

## Suprasegmental Effects Upon Segment Duration

John C. L. Ingram

University of Victoria

## 1.0 INTRODUCTION

It is well known that the duration of segments in running speech is substantially affected by a variety of suprasegmental phenomena that encode higher order syntactic and discourse features of the message.

Contrastive or emphatic stress is one way a speaker may direct the listener's attention to crucial items in the message, for such purposes as establishing a topic, to countermand a false inference he feels the listener may be inclined to draw on the basis of previous information given.

A possibly related but distinct phenomenon is the tendency for first mentioned or unfamiliar items to be spoken with deliberate stress. A speaker will momentarily retard his rate of speech and perhaps control articulation more carefully on bringing a novel lexical item into discourse. Upon subsequent mention, the speaker will return to normal rate of articulation (Coker 1973).

*Prepausal lengthening* occurs on words immediately preceding an actual or junctural pause. Actual pauses do not invariably occur at phrase, clause, or sentence boundaries to mark major syntactic constituents. Where such pauses may occur but do not, there is nevertheless often marking of the boundary by prepausal lengthening, i.e., lengthening of the syllables in the word just prior to the boundary. Prepausal lengthening is invariably accompanied by pitch inflection, to which it is probably mechanically linked (Lyberg 1979).

The constituents marked off by prepausal lengthening are variously referred to in the literature as 'phonemic clauses' (Boomer 1962), 'tone groups' (Crystal 1969) 'breath groups' (Lieberman 1967), or 'syntagma' (Kozhevnikov and Chistovich 1965). The diversity of terminology reflects diversity of conception as to the functional significance of units marked off by prepausal lengthening. Phoneticians like Lieberman have tended to regard them as natural units of production on an egressive airstream of finite capacity. Linguists tend to see the units marked by prepausal lengthening as there for the benefit of the listener, to aid in decoding syntactic structure. Kozhevnikov and Chistovich see them as the highest units of motor programming. There is possible truth in all these views.

In addition to the aforementioned systematic suprasegmental influence on segment duration, there are obviously paralinguistic features of individual speech style, fluency, and rate that may exercise an habitual or variable influence on speech production.

Word level or lexical stress pattern is another important influence on speech segment duration. The lexical stress pattern of a word is likely to be an important perceptual cue for its recognition. The term 'lexical' stress pattern may be a misnomer. Certainly, some portion of the internal stress pattern of words is predictable by rule, though this is a controversial matter.

Finally, there are well known phonological effects of segment combination, such as the lengthening of a vowel in a closed syllable before a voiced obstruent in English, the shortening of individual consonants in a consonant cluster, etc.

Given all these extrinsic influences on segment duration, it may sound unreasonable to assert that every phone has its own inherent duration. But experimental evidence clearly supports

such a notion. Temporal filtering or gating of individual speech segments can alter the phonemic identity of a sound. A labial glide may be converted into voiced bilabial stop by temporal compression of the formant transitions. Similarly, it is possible to convert [la] into [da], [sa] into [ta], etc. It is the temporal factor of abruptness of syllable onset that is being manipulated; a factor that has been found to be highly potent in the perceptual scaling of consonants (Ingram 1980). Duration is also an important perceptual cue for English vowel recognition (Klatt 1975).

In short, every phoneme (or, more accurately, small phonemic subclass) has an inherent duration which is important for distinguishing it from other targets. But in connected speech, there are a number of suprasegmental features which must be taken into account or 'normalized out' of the signal if phonemic recognition is to be achieved. From the restricted standpoint of phonemic recognition, these effects upon segment duration are simply sources of error. From the broader standpoint of extracting meanings from utterances however, they are vital independent sources of information.

The phonetic feature of segment duration would appear to suffer from a severe case of 'information overload'. Klatt (1976) has characterised the problem as a perceptual chicken and egg paradox. In order to detect the presence of suprasegmental factors affecting segment duration, the perceptual analyser would appear to need information on the phonemic identity of the segments involved. But in order to recognize the phonemic identity of a segment, the signal would appear to require normalization on the basis of suprasegmental information. Catch 22.

