



Hyperarticulation in Short Intonational Phrases in Three Australian Languages

Simone Graetzer¹, Janet Fletcher², John Hajek²

¹ Voice Biomechanics and Acoustics Laboratory, Michigan State University, U.S.A.

² Phonetics Laboratory, University of Melbourne, Australia

sgraetz@msu.edu, j.fletcher@unimelb.edu.au, j.hajek@unimelb.edu.au

Abstract

In Lindblom's Hyper- and Hypo-articulation (H & H) theory, speech varies between clear and less clear depending on the communicative context. Hyperarticulation is known to reflect prosodic boundary information and prosodic prominence or focus. The realisation of hyperarticulation appears to differ between languages. In this study of three Australian languages, it is asked whether, in pre-boundary position in short prosodic phrases, vowel lengthening tends to co-occur with acoustic evidence of hyperarticulation. Further, it is asked whether hyperarticulation is stronger in the pre-boundary syllable than in the post-boundary one. It is demonstrated that pre-boundary lengthening in short intonational phrases in Australian languages tends to co-occur with an increase in vowel peripherality.

Index Terms: hyperarticulation, vowel space area, Euclidean distance, prosodic boundary, acoustic phonetics

1. Introduction

According to Lindblom's Hyper- and Hypo-articulation theory [1], speech varies between clear and less clear. Prosodically induced hyperarticulation may be termed *localised hyperarticulation*, while the hyperarticulation that occurs in (especially, clear) laboratory speech may be termed *communicatively-driven hyperarticulation*. Both forms involve an increase in the perceptual saliency of the (hyperarticulated) speech material.

Localised hyperarticulation is known to occur in prosodically prominent locations. For example, in English, vowels in prosodically prominent syllables (e.g., syllables associated with a nuclear accent) lengthen and have more peripheral targets than vowels in weak prosodic contexts [2, 3]. Previous studies have also shown that in pre-boundary or final position, segment duration tends to correlate with acoustic parameters – including sharper or more peripheral vowel targets – indicating hyperarticulation, such that longer durations co-occur with greater hyperarticulation [2, 3]. In post-boundary (or domain-initial) position, temporal and spatial expansion can also occur [3].

Descriptions of prosody in Australian languages (e.g., [4]) suggest that they are head/edge prominence languages, in which both the left and right edges of constituents are marked for prosodic salience (after the prosodic typology outlined in [5]). Intonational pitch accents combine with phrase tones to mark the left and right edges of higher-level prosodic constituents, *i.e.*, Intonational Phrases (IPs). A range of earlier studies suggest that vowel quality is typically only marginally affected by the presence or absence of intonational accent, unlike languages such as English. Rather, post-intonational

accent word-medial consonants, in, for example, Warlpiri and Mawng, appear to carry important prosodic information. It has been argued that this is due to the fact that the full set of phonological consonant contrasts is realised only in word-medial position (e.g., [6]). Conversely, in acoustic phonetic studies of other Australian languages, pre-IP boundary vowel lengthening has been observed [4]. It has also been found that longer vowels in these languages have more peripheral vowel targets, which contributes to the analysis of many of these languages as head/edge prominence languages. In other words, these languages tend to have strong domain-initial and -final pitch patterns associated with higher level prosodic structure, but evidence to date suggests that supra-glottal cues to prosodic structure are more evident either on post-accent medial consonants or domain-final vowels.

The goal of the present study was to examine potential localised hyperarticulation in highly controlled speech material in three Australian languages, and, in particular, to determine how post-boundary stressed and pre-boundary unstressed vowels are realised durationally and spectrally in short (disyllabic) phrases. In this study, the languages Gupapuyngu, Warlpiri and Burarra were chosen for analysis due to the number of speakers available, to the taxonomic differences (Pama-Nyungan vs. non-Pama-Nyungan), and the vowel and consonant inventory differences (see Section 2). Another important factor determining the choice is that Burarra is one of few Australian languages in which vowels tend to centralise in non-prosodically prominent locations in words [4]. It was important to consider three distinct languages to determine whether the typological observations noted are typical of all Australian languages.

