AN ELECTROPALATOGRAPHIC, KINEMATIC, AND ACOUSTIC ANALYSIS OF SUPRALARYNGEAL CORRELATES OF WORD-LEVEL PROMINENCE CONTRASTS IN ENGLISH

Jonathan Harrington[1], Mary E. Beckman[2], Janet Fletcher[3] , Sallyanne Palethorpe[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, Victoria, Australia.

ABSTRACT

This study examines the phonetic characteristics of primary versus secondary stress on the first syllables of the surname ‘Wheateron’ and related adjective ‘Wheateresque’ in post-nuclear, deaccented position in a dialogue produced 40 times by 3 Australian English talkers. Synchronised acoustic, electromagnetometer, and electropalatographic recordings were analysed. One subject had a higher F0 in the primary stressed syllable. The other two had a longer acoustic duration for the syllable’s voiced portion, corresponding to a longer lip closing movement. One of these two also had a larger and faster lip opening movement into the vowel. Taken together, the results show that primary versus secondary lexical stress may be differentiated even when accent contrasts are neutralised, although the differences are inconsistent across talkers and small by comparison to those that have been shown to characterise the accented-unaccented contrast.

1. INTRODUCTION

Although there have been many experimental studies of stress contrasts in English, the phonetic basis of the distinction between lexically stressed and unstressed syllables remains controversial. A widely held view, which has its origins in classic experimental studies by Fry [1], Lieberman [2], and Ladefoged [3], is that stressed syllables can be defined in terms of their longer durations, their larger peak or mean amplitude values, and their higher fundamental frequencies or more extreme F0 movements. However, any such definition is problematic because the majority of the classic studies and more recent experiments based on this definition of stress do not adequately control for well-attested prominence contrasts at two other levels of the stress hierarchy. Pronounced F0 effects are almost certainly an artifact of the failure to control for a higher utterance-level contrast between accented and unaccented syllables (as in the third syllables of ‘automatic’ versus ‘automated’ produced with citation-form intonation), and any duration and amplitude differences could reflect a comparable failure to isolate ‘true’ lexical stress from a lower, foot-level contrast beween heavy and light syllables (as in the third syllables of ‘automated’ versus ‘automata’).

In the present experiment, we are concerned with establishing whether there are acoustic and articulatory differences between syllables that do not differ at either of these two other levels of stress contrast. In traditional terms, the target syllables should differ in whether they have primary or secondary lexical stress, and they should be elicited in intonational positions where the utterance-level prosodic contrast between accented and unaccented syllables is neutralised. Earlier experiments by Huss [4] showed that the lexical stress contrasts in noun-verb minimal pairs of the ‘digest’ kind are perceptually neutralised when the target words occurred in post-nuclear, deaccented position. This suggests that primary stress is just a by-product of the association of one of the heavy syllables in the word to the superordinate level at which accent contrasts are represented, as in the following metrical grid representation of the citation-form utterance:

```
accent        *                 *
heavy         *   *         *   *
              digest (noun)  digest (verb)
               |                 |
intonation     H* L- L%          H*L-L%
```

The alternative hypothesis, that the contrast between primary and secondary lexical stress is independent of the intonational contrast, may require an additional level in the above grid representation between ‘accent’ and ‘heavy’ to mark a syllable that has primary lexical stress as the prosodic head of its word even when it is produced in a deaccented context.

Under the second hypothesis, what sorts of the phonetic differences might we expect to find? One possibility is that (appearances to the contrary), the principal (tonal) markings of utterance-level prominence are not neutralised. Since a pitch accent is by definition associated to the primary stressed syllable of an accented word, there are fundamental frequency differences between accented and unaccented syllables. Could a reduced echo of the salient pitch peak of the nuclear accented syllable in the citation form utterance of the word carry over to the primary/secondary lexical stress contrast in deaccented words? Another possibility is that the phonetic properties of ‘pure’ lexical stress are similar to the non-tonal markers that many talkers use redundantly to mark the utterance-level contrast between accented and deaccented syllables. A number of studies have found that the distance through which the articulators are displaced often is greater in accented syllables and this greater displacement sometimes is accompanied by an increase in the duration of the accented syllable and/or an increase in the peak velocity of the movement. These effects can make accented vowels somewhat more peripheral in the formant plane (de Jong, [5]; Harrington, Fletcher, Beckman [6]), and in accented syllables with open vowels, they can contribute to a boost in acoustic energy relative to equivalent deaccented vowels (Harrington, Beckman, Fletcher, and Palethorpe [7]).

