{yl} - \text{TD}_{ys_i})^2}$$

A challenge of analysing articulatory data across participants is that differences in tongue shape, vocal tract size and sensor placement lead to cross-participant differences in constriction location that may not be linguistically meaningful. For example, a retraction of the TD sensor to 30 mm behind the front teeth (maxillary occlusal plane) may result in the production of a front vowel for one participant, or a back vowel for another participant, depending on the size and shape of each participant's vocal tract (Blackwood Ximenes et al. 2017). To compare across participants, Euclidean distance measures were normalised through *z*-scoring (TargDiffz), as outlined by Lobanov (1971). Lobanov's (1971) method was originally applied to vowel formants, however recently it has been applied to normalisation of EMA sensor positions (Shaw et al. 2016, Blackwood Ximenes et al. 2017).

Lindblom's (1963) target undershoot account would predict that long–short pairs with larger durational differences would also exhibit larger vowel quality differences. We therefore also included the difference between the time to target attainment of long and short vowels as a variable in our models examining vowel quality across the three long–short pairs. Difference in time to target attainment (ms; TimeTargDiff) was calculated for each participant between the average time to target of each of the three long vowels [AverageTimetoTarg_l] (GONS to MAXC in Figure 3) and time to target of individual tokens of their short equivalents [TimetoTarg_{xs_i}].

Lip protrusion was used as a measure of lip rounding in the present study, in line with previous work by Blackwood Ximenes et al. (2017). Degree of lip protrusion of /o:/ and /ɔ:/ was calculated based on the average horizontal position of the UL and LL sensors (Figure 2 above) measured at the target of the lingual gesture of the vowel (MAXC; Figure 3). This average horizontal position was *z*-transformed by participant.

2.6.1 Data exclusion

A total of 816 target words were elicited (12 target words × 10 repetitions × 6 participants) + (12 items × 8 repetitions × 1 participant). In six coronal context tokens there were more than two velocity peaks on the TD sensor trajectory between the maximum constrictions of the onset and coda consonant: these tokens were excluded from further analysis. Seven further items were excluded due to mispronunciation and sensor tracking errors, leaving a total of 803 analysed target words.

2.7 Statistical analysis

Statistical tests were applied in R using the *lme4* (Bates et al. 2015) and *lmerTest* (Kuznetsova, Brockhoff & Christensen 2017) packages.

For the dependent variables of (i) acoustic duration (AcDur), (ii) gesture duration (GDur), (iii) distance between articulatory targets (TargDiffz), (iv) lip protrusion of /o:/ and /ɔ:/ (LPz), (v) proportionate formation interval duration (FI%), (vi) proportionate gesture nucleus duration (GN%) and (vii) proportionate release interval duration (RI%), we constructed linear

mixed effects regression models with independent variables of vowel length (long, short), vowel pair (/i:-ɪ/, /e:-ɛ/, /o:-ɔ/) and consonant context (labial – /pVp/, coronal – /tVt/) with a three way interaction. When exploring TargDiffz time to vowel target was also included as a potential predictor.

To find the optimal model for each dependent variable, we explored top down, step-wise model building strategies, where a model was compared with another model one order less complex, using log-likelihood ratios. Final models only included main effects and interactions that significantly improved model fit ($p > .050$). Participant differences were modelled using random intercepts for participant and repetition. In cases where a full random-effects structure resulted in model convergence issues or a singular fit, the random effect with the lowest variance was removed; this is in line with recommendations by Barr et al. (2013) and Bates et al. (2015). The random components of models were not of further interest and are not reported.

P-values for main effects were obtained through maximum likelihood tests with Satterthwaite approximations to degrees of freedom (Kuznetsova et al. 2017). Because the variable vowel pair had three levels (/i:-ɪ/, /e:-ɛ/ and /o:-ɔ/), we also conducted individual pairwise least-mean squares regression analysis (with Holm-Bonferroni corrections) using the *emmeans* package (Lenth 2019). This facilitated the comparison of the main effect of vowel pair, and interactions between vowel length and vowel pair and consonant context and vowel pair. For pairwise analysis, factors were coded as: vowel length: LONG = 0 and consonant context: LABIAL = 0. For vowel pair analysis: /i:-ɪ/ = 0. For the comparison between /e:-ɛ/ and /o:-ɔ/, /e:-ɛ/ = 0. Full summaries of all linear mixed effects models are provided in Appendix Tables A2–A8.

Euclidean distance measures are an incomplete measure of vowel target similarity as they fail to take into account distribution differences across the different vowels. Two vowel pairs may exhibit a similar distance between their centroids but due to different overall distributions of individual token values may have vastly different degrees of overlap (Warren 2017). To overcome issues of different distributions across different vowels, Pillai-Bartlett scores have been used to examine spectral overlap in ongoing vowel mergers in acoustic literature (Hay, Warren & Drager 2006, Hall-Lew 2010, Nance 2011, Wong 2012, Havenhill 2015). The Pillai-Bartlett score is one of the test statistics of MANOVAs. The higher the value of the Pillai-Bartlett score, the greater the difference between the two analysed distributions with respect to the dependent variables of the MANOVA (Hay et al. 2006, Hall-Lew 2010). Three MANOVA models were constructed (one for each vowel pair), with dependent variables of *z*-transformed TD fronting (TD_{xz}) and TD height (TD_{yz}) with the following equation:

$$(TD_{xz}, TD_{yz}) \sim \text{vowel length} \times \text{consonant context}$$

Finally, because speech rate was not actively controlled during this experiment, we wished to determine whether differences in speech rate contributed to the observed differences in measured variables. We measured the onset of the target word to the onset of the target word in the following trial (token-to-token duration) as an approximation for speech rate. Token-to-token duration was a poor predictor in all the models analysed in this study, and as such was removed from all models during the model selection process. Appendix Figure A1 shows the correlation between token-to-token duration and the dependent variables analysed in this study.

3 Results

3.1 Vowel durations

3.1.1 Acoustic duration

We first wished to confirm that participants in the present study produced short vowels with a shorter acoustic duration than their long equivalents, in line with previous studies of vowel

Table 2 Mean acoustic durations (ms, AcDur, Figure 3), gesture durations (ms, GDur, Figure 3) and proportionate durations of formation intervals, gesture nuclei and release intervals for all vowels averaged across participants. Standard deviations in parentheses. Formation interval (FI), gesture nucleus (GN) and release interval (RI) durations expressed as a proportion of total vowel gesture durations (GDur).

Vow	Cons	Count	Total duration (ms)		Interval duration (ms)			Interval duration (%)		
			AcDur	GDur	FI	GN	RI	FI%	GN%	RI%
/i:/	all	137	106 (20)	424 (89)	175 (49)	104 (45)	149 (51)	42 (9)	23 (8)	36 (9)
	lab	70	106 (22)	467 (76)	205 (39)	120 (42)	154 (50)	44 (8)	24 (7)	33 (8)
	cor	67	106 (19)	379 (79)	141 (36)	87 (42)	143 (52)	40 (8)	22 (8)	39 (9)
/ɪ/	all	135	74 (21)	381 (97)	155 (50)	69 (25)	152 (60)	40 (9)	18 (5)	41 (10)
	lab	69	72 (19)	419 (71)	188 (39)	81 (22)	161 (60)	40 (7)	20 (6)	40 (9)
	cor	66	74 (20)	341 (104)	138 (58)	52 (20)	139 (57)	41 (10)	17 (4)	42 (10)
/e:/	all	137	160 (26)	436 (104)	171 (48)	114 (49)	150 (63)	40 (10)	26 (9)	34 (9)
	lab	68	155 (27)	491 (89)	174 (50)	143 (50)	175 (61)	36 (9)	29 (9)	35 (9)
	cor	69	164 (25)	383 (87)	169 (45)	85 (25)	125 (55)	44 (8)	23 (7)	32 (9)
/ɛ/	all	133	91 (22)	386 (88)	169 (47)	66 (36)	149 (64)	45 (12)	17 (8)	38 (11)
	lab	66	88 (21)	429 (73)	173 (47)	87 (38)	169 (57)	41 (11)	21 (8)	39 (9)
	cor	67	95 (22)	344 (81)	165 (46)	47 (18)	130 (65)	49 (11)	14 (6)	37 (12)
/o:/	all	134	154 (25)	430 (101)	161 (43)	126 (61)	145 (59)	39 (12)	28 (10)	33 (8)
	lab	67	147 (23)	510 (57)	160 (43)	169 (55)	183 (51)	32 (8)	33 (10)	35 (8)
	cor	67	158 (26)	352 (67)	162 (44)	82 (27)	107 (39)	46 (10)	24 (7)	30 (7)
/ɔ/	all	133	93 (21)	404 (97)	167 (46)	77 (35)	162 (64)	42 (10)	19 (7)	39 (9)
	lab	67	86 (19)	460 (68)	172 (45)	100 (32)	192 (54)	37 (9)	22 (7)	41 (8)
	cor	66	99 (21)	347 (90)	163 (47)	53 (18)	132 (60)	47 (9)	16 (5)	37 (10)
/i:–ɪ/		272	90 (26)	403 (97)	166 (50)	88 (41)	150 (55)	41 (9)	21 (7)	38 (10)
/ɛ:–ɛ/		270	126 (42)	411 (99)	170 (47)	90 (49)	150 (63)	42 (11)	22 (9)	36 (10)
/o:–ɔ/		267	122 (38)	430 (100)	164 (45)	101 (56)	153 (62)	40 (11)	22 (9)	36 (9)
Labial		407	1090 (38)	463 (79)	174 (46)	119 (53)	174 (57)	38 (10)	25 (9)	37 (9)
Coronal		402	117 (40)	358 (86)	160 (46)	68 (31)	127 (56)	44 (10)	19 (8)	36 (10)
Long		407	140 (34)	430 (98)	168 (47)	116 (54)	148 (59)	40 (10)	26 (9)	34 (9)
Short		402	86 (23)	391 (94)	166 (47)	71 (34)	155 (63)	42 (11)	18 (7)	39 (10)

