Onset - Vowel Articulatory Coordination - voiceless stops and vowel length

Louise Ratko[1], Michael Proctor[1], Felicity Cox[1]

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

ABSTRACT

Articulatory coordination in English phonemic vowel length contrasts has not been systematically explored. Australian English (AusE), which has a true phonemic vowel length contrast in the open vowels /ɐː-ɐ/ (as in *tart-tut*), provides a natural control case for exploring articulatory vowel length. Intergestural coordination was examined using Electromagnetic Articulography in stop-initial syllables differing in vowel length produced by five speakers of Australian English. Constriction formation of the supraglottal consonant gesture in the onset was synchronous with the nuclear vowel gesture irrespective of vowel length. Constriction release of the onset consonant was delayed, and VOT lengthened, in syllables containing a phonemically long vowel (*tart*), compared to syllables with short vowels of the same quality (*tut*). The findings are consistent with a model of syllable structure in which there is independent control of constriction formation and constriction release in onset gestures.

Keywords: vowel length; gestural coordination; Australian English; syllable structure; articulography

1. INTRODUCTION

Languages exploit different patterns of coordination of glottal and supraglottal gestures to create phonological contrasts between onset stops [5]. In English, both aspirated and 'voiced' onset stops contain two gestures: a supraglottal constriction gesture and a glottal abduction gesture [3, 4]. Aspirated onset stops are characterised by a delay between release of the supraglottal constriction and the offset of the glottal abduction gesture (VOT ~70 ms). 'Voiced' onset stops have greater synchrony of supraglottal and glottal gestures (VOT ~ 46 ms) [11].

An assumption of most models of syllable structure is that intrinsic timing in the onset is independent of other elements in the syllable [7, 12, 20]. In a simple CV syllable, for example, onset duration and intergestural coordination is primarily dependent upon the intrinsic phonological and (language-specific) voicing properties of the initial consonant. In Task Dynamic models [e.g. 3, 4], the duration of both glottal and supraglottal gestures of onset consonants is specified by a gestural stiffness that determines the global duration of the gesture.

However, observations from Canadian English [15], American English [16] and German [19] indicate that VOT is also influenced by vowel context: increased VOT was observed for stops preceding phonemically long vowels. This suggests the offset of the glottal abduction gesture is progressively delayed relative to the offset of the supraglottal constriction as a function of increasing vowel duration. This is inconsistent with models of syllable structure in which onset and vowel timings are independent and controlled by monolithic timing parameters [3, 4].

Studies of German [8, 9] show that vowel length is linked to changes in intergestural VC coordination in stressed syllables. Long vowel gestures are less overlapped with following coda consonant gestures than short vowel gestures [8, 9]. This suggests the increase in VOT could also arise from a reduction in overlap between the onset stop and the following vowel. However, the effect of phonological vowel length on the gestural coordination and durations of CV sequences has not been systematically explored.

Non-rhotic AusE is a language of interest with respect to these issues because of the nature and properties of its phonological vowel length contrasts. The long-short vowel pair /ɐː-ɐ/ (as in *cart-cut*) displays minimal spectral and articulatory difference [6] providing a natural control case for examining the relationship between onset stop gestures and phonemic vowel length.

The goals of this study are to:

1) Examine the gestural organisation of onset voiceless stops in Australian English, using articulography.
2) Determine how onset voiceless stop coordination varies with respect to phonemically long and short following vowels.

2. METHODS

2.1. Participants

Five monolingual Australian English speakers (two males; age: 18-24, mean 21 years) were recruited at Macquarie University, Sydney NSW.

2.2 Experimental Materials

The corpus contained four monosyllables contrasting long (/ɐː/) and short (/ɐ/) vowels in labial /pVp/ and coronal /tVt/ contexts: *'parp'*, *'tart'*, *'pup'*, and *'tut'*.

Target items were produced in the carrier phrase *'See CVC heat'* to control tongue and lip position prior to and after target production. Stimuli were presented via computer screen in ten randomised blocks of 18 items. We selected four of the items per block (*'parp', 'tart', 'pup', 'tut')* for this analysis.

