THE CONTRIBUTION OF SCHWA VOWELS TO THE PROSODIC ACCENT CONTRAST IN AUSTRALIAN ENGLISH

Sallyanne Palethorpe[1], Mary E. Beckman[2], Janet Fletcher[3], Jonathan Harrington[1]
1. Speech Hearing and Language Research Centre, Macquarie University, Sydney, Australia.
2. Department of Linguistics, Ohio State University, Columbus, Ohio, U.S.A.
3.Department of Linguistics, University of Melbourne, Australia.

ABSTRACT

Previous studies of the supralaryngeal correlates of prosodic accentuation in English have shown both an increase in the paradigmatic contrast between phonetic segments and a magnification of the syntagmatic contrast between abutting consonants and vowels in accented words. We analysed tongue and lip movement data from four Australian English talkers in dialogues that elicited accented and deaccented productions of trochaic /bVbə/ words. The results showed greater lip aperture magnitudes and peak velocities in the opening and closing gestures of the initial /bV/ syllables and in the opening gestures of the final weak /bə/ syllables. The distance of the rhythmically strong vowels to the centre of the vowel space was generally greater for accented words in both tongue position and formant spaces. The results suggest an enhanced syntagmatic contrast between vowels and consonants in both strong and weak syllables and a marginally greater paradigmatic contrast between the vowel targets in strong syllables.

1. INTRODUCTION

In all languages, one of the functions of prosody is to convey variations in relative prominence among elements in an utterance, and in English, one kind of prominence variation is achieved by marking some words as accented. This accentual marking is characterised by pitch changes that can be accounted for phonologically by the association of one of a number of pitch accents to the syllable that bears primary lexical stress – that is, the rhythmically strongest syllable of the accented word. These pitch changes are especially marked when the accented word is in nuclear accented position because of the juxtaposition of the pitch accent with the following phrase boundary tones.

The results of various studies are consistent with a long-standing idea that accented syllables are marked not just intonationally, but also by being louder and longer than unaccented ones [1]. The evidence comes not only from acoustic and psychophysical analyses of the accented/unaccented contrast [2,3], but also from kinematic investigations. For example, there are a number of studies (e.g., [4-6]) that have shown lower jaw positions and a more open vocal tract in accented syllables with open vowel nuclei: the vowel is therefore likely to be louder absolutely because of the greater radiation of acoustic energy from the more open vocal tract over a longer interval of loudness summation [7]. These kinds of data underlie the model of accent as sonority expansion developed in Beckman, Edwards and Fletcher [8]. In this model, vowels as a class have inherently higher 'sonority' than consonants (i.e. inherently more acoustic energy concentrated in psychophysically effective frequency regions), but accented vowels are even more sonorous. In an analysis of the accent contrast in high vowels, Harrington, Fletcher and Beckman [9] extended this theory by emphasising the greater sonority *rise* from a tautosyllabic consonant to a following vowel in accented syllables. They pointed out that the consonant's inherently lower sonority is also exaggerated under accent (e.g. the jaw is higher in a [b] before an accented vowel), and suggested that the resulting greater sonority rise from the consonant to the vowel may also be a clearer cue to the beginning of the accented syllable (which in English is often also the edge of the word).

The theory that some of the tautosyllabic consonants and vowel features are exaggerated in opposite directions in accented words can be thought of as a *syntagmatic* enhancement, because a vowel in an accented syllable is presumed to be exaggerated along some dimensions *in relation to the consonant that precedes it*. The theory in de Jong [11] on the other hand, which is based on Lindblom's hypoarticulate-hyperarticulate (H&H) continuum of speech production [12], is founded more directly on enhancing the *paradigmatic* contrasts between phonetic segments. For example, de Jong has shown that vowel durational differences that are relevant to the voicing distinction in postvocalic stops are magnified in 'stressed' syllables [10]. In a separate study, he also showed that the back vowel [ʊ]is more retracted in nuclear accented position thereby increasing its articulatory distance from the other vowels of the language [11].

In the present study, we investigated further the syntagmatic and paradigmatic enhancements in accented words by examining whether certain articulatory transitions between consonants and vowels are magnified (the syntagmatic theory), and by analysing whether there is an expansion of the articulatory vowel space (the paradigmatic theory) in accented words. Since accent serves to highlight a word relative to other words of an utterance, we decided to look beyond the rhythmically strong syllable that is the target of the intonational event, to see whether there are observable effects on a neighbouring unstressed syllable within the target accented words.

