Evidence for a syllable-based model of speech timing

W. N. Campbell *
Edinburgh University, Centre for Speech Technology Research
80 South Bridge, Edinburgh, EH1 1HN UK,
Telephone 031-225-8883, Fax 031-226-2730

Abstract

To test whether a measure of lengthening can be found which applies to segments uniformly within the syllable, instead of differentially with respect to segment type or position in the syllable, the durations of each segment in the SCRIBE database were converted to standard normal form, with a mean of zero and a variance of 1 for each phoneme, by subtracting the observed means and expressing the residuals in terms of their standard deviations. These normalised values were taken to represent the amount of lengthening or compression undergone by each segment relative to its elasticity as described by the distribution of observed durations of similar segments,

Campbell 1990 [1] showed that a constant can be found for each syllable that uniformly describes the lengthening or compression applied to segments in long and short syllables in terms of this elasticity. This paper explores the extent to which this uniform difference applies in the normal case, with particular reference to the lengthening undergone by vowels preceeding voiced or voiceless plosives, and shows that stress-induced lengthening may be differentiated from intonation-phrase-final lengthening and phonetically-motivated lengthening.

1 Introduction

Fant, Kruckenberg and Nord (1989 p.14 [3]) have found that listeners tend not to perceive intonation-phrase-final lengthening in the same way as they perceive stressed lengthening in Swedish.

Edwards and Beckman 1989 [2] in a study of sonority in the syllable found three different relationships between articulation and acoustic syllable duration. They showed that the increased acoustic duration resulting from intonation-phrase-final position was achieved by disproportionate lengthening of the part of the syllable after the sonority rise, with the lengthening affecting both the opening and closing mandibular gestures but with a greater effect on the latter. This is in contrast to lengthening of nuclear accents and monosyllabic words, in which increased syllable duration was achieved by a different, more symetrical sonority profile and a longer vowel period duration.

Taking the syllable as a necessary unit for describing stress patterns and temporal correlates such as the alignment of pitch accents, they argue that a complete model of timing must provide a quantitative description of the interactions between such suprasegmental levels as stress and

*This work was supported by the Information Engineering Directorate/Science and Engineering Research Council as part of the IED/SERC Large Scale Integrated Speech Technology Demonstrator Project (SERC grants D/29604, D/29628, F/10309, F/10316, F/70471) in collaboration with Marconi Speech and Information Systems and Loughborough University of Technology.

phrasing and the rise and fall of sonority within the syllable. They suggest that the rhythmic effects of various prosodic contrasts are not adequately represented in terms of millisecond values or durational ratios for acoustic segments.

Campbell 1990 [1] proposed a syllable-based measure of lengthening of segments, employing a concept of elasticity, that shows similar results when applied to stressed versus phrase-final syllables, and I shall argue here from results based on that measure that the lengthening and compression applied to segments in syllables is also of three kinds; stress-induced, phrasal, and phonetic.

Several researchers have followed House & Fairbanks 1953 [4] in reporting a lengthening of vowels before voiced plosives, If lengthening is uniform within the syllable, as has been shown for the subgroups of long and short syllables, then this differential effect that may serve to help cue the voicing distinction of the following plosive, may operate at a phonetic level separate from the prosodic levels mentioned above.

2 Elasticity

Different phonemes show different distribution densities with respect to duration. Individual segments contribute to these densities according to the compression or expansion undergone in each particular situation. To standardise the measure of any deviation from a typical or average duration it can be expressed relative to the standard deviation of all segments of the corresponding phoneme class.

Segment durations measured in a 200-sentence phonetically balanced database (for details see [1]) were converted to standard normal form, with each phoneme distribution having a zero mean and an sd and variance of 1.

$$z_{segment} = (observed\,duration_{segment} - \mu_{phoneme})/\sigma_{phoneme} \quad (1)$$

The data was divided into four subsets; segments in syllables in sentence-final position formed one group then, taking the plus-minus one sd cutoff of the segment scores averaged within each syllable as criterial, the remainder were divided according to membership of syllables in long, medial and short classes and comparisons were performed of relative compression and expansion of different segments with regard to position in the syllable and in the utterance.

3 Results

Results show that segments in sentence-final position undergo greater lengthening in the ryme than in the onset, whereas segments that are lengthened sentence-internally, for stress and rhythmic reasons, are done so uniformly within the syllable. Segments in short syllables were similarly found to be shortened uniformly, regardless of position in the syllable and of vocoid/contoid distinction.

	long syllables:			short syllables:			sentence-final syllables:		
	mean	sd	n	mean	sd	n	mean	sd	n
onset	1.56	0.93	102	-1.22	0.56	99	0.24	0.98	149
peak	1.47	1.16	187	-1.22	0.51	170	1.09	1.25	245
coda	1.03	1.08	87	-1.12	0.38	37	1.14	1.21	242
medial	1.48	0.92	63	-1.26	0.55	37	0.48	0.96	83

3.1 Final lengthening

Means and standard deviations were compared for words, syllables, and segments in sentence-final position, and for the component parts of sentence-final syllables.