The major research question was as follows: is there evidence of greater hyperarticulation of vowels in the first or second syllable of isolated disyllabic Australian words? Given findings for English and a number of different languages that segments are lengthened preceding prosodic boundaries [4, 7, 8], pre-boundary vowel lengthening is predicted. Additionally, given previous findings for Australian languages of vocalic pre-boundary lengthening, and greater peripherality in longer vowels than shorter ones, it is predicted that there is more hyperarticulation of vowels in pre-boundary than post-boundary position. If these predictions are confirmed, the study will support the claim that these languages are indeed head/edge prominence languages.

2. Methods

The subjects were 9 adult female speakers of Burarra, Gupapuyngu, and Warlpiri (3 speakers of each language). Detailed information about the three languages and the data collection process is provided elsewhere [9]. Gupapuyngu and

Warlpiri are members of the Pama-Nyungan family of Australian languages, while Burarra is non-Pama-Nyungan. Burarra and Gupapuyngu are spoken in Arnhem Land, and Warlpiri is spoken to the NW of Alice Springs. These languages have many place and few manner contrasts, and small vowel inventories: 3 vowels with a length contrast in Gupapuyngu and Warlpiri /i a u/, and 5 vowels in Burarra /i ɛ a o u/, although the distribution of vowels in word-final position is limited (Table 1).

The production of a list of isolated words by the speakers involved in the present study led to the words receiving higher-order prominence. In isolated disyllabic words, which are arguably realised as full IPs, both the post-boundary syllable (first position) and the pre-boundary syllable (second position) are strengthened.

Recordings were made in the field. Participants were asked to say real words (three times) without a carrier phrase at a self-selected normal rate. The word lists were intended to capture all consonant contrasts in each language. Real rather than nonsense words were elicited for practical and cultural reasons. No instructions were given as to the carefulness of the speech.

Following manual labelling of the relevant vowels for each speaker, a prosodic annotation was undertaken. It was assumed that the vowel carrying a sharp rise in fundamental frequency (*fo*) to a peak somewhere in or near the syllable rhyme was accentually prominent. In this data set, the main stress was word-initial. Words were uttered in isolation, and hence, as mentioned, were associated with phrasal prosodic prominence.

Emu 2.3.0 [10] and R 3.1.2 [11] were used for signal processing and statistical analysis. Detailed information about the relevant segmentation and formant estimation methods is reported elsewhere [12], [13].

Table 1. Vowel distribution (counts of observations in the data set), where position 1 indicates the vowel of the first syllable, and 2, the second.

	Position	a	i	u
Burarra	1	150	65	97
	2	422	23	5
Gupapuyngu	1	152	48	110
	2	281	108	165
Warlpiri	1	219	102	124
	2	199	110	136

Formant estimation at the midpoint of each vowel was carried out by means of a Linear Predictive Coding algorithm. Tracking errors involving portions of the formant track being mistakenly associated with higher or lower formants, were manually corrected. Using Bonferroni-adjusted outlier tests in the *car* package (in R), the largest absolute studentised residuals in the *F1* and *F2* data, which were few in number, were identified and removed, given the sensitivity of the vowel space area calculations to be performed. *F1* frequencies and vowel duration were analysed using non-parametric Kruskal-Wallis rank sum tests, with the position of the vowel within the word (hereafter *position*) as the independent variable. The *p* values were adjusted using the Benjamini-Hochberg False Discovery Rate method.