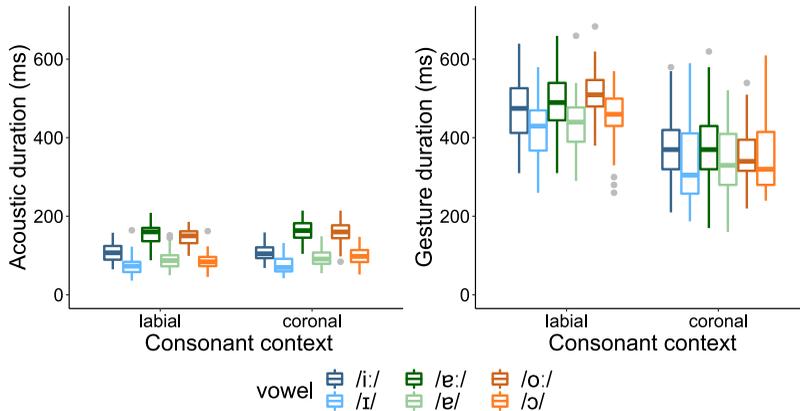


Figure 4 (Colour online) Grand mean acoustic (left) and gesture durations (right) of /i:-ɪ/, /e:-ɐ/, /o:-ɔ/ in labial (/pVp/) and coronal (/tVt/) consonant contexts. Mean durations (ms) calculated from all vowels produced by all participants in each consonant context. Acoustic duration = acoustic onset to acoustic offset (AcDur, see Figure 3), gesture duration = vowel gesture onset to vowel gesture offset (GDur, see Figure 3).

length in AusE (Cox 2006, Elvin et al. 2016). A linear mixed effects model was constructed using the method described in Section 2.7 with the following equation:

$$\text{AcDur} \sim \text{Vlength} + \text{Vpair} + \text{Conscontext} + (\text{Vlength} \times \text{Vpair}) \\ + (\text{Conscontext} \times \text{Vpair}) + (\text{Vlength}|\text{participant})$$

A full model summary is provided in Appendix Table A2.

Mean acoustic duration of short vowels was 62% of the mean acoustic duration of long vowels ($F(796) = 1720.9, p < .001$; Table 2, Figure 4).

On average, /ɪ/ was 70% the acoustic duration of /i:/, /ɐ/ was 57% the acoustic duration of /e:/ and /ɔ/ was 61% the acoustic duration of /o:/. There was a vowel length \times vowel pair interaction ($F(795) = 67.3, p < .001$) indicating that the magnitude of difference between long and short vowels differed by vowel pair. The acoustic duration difference between /i:/ and /ɪ/ was smaller than the acoustic duration difference between /e:/ and /ɐ/ ($\beta = -35$ ms, $t(802) = -11.0, p < .001$) and the acoustic duration difference between /o:/ and /ɔ/ ($\beta = -27$ ms, $t(796) = -8.5, p < .001$). The acoustic duration difference between /e:/ and /ɐ/ was also greater than the acoustic duration difference between /o:/ and /ɔ/ ($\beta = 8$ ms, $t(796) = 2.5, p = .012$).

Overall, coronal context vowels exhibited longer acoustic durations than labial context vowels ($F(796) = 30.2, p < .001$). However the magnitude of the effect of consonant context differed across the three vowel pairs ($F(796) = 7.2, p < .001$). The acoustic duration of /i:-ɪ/ did not differ between labial and coronal contexts ($p = .765$). The acoustic duration of /e:-ɐ/ was longer in coronal than in the labial context ($\beta = 8.3$ ms, $t(802) = 3.7, p = .001$), this was also the case for /o:-ɔ/ ($\beta = 12.6$ ms, $t(802) = 5.6, p < .001$). The acoustic duration of /e:-ɐ/ and /o:-ɔ/ increased to a similar extent in the coronal context ($p = .345$).

Our results regarding vowel length are therefore congruent with prior acoustic studies of vowel length in AusE.

3.1.2 Gesture duration

Our first prediction was that lingual gestures of short vowels should be shorter than those of long vowels (Cox 2006, Elvin et al. 2016). To test this, a linear mixed effects model was constructed using the method described in Section 2.7 with the following equation:

$$\text{GDur} \sim \text{Vlength} + \text{Vpair} + \text{Conscontext} + (\text{Vlength} \times \text{Conscontext}) \\ + (\text{Conscontext} \times \text{Vpair}) + (1|\text{participant}) + (1|\text{repetition})$$

A full model summary is provided in Appendix Table A3.

Mean duration of short vowel gestures was 90% of the mean duration of long vowel gestures ($F(786) = 59.3$, $p < .001$; Table 2, Figure 4). The magnitude of difference between long and short vowels differed across labial and coronal contexts ($F(787) = 7.2$, $p = .007$). The difference in gesture duration between long and short vowels was smaller in coronal than in the labial context ($\beta = 28$ ms, $t(787) = 2.7$, $p = .048$).

Vowel gesture durations were shorter in the coronal than the labial context ($F(787) = 405.7$, $p < .001$). The gesture duration of all vowel pairs were shorter in the coronal than in the labial context, but the magnitude of the effect of consonant context on vowel gesture duration differed across vowel pairs ($F(787) = 8.7$, $p < .001$). Consonant context had the largest effect on the gesture duration of /o:-ɔ/. The gesture duration of /o:-ɔ/ shortened to a greater extent in the coronal context than the gesture duration of both /i:-i/ ($\beta = -48.2$ ms, $t(787) = 3.7$, $p = .001$) and to a greater extent than /e:-e/ ($\beta = -38.2$ ms, $t(787) = 3.0$, $p = .019$). /i:-i/ and /e:-e/ shortened to a similar extent in the coronal (compared to the labial) context ($p = .544$).

3.2 Articulatory targets

Our second prediction posited that /e:-e/ will exhibit the most similar articulatory targets of the three vowel pairs, whereas /i:-i/ and /o:-ɔ/ will have less similar pairwise articulatory targets. Our third prediction posited that vowel duration and vowel quality would exhibit a trading relationship in AusE, such that vowel pairs with the largest acoustic duration difference would exhibit the smallest difference in target quality and vice versa. To determine this, the similarity in articulatory target tongue dorsum positions were compared for the three long-short vowel pairs produced in labial (/pVp/) and coronal (/tVt/) contexts (Table 3, Figure 5), using the method illustrated in Section 2.6. The z -transformed absolute Euclidean distance between the targets of long and short vowel pairs (TargDiffz) was modelled using the method described in Section 2.7, with the following equation:

$$\text{TargDiffz} \sim \text{VPair} + \text{Conscontext} + (1|\text{participant})$$

The duration difference between time to long and short vowel target (TimeTargDiff) did not improve model fit ($p = .536$), so was not included in the present model. A full model summary is provided in Appendix Table A4.

TargDiffz differed by vowel pair ($F(376) = 31.8$, $p < .001$). TargDiffz was greater between /i:/ and /i/ than between /e:/ and /e/ ($\beta = -0.7$, $t(372) = -3$, $p < .001$) but smaller than TargDiffz between /o:/ and /ɔ/ ($\beta = 0.3$, $t(372) = 2.4$, $p = .017$; Table 3, Figure 5). The TargDiffz between /e:/ and /e/ was also smaller than the TargDiffz between /o:/ and /ɔ/ ($\beta = 1.4$, $t(372) = 7.9$, $p < .001$). Vowels produced in the coronal context had larger TargDiffz than vowels produced in the labial context ($F(376) = 5.9$, $p = .016$).

Overlap between the distributions of long and short vowel targets was also compared using Pillai-Bartlett scores. Pillai-Bartlett scores are shown in Table 3. Lower Pillai scores indicate more overlap between two distributions. /e:-e/ exhibited the lowest Pillai-Bartlett scores (0.24) of the three vowel pairs, while /i:-i/ (0.47) and /o:-ɔ/ (0.48) exhibited similar scores.

Table 3 Mean Euclidean distances (TargDiff) between articulatory targets (MAXC, 3) of the three long-short vowel pairs and Pillai-Bartlett scores. TargDiff (mm) calculated from all vowels produced by all participants in each consonant context (lab = labial, cor = coronal). TargDiffz (z-transformed) are Euclidean distances z-transformed by participant. Pillai-Bartlett scores represent degree of overlap between two distributions. Lower values indicate more overlap between two distributions. All values averaged across participants. Standard deviations in parentheses.

		Count	TargDiff (mm)		TargDiffz (z-transformed)		Pillai-Bartlett
/i:–ɪ/	all	251	3.24	(1.64)	1.39	(0.98)	0.47
	lab	125	3.20	(1.66)	1.35	(0.97)	0.46
	cor	126	3.28	(1.63)	1.44	(1.00)	0.50
/e:–ɐ/	all	254	2.48	(1.15)	0.82	(0.70)	0.24
	lab	125	2.17	(1.02)	0.74	(0.60)	0.16
	cor	129	2.78	(1.21)	0.89	(0.78)	0.36
/o:–ɔ/	all	255	3.80	(1.55)	1.97	(1.54)	0.48
	lab	127	3.71	(1.88)	1.88	(1.66)	0.47
	cor	128	3.89	(1.82)	2.05	(1.40)	0.52

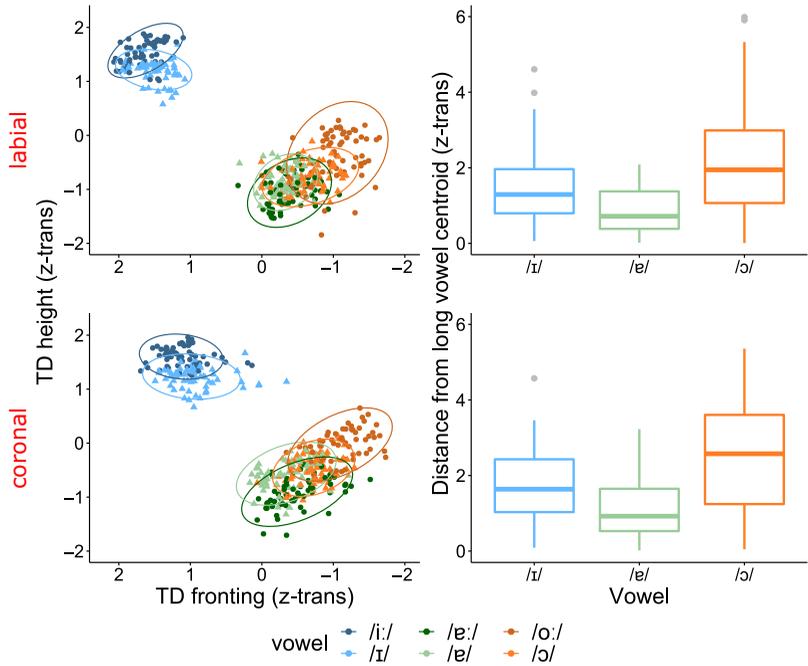


Figure 5 (Colour online) Left: TD sensor position at articulatory target (MAXC, 3) of /i:–ɪ/, /e:–ɐ/ and /o:–ɔ/ in labial (/pVp/) and coronal (/tVt/) consonant contexts. TDx and TDy z-transformed by participant. Right: Euclidean distance from long vowel centroid to short vowel articulatory target in labial (/pVp/) and coronal (/tVt/) consonant contexts. TDx and TDy z-transformed by participant.

