THE ROLE OF CONTRAST MAINTENANCE IN THE TEMPORAL STRUCTURE OF THE RHYME

Felicity Cox[a,b,c], Sallyanne Palethorpe[a,c], Kelly Miles[a,b,c]

[a]Centre for Language Sciences, [b]Department of Linguistics, [c]ARC Centre of Excellence in Cognition and its Disorders, Macquarie University
Felicity.Cox@mq.edu.au; Sallyanne.Palethorpe@mq.edu.au; Kelly.Miles@mq.edu.au

ABSTRACT

We explore the relationship between temporal elements of the VC rhyme as a function of inherent vowel length and coda voicing in Australian English. Our findings for single monosyllabic words suggest that temporal organisation of the rhyme may be modulated by the strength of the durational contrast associated with long/short vowel pairs. Vowels such as /ɐː/ vs /ɐ/ which contrast by duration alone in Australian English are compared to vowels /iː/ vs /ɪ/ for which duration is just one of the important contrastive cues. In contexts where coda voicing has the potential to disrupt length contrast, such as when the long vowel /ɐː/ occurs in a rhyme with a voiceless coda, we found that vowel length was maximised at the expense of coda closure duration. We propose that the requirement of contrast maintenance may place constraints on competing temporal influences and affect the organisation of elements within the syllable rhyme.

Keywords: duration, rhyme, Australian English, vowel length, contrast.

1. INTRODUCTION

In English, vowel duration is exploited phonologically for the purpose of separating long/ short vowel classes [16, 18] and it also makes an important contribution to the differentiation of obstruent coda voicing within the rhyme [e.g. 12, 18, 24, 25]. In addition, vowel duration has an essential function in the realisation of prosody, in particular, for signalling focus accent and also in the implementation of boundary strength [32, 33]. Thus vowel duration requires careful management so that the competing demands from the phonology are phonetically realised to ensure that the phonological stakeholders are successful accommodated.

In English there is an interesting set of covarying temporal relationships that exist between the subcomponents of the syllable rhyme. Regardless of coda voicing, long vowels occur with shorter codas than short vowels [10, 23, 38]. When coda voicing is considered, voiced obstruents are associated with longer preceding tautosyllabic vowels than voiceless obstruents [e.g. 6, 12, 14, 19, 20, 22, 24, 25] and closure duration of stop codas is longer for voiceless stops than voiced stops [e.g. 19, 22, 34]. Thus there is a trading relationship between the vowel and the coda such that long codas are paired with shorter vowels and vice versa. According to Solé [28], these interactions within the syllable rhyme are under the speaker's control as they represent language specific phonological characteristics rather than automatic effects (see also [5]).

Some researchers suggest that covariation can be considered a listener-oriented enhancing strategy to ensure the integrity of phonological contrasts [14, 17, 35]. Therefore, we might expect varying levels of contrastiveness to affect temporal relationships within the rhyme differently. The aim of this paper is to explore this proposition by examining cue weighting as an additional factor in the constellation of influences from the phonology which affect the relationships between durational components in the syllable rhyme. Our focus is Australian English (AusE) which has vowel pairs that exhibit varying degrees of durational contrast.

1.1. Vowel Length in Australian English

Like most varieties of English, AusE displays phonologically long and short vowels [8]. Generally the quality features are considered primary in vowel identification [37]. However, in non-rhotic AusE two pairs of vowels contrast by duration alone: /ɐː/ vs /ɐ/ (*heart/hut)* and /eː/ vs /e/ (*shared/shed*) [7]. Durational contrast in the absence of spectral differentiation for /ɐː, ɐ/ (*heart/hut*) has been confirmed by xray data [1] and also in several acoustic studies [e.g. 2, 7, 13, 37].