2.2 Data Acquisition

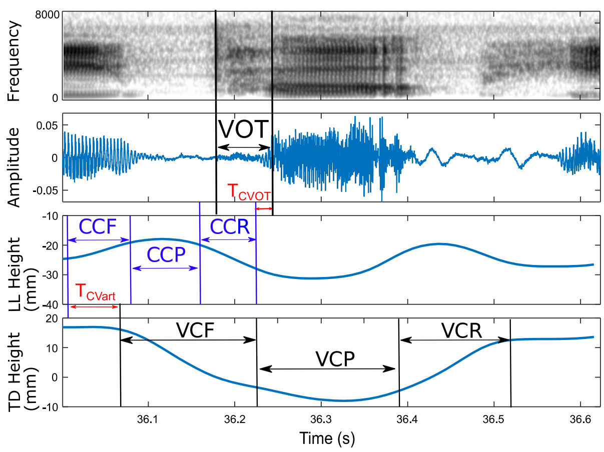
Speech movements were tracked at a sampling rate of 100 Hz using a Northern Digital Inc. Wave Electro-magnetic Articulography (EMA) system. Sensors were located on the mid-sagittal plane on key articulators including the lower lip (LL), tongue tip (TT) and tongue dorsum (TD). Head movement was corrected using three reference sensors located on the participants' left and right mastoids and nasion. Vertical sensor displacement was expressed relative to participants' occlusal plane; horizontal displacement with respect to the rear of the upper incisors. Articulographic data were synchronised with companion speech audio recorded using a shotgun microphone at a sampling rate of 22,050 Hz.

2.3. Data analysis

Ten repetitions of four syllables were produced by four participants. One participant completed only eight repetitions, providing 192 onsets, 98 long vowels and 98 short vowels in total for analysis.

Gestural landmarks were semi-automatically located using the *findgest* algorithm in Mview [17]. Three intervals were demarcated in each supraglottal stop gesture (C) and each vowel gesture (V) (Fig. 1): i) Constriction Formation (CF) = Gesture onset to Nucleus Onset; ii) Constriction Plateau (CP) = Nucleus Onset to Nucleus Offset; iii) Constriction Release (CR) = Nucleus Offset to Gesture Offset.

Figure 1: Articulatory and acoustic measurements: *'parp'*. Top to bottom: 1) Spectrogram, 2) Wave-form, 3) LL vertical displacement, 4) TD vertical displacement.



Total gesture durations (C and V) were calculated by summing durations of the three sub-gestural intervals. Onset stop VOT (Fig. 1, 2nd row) was measured using the spectrogram and waveform from beginning of the stop release burst to the onset of voicing for the vowel. The onset of vowel voicing was used as an approximation for the end of the C glottal abduction gesture which could not be directly measured using EMA. Two measures of intergestural timing were calculated from these landmarks within the onset and nuclear gestures:

- T_{CVOT} = C Gestural Offset – VOT Offset
- T_{CVart} = C Gestural Onset – V Gestural Onset

3. RESULTS

3.1 VOT and gesture durations in CV

We constructed linear mixed effects models to test for main effects of Phonemic Vowel Length (short/long) on VOT and on the durations of C, CCF, CCP, CCR, V, VCF, VCP and VCR. The fixed effects structure for all models included main effects of Phonemic Vowel Length (short /ɐ/ as intercept) and Consonant Type (labial/coronal) with random intercepts of speaker. We excluded a Phonemic Vowel Length × Consonant Type interaction when it did not improve model fit. An interaction between Phonemic Vowel Length × Consonant Type improved model fit for VCP, so was included for this variable, but information about the interaction is not reported here. Modelling was conducted using the *lme4* package in R [1]. *p*-values were obtained through maximum likelihood tests with Satterthwaite approximations to degrees of freedom [10].

VOT was significantly longer (β = 11.6 ms) for stops preceding long vowels (F = 42.9, p < .001). C supraglottal gestures were also significantly longer (β = 14.3 ms) preceding long vowels (F = 12.3, p < .001). The three intervals within the C supraglottal gesture were also examined (Fig. 2).

The Constriction Formation interval (CCF) was not significantly longer for stops preceding long vowels (β = 1.21 ms, F = 0.17, p = .686). Constriction Plateau (CCP) was shorter for stops preceding long vowels (β = -1.66 ms) but once again this difference was not significant (F = 0.28, p = .600). Finally, Constriction Release (CCR) was found to be significantly longer for stops preceding long vowels (β = 11.9 ms, F = 15.5, p < .001).

Long vowel gestures had a significantly greater duration than short vowel gestures (β = 74.7 ms, F = 73.4, p < .001). The three intervals within the vowel gesture were also examined (Fig. 3). The Constriction Formation (VCF) of long vowel gestures was significantly longer than for short vowel gestures (β = 18.6 ms, F = 11.4, p < .001). The Constriction Plateau

(VCP) of long vowel gestures was significantly longer than for short vowels ($\beta = 57.9$ ms, F = 195.4, $p < .001$). Constriction Release (VCR) durations of long vowel gestures were not significantly longer than VCRs of short vowel gestures ($\beta = 10.6$ ms, F = 1.9, $p = .168$).