3.3 Lip rounding differences between /o:/ and /ɔ/

Our fourth prediction posited that lip rounding should be greater for /o:/ than /ɔ/. To investigate this, we compared lip protrusion of /o:/ and /ɔ/. Lip protrusion was calculated as the average horizontal position of the UL and LL sensors at the lingual target of the two

vowels z -transformed across participants. Differences in lip protrusion between /o:/ and /ɔ/ was modelled using the method described in Section 2.7 using the following equation:

$$\text{LPz} \sim \text{Vlength} + \text{ConsContext} + (1|\text{participant}) + (1|\text{repetition})$$

A full model summary is provided in Appendix Table A5.

Overall, /o:/ was produced with more lip protrusion than /ɔ/ ($F(198) = 143.4, p < .001$; Figure 6). Z -transformed lip protrusion was also greater for coronal than labial tokens ($F(198) = 10.3, p = .002$), suggesting greater lip rounding for /o:/ than /ɔ/.

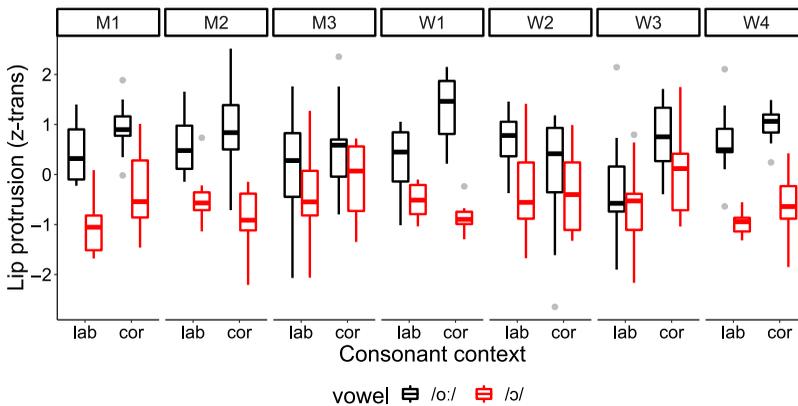


Figure 6 (Colour online) By-participant lip protrusion at target of /o:/ and /ɔ/ in labial and coronal contexts. Averaged across UL and LL sensors and repetitions. Lip protrusion z -transformed by participant. Greater lip protrusion indicates a greater degree of rounding.

3.4 Interval durations

Our final prediction was that, in line with acoustic studies (Watson & Harrington 1999, Cox 2006, Elvin et al. 2016), long and short vowel gestures will be characterised by different dynamic articulatory patterns. Short vowels should exhibit a proportionately shorter period of articulatory stability around their mid-points and proportionately longer articulatory transitions to following consonants. However, the long vowel /i:/ should be characterised by a lengthy phonological onglide as is characteristic of AusE (Cox 2006). To determine whether this was the case, proportionate durations of three intervals within each vowel gesture were compared across the three long–short vowel pairs using a linear mixed effects model constructed using the method described in Section 2.7. Full model summaries are provided in Appendix Tables A6–A8. Absolute and proportionate durations of the three sub-gestural intervals are presented in Table 2 and Figures 7 and 8. However, statistical analyses were undertaken only for the proportionate formation interval, gesture nucleus and release interval durations (FI%, GN% and RI%, respectively).

Figure 7 compares tongue dorsum displacement throughout the vowel gesture for each pair of vowels. For each vowel, TD displacement with respect to dorsal location at the (i) vowel gesture onset (GONS) is tracked at three gesture landmarks: (ii) Nucleus onset (NONS), (iii) Nucleus offset (NOFFS), and (iv) Gesture offset (GOFFS). At each landmark, displacement is calculated as mean Euclidean distance in the midsagittal plane between TD_{xy} and TD_{xy} at GONS. Timing of landmarks is expressed as a proportion of total vowel gesture duration (GDur).

We discuss results for each of the three intervals separately below.

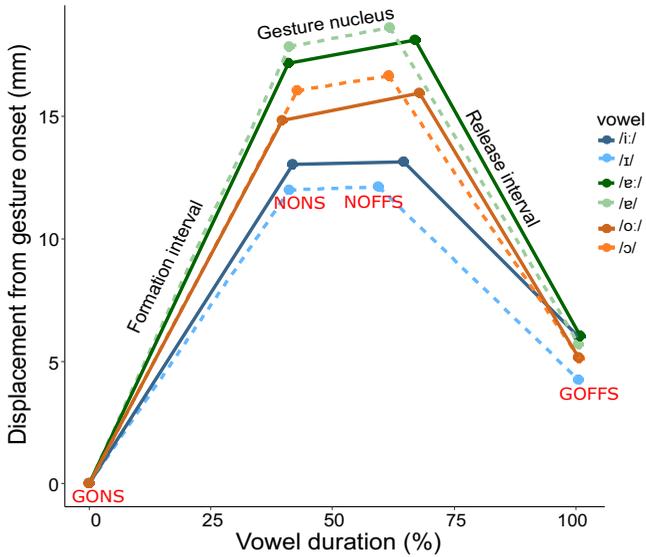


Figure 7 (Colour online) Euclidean TD displacement throughout vowel gesture for /i:–ɪ/, /e:–ɛ/ and /o:–ɔ/. TD displacement measured from origin (gons; Figure 3). Displacement measured at four landmarks: (i) gesture onset (GONS), (ii) nucleus onset (NONS), (iii) nucleus offset (NOFFS), and (iv) gesture offset (GOFFS). Gesture landmarks located using criteria illustrated in Figure 3. Vowel duration expressed as a proportion of total vowel gesture duration (GDur, Figure 3). Mean displacement (mm) calculated from all vowels produced by all participants in both consonant contexts.

3.4.1 Formation interval

A linear mixed effects model was constructed using the method described in Section 2.7 with the following equation:

$$FI\% \sim Vlength + Vpair + Conscontext + (Vlength \times Vpair) + (Conscontext \times Vpair) + (Vlength \times Conscontext) + (1|participant)$$

A full model summary provided in Appendix Table A6.

As shown in Figure 8, vowel length conditioned FI% of the three vowel pairs differently across the two consonant contexts ($F(796) = 6.0, p = .003$).

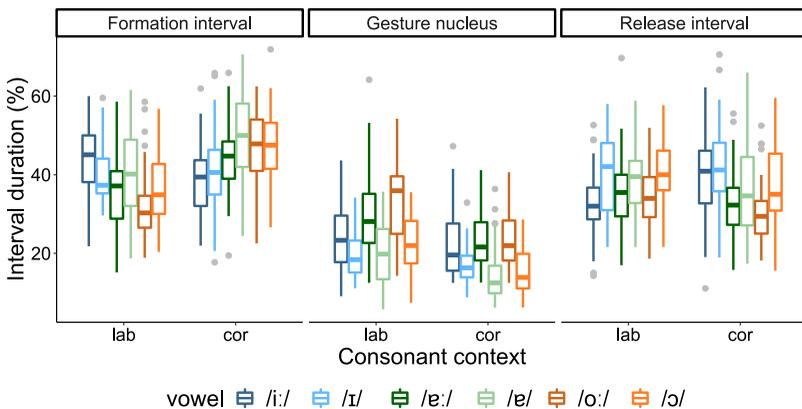


Figure 8 (Colour online) Formation interval (FI), gesture nucleus (GN) and release interval (RI) of /i:–ɪ/, /e:–ɛ/ and /o:–ɔ/. Durations expressed as proportion of entire vowel gesture duration (GDur). Intervals determined as shown in Figure 3.

In the labial context, /i:/ had a longer FI% than /ɪ/ ($\beta = -4\%$, $t(796) = -2.6$, $p = .030$). While in the coronal context, FI% of /i:/ did not differ from FI% of /ɪ/ ($p = .300$). FI% of /e:/ was shorter than FI% of /ɛ/ in both labial ($\beta = 5\%$, $t(807) = 3.2$, $p = .008$) and coronal contexts ($\beta = 5\%$, $t(807) = 3.1$, $p = .009$). The magnitude of difference between /e:/ and /ɛ/ did not differ across labial and coronal context ($p = .999$). FI% of /o:/ was shorter than FI% of /ɔ:/ in the labial context ($\beta = 5\%$, $t(807) = 3.4$, $p = .004$), but did not differ in the coronal context ($p = .898$). In the labial context, the magnitude of difference in FI% between /e:/ and /ɛ/ did not differ from the magnitude of difference between /o:/ and /ɔ:/ ($p = .858$).

3.4.2 Gesture nucleus

A linear mixed effects model was constructed using the method described in Section 2.7 with the following equation:

$$\text{GN\%} \sim \text{Vlength} + \text{Vpair} + \text{Conscontext} + (\text{Vlength} \times \text{Vpair}) \\ + (\text{Conscontext} \times \text{Vpair}) + (1|\text{participant})$$

A full model summary is provided in Appendix Table A7.