By position in final syllables:

	mean	sd	n
onset	0.33	0.97	232
peak	1.09	1.25	245
coda	1.14	1.21	242

Segments in sentence-final

	words:	syllables:	segments:	vowels:	& all w-f vowels:
mean	0.5966	0.8247	1.337	1.629	-0.1409
sd	1.232	1.3157	1.4395	1.416	0.9113
n	1154	770	200	49	811

The difference in means between segments in sentence-final words and sentence-final syllables is significant ($t = 3.89$, df 1876, $p < 0.001$) as is that between segments in sentence-final syllables and in absolute sentence-final position ($t = 4.09$, df 966, $p < 0.001$). The mean for the subgroup of vowels in sentence-final position is significantly higher ($t = 4.26$, df 441, $p < 0.001$) than that of the peak subgroup as a whole in which they are included and on which they presumably therefore have an inflationary effect. That this is not simply attributable to a lengthening of vowels in open syllables can be seen from the result for the group of word-final vowels as a whole, which is significantly less than zero.

The difference between means for onset and peak segments is significant ($t = 7.08$, df 392, $p < 0.001$), but no significance was found in the difference between peak and coda segments ($t = 0.448$, df 485, ns).

3.2 Phonetic lengthening

Values of lengthening were determined for vowels followed by voiced and unvoiced plosives in the following contexts;

	unvoiced:			voiced:			t-test:		
	mean	sd	n	mean	sd	n	t	df	p
all syllables	-0.022	0.89	374	0.227	0.98	203	3.09	575	<0.01
sent-internal sylls	-0.105	0.83	345	0.125	0.84	190	3.05	533	<0.01
sent-final sylls	0.959	1.04	29	1.714	1.53	13	1.87	40	n.s.
long syllables	0.703	0.65	176	0.844	0.84	114	1.59	288	n.s.
short syllables	-0.667	0.48	198	-0.564	0.39	89	1.78	285	n.s.
across wrd boundary	-0.313	0.86	125	-0.308	0.86	112	0.29	235	n.s.

There is a significant 0.2 standard deviation difference observable in the data for all syllables before factorisation and for sentence-internal syllables in general, but the difference ceases to be significant (at less than 5% chance) when we examine the groups of long or short syllables individually. The difference appears to be large in intonation-phrase-final sylllables however, and in this case the lack of statistical significance may be due to the small number of tokens of voiced phrase-final plosives in the data.

4 Discussion

From the results for phrase-final contexts, it seems clear that lengthening increases with proximity to the boundary rather than being limited to the segment immediately preceeding it. This agrees with the articulatory data showing an overall lengthening of the final syllable with the lengthening increasing throughout. The onset segments in phrase-final syllables are significantly different in

lengthening from those in the coda, and in this respect particuarly contrast with onset segments lengthened sentence-internally.

The third form of lengthening examined here, associated with voicing status of the following plosive, appears to be in a complementary distribution with stress-induced lengthening and compression. Its effect is diminished in long or short syllables, possibly because in the former the articulation is strong and disambiguation unneccessarry, and in the latter, reduced and by implication less important to the utterance as a whole. Only in the intermediate cases does it seem to have an effect. That this effect is syllable-internal, and therefore presumably phonetically motivated, is shown by the results for open syllables where the onset of the following word is a plosive; even with the considerable coarticulation across word boundaries that takes place in fluent continuous speech, no lengthening is found.

5 Conclusion

Three forms of lengthening have been examined here; that related to stress and rhythmic constraints, phrase-final lengthening applying to words and syllables in sentence-final position, and phonetically motivated lengthening applying within the syllable and triggered by voicing status of the following plosive.

If lengthening is measured in terms of deviation from a norm, then expansion and contraction can be shown to take place uniformly within a syllable in the case of phrase-internal, stress-induced changes, but asymetrically, increasing in strength with proximity to the boundary in the case of phrase-final syllables, and sporadically, being overridden by stronger effects when cueing phonetic differences.

If the three forms of lengthening are functionally different, serving separately to cue stress, phrasing, and phonetic contrasts, then the fact that there are also articulatory differences accords well with these results.

References

[1] W. N. Campbell *Normalised Segment Durations in a Syllable Frame*, in Proc. ESCA Tutorial and Research Workshop on Speech Synthesis, Autrans France. 1990.

[2] J. Edwards and M. Beckman *Articulatory Timing and the Prosodic Interpretation of Syllable Duration* pp 1576 - 174 in Phonetica 45, 1988.

[3] G. Fant & A. Kruckenberg, *Preliminaries to the Study of Swedish Prose Reading and Reading Style*, STL-QPSR 2/1989 Stockholm, 1989.

[4] A. House & G. Fairbanks *The influence of consonant environment upon the secondary acoustical characteristics of vowels* J. Acoustic Soc. Am 25, pp 105 - 113 1953.

[5] 216zD. H. Klatt *Linguistic uses of segment duration in English* J. Acoustic Soc. Am 59, pp 1208 - 1221 1976.