In contrast, for /iː, ɪ/ (*heat/hit*), the length difference is accompanied by a set of additional contrastive cues. /iː/ is phonetically higher and more fronted [9] and exhibits onglide, described in [15] as 'delayed target' [e.g. 2, 7, 11, 37]. The onglide of /iː/ gives the vowel a diphthongal quality. Watson and Harrington [37] showed that excluding duration from a classification task based on modelling from discrete cosine transform coefficients returned 92.9% and 93.5% correct classification for /iː/ but only 67.6% and 77.4% classification for /ɐː/. Their analyses confirm that duration is not as crucial for the /iː/ vs /ɪ/ classification but is vital for /ɐː/ vs /ɐ/. Therefore, in AusE the functional weight of contrastive duration varies across different vowel pairs. It is heavily weighted for pairs like /ɐː/ vs /ɐ/ but is less important for pairs like /iː/ vs /ɪ/ where the pressure of contrast is shared by features like duration, F1/F2 spectral characteristics and onglide.

Solé and Ohala [29] suggest that the phonological tense-lax contrast is under speaker control and is manipulated under various conditions. They found that tense vowels lengthen more than lax vowels at slower speaking rates, 'presumably in order to maintain a constant perceptual distance across changes in global timing' ([29] p.634). Analogously, data from [10] indicates that long vowels lengthen more than short vowels in voiced contexts, supporting the notion of language-specific control at the syllable level. These findings suggest that the nature of the phonemic contrast is important in the organisation of phonetic timing in different contexts [29].

Our specific aim in this paper is to examine how the patterns of timing required for the realisation of voicing and vowel length/coda timing relationships interact with variously weighted phonological durational contrasts.

The phonologised voicing vowel length effect supports coda contrast but may be at odds with vowel contrast requirements. We predict that /ɐː/ vs /ɐ/ (i.e. vowels that rely heavily on duration for contrast) will be affected more strongly than /iː/ vs /ɪ/ by environments that have the potential to compromise the contrast (i.e. when a long vowel occurs in voiceless contexts).

2. METHOD

2.1. Data

Data were extracted from the isolated wordlist task from the AusTalk corpus [4]. For AusTalk, speakers were recorded completing a range of spontaneous and scripted tasks including reading a set of 322 isolated words on 3 separate occasions [4]. Words were separately randomised within each list. Here we selected words containing the vowels /iː, ɪ, ɐː, ɐ/ in hVt and hVd contexts from each of the three sessions. Table 1 shows the words and number of items used in the analysis. Note that for *heed* there are fewer examples because some speakers in the AusTalk corpus only produced two repetitions of *heed* rather than three. Other omissions are due to errors in production, poor quality audio or failure to release the stop such that closure duration could not be determined. 673 tokens were analysed.

2.2. Speakers

Data were selected from 36 speakers (mean age: 26.69 (*SD* 4.65) who had completed all of their school education in Sydney: 17 females (age: *M*=25.76) and 19 males (age: *M* =27.53).

Table 1: Number of items per word.

Vowel	Word	Number	Word	Number
/iː/	heat	90	heed	62
/ɪ/	hit	86	hid	90
/ɐː/	heart	85	hard	87
/ɐ/	hut	84	hud	89

2.3. Acoustic Analysis

Audio data were first processed by the MAUS automatic aligner [26] using an AusE model. Automatically generated Praat textgrids [3] were hand corrected and further augmented to identify the following acoustic events using criteria established in [27]: (1) vowel onset and offset, (2) start and end of irregular pitch periods, (3) voice bar, (4) burst.

The resulting textgrids were imported into Emu (http://emu.sourceforge.net/) for formant checking and data extraction in R (www.r-project.org) [31]. Formants were automatically tracked using ESPS/Waves (12th order LPC analysis with a 25 ms. raised cosine window and a frame shift of 5 ms). Formant data were speaker normalised [21]. F1 and F2 at vowel target, vowel duration, consonant closure duration, and duration of irregular pitch period within the vowel were extracted. This analysis is restricted to the vowel duration and consonant closure duration.

Figure 1 shows data from male speakers' hVt words to illustrate the F1/F2 spectral cues associated with the long/short vowel pair differentiation. The top panel shows F1/F2 normalised data to illustrate the spectral differentiation that pertains for /iː, ɪ/ as opposed to the closer spectral relationship between /ɐː, ɐ/. The bottom panel illustrates time-normalised trajectories of F1 and F2 separately for /iː/ and /ɪ/ showing the characteristic onglide for /iː/.