2. METHOD

Materials. We constructed a dialogue (Table I) to elicit stress contrasts at various levels of the prominence hierarchy. The target syllable 'Wheat' has – in traditional terminology – primary lexical stress in the surname 'Wheateron' but only secondary stress in the adjective 'Wheateresque' derived from the surname. The dialogue place the target words in each of Prenuclear accented, Nuclear accented, and post-nuclear Deaccented contexts, distinguished by the P, N, and D subscripts in Table I. Here, our interest is in the D1 and D2 contexts – i.e. in postnuclear position, where accent level contrasts are neutralised, isolating the contrast between primary and secondary lexical stress. In order to match the preceding context as closely as possible, we only analysed the deaccented words following 'Kate' and late'.

Kate Wheateron's plays are so famous that some people write Wheateresque_{N2}. You know – after the playwright Wheateron_{N1}.

That's Pete Wheateron_{D1}. Not Kate Wheateron_{D1}.

Great Wheateresque_{P1} performance, the critics said. One said mid Wheateresque_{D2}. But I'd say late Wheateresque_{D2}.

Table I: Dialogue with target syllable 'Wheat' exemplifying primary lexical stress (1) versus secondary lexical stress (2) in three intonational contexts.

Recordings. Three female talkers of Australian English each read the dialogue 40 times. Recordings were made in a sound-treated room in the Speech Hearing and Language Research Centre, Macquarie University, using the MOVETRACK magnetometer and the EPG3+ Reading electropalatograph. The simultaneously recorded acoustic and articulatory (EPG and magnetometer) data were digitized directly to a SUN workstation at 20 kHz and 500 Hz respectively. For the magnetometer signals, four sets of transducer coils were attached to the midpoint of the upper and lower lips on the vermilion border, to the chin (to register jaw position), and to the surface of the tongue dorsum about 1.75 cm behind the tongue-tip. The horizontal and vertical positions of the receiver coils were recorded relative to fixed transmitters mounted on a helmet behind and above the head and rotated prior to analysis to make the x-axis parallel to the occlusal plane. A derived signal, 'lip aperture', was calculated by subtracting the vertical movement of the lower lip coil from that of the upper lip coil, and a 'velocity' was calculated for it (and each other magnetometer trace of interest) by taking the difference between adjacent frame values and smoothing. The ESPS/Waves+ system was used for acoustic segmentation and labelling and to compute the fundamental frequency and formant frequencies. The automatically tracked formants were checked for accuracy and hand corrections were made. All subsequent analyses, including the labelling of events in the magnetometer traces and EPG contact patterns, and the calculation of lip aperture and magnetometer trace velocities, was carried out in the EMU speech database analysis system.

Acoustic measures. We marked six acoustic events in the target 'Wheat' syllable: the acoustic onset of [w] as judged from the onset of periodicity in the waveform; the acoustic vowel target of [i:], based on the time at which F2 reached a maximum value; the beginning of the closure of the following [t]; the end of that closure; the time at which F0 attained a maximum value in the voiced part of 'Wheat'; and the time at which the RMS contour (calculated from a rectangular window of 20 ms and a frame shift of 10 ms) peaked. Using these marks, we calculated the acoustic durations of the target syllable and its voiced and voiceless subparts, and we extracted the peak and average F0 values in the voiced section, as well as the peak dB-RMS and mean dB-RMS in the same interval i.e. from the acoustic onset of [w] to the acoustic onset of [t]. We also extracted the F2 value at the onset of the [w] and at its peak in [i:], and calculated the difference as an estimate of the magnitude of the acoustic effects of the tongue fronting and lip spreading gestures.

Articulatory measures. Two articulatory events were marked using the EPG contact patterns: the time of maximum closure in the [t] of the preceding word 'Kate/late'; and in the [t] of the target syllable 'wheat'.[1] We defined the articulatory duration of the syllable as the interval between these points. Eight other articulatory events were marked at local extrema for four magnetometer traces which are affected in opposite directions by the articulations of [w] and [i:]. Lip aperture is at a minimum in [w] and a maximum in [i:], and since the lower lip is coupled to the jaw, jaw-Y is at a maximum in [w] and a minimum in [i:]. The tongue-X trace is at a maximum (maximally retracted) in [w] and a minimum in [i:], whereas lip-X is at a minimum (maximally protruded) in [w] and a maximum in [i:]. (For both segments, the associated extrema for the four traces were at the same or adjacent frames.) We defined the *magnitude* of the *opening gesture* on each of these traces as the absolute difference (mm) between the values of the displacement extrema associated with [w] and [i:]. The opening gesture also has a *duration* (ms) and *peak velocity* (mm/s). Although the jaw-Y trace did not consistenly peak in