From the viewpoint of production, the central problem of accounting for segment duration would appear to lie in understanding the control mechanisms operating at different levels of

linguistic organization and how they mutually influence one another in running speech. The neuromuscular control systems are not of course amenable to direct investigation, but through the careful study of output characteristics, under controlled conditions, where the relevant linguistic sources of variance are systematically manipulated in the task set for the speaker, it should be possible to gain information on the separate and joint influence of the various factors affecting segment duration mentioned above.

Quantitative information on the temporal patterning of speech, and how the various factors known to affect segment duration interact, would seem to be a logical basis from which to begin to construct theoretical models of the speech production process. Interaction effects are of particular interest because they may place conditions of competing demand on the speech mechanism and its behaviour under such conditions may be particularly revealing of underlying control mechanisms.

In the following preliminary study, subjects were presented with short passages to be read aloud. The passages were constructed so as to control for the segmental composition of the spoken material while attempting to obtain systematic variation of three suprasegmental features:

1. Prepausal lengthening
2. Contrastive stress
3. Lexical stress pattern

Only performances which met criteria of prosodic adequacy were analyzed for segment duration.

## 2.0 METHOD

### 2.1 Materials

The test passages constructed for the study are given below.

The familiar noun - verb paradigm (*pérmít* - *perμίt*) was used to generate the lexical stress pattern and provide a constant segmental environment against which to evaluate the durational effects of the other two suprasegmental factors, contrastive stress and pre-pausal lengthening. The word *perμίt* was chosen from among other candidates for its relative ease of segmentation with the sonograph.

#### Test Passages

- 1) *John has lost his visa. He knows that the border officials will not listen to arguments. Nor will bribery get him across that border, or personal connections. Only the permit permits him to go. That's all that will do.*
- 2) *The house needs painting and John's wife does not want him to go fishing. He argues that his fishing licence expires next week. But she argues that the licence does not require him to go fishing. It does not oblige him to leave. The permit permits him to go. That's all it says.*
- 3) *If he refuses this person a licence there will be a noisy appeal. But to allow the licence will create an exception that is unfair to the rest who applied. Should he refuse or allow it? He permits the permit. That's what he does.*
- 4) *Our popular building inspector encourages every applicant, no matter how crazy the scheme. But by careful use of the by-laws, he can often reject their plans. He says his job is not to approve the person but the plan. He permits the permit. That's all he does.*

The strategy used for invoking contrastive stress ('accent' may be a more accurate term, but has less general currency) is aptly described by Bolinger (1961):

...two or more items are counterbalanced and a preference is indicated for some member or members of the group.

The test passages possess a relatively uniform discourse structure. The first two sentences establish a context and set up a contrast which is explicitly stated in the sentence preceding the target, invoking focus on either the noun or verb. An attempt was made to build variety and plausability into the 'stories', to distract attention from the odd and repetitious collocation of items in the target sentence.

## 2.2 Subjects

Subjects for the experiment were 6 male, native born speakers of Australian English, one of whom was the author. All were college instructors, between 25 and 45 years of age. One subject, RM, had professional broadcasting experience, a factor which turned out to be quite significant.

## 2.3 Procedure

Recordings were made in a quiet but not sound-proofed room. Subjects were instructed to:

'Read the following passage through silently and decide how it should best be read to bring out its meaning clearly. Then read the passage aloud in your natural speaking voice. Try to read the passage as you might say it in casual speech. Do not try to read it in an exaggerated or overly formal manner.'

The passages were typed on cards and presented to the subjects in partially randomized order such that no two successive passages contained the same basic target sentence. The set of passages was read through 4 times to provide more stable performance measures and permit statistical analysis of individual subject performances.

## 2.4 Analysis

Broad band sonagrams were made of the 96 (6x4x4) target sen-

tences using a Kay 6061B Sonograph with standard 300 Hz bandpass filter setting. The key words (*pérmít* - *pérmít*) were identified in the sonagrams of target sentences and segmented according to established acoustic features: noise bursts, silent periods, voicing striations, formant structure. A sample segmentation of the target sentence:

*The permit permits him to go.*

is given in Figure 2.1. Segment durations were measured by eye with a special ruler for sonograph displays calibrated in 10 msec. intervals, permitting a reliable resolution of approximately 3 - 5 msec.