Short vowels had shorter GN% than long vowels ($F(796) = 236.7$, $p < .001$; Figure 8, Table 2). There was also a vowel length \times vowel pair interaction, indicating that the magnitude of difference between long and short vowel GN% differed across the three vowel pairs ($F(796) = 8.9$, $p < .001$). /i:-ɪ/ had the smallest pairwise difference in GN% of the three pairs. The difference in GN% between /i:/ and /ɪ/ was smaller than the difference between /e:/ and /ɛ/ ($\beta = -4\%$, $t(796) = -3.2$, $p = .007$), also smaller than the difference in GN% between /o:/ and /ɔ:/ ($\beta = -5\%$, $t(796) = -3.9$, $p = .001$). The difference in GN% between /e:/ and /ɛ/ was equivalent to the difference between /o:/ and /ɔ:/ ($p = .693$). GN% was also shorter for coronal context than labial context vowels ($F(796) = 125.3$, $p < .001$). However, there was also a consonant context \times vowel pair interaction, indicating that consonant context conditioned GN% of the three vowel pairs to different extents ($F(796) = 8.9$, $p < .001$; Figure 8, Table 2). While the GN% of all three vowel pairs decreased in the coronal compared to labial contexts, GN% of /i:-ɪ/ differed less across labial and coronal contexts than /e:-ɛ/ ($\beta = -4\%$, $t(804) = -3.6$, $p = .001$) and /o:-ɔ:/ ($\beta = -6\%$, $t(804) = -5.1$, $p < .001$). GN% of /e:-ɛ/ and /o:-ɔ:/ decreased by a similar extent in the coronal (compared to the labial) context ($p = .140$).

3.4.3 Release interval

A linear mixed effects model was constructed using the method described in Section 2.7 with the following equation:

$$\text{RI\%} \sim \text{Vlength} + \text{Vpair} + \text{Conscontext} + (\text{Conscontext} \times \text{Vpair}) + (1|\text{participant})$$

Full model summary provided in Appendix Table A8.

Short vowels had greater RI% than long vowels ($F(796) = 70.8$, $p < .001$; Figure 8, Table 2). Overall, there was no difference in RI% across labial and coronal contexts ($p = .237$). However, there was a consonant context \times vowel pair interaction, indicating that the effect of consonant context on RI% differed across vowel pairs ($F(796) = 19.5$, $p < .001$; Figure 8). RI% of /i:-ɪ/ was longer in the coronal context than labial context ($\beta = 5\%$, $t(802) = 4.3$, $p < .001$). While RI% of /e:-ɛ/ tended to be shorter in the coronal context ($\beta = -2\%$, $t(802) = -2.2$, $p = .083$). RI% of /o:-ɔ:/ was also shorter in the coronal than the labial context ($\beta = -4\%$, $t(802) = -4.1$, $p < .001$). Both /e:-ɛ/ and /o:-ɔ:/ decreased to a similar extent in the coronal (compared to the labial) context ($p = .346$).

3.5 Summary of main findings

This study compared the lingual articulatory properties of three long–short vowel pairs /i:–ɪ/, /e:–ɛ/ and /o:–ɔ/ in two symmetrical consonant contexts /pVp/ and /tVt/. The main findings of this study are:

- acoustic duration⁴ of short vowels was 62% the acoustic duration of long vowels
- gesture duration (measured as GDur) of short vowels was 90% that of long vowels
- /e:–ɛ/ had the greatest pairwise difference in acoustic duration, while /i:–ɪ/ had the smallest pairwise difference in acoustic duration
- the difference between long and short vowel gesture durations was larger in the labial than in the coronal context
- /e:/ and /ɛ/ were produced with the most similar mean articulatory targets and the most overlapping long–short distributions; /o:/ and /ɔ/ were produced with the least similar articulatory targets
- contrasts between long and short vowel FI% and GN% were reduced in coronal compared to labial contexts
- FI% was longer for /i:/ than for /ɪ/, but longer for /ɛ/ and /ɔ/ than /e:/ and /o:/ in the labial context
- GN% was shorter for short vowels, compared to their long equivalents
- pairwise difference in GN% was smallest between /i:–ɪ/, and equivalent between /e:–ɛ/ and /o:–ɔ/
- RI% was longer for short vowels, compared to their long equivalents

4 Discussion

The aim of this study was to investigate lingual articulation of vowel length contrasts in AusE, building on previous, largely acoustic description of AusE vowel contrasts. These data provide an articulatory characterisation of some key aspects of vowel length contrasts in AusE, revealing new insights into AusE production kinematics.

4.1 Gesture duration

We first explored the impact of contrastive vowel length on vowel gesture durations. Our first prediction was that in line with acoustic durations, the duration of short vowel gestures would be shorter than those of long vowel gestures, but that the duration difference between long and short vowels should be reduced in the articulatory domain. Our results confirmed this prediction. On average short vowels were 62% the acoustic duration of long vowels, in line with previous acoustic studies of AusE vowel length (Cox 2006, Elvin et al. 2016), while short vowel gestures were 90% the duration of long vowel gestures. Hertrich & Ackermann (1997) have speculated that the discrepancy between acoustic duration and gesture durations indicates that phonological vowel length contrast is not produced as a difference in either the duration of laryngeal activity (vowel voicing) or supralaryngeal (lips, tongue, jaw) movement alone, but rather reflects a vowel length dependent difference in the coordination of laryngeal and supralaryngeal gestures.

The gesture durations of coronal context vowels were 77% the duration of labial context vowels. This result is congruent with findings that show coronal consonants constrain the production of following vowels to a greater degree than labial consonants (Recasens, Pallarès & Fontdevila 1997, Sussman et al. 1997, Löfqvist 1999, Fowler & Brancazio 2000, Recasens 2002, Harrington et al. 2011, Harrington, Kleber & Reubold 2011). Due to the relative independence of the lips and tongue dorsum, vowel gestures in labial contexts can begin earlier

⁴ See Figure 3 for explanation of durations and landmarks.

than those in coronal contexts, which can result in a longer duration as has been observed in this study. There was also a smaller difference between the duration of long and short vowel gestures in the coronal than in the labial context. Across the two consonant contexts, the duration of short vowel gestures was less conditioned by consonant context than the duration of long vowel gestures, resulting in a reduction of contrast in the coronal context. This finding is consistent with general observations that the duration of short vowels is more stable than the duration of long vowels across different speech rates, prominences and phonetic contexts (Klatt 1973, Port 1981, Gopal 1990, Fletcher et al. 1994, Hoole, Mooshammer & Tillmann 1994, Hoole & Mooshammer 2002, Jong & Zawaydeh 2002, Mooshammer & Fuchs 2002, Hirata 2004, White & Mády 2008, Nakai et al. 2009, Beňuš 2011, Cox & Palethorpe 2011, Cox et al. 2015, Peters 2015, Penney et al. 2018).

However, the shorter gesture duration of coronal context vowels contradicts our acoustic duration results, where coronal context vowels exhibited a longer acoustic duration than labial context vowels. While this has also been observed in other acoustic studies of English (House & Fairbanks 1953, Lehiste & Peterson 1961, Port 1981), the discrepancy between acoustic and gesture durations once again suggests that the relationship between acoustic and articulatory landmarks in vowel production are sensitive to factors such as vowel length and consonant context.

4.2 Articulatory target similarity

In Section 3.2, we compared articulatory targets of long and short vowel pairs in AusE. Acoustic studies of AusE have shown that long–short vowel pairs differ in the degree of spectral similarity, with /e:–e/ the least spectrally differentiated and /o:–ɔ/ the most spectrally differentiated (Bernard 1970, Harrington & Cassidy 1994, Watson & Harrington 1999, Cox 2006, Cox et al. 2014, Elvin et al. 2016, Cox & Fletcher 2017). Our second prediction was that although all long–short vowel pairs should be produced with similar articulatory targets, the degree of similarity should differ by vowel pair. Of the three long–short vowel pairs, /e:–e/ was predicted to exhibit the most similar articulatory targets, while /o:–ɔ/ was predicted to be realised with the least similar articulatory targets. This was indeed the case, /e:–e/ had the shortest Euclidean distance between vowel targets and the most overlapping distributions of the three vowel pairs (Figure 5). /o:–ɔ/ had the largest Euclidean distance between long and short targets and the least overlapping distributions. However, Euclidean distance values were also highly variable for /o:–ɔ/ (Figure 5), which may indicate participant-specific strategies for the production of this pair, with some participants producing the pair with more articulatorily distinct targets than others (Appendix Figure A2).

There was a larger difference in long and short vowel target quality in coronal than in the labial context (Figure 7). Prior studies have suggested that short vowels may be more coarticulated with following consonants than their long equivalents (Hoole & Mooshammer 2002), resulting in short vowels exhibiting more target quality variation across consonant contexts than their long equivalents. Future studies should examine interactions between consonant context and articulation of AusE vowels in more detail.

We also observed that /o:/ exhibited greater lip protrusion than /ɔ/, suggesting that /o:/ is more rounded than /ɔ/. This is congruent with Blackwood Ximenes et al.'s (2017) observations of lip rounding differences between /o:/ and /ɔ/ in three speakers of AusE. Blackwood Ximenes et al. (2017) have suggested that differences in lip rounding between /o:–ɔ/ may also contribute to F1 and F2 differences between the pair, raising and retracting /o:/ in the acoustic space relative to /ɔ/ independent of lingual adjustments. This may also be the case here; however, the tongue dorsum position of /o:/ was still higher and retracted compared to /ɔ/. There was also variation in the degree of lip protrusion differences across participants. M3 and W3 produced /o:–ɔ/ with overlapping lip protrusion values. This once again highlights potential speaker-specific strategies in the production of these vowels. Although more

research is needed to determine whether overlapping lip protrusion and/or tongue dorsum postures between /o:/ and /ɔ/ are reflected in overlapping F2 values in these speakers.