---

[1]Degree of closure was estimated using the total number of contacted electrodes in the first three rows, corresponding approximately to the dental to post-alveolar region. None of the talkers ever had a strongly articulated closure in either [t], and sometimes there was no EPG contact at all because the closure was produced as a glottal (rather than an alveolar) stop. The tokens (out of a total of 80 tokens per talker) in which the [t] of 'Kate/late' could not be labelled were: 0 for SPK1: 26 for SPK2; 1 for SPK3. Similarly for 'Wheat': 5 for SPK1: 2 for SPK2: 0 for SPK3.

the following [t], aveolar consonants characteristically are associated with high jaw positions. We therefore also used the jaw-Y trace together with the EPG contact to define a *closing gesture*. The closing gesture *magnitude* was the difference between the value of jaw-Y at its minimum in [i:] and its value at the time of the maximum closure in [t] as determined from EPG contact patterns, and the closing gesture *duration* was the interval (ms) between these two time points. Another measure of closing gesture force was the total number of electrodes contacted at the point of maximum closure in the following [t].

Statistics. We assessed differences on each of these measures between the primary-stressed target syllables in tokens of ‘Wheateron’ and the secondary-stressed syllables in tokens of ‘Wheateresque’ using a two-level ANOVA function (which is equivalent to applying a t-test). When results are reported below as significant, the criterion is $p < 0.01$ in all cases.

3. RESULTS

Fundamental frequency. SPK2 had a significantly higher peak F0 in ‘Wheateron’ (mean = 188 Hz) than ‘Wheateresque’ (mean = 179 Hz), but there were no differences in either F0 measure for either of the other two subjects.

Duration. Table II shows the articulatory duration of the target syllable (between the points of maximum alveolar contact in the preceding and the following [t] closures), and the acoustic durations of the syllable and of each of its voiced and voiceless subparts. Subjects SPK1 and SPK3 (who had no F0 differences) showed significantly longer durations in the primary stressed syllable for all (or nearly all) intervals that included the voiced portion of the target syllable. Subject SPK2 showed a significantly longer following [t] closure, consistent with one strategy that talkers use to set off accented syllables, but no other differences.

		SPK1	SPK2	SPK3
EPG	1str	258.7	217.2	208.6
-t wheat	2str	234.0	208.3	193.0
	sig	*24.7		*15.6
Acoustic	1str	121.6	97.1	94.3
[wi:t]	2str	108.9	96.0	86.8
	sig	*12.7		
	1str	95.7	78.5	67.2
[wi:]	2str	85.4	77.0	53.3
	sig	*10.3		*13.9
		94.3	79.8	59.1
[t]		86.8	73.4	64.6
			* 6.4	

Table II: Average durations (ms) of four target intervals in the primary (1str) versus secondary (2str) stress contexts, and significant differences (sig=*), for each of the three subjects.

Opening gesture effects. SPK2 had no significant effects on any of the measures of opening gesture size. Neither of the other two subjects showed a significant difference in the magnitude of the tongue-fronting gesture as evident in the tongue-Y trace, or of the lip-spreading gesture as evident in the lip-X trace, but each showed some effect on the [wi:] opening gesture as assessed via the lip-aperture and jaw-Y traces (Table III). However, the kinematic parameter affected was not consistent across the two subjects or the two articulators. SPK1 showed a significantly larger (but not longer or faster) lip opening movement and a significantly longer (but not larger) jaw lowering movement. SPK3 had no durational effect, but a significantly larger and faster jaw lowering movement and a faster (but not larger) lip opening. In keeping with the lack of any effect of lexical stress on the tongue-X fronting movement, neither subject showed a significant difference in F2 value at the [i:] target. Counter-intuitively, the F2 value at the acoustic onset of the [w] was significantly lower for SPK3 (rather than for SPK1, who showed the larger lip opening movement), so that the magnitude of the F2 rise from [w] to [i:] in her ‘Wheateron’

		LIP		**JAW**	
		1str	**2str**	**1str**	**2str**
SPK1	**mag**	9.07*	8.23	-6.43	-5.86
	dur	100.42	97.94	127.2*	109.8
	pv	170	170	-80	-90
SPK3	**mag**	7.22	6.73	-4.50*	-4.0
	dur	115.56	114.27	106.8	106.0
	pv	150*	130	-80*	-70

Table III: Magnitude (mm), duration (ms) and peak velocity (mm/s) of the lip-aperture and jaw opening gestures in ‘wheat’ of ‘wheateron’ (1str) and ‘wheateresque’ (2str). An asterisk next to the ‘1str’ values indicates a significantly greater value on these measures in the primary stressed context.