2. METHOD

2.1 Talkers and materials.

Four talkers of Australian English (2 female, 2 male), with no known speech or hearing deficiencies participated in the experiment. (One of the subjects, SP, is also an author of this paper; the other subjects had no training in phonetics). The talkers' accents were intermediate between Cultivated Australian, which bears the greatest resemblance to British English Received Pronunciation, and General Australian, which is spoken by the majority of the population (see [13] for further details on the acoustic characteristics of Australian English vowels). The speech materials were dialogues constructed to induce nuclear accented or deaccented renderings of nine target bisyllabic words, which provided different phonetic contexts for the accented-deaccented contrast. One of the dialogues is shown below. The underlined words are nuclear accented; any word following the nuclear accented word is deaccented:

A. This is Hector Beeber.
B. Do you have a nameplate for him? What is his name?
A. His name is Hector Beeber.
B. How shall we write the first part?
A. Well, we can have Hector Beeber. Or Doctor Beeber.

Care was taken in constructing the dialogue to ensure that the target words that differed in their accentual pattern were in an equivalent phrasal position. The other eight dialogues were identical except for the target word, which varied in the vowel nucleus of the strong syllable (tense high, lax high, or tense low – e.g. 'beeber/bibber/barber') and in its consonantal context (bilabial, alveolar, velar – e.g., 'beeber/deeder/geeger'). There were 12 repetitions of each dialogue resulting in 12 (dialogue repetitions) x 9 (target word types) = 108 dialogues (i.e., 216 accented and 216 deaccented word tokens per talker). An attending experimenter (a phonetician with considerable ear training in English prosody) verified that each dialogue token had been produced with the intended accentuation pattern and prosodic phrasing, or asked the subject to repeat the dialogue if there were any errors. The present study is restricted to an analysis of the three bilabial contexts. Australian English is non-rhotic and the strong vowel nuclei in 'beeber/bibber/barber' have qualities close to [iː ɪ ɐː] respectively [13].

2.2. Recording conditions

Subjects were recorded in a sound-treated room in the Speech Hearing and Language Research Centre, Macquarie University, using the MOVETRACK magnetometer system. Transducer coils were attached to various articulatory landmarks to record four sets of vertical and horizontal positions. The first coil was placed at a fixed 1.75 cm back from the tip of the tongue. The three remaining transducers were then placed at the midpoint of the upper and lower lips on the vermilion border and on the chin (to register jaw position). The x- and y-axis values of the receiver coils were measured relative to fixed transmitters mounted on a helmet behind and above the head. Since the helmet was in a fixed position on the head, and since the transmitters were fixed on the helmet, there was no need to correct the data further for head movements. Prior to any further analysis, all the kinematic data were rotated such that the horizontal (X) axis was parallel to the occlusal plane for each subject. The rotations to reposition the X axis in this way ranged between 16^{o} (JP) and 11^{o} (SP) in an anticlockwise direction.

2.3 Digitisation, segmentation, acoustic labelling.

The acoustic and kinematic data were digitized directly to a SUN workstation at 20 kHz and 500 Hz respectively. The ESPS/Waves+ system was used for acoustic segmentation and labelling and to compute the formant frequencies. The first four formant centre frequencies and their bandwidths were automatically tracked using the default settings (12th order LPC analysis, cosine window, 49 ms frame size, and 39 ms frame shift). The automatically tracked formants were checked for accuracy and hand corrections were made. The acoustic segmentation and labelling consisted of marking the phonetic boundaries and labels within each target word. The acoustic vowel target was also marked at the second formant, or occasionally the third formant, frequency maximum in the tense and lax high front vowels and at the F1 maximum in the open vowels. All subsequent graphical and statistical analyses, including some kinematic labeling, were carried out using the EMU system for speech database analysis [15].

2.4 Data extraction and analysis.

An estimated lip aperture trajectory was obtained by subtracting the displacements of the transducer attached to the lower lip from that attached to the upper lip. The resulting trajectory, together with the four kinematic marks that were placed on each token (at the minima associated with the two [b] segments and the lip aperture maxima associated with the strong vowel and final schwa), are shown in Fig. 1.