4.3 Trade-offs between acoustic duration and articulatory target

In languages such as Japanese, Swedish and Thai, acoustic vowel duration and spectral quality have a trading relationship as cues to vowel length (Delattre 1962, Hadding-Koch & Abramson 1964, Lehnert-LeHouillier 2010); that is, the more differentiated the acoustic targets of a long–short vowel pair, the less listeners rely on durational cues and vice versa (Hadding-Koch & Abramson 1964). In line with these studies, our third prediction was that the vowel pair with the largest pairwise difference in duration would have the most similar articulatory targets and vice versa. Our results partially support this prediction. /e:–e/ had the largest pairwise difference in acoustic duration, and the most similar articulatory targets of the three vowel pairs. This is consistent with prior studies that have shown that vowels in this pair differ primarily in acoustic duration, and have largely overlapping acoustic targets (Bernard 1970, Cochrane 1970, Harrington & Cassidy 1994, Watson & Harrington 1999, Cox 2006, Elvin et al. 2016). However, /i:–i/ had the smallest pairwise difference in acoustic duration, but /o:–ɔ/ had the least similar articulatory targets. This result is not unexpected for two reasons. First, /o:–ɔ/ can be differentiated by acoustic target quality alone, independent of durational information (Watson & Harrington 1999). Second, while duration is important for differentiating /i:/ and /ɪ/ in AusE, there are also dynamic formant differences (namely /i:/’s prolonged acoustic onglide) that also serve to further differentiate /i:/ from /ɪ/ (Harrington & Cassidy 1994, Harrington et al. 1997, Watson & Harrington 1999, Cox 2006, Cox et al. 2014, Cox et al. 2015). The previously observed trade-off between acoustic duration and spectral quality as cues to vowel length (Delattre 1962, Hadding-Koch & Abramson 1964, Lehnert-LeHouillier 2010) may rather be a trade-off between durational and non-durational cues to vowel length contrast, with the dynamic differences between /i:/ and /ɪ/ contributing to this trading relationship.

Our findings also challenge a purely physiological account of vowel quality differences between long and short vowels, such as that proposed in Lindblom’s (1963) target undershoot model. First, in a target undershoot account, we would expect the vowel pair with the largest durational differences to exhibit the largest vowel quality differences. However, this was not the case for either acoustic duration or gesture duration. As mentioned above, /e:–e/ had the largest difference in acoustic duration, and the smallest pairwise difference in vowel quality. In terms of gesture duration, the difference between long and short vowels was similar across the three vowel pairs. Furthermore, in a target undershoot account, we would predict that the difference in duration to the time of gestural target would be a predictor of differences in target quality (TargDiffz). Our results do not support this account. Difference in time to target was not a significant predictor of vowel quality differences across our three vowel pairs.

4.4 Kinematic differences

Finally, we examined dynamic kinematic differences between long and short vowel gestures. Previous acoustic production studies have found differences in the formant dynamics of long and short AusE vowels suggesting differences in articulatory kinematics (Bernard 1970, Cochrane 1970, Watson & Harrington 1999, Cox 2006). We predicted that short vowel gestures would have a proportionately shorter period of articulatory stability around their mid-points and proportionately longer articulatory transitions to following consonants than their long equivalents. In our comparison of these intervals in long and short vowel gestures, short vowel gestures indeed had proportionately shorter gesture nuclei and proportionately longer release intervals than long vowel gestures (Figures 7 and 8). These data provide the first articulatory evidence supporting acoustic studies that have found AusE short vowels

to have a proportionately shorter acoustic steady-states and proportionately longer acoustic offglides than long vowels (Cox 2006).

We also posited that /i:/ would exhibit a prolonged phonological onglide as is characteristic of AusE (Harrington et al. 1997, Cox 2006, Cox et al. 2014). Our results generally confirmed this, with /i:/ exhibiting the longest proportionate formation interval of the long vowels. However, proportionate formation interval of /i:/ was only significantly longer than /ɪ/ in the labial context. The shortening of /i:/ in the coronal context, may be due to the articulatory requirements of coda /t/ on the /i:/ gesture (Recasens et al. 1997, Sussman et al. 1997, Recasens 2002). Several studies have noted that high front vowels in syllables containing coronal consonants, exhibit more retracted acoustic and articulatory targets than those produced in other consonantal contexts (Stevens & House 1963; Schouten & Pols 1979a, b; Sussman et al. 1997; Strange & Bohn 1998; Hoole 1999; Nearey 2013). In English, coronal consonants exhibit higher coarticulatory resistances than surrounding vowels, with the targets of surrounding vowels compromised to reach the desired articulatory goal of the coronal consonant (Sussman et al. 1997, Hoole 1999). In the production of coda /t/, the tongue dorsum must be sufficiently retracted for the tongue tip to be raised for alveolar closure (Hoole 1999).

As shown in Figure 5, /i:/ and /ɪ/ are sometimes produced with a more retracted TD posture in the coronal context, supporting these observations. In the production of /i:/ in the coronal context, the proportionately later achievement of vowel target, due to prolonged onglide, is antagonistic to the required retracted position necessary for production of the coronal coda. Therefore, speakers may shorten the duration of onglide in /t/ final syllables to allow earlier tongue dorsum retraction for coda /t/ closure. As no such constraint is placed on /i:/ in labial final syllables, the phonological onglide is present. More research into production of /i:/ in non-symmetrical consonant contexts may further illuminate the relative contribution of onset–vowel and vowel–coda organisation in the production of onglide in AusE.

/e:/ had a shorter proportionate formation interval than /ɛ/ in both the labial and coronal context, while /o:/ had a shorter proportionate formation interval than /ɔ/ in the labial context (Figures 6 and 7). This result is largely inconsistent with previous acoustic studies of vowel length in not only AusE but also American English and German, which have found no significant difference in proportionate acoustic onglide between long and short vowels (Lehiste & Peterson 1961, Strange & Bohn 1998, Cox 2006). However, Lehiste & Peterson (1961) descriptively reported that short/lax vowels in American English (excluding /ɪ/) had proportionately longer acoustic onglides than their long equivalents. The proportionately longer formation interval of short /ɛ/ and /ɔ/ reported here, are consistent with general observations that vowels of shorter durations exhibit proportionately longer transitions from and to surrounding phonemes (Gay 1981, Soli 1982, Van Summers 1987). The discrepancy between prior acoustic and current articulatory results may arise from, as discussed above, potential differences in laryngeal–supralaryngeal coordination in long and short vowel gestures (Hertrich & Ackermann 1997). If a larger proportion of short vowels is concealed by preceding consonant aspiration than long vowels, it may mask articulatory differences in onglide in the acoustic domain.

We also found that the magnitude of gesture nucleus durations differed by vowel pair, with /i:–ɪ/ exhibiting the smallest pairwise difference in proportionate gesture nucleus duration of the three vowel pairs. This appears to be the result of shorter proportionate gesture nucleus duration of /i:/ compared to the other two long vowels (Table 2), driven by the presence of onglide in /i:/. Coronal context vowels also exhibited a proportionately shorter gesture nucleus duration than labial context vowels. The shortened gesture nucleus duration in the coronal context once again may be due to the relatively greater coarticulatory influence of /t/ on vowels (Recasens et al. 1997, Recasens 2002).

The effect of consonant context on proportionate release interval duration also differed across vowel pairs. Release intervals were longer in the coronal context than labial context for /i:–ɪ/ but were shorter in the coronal than labial context for /ɛ:, ɛ, o:, ɔ/. These patterns

appear to be in a trading relationship with formation interval duration, although the exact mechanism behind this requires further investigation.

4.5 Future directions

There are some limitations to this study. First, we did not investigate articulatory control mechanisms that may underlie durational and vowel quality differences such as stiffness and velocity. This is primarily because speech rate was not actively controlled in this study. Speech rate also conditions the durational, spatial and kinematic properties of vowels (Ostry & Munhall 1985, Kroos et al. 1997, Hoole & Mooshammer 2002, Beňuš 2011). In particular, changes in duration due to variation in speech rate may be implemented through adjustments in gestural stiffness (the ratio of velocity to displacement) or adjustments in only velocity (Gay 1981, Byrd & Tan 1996, Shaiman 2001). These mechanisms are not mutually exclusive and may also interact with the implementation of vowel length (Kroos et al. 1997, Hoole & Mooshammer 2002). Future speech-rate controlled studies should investigate differences in stiffness, velocity and intergestural overlap between long and short vowels and how these can be understood within mass-spring implementations of Task-Dynamics (Saltzman & Kelso 1987, Saltzman & Munhall 1989, Hawkins 1992, Turk & Shattuck-Hufnagel 2020).

We also did not investigate differences in the intergestural organisation of syllables containing long vs. short vowels in AusE. In German, research suggests that short vowels are more overlapped with following coda consonants than long vowels (Hertrich & Ackermann 1997, Kroos et al. 1997, Hoole & Mooshammer 2002). This may also be the case in AusE, but requires further investigation to confirm.

Acoustic target and dynamic acoustic data were not directly compared to articulatory target and articulatory kinematic data in this study. In this study we found discrepancies between relative acoustic and relative gestural duration measures, with short vowels ~ 62% the acoustic duration of long vowels, but ~ 90% the gestural duration of long vowels. This is similar to prior studies investigating this relationship in German vowel length contrast (Hertrich & Ackermann 1997). This suggests that the timing relationship between the larynx and the supralaryngeal articulators in vowel production may differ between long and short vowels, however this requires further empirical examination.

We examined only the lingual articulation of vowels, and only using data from a single lingual sensor. Differences in vowel identity arise due to differences in overall vocal tract shape (Stevens & House 1955, Chiba & Kajiyama 1958, Lindblom & Sundberg 1971, Fant 1980), which is dependent on the coordinated placement of the tongue with respect to the jaw and lips (Lindblom & Sundberg 1971, Hoole & Mooshammer 2002). More detailed articulatory characterisation of vowels should examine the entire vocal tract. Future studies would benefit from rtMRI imaging technologies which offer high spatial and temporal resolution imaging of the vocal tract (Zhu et al. 2013, Lingala et al. 2017).

Finally, perceptual studies are also needed to examine how duration and target quality are used by listeners to cue long versus short vowels. Investigation of participant-specific trading relationships between duration and vowel quality in the production of vowel length contrasts may also provide further insight into the representation and implementation of vowel length.