2.4. Statistical Analysis

Multilevel modeling (SPSS, Version 22.0, mixed procedure) was used in these analyses. The fixed factors were voicing (voiced, voiceless), length (long, short) and vowel pair (/iː, ɪ/, /ɐː, ɐ/). Speaker was included as a random factor. It was not possible to conduct three-way tests of significance as the dataset would not have provided sufficient power for such analyses.

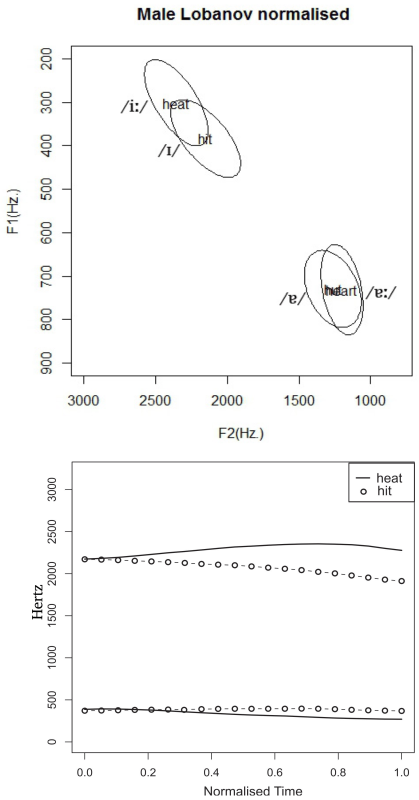

Figure 1: Male speakers' hVt words. Top Panel: Vowels in the F1/F2 plane - 95% CI ellipses. Bottom Panel: F1 and F2 trajectories for vowels in *heat* and *hit* over normalised time.

3. RESULTS

3.1. Vowel Duration

Vowel duration was examined for each pair of vowels separately to compare the factors voicing and inherent vowel length. Results show main effects for voicing, length and a voicing by length interaction in both vowel pairs. In all cases the effects reached significance beyond $p<0.0001$. For /iː, ɪ/: [length: $F(1,324)=703.65$, voicing: $F(1,324)=340.74$, length X voicing: $F(1,324)=133.92$]. For /ɐː, ɐ/: [length: $F(1,341)=1269.06$, voicing: $F(1,341)=123.40$, length X voicing: $F(1,341)=18.04$]. Figure 2 illustrates these effects (solid bars) for short (left panels) and long vowels (right panels). The voiced context is represented in the top panels and voiceless in the bottom panels.

As expected, long vowels were significantly longer than short vowels and vowels before voiced consonants were significantly longer than those before voiceless consonants. Importantly, there was a larger difference between the voiced and voiceless contexts for long vowels compared to short vowels.

3.2. Closure Duration

Closure duration was compared across length and voicing conditions separately for each vowel pair. Results show main effects for voicing and length in both vowel pairs but a voicing by length interaction was only found for the /iː, ɪ/ pair. In all cases the effects reached significance beyond $p<0.0001$. For /iː, ɪ/: [length: $F(1,324)=16.84$, voicing: $F(1,324)=38.29$, length X voicing: $F(1,324)=13.28$]. For /ɐː, ɐ/: [length: $F(1,341)=60.33$, voicing: $F(1,341)=24.88$]. Short coda closures were associated with long vowels, voiced codas were shorter than voiceless codas and for /iː, ɪ/, the coda closure in the long vowel context differed according to voicing more than codas in the short vowel context. Figure 2 illustrates these effects (striped bars) showing, in particular, reduction in the voiceless coda closure following /ɐː/ compared to /iː/.

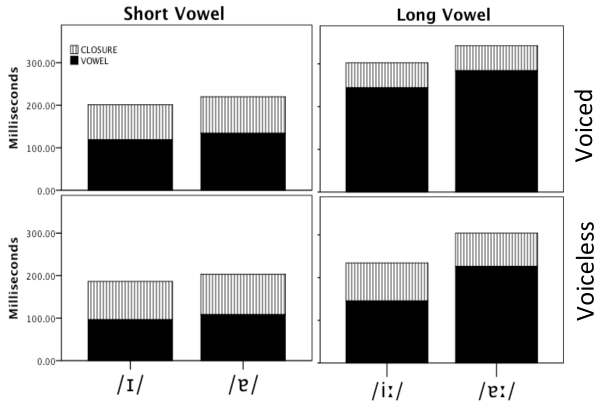

Figure 2: Vowel and coda closure absolute durations. Left panels: short vowels; Right panels: long vowels. Top panels: voiced context; Bottom panels: voiceless context.