Figure 2: Durations (ms) of three gestural intervals in consonants produced before long (black) and short (orange) vowels. L-to-R: constriction formation (CCF); constriction plateau (CCP); constriction release (CCR).

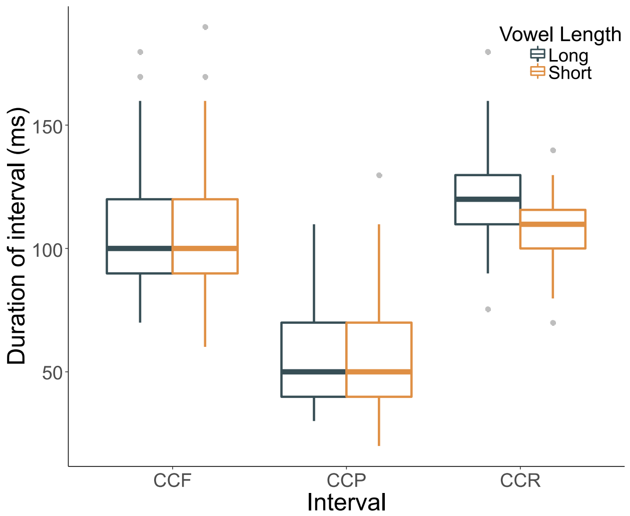
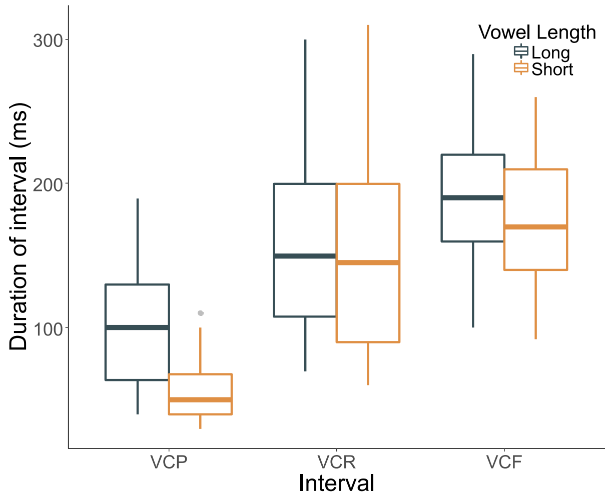


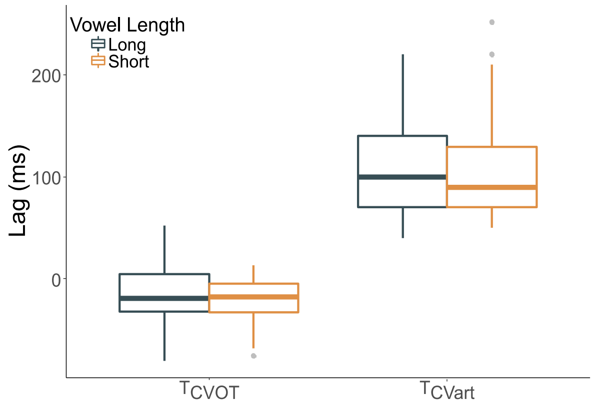
Figure 3: Durations (ms) of three gestural intervals in long (black) vs. short (orange) vowels. L-to-R: constriction formation (VCF); constriction plateau (VCP); constriction release (VCR).



3.2 Intergestural timing in CV

Intergestural coordination was examined through analysis of two lags. First, T_{CVOT} measured the delay between the offset of the supraglottal C gesture and the offset of VOT; the acoustic cue to the offset of the C glottal abduction gesture. Higher lag values indicate a larger delay in the offset of the glottal abduction gesture relative to the offset of the C supraglottal abduction gesture. T_{CVOT} was modelled as a linear function of Phonemic Vowel Length (/ɐ/ as intercept) and Consonant Type, with random intercepts for speaker. Phonemic Vowel Length × Consonant Type interaction was not found to improve model fit so was not included. Phonemic vowel length did not have a significant effect on T_{CVOT} ($\beta = -0.7$ ms, F = 0.0, $p = .833$; Fig. 4).

Figure 4: Time lags (ms) between onset C and nuclear V gestures in long (black) and short (orange) vowels. Left: supraglottal constriction release to VOT onset; Right: onset of C constriction formation to vowel constriction formation.