4.6 Conclusions

This study has systematically examined articulatory differences between long and short vowels in AusE. Long vowels were characterised by different temporal, spatial and dynamic kinematic properties compared to their short equivalents. Our results suggest that vowel duration and vowel quality may be actively and independently controlled to realise vowel length contrasts in AusE. Our results also highlight discrepancies between acoustic and articulatory measures of vowel duration, raising questions about the relationship between these two ways

of measuring durational contrast. These data reveal the importance of studying vowel production in both the acoustic and articulatory domains to more fully understand the representation and implementation of vowel contrasts.

Acknowledgments

This research was supported in part by Australian Research Council Award DE150100318 and Australian Research Council award FT180100462. Parts of this study were presented at the 16th Conference on Laboratory Phonology (LabPhon16), Universidade de Lisboa, Lisbon, 19–22 June 2018, and the 17th Australasian International Conference on Speech Science and Technology (SST2018), University of New South Wales, Sydney, 4–7 December 2018.

Appendix. Additional materials

Table A1 Total displacement (mm) of TB (DispTB) and TD (DispTD) sensor during vowel gesture articulation by participant and vowel pair. Higher values indicate greater displacement during vowel gesture. Sensor with greater displacement was chosen as the target sensor.

Participant	/i:–ɪ/		/e:–e/		/o:–ɔ/	
	DispTB	DispTD	DispTB	DispTD	DispTB	DispTD
W1	19.80	20.00	22.88	33.44	21.73	30.03
W2	30.70	32.11	33.44	35.35	29.07	29.34
W3	26.03	28.45	25.28	28.89	28.68	31.00
W4	30.70	34.06	35.20	42.56	32.19	34.93
M1	25.61	27.29	36.58	39.20	38.13	41.35
M2	24.03	28.50	21.60	29.32	22.64	29.97
M3	19.61	22.10	28.30	30.03	32.41	36.59

Table A2 Results of the mixed model analysis to test vowel acoustic duration (AcDur).

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
Vlength	574940.91	574940.91	1	796	1720.89	< .001
Vpair	201252.60	100626.30	2	796	301.19	< .001
Conscontext	10099.17	10099.17	1	796	30.23	< .001
Vlength:Vpair	44960.24	22480.12	2	796	67.29	< .001
Conscontext:Vpair	4797.38	2398.69	2	796	7.18	< .001

Formula = AcDur ~ Vlength + Vpair + Conscontext + (Vlength × Vpair) + (Conscontext × Vpair) + (Vlength | participant)

Table A3 Results of the mixed model analysis to test vowel gesture duration (GDur).

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
Vlength	325239.96	325239.96	1	786	59.29	< .001
Vpair	35797.11	17898.56	2	787	3.26	.039
Conscontext	2225473.88	2225473.88	1	786	405.68	< .001
Vlength:Conscontext	39459.02	39459.02	1	787	7.19	.007
Conscontext:Vpair	94877.37	47438.69	2	787	8.65	< .001

Formula = GDur ~ Vlength + Vpair + Conscontext + Vlength × Conscontext + Conscontext × Vpair + (1 | participant)

Table A4 Results of the mixed model analysis to test z-transformed Euclidean distance between long and short vowel targets (TARGDIFFz).

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
Vpair	119.27	59.63	2	376	31.81	< .001
Conscontext	10.99	10.99	1	376	5.86	.016

Formula = TARGDIFFz ~ Vpair + Conscontext + (1 | participant)

Table A5 Results of the mixed model analysis to test z-transformed lip protrusion differences (LPz) between /o:/ and /ɔ/.

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
Vlength	80.00	80.00	1	198	143.36	<.001
Conscontext	5.72	5.72	1	198	10.25	<.001

Formula = LPz ~ + Vlength + Conscontext + (1 | participant) + (1 | repetition)

Table A6 Results of the mixed model analysis to test proportionate formation interval duration (FI%).

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
Vlength	1033.55	1033.55	1	796	13.96	< .001
Vpair	607.71	303.85	2	796	4.10	.017
Conscontext	7718.36	7718.36	1	796	104.27	< .001
Vlength:Vpair	1059.64	529.82	2	796	7.16	< .001
Vlength:Conscontext	14.03	14.03	1	796	0.19	.663
Conscontext:Vpair	8361.09	4180.54	2	796	56.47	< .001
Vlength:Conscontext:Vpair	891.84	445.92	2	796	6.02	< .001

Formula = FI% ~ Vlength + Vpair + Conscontext + (Vlength × Vpair) + (Vlength × cons.context) + (Conscontext × Vpair) + (Vlength × Conscontext × Vpair) + (1 | participant)

Table A7 Results of the mixed model analysis to test proportionate gesture nucleus duration (GN%).

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
Vlength	11323.14	11323.14	1	796	236.65	< .001
Vpair	1259.78	629.89	2	796	13.16	< .001
Conscontext	5997.12	5997.12	1	796	125.34	< .001
Vlength:Vpair	855.74	427.87	2	796	8.94	< .001
Conscontext:Vpair	1314.92	657.46	2	796	13.74	< .001

Formula = GN% ~ Vlength + Vpair + Conscontext + Vlength × Vpair + Conscontext × Vpair + (1 | participant)

Table A8 Results of the mixed model analysis to test proportionate release interval duration (RI%).

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
Vlength	5532.53	5532.53	1	796	70.77	< .001
Vpair	1366.48	683.24	2	796	8.74	< .001
Conscontext	109.45	109.45	1	796	1.40	.237
Conscontext:Vpair	3050.35	1525.18	2	796	19.51	< .001

Formula = RI% ~ Vlength + Vpair + Conscontext + Conscontext × Vpair + (1 | participant)

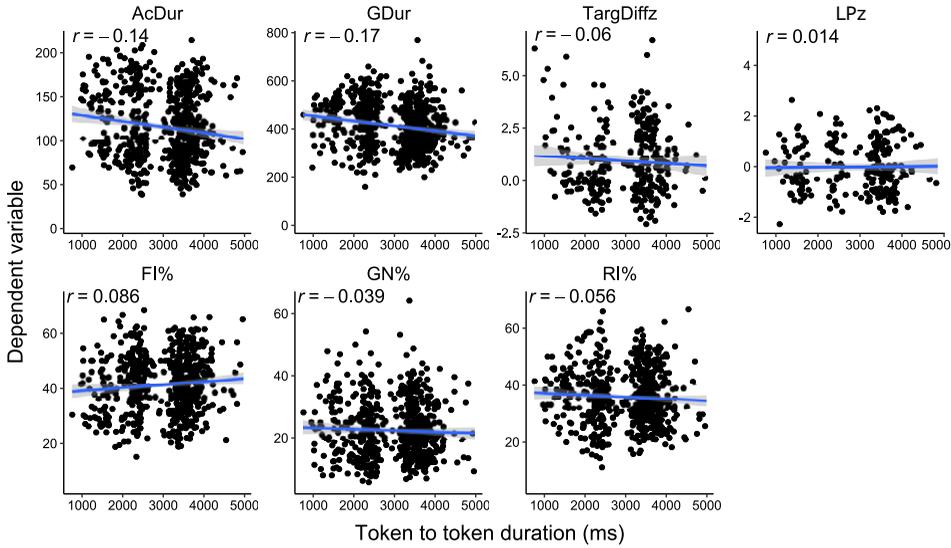


Figure A1 (Colour online) Correlation between token-to-token duration and dependent variables analysed in this study. Token-to-token duration used as an approximation for global speech rate. Left to right: AcDur = acoustic duration (ms), GDur = gesture duration (ms), TargDiffz = z-transformed euclidean distance between long and short vowel targets, LPz = z-transformed lip protrusion for /o:-ɔ/, FI% = proportionate formation interval duration, GN% = proportionate gesture nucleus duration, RI% = proportionate release interval duration. Correlation coefficient (*r*) provided for each variable.

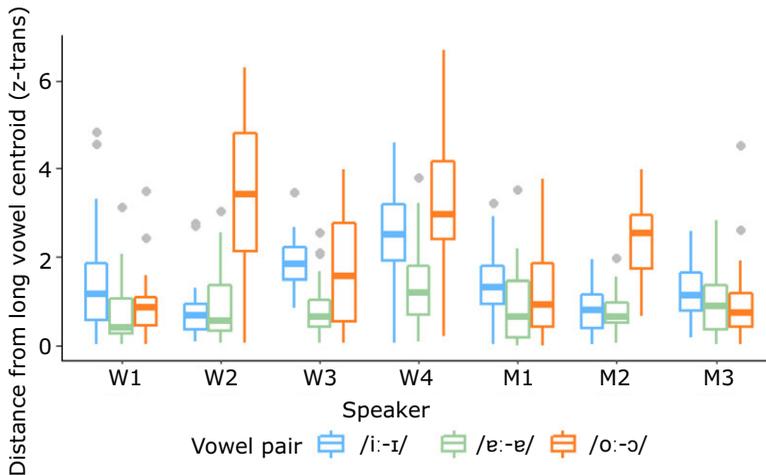


Figure A2 (Colour online) Euclidean distance from long vowel centroid to short vowel articulatory target by vowel pair and participant. TDx and TDy z-transformed by participant.

References

Barr, Dale J., Roger Levy, Christoph Scheepers & Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 44, 255–278.

Bates, Douglas, Martin Maechler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using *lme4*. *Journal of Statistical Software* 67(1), 1–48.

Bell-Berti, Fredericka & Katharine S. Harris. 1981. A temporal model of speech production. *Phonetica* 38(1–3), 9–20.