3.3. C/V Ratio – Contrast Variation

In order to make comparisons between the pairs of vowels we calculated the closure/vowel duration ratio (C/V ratio) as a metric to reflect the syllable composition. Values approaching 1 indicate close correspondence in duration between the vowel and the closure. Lower values indicate a proportionately longer vowel than closure. We examined long and short vowels separately and compared voicing and vowel type as fixed factors, with speaker as a random factor. Figure 3 illustrates the results.

For short vowels (left panel) the only effect to reach significance was voicing [$F(1,345)=52.40$, $p<.0001$] with voiceless rhymes having larger ratios indicating that the vowel and closure comprise a relatively equal division across the syllable (*hit* 0.98 and *hut* 0.94). In the voiced short vowel case, lower ratios indicate that a larger portion of the syllable was dedicated to the vowel (*hid* 0.72 and *hud* 0.67).

In general, the long vowels occupy a greater proportion of the syllable than short vowels. This is reflected in the lower C/V ratios. For long vowels a complex set of relationships was found. Significant effects were beyond $p<0.0001$ for voicing [$F(1,320) =196.11$], vowel [$F(1,320) =64.73$] and a voicing X vowel interaction [$F(1,320) =38.23$]. In the voiced context long vowels displayed the smallest ratios (*heed* 0.24 and *hard* 0.21) indicating that the vowel dominates the syllable. This context is facilitatory for both vowel contrast and coda voicing because the long vowel is required for the voiced context and voiced codas require short closures [17].

The interaction between voicing and vowel shows that there is a larger difference in C/V ratio between /iː/ and /ɐː/ in the voiceless compared to the voiced context. /iː/ displays a larger ratio (*heat* 0.63) than /ɐː/ (*heart* 0.36). In other words, in the voiceless context where vowels are typically shortened, /ɐː/ retains considerable length relative to the closure. Conversely, /iː/ does not exhibit the same effect and undergoes shortening as would be expected in the voiceless context (Figure 2 illustrates). These results might suggest some resistance to shortening in the case of /ɐː/ but not /iː/.

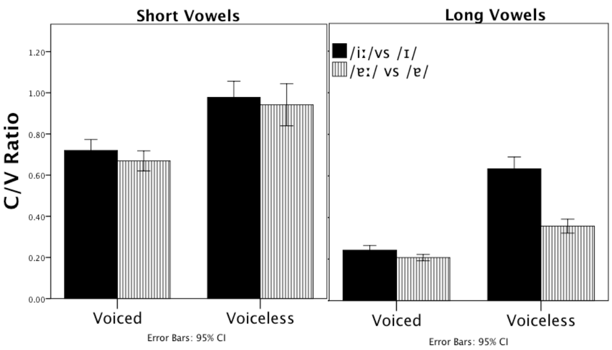

Figure 3: The C/V ratios according to voicing and length (left panel: short vowels, right panel: long vowels). /iː, ɪ/ (solid bars) /ɐː, ɐ/ (striped bars).

4. DISCUSSION

The results confirm the trading relationship between the vowel and coda, previously discussed in [38, 10] whereby, within voicing class, inherently short vowels are associated with longer codas than inherently long vowels. We also found the expected relationship between coda voicing and preceding vowel length and confirmed that long vowels lengthen more than short vowels in the voiced context. This result could reflect a constraint on short vowel lengthening which preserves the short/ long vowel contrast. Long vowels, however, are free to lengthen in the voiced context. The results also support our prediction that the degree of shortening licenced for /ɐː/ in the voiceless context would be less than /iː/ because length is more crucial for /ɐː/.