Linear mixed effects models (LMEs) were fit by restricted maximum likelihood (REML) with response variables

comprising Euclidean distances and vowel space area (VSA) parameters. Random effects terms were chosen on the basis of variance explained. Models were selected on the basis of the Akaike information criterion and the results of likelihood ratio tests and were built using *lme4*, *lmerTest* and *multcomp* R packages [11]. The *p* values for these multiple comparisons were adjusted using the single-step method, in which values are computed from the joint normal or *t* distribution of the *z* statistics. The LME output includes the estimates of the fixed effects coefficients, the standard error (*SE*) associated with the estimate, the degrees of freedom, *df*, the test statistic, *t*, and the *p* value. The Satterthwaite method is used to approximate the *df* and calculate *p* values.

The VSA method of triangular vowel space area calculation is defined as follows by Sapir *et al.* [14] amongst others:

$$VSA = \text{abs}\left(\frac{F1i \cdot (F2a - F2u) + F1a \cdot (F2u - F2i) + F1u \cdot (F2i - F2a)}{2}\right) \quad (1)$$

Where *F1* indicates formant 1, *F2*, formant 2, and *i*, *a*, *u* refer to the three vowel qualities. Data visualisation of the VSA was performed using the convex hull estimation (or *chull*) method, which involves the calculation of the area defined by the subset of points lying on the convex hull of the set of *F1* and *F2* frequencies in two dimensional space. A custom function was written in R for both area calculation methods.

Mel transformation of *F1* was performed by means of the following equation, where *f* is the original value in Hertz and *m* is the output in mel:

$$m = 1127.01048 \cdot \log_e\left(1 + \left(\frac{f}{700}\right)\right) \quad (2)$$

Euclidean distances – straight line distances between points *p* and *q* – were calculated as a measure of the magnitude of dispersion in the vowel space according to the equation given below:

$$d(p, q) = \sqrt{\sum_{i=1}^n (q_i - p_i)^2} \quad (3)$$

where *d(p, q)* denotes the distance from *p* to *q*. In the present context, the two-dimensional space is the *F2* x *F1* plane, *p* is /a/, and *q* is the associated grand centroid (of the triangle formed by /a i u/), per speaker and vowel position.

As all three vowel systems are relatively small (with either three or five vowel qualities), the issue of an effect of inventory size on vowel variability or dispersion is not addressed here.

3. Results

3.1. Vowel duration in short vowels

For statistical purposes, the null hypothesis (H0) for duration is that the location parameters of the distribution of vowel duration (ms) are the same in each position for each language. After the exclusion of phonologically long vowels, which occurred only in the first position and tend to occur infrequently in Gupapuyngu and Warlpiri, Kruskal-Wallis tests per language of vowel duration by position (with Benjamini-Hochberg corrected *p* values [15]) indicated that the duration was longer in each language in the second position at *p*<0.0001. As shown in Figure 1, the means were 164 and 201ms in Burarra, 147 and 273ms in Gupapuyngu, and 138 and 184ms in Warlpiri, in first and second position, respectively. H0 was therefore rejected for all languages. In

sum, final position is associated with significant pre-boundary lengthening, particularly in Gupapuyngu.

3.2. Triangular vowel space area

The null hypothesis is that there is no difference in triangular VSA between the two positions, and that there is no interaction of position and language. A LME model with triangular VSA as the response variable and the interaction term vowel position and language (and subject as the random effect) indicated that VSA was smaller in the first position, and that there was no difference in this pattern between languages (*i.e.*, no observable interaction). In the case of Burarra, while there is a low number of /i u/ tokens in the second position, this low number would likely lead to an underestimation of the area for the second position, and hence is unlikely to be responsible for the result.

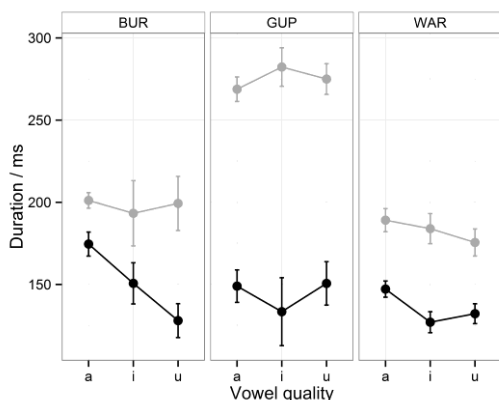


Figure 1: Vowel duration means with associated 95% confidence intervals (*BUR*: Burarra, *GUP*: Gupapuyngu, *WAR*: Warlpiri) where position 1 (post-boundary position) is shown in black, and position 2 (pre-boundary position) in grey.