- Beňuš, Štefan. 2011. Control of phonemic length contrast and speech rate in vocalic and consonantal syllable nuclei. *The Journal of the Acoustical Society of America* 130(4), 2116–2127.
- Bernard, John. 1967. On nucleus component durations. *Language and Speech* 13(2), 89–101.
- Bernard, John. 1970. A cine X-ray study of some sounds of Australian English. *Phonetica* 21(3), 138–150.
- Blackwood Ximenes, Jason A. Shaw & Christopher Carignan. 2016. Tongue positions corresponding to formant values in Australian English vowels. *Proceedings of the 16th Australasian International Conference on Speech Science and Technology*, Sydney, Australia, 109–112.
- Blackwood Ximenes, Jason A. Shaw & Christopher Carignan. 2017. A comparison of acoustic and articulatory methods for analyzing vowel differences across dialects: Data from American and Australian English. *The Journal of the Acoustical Society of America* 142(1), 363–377.
- Browman, Catherine P. & Louis Goldstein. 1990. Tiers in articulatory phonology with some implications for casual speech. In John Kingston & Mary E. Beckman (eds.), *Papers in Laboratory Phonology I: Between the grammar and physics of speech*, 1–30. Cambridge: Cambridge University Press.
- Byrd, Dani & Cheng C. Tan. 1996. Saying consonant clusters quickly. *Journal of Phonetics* 24(2), 263–282.
- Chen, Matthew. 1970. Vowel length variation as a function of the voicing of the consonant environment. *Phonetica* 22(3), 129–159.
- Chiba, Tsutomu & Masato Kajiyama. 1958. *The vowel: Its nature and structure*. Tokyo: Phonetic Society of Japan.
- Chitoran, Ioana, Louis Goldstein & Dani Byrd. 2002. Gestural overlap and recoverability: Articulatory evidence from Georgian. In Carlos Gussenhoven & Natasha Warner (eds.), *Papers in Laboratory Phonology 7*, 419–448. Berlin: de Gruyter.
- Cochrane, George R. 1970. Some vowel durations in Australian English. *Phonetica* 22(4), 240–250.
- Cox, Felicity. 2006. The acoustic characteristics of /hvd/ vowels in the speech of some Australian teenagers. *Australian Journal of Linguistics* 26(2), 147–179.
- Cox, Felicity & Janet Fletcher. 2017. *Australian English pronunciation and transcription*. Melbourne: Cambridge University Press.
- Cox, Felicity & Sallyanne Palethorpe. 2007. Australian English. *Journal of the International Phonetic Association* 37(3), 341–350.
- Cox, Felicity & Sallyanne Palethorpe. 2011. Timing differences in the VC rhyme of standard Australian English and Lebanese Australian English. *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS XVII)*, Hong Kong, China, 528–531.
- Cox, Felicity, Sallyanne Palethorpe & Samantha Bentink. 2014. Phonetic archaeology and 50 years of change to Australian English /i/. *Australian Journal of Linguistics* 34(1), 50–75.
- Cox, Felicity, Sallyanne Palethorpe & Kelly Miles. 2015. The role of contrast maintenance in the temporal structure of the rhyme. *Proceedings of the 18th International Congress of the Phonetic Sciences (ICPhS XVIII)*, Glasgow, UK, 1–4.
- Davidson, Lisa. 2004. *The atoms of phonological representation: Gestures, coordination and perceptual features in consonant cluster phonotactics*. Ph.D. dissertation, Johns Hopkins University.
- Delattre, Pierre. 1962. Some factors of vowel duration and their cross-linguistic validity. *The Journal of the Acoustical Society of America* 34(8), 1141–1143.
- Elvin, Jaydene, Daniel Williams & Paola Escudero. 2016. Dynamic acoustic properties of monophthongs and diphthongs in Western Sydney Australian English. *The Journal of the Acoustical Society of America* 140(1), 576–581.
- Fant, Gunnar. 1980. The relations between area functions and the acoustic signal. *Phonetica* 37(1–2), 55–86.
- Fletcher, Janet, Jonathan Harrington & John Hajek. 1994. Phonemic vowel length and prosody in Australian English. *Proceedings of the 5th Australasian International Conference on Speech Science and Technology*, Perth, Australia, 656–661.
- Fletcher, Janet & Andrew McVeigh. 1993. Segment and syllable duration in Australian English. *Speech Communication* 13(3–4), 355–365.
- Fowler, Carol A. & Lawrence Brancazio. 2000. Coarticulation resistance of American English consonants and its effects on transconsonantal vowel-to-vowel coarticulation. *Language and Speech* 43(1), 10–41.

- Fowler, Carol A. & Jonathan Housom. 1987. Talkers' signalling of "New" and "Old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language* 26(1), 489–504.
- Gafos, Adamantios I. 2002. A grammar of gestural coordination. *Natural Language & Linguistic Theory* 20(2), 269–337.
- Garcia, Damien. 2010. Robust smoothing of gridded data in one and higher dimensions with missing values. *Computational Statistics and Data Analysis* 54(4), 1167–1178.
- Gay, Thomas. 1981. Mechanisms in the Control of Speech Rate. *Phonetica* 38(13), 148–158.
- Gopal, H. S. 1990. Effects of speaking rate on the behavior of tense and lax vowel durations. *Journal of Phonetics* 18(4), 497–518.
- Gussenhoven, Carlos. 2007. A vowel height split explained: Compensatory listening and speaker control. In Jennifer Cole & José Hualde (eds.), *Papers in Laboratory Phonology 9*, 145–172. Berlin: de Gruyter.
- Hadding-Koch, Kerstin & Arthur S. Abramson. 1964. Duration versus spectrum in Swedish vowels: Some perceptual experiments. *Studia Linguistica* 18(2), 94–107.
- Hall-Lew, Lauren. 2010. Improved representation of variance in measures of vowel merger. *Proceedings of 159th Meeting Acoustical Society of America*, Baltimore, MD, 1–10.
- Harrington, Jonathan & Stephen Cassidy. 1994. Dynamic and target theories of vowel classification: Evidence from monophthongs and diphthongs in Australian English. *Language and Speech* 37(4), 357–373.
- Harrington, Jonathan, Felicity Cox & Zoe Evans. 1997. An acoustic phonetic study of broad, general and cultivated Australian English vowels. *Australian Journal of Linguistics* 17(2), 155–184.
- Harrington, Jonathan, Philip Hoole, Felicitas Kleber & Ulrich Reubold. 2011. The physiological, acoustic, and perceptual basis of high back vowel fronting: Evidence from German tense and lax vowels. *Journal of Phonetics* 39(2), 121–131.
- Harrington, Jonathan, Philip Hoole & Ulrich Reubold. 2012. A physiological analysis of high front, tense–lax vowel pairs in Standard Austrian and Standard German. *Italian Journal of Linguistics* 24(1), 149–173.
- Harrington, Jonathan, Felicitas Kleber & Ulrich Reubold. 2011. The contributions of the lips and the tongue to the diachronic fronting of high back vowels in Standard Southern British English. *Journal of the International Phonetic Association* 41(2), 137–156.
- Havenhill, Jonathan. 2015. An ultrasound analysis of low back vowel fronting in the Northern cities vowel shift. *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS XVIII)*, Glasgow, UK, 1–4.
- Hawkins, Sarah. 1992. An introduction to Task Dynamics. In Gerry Docherty & D. Robert Ladd (eds.), *Papers in Laboratory Phonology II: Gesture, segment and prosody*, 9–25. Cambridge: Cambridge University Press.
- Hay, Jen, Paul Warren & Katie Drager. 2006. Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics* 34(4), 458–484.
- Hertrich, Ingo & Hermann Ackermann. 1997. Articulatory control of phonological vowel length contrasts: Kinematic analysis of labial gestures. *The Journal of the Acoustical Society of America* 102(1), 523–536.
- Hillenbrand, James, Michael J. Clark & Terrance M. Nearey. 2001. Effects of consonant environment on vowel formant patterns. *The Journal of the Acoustical Society of America* 109(2), 748–763.
- Hirata, Yukari. 2004. Effects of speaking rate on the vowel length distinction in Japanese. *Journal of Phonetics* 32(4), 565–589.
- Hoole, Philip. 1999. On the lingual organization of the German vowel system. *The Journal of the Acoustical Society of America* 106(2), 1020–1032.
- Hoole, Philip & Christine Mooshammer. 2002. Articulatory analysis of the German vowel system. In Peter Auer, Peter Gilles & Helmut Spiekermann (eds.), *Silbenschnitt und Tonakzente*, 129–152. Tübingen: Niemeyer.
- Hoole, Philip, Christine Mooshammer & Hans G. Tillmann. 1994. Kinematic analysis of vowel production in German. *Proceedings of the 3rd International Conference on Spoken Language Processing*, Yokohama, Japan, 53–56.