The results for closure duration after /iː/ showed longer closures in the voiceless context compared to the voiced (as expected) whereas for /ɐː/ there was not the same increase in closure in this context. These results raise the question of whether the compromise of closure duration is a strategy to accommodate the vowel length required for /ɐː/ to retain its contrast with /ɐ/.

The C/V ratio allowed us to directly compare the impact of different vowel types on temporal management. Results show that vowel type was influential in the temporal organisation of the syllable. When /ɐː/ occurred in an environment that had the potential to reduce the salience of the contrast (i.e. voiceless), adjustments were made within the syllable to supports the integrity of the contrast. For /ɐː/, vowel length was maintained at the expense of closure. In the case of /iː/ which has a range of supporting cues to its identity, the voiceless context did not have the same impact, possibly because other cues are available to ensure contrast integrity. Luce and Charles-Luce [22] also found a difference in CV ratio for high vs low vowels but this difference varied according to sentential context, motivating further investigation.

For /ɐː/, maintaining vowel length could benefit the vowel contrast but may be detrimental to the coda. If coda closure is compromised and the preceding vowel does not behave in the predicted fashion (i.e. we expect it to be short before a voiceless stop) how is the stop voicing contrast realised? Coda voicing is cued by a constellation of features including closure voicing, preceding vowel F0 and F1 offset and burst features [16, 30, 36]. These factors require examination before a clear picture can emerge of how contrast is implemented in the face of competing demands.

The results reported here are specific to just two sets of vowels in highly restricted environments. However, the findings raise the question of how language-specific cue weighting contributes to the phonetic implementation of phonology within the rhyme. A larger set of vowels in various segmental and prosodic contexts will provide a more complete picture of the extent to which phonetic implementation is affected by cue weighting and to explore how listeners use this variable information.

5. CONCLUSION

We found that rhymes containing long/short vowel pairs varying in the relative weighting of contrastive duration were organised differently with respect to vowel-coda timing relationships within the rhyme. These findings raise further questions about how speakers manage contractiveness in the face of the competing demands that the phonology places on temporal structure.

6. REFERENCES

[1] Bernard, J. 1970. A cine-X-ray study of some sounds of Australian English. *Phonetica* 21, 138-150.

[2] Bernard, J. 1970. On nucleus component durations. *Language & Speech* 13, 89-101.

[3] Boersma, P., Weenink, D. 2009. Praat: doing phonetics by computer (Version 5.1. 05) Retrieved May 1, 2009.

[4] Burnham, D., et al. 2009. A blueprint for a comprehensive Australian English auditory-visual speech corpus. In Haugh M. et al. (eds), *Selected Proceedings of the 2008 HCSNet Workshop on Designing the Australian National Corpus*, 96-107.

[5] Caisse, M. 1988. *Modeling English vowel durations* (Doctoral dissertation), University of California, Berkeley Dec. 1988.

[6] Chen, M. 1970. Vowel length variation as a function of the voicing of the consonant environment. *Phonetica* 22, 129-159.

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

[8] Cox, F., Palethorpe, S. 2007. Australian English. *JIPA* 37, 341-350.

[9] Cox, F., Palethorpe, S. 2008. Reversal of short front vowel raising in Australian English. In *INTERSPEECH* 342-345.

[10] Cox, F. Palethorpe, S. 2011. Timing differences in the VC rhyme of Standard Australian English and Lebanese Australian English, Proceedings of the 17th ICPhS, Hong Kong, 17-21 August.

[11] Cox, F., Palethorpe, S., Bentink, S. 2014. Phonetic Archaeology and 50 Years of Change to Australian English /iː/. *AJL* 34, 50-75.

[12] De Jong, K. 2004. Stress, lexical focus, and segmental focus in English: patterns of variation in vowel duration. *J Phon* 32, 493-516.

[13] Fletcher, J., McVeigh, A. 1993. Segment and syllable duration in Australian English. *Speech Comm* 13, 355-365.

[14] Fowler, C. A. 1992. Vowel duration and closure duration in voiced and unvoiced stops: There are no contrast effects here. *J Phon* 20, 143-165.

[15] Harrington, J., Cox, F., Evans, Z. 1997. An acoustic phonetic study of broad, general, and cultivated Australian English vowels. *AJL* 17, 155-184.