Closing gesture effects. Other than the longer acoustic closure duration for the [t] in ‘Wheateron’ (see Table II above), SPK2 showed no significant effects of lexical stress on closing gesture size. Table IV shows the measures relevant for evaluating closing gesture effects for the other two subjects. Both showed longer and larger jaw-raising movements into the [t] that closed the syllable with primary lexical stress. Also, the sum of the EPG contacts at the maximum point of closure in the [t] was significantly greater in SPK1's 'Wheateron', further indicating hyperarticulation of the closing movement.

		1str	2str
SPK1	**mag**	3.71*	2.45
	dur	65.3*	54.4
SPK3	**mag**	0.72*	0.37
	dur	30.8*	19.3

Table IV: Magnitude (mm), duration (ms) and peak velocity (mm/s) of the jaw closing gestures in 'wheat' of 'wheateron' (1str) and 'wheateresque' (2str). An asterisk next to the '1str' values indicates a significantly greater value on these measures in the primary stressed context.

RMS amplitude. There were no significant differences on either of the RMS amplitude measures for SPK2. In keeping with their larger and/or longer jaw-opening and closing gestures movements, SPK1 had a significantly greater mean amplitude and SPK3 a significantly greater peak amplitude in 'Wheateron'.

4. DISCUSSION

Summarizing the results above, we can say that one of the subjects showed a small effect of lexical stress on the primary acoustic correlate of pitch accent – fundamental frequency – whereas the other two subjects showed small effects on other acoustic and articulatory measures of the sort that have been implicated as supralaryngeal correlates of accentual prominence in other studies of that higher-level stress contrast.

Since the intonational context of the target words in our current study was the low-pitched postnuclear tail, the higher F0 peak in SPK2's 'Wheateron' tokens might be interpreted as a small echo of the peak accent that characterises nuclear position in citation form intonation. It is reminiscent of the small post-focal accent in interrogatives with early narrow focus in the Neopolitan variety of Italian (D'Imperio, [9]).

The effects of primary stress that the other two talkers showed are very similar to effects that de Jong [5] and others have interpreted as localized hyperarticulation. In de Jong's and other researchers' studies, many (but not all) talkers have been shown to produce larger, longer, and/or faster jaw-opening and closing gestures in accented relative to unaccented syllables. Some speakers also produce larger, longer, and/or faster movements of other articulators relevant to the contrast between the accented vowel and any neighboring consonant segments. Also, the lips may be more approximated in rounded segments and the tongue more retracted (or fronted) in back (or front vowels and glides). When there are such hyperarticulation effects in the articulatory measures under study, they can be associated with small increases in accented vowel duration and RMS amplitude, and with more extreme formant movements between accented vowels and neighbouring consonants, or between accented syllables and adjacent unstressed syllables.

However, even more than the supralaryngeal effects that have been demonstrated for the higher-level accent contrast, the effects of primary lexical stress in the current study are small and inconsistent across talkers. They are similar to the small effects of the 'underlying' voicing contrast demonstrated in studies of incomplete neutralisation in German and Polish (e.g. Port and Dalby [8]). Thus, whether they should be incorporated into phonological representations such as the metrical grid above, depends on how one chooses to model the relationships among utterance perception, production, and the representation of phonological categories in the mental lexicon.

5. REFERENCES

1. Fry, D. "Experiments in the perception of stress." *Language & Speech* 1: 126-152, 1958.
2. Lieberman, P., Some acoustical correlates of word stress in American English," *Journal of the Acoustical Society of America* 32: 451-454, 1960.
3. Ladefoged, P., *Three Areas of Experimental Phonetics*, Oxford University Press, Oxford, 1967.
4. Huss, V., "English word stress in post-nuclear position," *Phonetica* 35: 86-105, 1978.
5. de Jong K., "The supraglottal articulation of prominence in English: linguistic stress as localized hyperarticulation," *Journal of the Acoustical Society of America* 97: 491-504, 1995.
6. Harrington, J., Fletcher, J., and Beckman, M.E., "Manner and place conflicts in the articulation of accent in Australian English," in *Papers in Laboratory Phonology V*, edited by M. Broe, Cambridge University Press, Cambridge, in press.
7. Harrington J., Beckman M.E. , Fletcher J. and Palethorpe S., "Competing hypotheses concerning the articulation of stress in English," J*ournal of the Acoustical Society of America* 99: 2494 (A), 1996.
8. Port, R. and O'Dell, M., "Neutralization of syllable-final voicing in German," Journal of Phonetics: 13, 455-471, 1985.
9. D'Imperio, M., "Narrow focus and focal accent in theNeapolitan variety of Italian," In A. Botinis, G. Kouroipetroglou, & G. Carayiannis, eds., *Intonation: Theory, Models and Applications* (Proceedings of an ESCA Workshop, 18-20 September 1997, Athens, Greece), pp. 87-90, 1997.