- House, Arthur S. 1961. On vowel duration in English. *The Journal of the Acoustical Society of America* 33(9), 1174–1178.
- House, Arthur S. & Grant Fairbanks. 1953. The influence of consonant environment upon the secondary acoustical characteristics of vowels. *The Journal of the Acoustical Society of America* 25(1), 105–113.
- Jenkins, James J., Winifred Strange & Thomas R. Edman. 1983. Identification of vowels in “vowelless” syllables. *Perception and Psychophysics* 34(5), 441–450.
- Jessen, Michael. 1993. Stress conditions on vowel quality and quantity in German. *Working Papers of the Cornell Phonetics Laboratory* 8, 1–27.
- Jong, Kenneth & Bushra Zawaydeh. 2002. Comparing stress, lexical focus and segmental focus: Patterns of variation in Arabic vowel duration. *Journal of Phonetics* 30(1), 53–75.
- Klatt, Dennis H. 1973. Interaction between two factors that influence vowel duration. *The Journal of the Acoustical Society of America* 54(4), 1102–1104.
- Klatt, Dennis H. 1976. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America* 59(5), 1208–1220.
- Kroos, Christian, Philip Hoole, Barbara Kühnert & Hans G. Tillmann. 1997. Phonetic evidence for the phonological status of the tense–lax distinction in German. *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München* 35, 17–25.
- Kuznetsova, Alexandra, Per B. Brockhoff & Rune H. B. Christensen. 2017. *LmerTest* package: Tests in linear-mixed effects models. *Journal of Statistical Software* 82(13), 1–26.
- Lehiste, Ilse. 1970. *Suprasegmentals*. Cambridge, MA: MIT Press.
- Lehiste, Ilse & Gordon E. Peterson. 1961. Transitions, glides, and diphthongs. *The Journal of the Acoustical Society of America* 33(3), 268–277.
- Lehnert-LeHouillier, Heike. 2010. A cross-linguistic investigation of cues to vowel length perception. *Journal of Phonetics* 38(3), 472–482.
- Lenth, Russell. 2019. *Emmeans*: Estimated marginal means, aka least-squares means [R package, Version 1.6.1].
- Lindau, Mona. 1978. Vowel features. *Language* 54(3), 541–563.
- Lindblom, Björn. 1963. Spectrographic study of vowel reduction. *The Journal of the Acoustical Society of America* 35(11), 1773–1781.
- Lindblom, Björn & Johan E. Sundberg. 1971. Acoustical consequences of lip, tongue, jaw and larynx movement. *The Journal of the Acoustical Society of America* 50(4B), 1166–1179.
- Lingala, Sajan G., Yinghua Zhu, Yoon-Chul Kim, Asterios Toutios, Shrikanth Narayanan & Krishna S. Nayak. 2017. A fast and flexible MRI system for the study of dynamic vocal tract shaping. *Magnetic Resonance in Medicine* 77(1), 112–125.
- Lobanov, Boris. 1971. Classification of Russian vowels spoken by different speakers. *The Journal of the Acoustical Society of America* 49(2B), 606–608.
- Löfqvist, Anders. 1999. Interarticulator phasing, locus equations, and degree of coarticulation. *The Journal of the Acoustical Society of America* 106(4), 2022–2030.
- Mády, Katalin & Uwe D. Reichel. 2007. Quantity distinction in the Hungarian vowel system: Just theory or also reality? *Proceedings of the 16th International Congress of the Phonetic Sciences (ICPhS XVI)*, Saarbrücken, Germany, 1053–1056.
- Meister, Einar, Stefan Werner & Lya Meister. 2011. Short vs. long category perception affected by vowel quality. *Proceedings of the 17th International Congress of the Phonetic Sciences (ICPhS XVII)*, Hong Kong, China, 1362–1365.
- Mooshammer, Christine & Susanne Fuchs. 2002. Stress distinction in German: Simulating kinematic parameters of tongue-tip gestures. *Journal of Phonetics* 30(3), 337–355.
- Mooshammer, Christine & Christian Geng. 2008. Acoustic and articulatory manifestations of vowel reduction in German. *Journal of the International Phonetic Association* 38(2), 117–136.
- Nakai, Satsuki, Sari Kunnari, Alice Turk, Kari Suomi & Riikka Ylitalo. 2009. Utterance final lengthening and quantity in Northern Finnish. *Journal of Phonetics* 37(1), 29–45.
- Nance, Claire. 2011. High back vowels in Scottish Gaelic. *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS XVII)*, Hong Kong, China, 1446–1449.

- Nearey, Terrance M. 2013. Vowel inherent spectral change in the vowels of North American English. In Geoffrey Stewart Morrison & Peter F. Assman (eds.), *Vowel inherent spectral change*, 49–85. Berlin: Springer.
- Nearey, Terrance M. & Peter F. Assmann. 1986. Modeling the role of inherent spectral change in vowel identification. *The Journal of the Acoustical Society of America* 80(5), 1297–1308.
- Nooteboom, Sieb G. & Gert J. N. Doodeman. 1980. Production and perception of vowel length in spoken sentences. *The Journal of the Acoustical Society of America* 67(1), 276–287.
- Northern Digital Inc. 2016. NDI Wavefront.
- Ostry, David J. & Kevin G. Munhall. 1985. Control of rate and duration of speech movements. *The Journal of the Acoustical Society of America* 77(2), 640–648.
- Penney, Joshua, Felicity Cox, Kelly Miles & Sallyanne Palethorpe. 2018. Glottalisation as a cue to coda consonant voicing in Australian English. *Journal of Phonetics* 66, 161–184.
- Peters, Sandra. 2015. *The effects of syllable structure on consonantal timing and vowel compression in child and adult speaker*. Ph.D. dissertation, Ludwig-Maximillan Universität.
- Peterson, Gordon E. & Ilse Lehiste. 1960. Duration of syllable nuclei in English. *The Journal of the Acoustical Society of America* 32(6), 693–703.
- Port, Robert F. 1981. Linguistic timing factors in combination. *The Journal of the Acoustical Society of America* 69(1), 262–274.
- Pycha, Anne & Delphine Dahan. 2016. Differences in coda voicing trigger changes in gestural timing: A test case from the American English diphthong /a/. *Journal of Phonetics* 56, 15–37.
- Ratko, Louise, Michael Proctor, Felicity Cox & Sean Veld. 2016. Preliminary investigations into the Australian English articulatory vowel space. *Proceedings of the 16th Australasian International Conference on Speech Science and Technology*, Sydney, Australia, 117–120.
- Recasens, Daniel. 2002. An EMA study of VCV coarticulatory direction. *The Journal of the Acoustical Society of America* 111(6), 2828–2841.
- Recasens, Daniel, Maria D. Pallarès & Jordi Fontdevila. 1997. A model of lingual coarticulation based on articulatory constraints. *The Journal of the Acoustical Society of America* 102(1), 544–561.
- Saltzman, Elliot & J. A. Scott Kelso. 1987. Skilled actions: A task-dynamic approach. *Psychological Review* 94(1), 84–106.
- Saltzman, Elliot & Kevin G. Munhall. 1989. A dynamical approach to gestural patterning in speech production. *Ecological Psychology* 1(4), 333–382.
- Schouten, M. E. H. & L. C. W. Pols. 1979a. Vowel segments in consonantal contexts: A spectral study of coarticulation – Part I. *Journal of Phonetics* 7(1), 1–23.
- Schouten, M. E. H. & L. C. W. Pols. 1979b. CV- and VC-transitions: A spectral study of coarticulation – Part II. *Journal of Phonetics* 7(3), 205–224.
- Shaiman, Susan. 2001. Kinematics of compensatory vowel shortening: Intra and interarticulatory timing. *The Journal of the Acoustical Society of America* 106(4), 89–107.
- Shaw, Jason, Wei-rong Chen, Michael Proctor & Donald Derrick. 2016. Influences of tone on vowel articulation in Mandarin Chinese. *Journal of Speech Language and Hearing Research* 59(6), S1566–S1574.
- Soli, Sigfred. 1982. Structure and duration of vowels together specify fricative voicing. *The Journal of the Acoustical Society of America* 72(2), 366–378.
- Stevens, Kenneth & Arthur S. House. 1955. Development of a quantitative description of vowel articulation. *The Journal of the Acoustical Society of America* 27(3), 484–493.
- Stevens, Kenneth & Arthur S. House. 1963. Perturbation of vowel articulations by consonantal context: An acoustical study. *Journal of Speech Language and Hearing Research* 6(2), 111–128.
- Strange, Winifred & Ocke-Schwen Bohn. 1998. Dynamic specification of coarticulated German vowels: Perceptual and acoustical studies. *The Journal of the Acoustical Society of America* 104(1), 488–504.
- Strange, Winifred, James J. Jenkins & Thomas L. Johnson. 1983. Dynamic specification of coarticulated vowels. *The Journal of the Acoustical Society of America* 74(3), 695–705.
- Strange, Winifred, Andrea Weber, Erika S. Levy, Valeriy Shafiro, Miwako Hisagi & Kanae Nishi. 2007. Acoustic variability within and across German, French, and American English vowels: Phonetic context effects. *The Journal of the Acoustical Society of America* 122(2), 1111–1129.

- Sussman, Harvey M., Nicola Bessell, Eileen Dalston & Tivoli Majors. 1997. An investigation of stop place of articulation as a function of syllable position: A locus equation perspective. *The Journal of the Acoustical Society of America* 101(5), 2826–2838.
- Tiede, Mark. 2005. MVIEW: Software for visualization and analysis of concurrently recorded movement data.
- Tomaschek, Fabian, Hubert Truckenbrodt & Ingo Hertrich. 2015. Discrimination sensitivities and identification patterns of vowel quality and duration in German /u/ and /ɔ/ instances. In Adrian Leemann, Marie-José Kolly, Stephan Schmid & Volker Dellwo (eds.), *Trends in phonetics and phonology: Studies from German speaking Europe*, 1–8. Frankfurt am Main & Bern: Peter Lang.
- Turk, Alice & Stefanie Shattuck-Hufnagel. 2020. *Speech timing: Implications for theories of phonology, phonetics, and speech motor control* (Oxford Studies in Phonology and Phonetics). Oxford: Oxford University Press.
- Umeda, Noriko. 1975. Vowel duration in American English. *The Journal of the Acoustical Society of America* 58(2), 434–445.
- Van Summers, W. 1987. Effects of stress and final-consonant voicing on vowel production: Articulatory and acoustic analyses. *The Journal of the Acoustical Society of America* 82(3), 847–863.
- Warren, Paul. 2017. Quality and quantity in New Zealand English vowel contrasts. *Journal of the International Phonetic Association* 48(3), 305–330.
- Watson, Catherine & Jonathan Harrington. 1999. Acoustic evidence for dynamic formant trajectories in Australian English vowels. *The Journal of the Acoustical Society of America* 106(1), 458–468.
- Wells, John C. 1982. *Accents of English*, vol. 3: *Beyond the British Isles*. Cambridge: Cambridge University Press.
- White, Laurence & Katalin Mády. 2008. The long and the short and the final: Phonological vowel length and prosodic timing in Hungarian. *Proceedings of the 4th International Conference on Speech Prosody*, Campinas, Brazil, 363–366.
- Wong, Amy. 2012. The lowering of raised thought and the low–back distinction in New York City: Evidence from Chinese Americans. *Selected Papers from NWAV 40: Special issue of University of Pennsylvania Working Papers in Linguistics* 18(2), 157–166.
- Zhu, Yinghua, Yoon-Chul Kim, Michael Proctor, Shrikanth Narayanan & Krishna S. Nayak. 2013. Dynamic 3-D visualization of vocal tract shaping during speech. *IEEE Transactions on Medical Imaging* 32(5), 838–848.