T_{CVart} measured the delay between the onset of the primary supraglottal gesture of the onset consonant and the dorsal gesture of the nuclear vowel. Higher T_{CVart} values indicate less overlap between C and V gestures, which would be expected if long vowels are less overlapped with onset consonants than short vowels. T_{CVart} was modelled as a linear function of Phonemic Vowel Length (/ɐ/ as intercept) and Consonant Type, with random intercepts for speaker. Phonemic vowel length did not have a significant effect on T_{CVart} ($\beta = 2.3$ ms, F = 0.2, $p = .68$; Fig. 4).

3.3 Summary of results

- VOT is 12 ms (16%) longer in stops before long vowels, compared to stops before short vowels.
- Supraglottal onset gestures are 14 ms (5%) longer in CV: syllables, compared to CV
- Constriction Release interval (CCR) is longer in supraglottal onset gestures (12 ms / 11%) preceding long vowels.
- Constriction Formation (CCF) and Plateau (CCP) intervals of onset gestures are unaffected by length of following vowel.
- Long vowel gestures are 75 ms (18%) longer than in short vowels.

- Constriction Formation interval (VCF) is 19 ms (10%) longer in long vowels.
- Constriction Plateau interval (VCP) is 58 ms (86%) longer in long vowels.
- Lag between offset of supraglottal and glottal gestures (T_{CVOT}) in onset stops unaffected by tautosyllabic vowel length.
- Lag between onsets of supraglottal onset gesture and tautosyllabic vowel gesture (T_{CVart}) unaffected by vowel length.

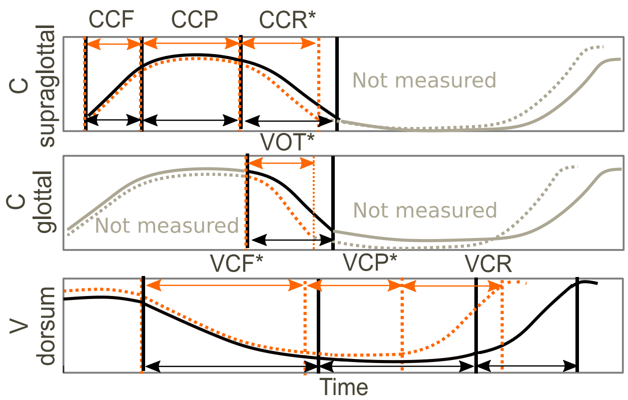


Figure 5: Schematic of kinematic trajectories for C supraglottal (top), C glottal (mid) and vowel (bottom) gestures in long (black) and short (orange) vowel conditions. C glottal estimated from VOT landmarks in 2.3. Asterisks mark intervals with significant effects of phonemic vowel length.

4. DISCUSSION

In this study, we investigated gestural organisation in voiceless onset stops preceding long and short AusE vowels. Consistent with previous research in other Germanic languages [15, 16, 19], we found longer VOT in voiceless stops preceding phonemically long vowels. We used two metrics to examine details of intergestural coordination to shed more light on the mechanisms underlying the observed lengthening of VOT before long vowels in Australian English.

First, we found that the relationship between the offset of the supraglottal stop gesture and the onset of vowel voicing was unaffected by vowel length. Differences in VOT therefore do not appear to be driven by a progressive delay in the offset of the glottal abduction relative to the offset of the supraglottal stop gesture in the long vowel condition.

Second, we examined the lag between the onset of the supraglottal stop gesture and the onset of the nuclear vowel gesture to determine if there were different patterns of intergestural coordination for CVs containing long vs short nuclear vowels. Contrary to studies of German VC rimes [8, 9] we did not find a change in intergestural coordination between long and short vowel conditions. This suggests that changes in VOT are unlikely to be the result of fundamental changes in the gestural coordination between onset stop gestures and nuclear vowel gestures. These findings are consistent with models of syllable structure that assume fixed intergestural relationships between consonant and vowel gestures in CV sequences [3, 4]

If the same fundamental patterns of gestural coordination govern the organisation of CV and CV: syllables in Australian English, the question remains as to why VOT differs as a function of tautosyllabic vowel length. As illustrated in Figure 5, both consonant gestures were found to be longer in onsets of syllables containing long vowels. However, there was not a linear rescaling of either the supraglottal C gesture or the V gesture with vowel length (the entire glottal abduction gesture of the onset stop could not be measured acoustically). Only the constriction release interval (CCR) of supraglottal C gestures was lengthened before long vowels; the formation (CCF) and plateau (CCP) intervals of onsets remained unchanged. Similarly, only the formation (VCF) and plateau (VCF) intervals of long vowel gestures were found to be longer than the equivalent intervals in short vowels; gestural release intervals (VCR) did not differ significantly with vowel length.