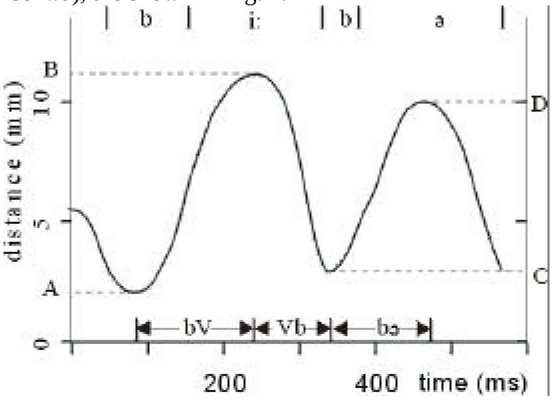

Fig. 1. A lip aperture trajectory for a 'beeber' token showing the three intervals /bV/, /Vb/ /bə/. The three absolute magnitudes that are associated with these intervals respectively are given by |B – A|, |C – B|, and |D – C|. The acoustic segments and boundaries are shown at the top of the display.

We calculated the absolute magnitude values of lip aperture change for each of the three intervals shown in Fig. 1. (These intervals will henceforth be referred to as /bV/, /Vb/, and /bə/). We also calculated the peak-velocity values from the first differenced lip aperture movement data for these same three intervals.

Tongue X and Y position values were extracted at the acoustic vowel target of [iː ɪ ɐː] in 'beeber/bibber/barber' respectively. For each talker separately, we calculated the mean tongue X and mean tongue Y values across these tongue X-Y vowel target data to give us the centroid of this two-dimensional space. We then calculated the Euclidean distance from the tongue X-Y target values to this centroid. We also extracted F1 and F2 formant frequency values at the acoustic vowel target in the same data and calculated the Euclidean distances from the acoustic vowel target to the formant centroid in an analogous way. This procedure follows that of Wright [14] in his acoustic analysis of vowels in low and high frequency words.

3. RESULTS

For both the syntagmatic and paradigmatic analyses, we applied an ANOVA separately to the data for each of the four talkers with Word-type (‘beeber’, ‘bibber’, ‘barber’) and Accent (‘accented’, ‘deaccented’) as the independent variables. For the syntagmatic analysis, the dependent variables are the six kinematic parameterisations discussed in 2.4.1 which are the magnitudes and peak velocities over the /bV/, /Vb/, and /bə/ intervals. For the paradigmatic analysis, the dependent variables are the Euclidean distances from the vowel target to the centroid in (i) the tongue X-Y space and (ii) the F1 x F2 formant space as discussed in 2.4.2. As expected, the main effect for Word-type was significant in all cases and will not be reported on further. When there were significant interactions on the Accent parameter, *post-hoc* pairwise *t*-tests (one-tailed) were performed on each word separately, using a Bonferroni alpha-adjusted *p* value.

3.1 Syntagmatic analysis.

A summary of the statistical results is shown in Table 1. For the word-initial stressed /bV/ interval, the lip aperture magnitude was significantly greater in accented words for JP (F(1,135)=72.9, p < 0.01), for MJ (F(1,134)=29.6 p<0.01), for SP (F(1,138)=97.7 p<0.01), and for TK (F(1,138)=55.9 p<0.01). There were significant interactions for JP, SP, and TK. *Post-hoc t*-tests showed that the magnitudes were significantly greater for accented words on all three word-types for JP and TK, but only on ‘beeber’ and ‘barber’ for SP. Peak-velocity of lip aperture chance over this opening movement was significantly greater in accented words for JP (F(1,135)=24.5, p<0.01), for MJ (F(1,134)=30.2, p<0.01), for SP (F(1,138)=12.0, p<0.01), and for TK (F(1,138)=88.9, p<0.01). There were significant interactions for all four subjects. *Post-hoc t*-tests showed a significantly greater peak velocity for accented words on all three word types (JP, TK), for ‘beeber’ and ‘bibber’ (MJ), and only for ‘ barber’ (SP).