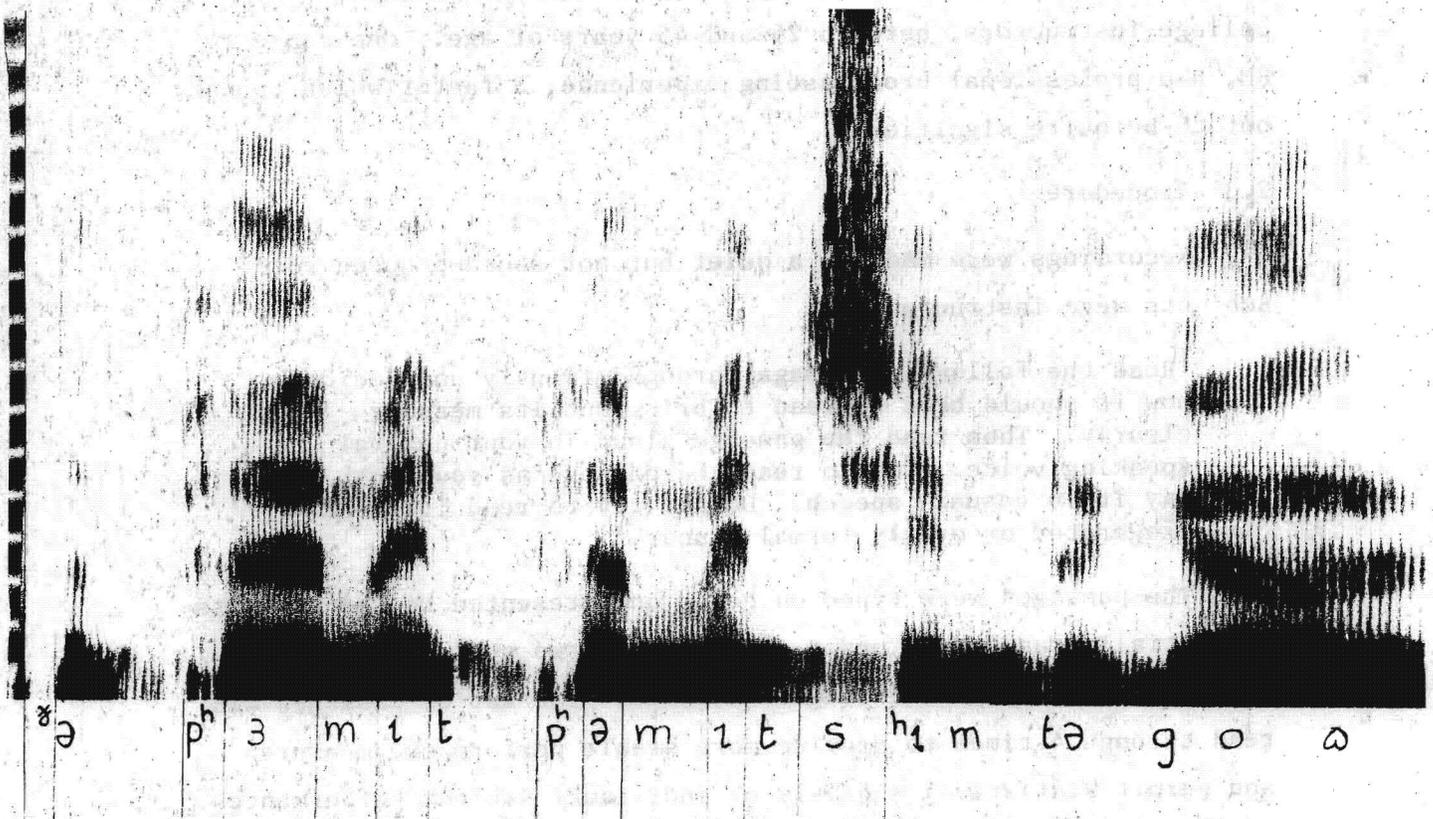


Figure 2.1 Measurement of Segment Duration. The target sentence is *The permit permits him to go.* Subject RK.