[16] Hillenbrand, J. M., Clark, M. J., Nearey, T. M. 2001. Effects of consonant environment on vowel formant patterns. *JASA* 109, 748-763.

[17] Kingston, J., Diehl, R.L. 1994. Phonetic knowledge. *Language* 70, 419-454.

[18] Klatt, D.H. 1976. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *JASA* 59, 1208-1221.

[19] Lisker, L. 1957. Closure duration and the intervocalic voiced-voiceless distinction in English. *Language* 33, 42-49.

[20] Lisker, L. 1986. "Voicing" in English: a catalogue of acoustic features signalling /b/ versus /p /in trochees. *Language & Speech* 29, 3-11.

[21] Lobanov, B.M. 1971. Classification of Russian vowels spoken by different speakers. *JASA* 49, 606-608.

[22] Luce, P. A., Charles-Luce, J. 1985. Contextual effects on vowel duration, closure duration, and the consonant/ vowel ratio in speech production. *JASA* 78, 1949-1957.

[23] Maddieson, I. 2010. Variation in co-variation: the search for explanatory principles. Iin Fougeron, C., Kühnert B., D'Imperio M., and Vallée N. (eds) *Laboratory Phonology 10*. Berlin:de Gruyter, 133-150.

[24] Port, R. F, Dalby, J. 1982. Consonant/vowel ratio as a cue for voicing in English. *Percept Psychophys* 32, 141-152.

[25] Raphael, L. J. 1972. Preceding Vowel Duration as a Cue to the Perception of the Voicing Characteristic of Word-Final Consonants in American English. *JASA* 51, 1296-1303.

[26] Schiel, F., Draxler, C. & Harrington, J. 2011. Phonemic Segmentation and Labelling using the MAUS Technique. *Workshop 'New Tools and Methods for Very-Large-Scale Phonetics Research'*. University of Pennsylvania, Philadelphia.

[27] Shattuck-Hufnagel, S., Demuth, K., Hanson, H., Stevens, K.N. 2011. Acoustic cues to stop-coda voicing contrasts in the speech of 2-3-year-olds learning American English. In: Clements G.N. & Ridouane R. (eds*), Where do features come from? The nature and sources of phonological primitives*. Elsevier: North-Holland Linguistic Series.

[28] Solé, M.J. 2007. Controlled and mechanical properties in speech: a review of the literature. In: Solé, M.J, Beddor, P. & Ohala M. (eds), *Experimental Approaches to Phonology*. Oxford: Oxford University Press, 302-321.

[29] Solé, M.J. & Ohala, J.J. 2010. What is and what is not under the control of the speaker. Intrinsic vowel duration. In Fougeron C., Kühnert B., D'Imperio M., and Vallée N. (eds.) *Papers in Laboratory Phonology 10*. Berlin: de Gruyter, 607-655.

[30] Steriade, D. 1997. Phonetics in phonology: The case of laryngeal neutralization. Ms. UCLA.

[31] R Core Team. 2013. R: A language and environment for statistical computing. R Foundation.

[32] Turk, A. E., Shattuck-Hufnage,l S. 2007. Multiple targets of phrase-final lengthening in American English words. *J Phon* 35, 445-472.

[33] Turk, A. E., White, L. 1999. Structural influences on accentual lengthening in English. *J Phon* 27, 171-206.

[34] Umeda, N. 1977. Consonant duration in American English. *JASA* 61, 846-858.

[35] van Dommelen, W. A. 1999. Auditory accounts of temporal factors in the perception of Norwegian disyllables and speech analogs. *J Phon* 27, 107-123.

[36] Wardrip–Fruin, C. 1982. On the status of temporal cues to phonetic categories: Preceding vowel duration as a cue to voicing in final stop consonants. *JASA* 71, 187-195.

[37] Watson, C. I., Harrington, J. 1999. Acoustic evidence for dynamic formant trajectories in Australian English vowels. *JASA* 106, 458-468.

[38] Yuen, I., Cox, F., Demuth, K. 2014. Three-year-olds' production of Australian English phonemic vowel length as a function of prosodic context. *JASA* 135, 1469-1479.