To validate the model, the analysis was re-run 100 times with parametric bootstrapping (using the *bootMer* function in *lme4* [11]). 95% confidence intervals indicated an effect of position such that VSA was smaller in the first position, although the population parameter could not be estimated precisely (Position 1: 2.5% -128687, 97.5% -738). There was no effect of language and no interaction between position and language (where the intervals spanned 0).

Table 2. LME model output for VSA (reference levels: position 2, Lang BUR).

Term	Estimate	SE	df	t
(Intercept)	204445	26326	6	<0.0001
Position 1	-68915	7695	64	<0.0001
Lang GUP	77802	37102	6	0.08
Lang WAR	-34920	37069	6	0.38
Position 1: Lang GUP	2266	12442	65	0.86
Position 1: Lang WAR	918	10516	64	0.93

3.3. F1 frequency in /a/

Given the need to control for differences in vowel distribution in the two positions, the null hypothesis regarding *F1* is that the location parameters of the distribution of /a/ vowel *F1* frequency are the same in each position for each language. It is assumed that the size of *F1* reflects lip aperture and jaw opening.

With the exception of Burarra speaker, MW, there was similar or greater *F1* variance in /a/ in the second position relative to the first for all speakers and languages. Kruskal-Wallis tests per language of *F1* by position (with Benjamini-Hochberg corrected *p* values) indicated that, while the first position was associated with lower *F1* values for all languages (see Figure 2), and for all speakers except Burarra speaker, KF, for whom there were fewer data points, the difference reached significance only for Gupapuyngu and Warlpiri (Burarra, $H = 1.25$, $df = 1$, $p = 0.26$; Gupapuyngu, $H = 42.97$, $df = 1$, $p < 0.0001$; Warlpiri, $H = 20.28$, $df = 1$, $p < 0.0001$).

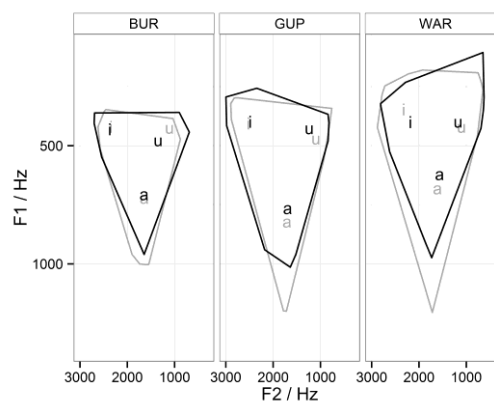


Figure 2: Vowel medians and VSA per language (*BUR*: Burarra, *GUP*: Gupapuyngu, *WAR*: Warlpiri) where position 1 is shown in black, and position 2 in grey. Polygons are fit using a convex hull estimation.

Results were very similar when the tests were applied to *F1* converted into mel per language. The *F1* means were 713 and 721Hz in Burarra, 763 and 837Hz in Gupapuyngu, and 608 and 661Hz in Warlpiri, in first and second position, respectively. Given that *F1* rises as the jaw opens, these results suggest more open vowels in the second position. H_0 in this case was rejected for Gupapuyngu and Warlpiri but retained for Burarra.

3.4. Euclidean distances for /a/

Once again, due to the need to control for differences in vowel distribution in the two positions, the Euclidean distance analysis was constrained to /a/ in each of the languages. The null hypothesis in this case is that there is no difference in Euclidean (straight line) distances between the grand centroid (considering all vowels) and /a/, calculated per speaker, between the positions, for each language.