For the following /Vb/ interval, the lip aperture magnitude was significantly greater in accented words for JP (F(1,135)=42.8 p<0.01), for MJ (F(1,134)=26.7 p=.000), for SP (F(1,138)=132.8 p<0.01), and for TK (F(1,138)=16.7 p<0.01). There were significant interactions for two subjects and *post-hoc t*-tests showed significantly greater lip aperture magnitudes in accented ‘beeber’ and ‘barber’ for JP, and in all three word types for SP. These /Vb/ lip aperture closing gestures have a peak velocity minimum which was significantly lower in accented words for JP (F(1,135)=38.9, p<0.01), for MJ (F(1,134)=18.3, p<0.01), for SP (F(1,138)=130.9, p<0.01), and for TK (F(1,138)=34.5, p<0.01). There were interactions for two subjects and *post-hoc t*-tests showed a significantly lower peak velocity for ‘barber’ only (JP) and for all three word types (SP).

For the opening movement into the word-final weak /bə/ syllable, again the lip aperture magnitude was significantly greater in accented words for all four subjects and there were no interactions. The details are as follows: JP: F(1,135)=6.9, p=.010; MJ: F(1,134)=31.9, p<0.01; SP: F(1,138)=81.6, p<0.01; TK: F(1,138)=12.0, p<0.01. In the /bə/ interval, the lip aperture peak velocity has a maximum value, and this was significantly greater for all four subjects in accented words and there were no interactions. The details are as follows. JP: F(1,135)=16.0 p<0.01; MJ: F(1,134)=24.9, p<0.01; SP: F(1,138)=82.7, p<0.01; TK: F(1,138)=62.3, p<0.01.

Subj.	/bV/		/Vb/		/bə/	
	mag.	p-v.	mag.	p-v.	mag.	p-v.
JP	ALL	ALL	i: ɐ:	ɐ:	ALL	ALL
MJ	ALL	i: ɪ	ALL	ALL	ALL	ALL
SP	i: ɐ:	ɐ:	ALL	ALL	ALL	ALL
TK	ALL	ALL	ALL	ALL	ALL	ALL

Table 1. A summary of the magnitude (*mag.*) and peak-velocity (*p-v.*) results over the three intervals /bV/, /Vb/, /bə/ shown separately for the four subjects. A cell marked ALL means that there were significant differences between accented and deaccented categories for all word types. A cell marked with any of [i: ɪ ɐ:] means there were significant differences between accented and deaccented words in the expected direction for ‘beeber/bibber/barber’ respectively.

3.2 Paradigmatic analysis.

A summary of the results is presented in Table 2. In the tongue X-Y space, the Euclidean distances from the vowel target position to the centroid were significantly greater in accented words for all four subjects. The details are as follows. JP: F(1,135)=14.9, p < 0.01; MJ: F(1, 134)=21.6, p < 0.01; SP: F(1,138)=146.9, p < 0.01; TK: F(1,138)=11.6, p < 0.01. There were significant interactions for three subjects and *post-hoc t*-tests showed significantly greater distances to the centroid in ‘bibber’ and ‘barber’ for MJ, in all three words for SP, and in ‘barber’ only for TK.

Subject	Tongue X-Y	F1 x F2
JP	ALL	i: ɪ
MJ	ɪ ɐ:	i: ɪ
SP	ALL	ALL
TK	ɐ:	NS

Table 2. A summary of the statistical analyses on the Euclidean distances in the tongue X-Y (*left*) and formant (*right*) spaces. A cell marked ALL means that there were significant differences between accented and deaccented categories for all word types. A cell marked with any of [i: ɪ ɐ:] means that there were significant differences for ‘beeber/bibber/barber’ respectively.

In the formant space, the Euclidean distances from the vowel target to the centroid of that space were significantly greater for three subjects and non-significant for TK. There were interactions for those three subjects that had significant main effects. *Post-hoc t*-tests showed significantly greater formant distances to the centroid in accented words on ‘beeber’ and bibber’ only (JP and MJ) and in all three word types (SP).

4. DISCUSSION

The analysis of transition magnitudes and velocities in the accented syllable confirms that there was a consistent syntagmatic effect: the difference in lip aperture value between the vowel and each of the two adjacent consonants was significantly greater in the accented syllable for all four talkers, and for many or all of the three vowel types. Because our corpus of materials included three different vowels in the target syllable, we were able to show that these effects are not limited to low vowels. Because we looked at all of the opening and closing movements in the disyllabic target words in our corpus, we also were able to see that the effect of accent on supralaryngeal articulation is not localized to the accented syllable. Both syllables in the word showed larger, faster transitions into the vowel. This is reminiscent of Maekawa's results showing an expansion of the vowel space in phrases under focus in standard Japanese [16], a language in which pitch accent placement is not constrained by rhythmic prominence or information focus in the ways that have led phoneticians to apply the term 'stress' to the intonational highlighting of words in English, Dutch, and German. However, because the unstressed syllable followed rather than preceded the target stressed syllable, we cannot tell whether the domain of the effect is the accented word as a whole, or whether it is limited to the weak sister syllable in the same stress foot as the accented strong syllable.