In Task Dynamic models of syllable structure, gestural duration is determined by intrinsic stiffness [3, 4, 14]. Changes in duration of sub-gestural intervals which are sensitive to overall syllable length, such as those observed in these onset consonants, suggest that additional local prosodic mechanisms may be involved. The patterns of intergestural coordination observed in these data are consistent with models of speech production in which there is independent timing control of consonant constriction formation and release components [2, 13, 18].

4. FUTURE DIRECTIONS

More data is required to determine how these patterns hold in larger populations of speakers, and in other varieties of English. Different types of onset consonants will shed more light on patterns of intergestural coordination and how they interact with short and long vowels in Australian English. Extended sensing techniques – including EGG and transglottal illumination – could provide richer direct information about glottal activity, informing more accurate models of intergestural timing in onsets. Manipulation of speech rate can be used to determine which timing properties of these syllables are intrinsic to their constituent gestures, and which are sensitive to more global principles of prosodic organisation.

5. REFERENCES

[1] Bates, D. 2010. lme4: Mixed-effects modelling with R. Springer.

[2] Browman, C. 1994. Lip aperture and consonant releases. Papers in laboratory phonology III: Phonological structure and phonetic Form: P. Keating (eds.), 331-353. Cambridge: Cambridge University Press.

[3] Browman, C. & Goldstein, L. 1986. Towards an articulatory phonology. *Phonology Yearbook* 3, 219-252.

[4] Browman, C., & Goldstein, L. 1988. Some notes on syllable structure in Articulatory Phonology. *Phonetica*, 45, 140-155.

[5] Cho, T. & Ladefoged, P. 1999. Variation and universals in VOT: evidence from 18 languages. *J. Phon*. 27, 207–229.

[6] Cox, F. 2006. The acoustic characteristics of /hVd/ vowels in the speech of some Australian teenagers. *Aust. J. Ling.* 26, 147-179.

[7] Hayes, B. 1995. *Metrical stress theory: Principles and case studies.* University of Chicago Press.

[8] Hertrich, I. & Ackermann, H. 1997. Articulatory control of phonological vowel length contrasts: Kinematic analysis of labial gestures, *J. Acoust. Soc. Am.* 102, 523-536.

[9] Hoole, P. & Mooshammer, C. 2002. Articulatory analysis of the German vowel system. *Silbenschnitt und Tonakzente* 1, 129-152.

[10] Kuznetsova, A., Brockhoff, P. & Christensen, R. 2017. LmerTest Package: Tests in Linear-Mixed Effects Models, *J. Stat. Softw.* 82, 1-26.

[11] Lisker, L. & Abramson, A. S. 1967. Some effects of context on voice onset time in English stops, *Lang. Speech* 10, 1-28.

[12] McCarthy, J. 1979. On stress and syllabification. *Linguist. Inq.* 10, 443-465.

[13] Nam, H. 2007. A competitive, coupled oscillator model of moraic structure: Split-gesture dynamics focusing on positional asymmetry. In Cole, J. & Hualde, J. I. (eds), New York: Mouton de Gruyter, 483-506.

[14] Nam, H., Goldstein, L. & Saltzman, E. 2009. Self organization of syllable structure: A coupled oscillator model. In Pellegrino, F. E. M. & Chitoran, I. (eds), *Approaches to phonological complexity.* Berlin: Mouton de Gruyter, 299-328.

[15] Nearey, T. M., & Rochet, B. L. 1992. Effects of place of articulation and vocalic context on the perception of VOT continua in French and English, *J. Acoust. Soc. Am.* 91, 2471–2472.

[16] Port, R. F. & Rotunno, R. 1979. Relation between voice-onset time and vowel duration. *J. Acoust. Soc. Am*. 66, 654–662.

[17] Tiede, M. 2005. MVIEW: software for visualization and analysis of concurrently recorded movement data, Haskins Laboratory.

[18] Tilsen, S. & Goldstein, L. 2012. Articulatory gestures are individually selected in production. *J. Phon.* 40, 764-779.

[19] Weismer, G. 1979. Sensitivity of voice-onset time (VOT) measures to certain segmental features in speech production. *J. Phon.* 7, 197-204.

[20] Zec, D. 1995. Sonority constraints on syllable structure. *Phonology* 12, 85-129.