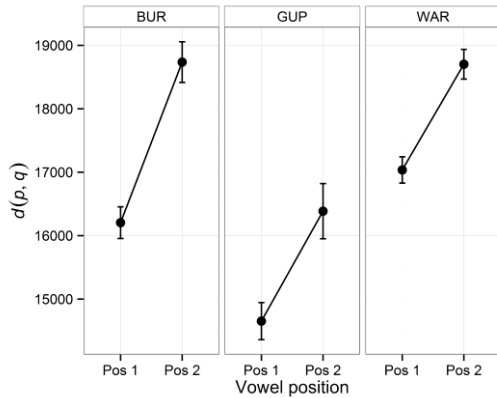


Figure 3: Euclidean distance means for /a/ with associated 95% confidence intervals per language (BUR: Burarra, GUP: Gupapuyngu, WAR: Warlpiri) and vowel position.

Linear regression models fit per language with Euclidean distances as the response variable and vowel position as the independent variable indicated that distances were lower in the first position (as shown in Figure 3), *i.e.*, the vowel in the second position was more open, in agreement with the results for $F1$ (with second position as the reference level: Burarra, $\hat{\beta} = -2588\text{Hz}$, SE = 270, df = 568, $p < 0.0001$; Gupapuyngu, $\hat{\beta} = -1609\text{Hz}$, SE = 299, df = 429, $p < 0.0001$; Warlpiri, $\hat{\beta} = -1649\text{Hz}$, SE = 156, df = 414, $p < 0.0001$). H_0 is therefore rejected for all languages.

4. Discussion

The results of this study of three Australian languages indicate that the spectral properties of vowels do not necessarily reflect patterns that are typically reported for higher-level prosodic prominence in other languages. While vowels in first position in disyllabic words/phrases in these languages are typically associated with the highest level of prosodic prominence (intonational accent), vowels in the second position (pre-boundary) are longer in duration, /a/ is more open (or, in Burarra, is more distant from the centroid), and, in general, vowels are often more peripheral in domain final position. Hence, pre-boundary vowels appear to be more strongly hyperarticulated than post-boundary intonationally accented vowels, certainly for Gupapuyngu and Warlpiri, and on most parameters also for Burarra.

Regarding cross-linguistic differences in the use of the vowel space, the differences between languages with regard to $F1$ frequency in /a/ may be associated with previous findings that Burarra vowels, unlike those of the other languages, are known to undergo some magnitude of vowel reduction in unstressed syllables (here, position 2) [16]. A possible explanation of the finding of greater lengthening of vowels in pre-boundary position in Gupapuyngu may be that vowels are regularly lengthened phonemically (but in non-final position) in this language, and so such phonetic lengthening is not unusual.

Notwithstanding these language differences, the results are consistent with previous findings that (1) pre-boundary lengthening tends to correlate with acoustic parameters reflecting potential localised hyperarticulation in this position and (2) the increased duration of the vowel results in better

achievement of the vowel target, *i.e.*, a more peripheral vowel in the $F2 \times F1$ space. Given that it has been documented elsewhere that the magnitude of the lengthening tends to increase as the strength of the following boundary increases (e.g., [3]), it is not surprising that in this data set – where a pre-final IP boundary was always present – there were large differences in vowel durations between the two vowel positions. These large differences are consistent with the claim that pre-boundary lengthening is an active process rather than a merely biomechanical one in these languages. The results may also reflect the earlier observation that Australian indigenous languages are largely head/edge prominence languages, with post-boundary prominence marked primarily by pitch (not examined in this study) and the right edges of constituents marked by longer segment durations [4, 5].

5. Conclusions

In this study, evidence is presented for the claim that pre-boundary lengthening in short intonational phrases in three Australian languages co-occurs with acoustic evidence of hyperarticulation, specifically, a tendency towards increased vowel peripherality, most clearly in Gupapuyngu and Warlpiri and to a lesser extent in Burarra. It can be argued on the basis of the reported results that there is greater hyperarticulation in pre-boundary than post-boundary vowels in short intonational phrases in these languages.