In addition to these syntagmatic effects, we also looked at a paradigmatic one. Because we looked at both tongue position and formant values, we were able to show that the distance to the centre of the vowel space is increased for rhythmically strong vowels in accented words relative to their deaccented counterparts in both an articulatory and an acoustic vowel space. This expansion of the vowel space suggests an enhancement of the vowel paradigmatic contrast: since accented vowel tokens are generally further away from the centre of the vowel space, it is likely that they are also more distance from, and less confusable with, each other in accented words.

In summary, our study provides clear evidence that consonants and vowels are syntagmatically enhanced relative to each other in both strong and weak syllables, and that the vowel space is adjusted in such a way that the paradigmatic contrast between contrasting vowels is also magnified. We are currently extending this study to an analysis of a greater range of vowel contrasts beyond the two long and one short vowel analysed in this study, as well as to an investigation of whether there are articulatory adjustments to schwa vowels in accented words when they occur in word-initial, as well as word-final position.

5. ACKNOWLEDGMENTS

This research was supported by an Australian Research Council large grant A79700281 (1997-1999) to the authors.

6. REFERENCES

[1] Fry, D. 1958. Experiments in the perception of stress. *Language and Speech* 1, 126-152.

[2] Campbell, N., & Beckman, M. E. 1997. Stress, prominence, and spectral tilt. In Botinis, A., Kouroupetroglou, G, and Carayannis, G. (eds.), *Proceedings of the ESCA Workshop on Intonation.*

[3] Sluijter, A. M. C., van Heuven, V. J., and Pacilly, J. J. A. 1997. Spectral balance as a cue in the perception of linguistic stress. *J. Acoust. Soc. Am*. 101, 503-513.

[4] Edwards, J., Beckman, M.E., and Fletcher, J. 1991. The articulatory kinematics of final lengthening. *J. Acoust. Soc. Am*. 89, 369-382.

[5] Harrington J., Fletcher J., Roberts C. 1995. An analysis of truncation and linear rescaling in the production of accented and unaccented vowels. *Journal of Phonetics*, 23, 305-322.

[6] Summers, W.V. 1987. Effects of stress and final-consonant voicing on vowel production: articulatory and acoustic analyses. *J. Acoust. Soc. Am*. 82, 847-863.

[7] Lindblom, B. and Sundberg, J. 1971. Acoustical consequences of lip, tongue, jaw, and larynx movement. *J. Acoust. Soc. Am*., 50, 1166-1179.

[8] Beckman, M.E., Edwards, J., and Fletcher, J. 1992. Prosodic structure and tempo in a sonority model of articulatory dynamics. In Docherty, G.J. and Ladd, D.R. (eds.), *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*. Cambridge: Cambridge University Press..

[9] Harrington, J., Fletcher, J., and Beckman, M.E. In press. Manner and place conflicts in the articulation of accent in Australian English. In Broe, M. (ed.), *Papers in Laboratory Phonology V*. Cambridge: Cambridge University Press.

[10] de Jong, K. 1991. An articulatory study of vowel duration changes in English. *Phonetica*, 48, 1-18.

[11] de Jong K. 1995. The supraglottal articulation of prominence in English: linguistic stress as localized hyperarticulation. *J. Acoust. Soc. Am.*, 97, 491-504.

[12] Lindblom, B. 1990. Explaining phonetic variation: a sketch of the H&H theory. In Hardcastle, W.J. and Marchal, A. (eds.). *Speech Production and Speech Modelling.* Dordrecht: Kluwer.

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

[14] Wright, R. In press. Factors of lexical competition in vowel articulation. *Laboratory Phonology VII*. Cambridge: Cambridge University Press.

[15] Cassidy, S. and Harrington, J. http://www.shlrc.mq.edu.au/emu/

[16] Maekawa, K. 1997. Effects of focus on duration and vowel formant frequency in Japanese. In Y. Sagisaka, N. Campbell, & N. Higuchi, (eds.). *Computing Prosody: Computational Models for Processing Spontaneous Speech*. New York: Springer.