6. Acknowledgements

The authors would like to thank Andrew Butcher for access to the data set and the subjects who consented to being recorded. This work was funded in part by an Australian Postgraduate Award granted to the first author.

7. References

- [1] B. Lindblom, "Spectrographic study of vowel reduction," *J. Acoust. Soc. Am.*, vol. 35, pp. 1773-1781, 1963.
- [2] J. Edwards, M. E. Beckman, and J. Fletcher, "The articulatory kinematics of final lengthening," *J. Acoust. Soc. Am.*, vol. 89, pp. 369-382, 1991.
- [3] C. Fougerson and P. Keating, "Articulatory strengthening at edges of prosodic domains," *J. Acoust. Soc. Am.*, vol. 101, pp. 3728-3740, 1997.
- [4] J. Fletcher and A. R. Butcher, "Sound patterns of Australian languages," in *The Languages and Linguistics of Australia: A Comprehensive Guide (The World of Linguistics)*, vol. 4, H. Koch and R. Nordlinger, Eds. Berlin: Mouton de Gruyter, pp. 89-132, 2014.
- [5] S. -A. Jun, "Prosodic typology: by prominence type, word prosody, and macro-rhythm," in *Prosodic Typology II: The Phonology of Intonation and Phrasing*, S. -A. Jun, Ed. Oxford University Press, Oxford, pp. 520-539, 2014.
- [6] A. R. Butcher and J. Harrington, "An acoustic and articulatory analysis of focus and the word/morpheme boundary distinction in Warlpiri," *6th Int. Seminar on Speech Production, Proceedings*, Macquarie University, Sydney, 2003.
- [7] D. H. Klatt, "Vowel lengthening is syntactically determined in a connected discourse," *J. Phon.*, vol. 3, pp. 129-140, 1975.
- [8] T. Cho, "Language effects on timing at the segmental and suprasegmental levels," in *The Handbook of Speech Production*, M. A. Redford, Ed. John Wiley and Sons, Oxford, pp. 505-529, 2015.
- [9] S. Graetzer, "An acoustic study of coarticulation: Consonant-vowel and vowel-to-vowel coarticulation in four Australian languages," Ph.D. dissertation, School of Lang. and Ling., University of Melbourne, Victoria, 2012.

- [10] S. Cassidy and J. Harrington, "Multi-level annotation in the Emu speech database management system," *Sp. Comm.*, vol. 33, no. 1-2, pp. 61-77, 2001.
- [11] R Development Core Team; R: a language and environment for statistical computing [Online]. Available: <http://www.R-project.org>
- [12] S. Graetzer, J. Fletcher, and J. Hajek, "Prosodic effects on vowel spectra in three Australian languages," *7th Int. Conf. on Speech Prosody, May 20-23, Dublin, Ireland, Proceedings*, 2014.
- [13] S. Graetzer, J. Fletcher, and J. Hajek, "Locus Equations and coarticulation in three Australian languages," *J. Acoust. Soc. Am.*, vol. 137, 806-821, 2015.
- [14] S. Sapir, L. O. Ramig, J. L. Spielman, and C. Fox, "Formant centralization ratio (FCR): A proposal for a new acoustic measure of dysarthric speech," *J. Speech Lang. Hear. Res.*, vol. 53, no. 1, pp. 114-125, 2010.
- [15] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *J. R. Statist. Soc. B*, vol. 57, no. 1, pp. 289-300, 1995.
- [16] Butcher, A. R., "Australian Aboriginal languages: consonant salient phonologies and the 'place-of-articulation imperative'", in J. M. Harrington and M. Tabain [Eds], *Speech Production: Models, Phonetic Processes and Techniques*, Psychology Press, 187-210, 2006.