Measurements of the segment durations in the key words were stored on disk file for statistical analysis. The variable transformation facility provided by the *Statistical Package for the Social Sciences (SPSS)* permitted easy recombination of segments into syllabic or whole word units.

### 3.0 RESULTS

#### 3.1 Individual differences in subjects' performance

Cooper (1976) has described the sentence reading procedure as a '...controlled yet relatively natural speaking situation'. The evidence of this experiment suggests that the fluent reading of sentences in context, requiring the correct placement of a marked prosodic feature evoked by discourse cues, requires considerable skill that only a minority of people may possess. Both impressionistically, and on the basis of a crude objective index of fluency (Table 3.1), it was found that subjects varied substantially in their performance.

'Fluency' is a complex attribute, composed in unknown ratio of such factors as 'evenness of tempo', frequency of false starts, filled pauses, the 'normal' realization of syntactically or semantically anticipated (not necessarily required) pitch inflections, etc. As a practical expedient to the problem of assessing the fluency of the subjects' performance, a simple index based on the consistency of reading rate was used. It was reasoned that fluent readers would have smaller differences in total reading time over the same passages on successive readings than less fluent readers. Hesitations, changes in speech rate, false starts, would be expected to increase the variability as well as the overall reading time.

The standard deviation of the total reading time was calculated for each passage over the four readings. This was then divided

by the mean reading time for a given passage in order to take into account base differences between subjects in reading rate. The rank ordering of subjects on this measure agreed with the author's impressionistic ratings of the subjects' reading fluency.

Subject	Passage					X	
	1	2	3	4			
RM	05	03	02	04	3.50		Most Fluent
JI	03	04	05	03	3.75		
CL	02	06	07	07	5.50		
RK	05	07	09	06	6.75		
BM	17	08	07	07	9.75		
GT	02	03	17	20	10.50		Least Fluent

$$\text{Fluency Index} = \frac{\text{Standard Deviation Reading Time}}{\text{Mean Reading Time}} \times \frac{100}{1}$$

Table 3.1 Fluency index of subjects' reading of the four passages

The fluency index also correlated with the author's judgments as to the number of errors of contrastive stress placement (including failures to perceptibly highlight the target word) that subjects made (Table 3.2). One subject (CL) was in fact quite consistent in his misplacement of contrastive stress on passages 2 and 4.

Only two subjects consistently read the passages 'correctly' - their author (JI), and RM, who was regular professional broadcasting experience. Consequently, only the data from these two subjects has been used for the subsequent analysis of durational effects, reported below.

Subject	Passage				TOTAL
	1	2	3	4	
JI	0	0	0	0	0
RM	0	0	0	1	1
RK	0	3	0	2	5
GT	0	2	1	3	6
BM	1	2	2	1	6
CL	0	4	1	4	9

Table 3.2 Errors in placement of emphatic stress

3.2 Suprasegmental effects on word duration

The effect on word duration of the presence or absence of contrastive stress (STRESS), of the grammatical class membership of the key word (WORD CLASS), and the position of occurrence of the word in the target sentence (POSITION), was assessed by a 3-way analysis of variance (ANOVA) (2x2x2 design). The factor of 'sentence position' refers to the ordering of the noun and verb in the target sentence:

- (i) The permit      permits him to go.
  - (ii) He permits the permit.
- POS.1            POS.2

Note that the key lexical item *permit*, may function (depending on lexical stress) as a noun or a verb, may occur in the first or second position in the target sentence, and take or not take contrastive stress. However, in interpreting the results of the ANOVA in a linguistically meaningful way, it is necessary to step outside the factorial framework of the ANOVA design. In particular it is obvious that POSITION has quite different linguistic signifi-

cance in the target sentences (i) and (ii) above.

Durational effects were analysed separately and jointly for the two subjects RM and JI. Apart from a POSITION x STRESS interaction which was significant for JI, but merely a non-significant trend in the same direction in the case of RM (see Appendix A), almost identical results were obtained for both subjects, allowing for base differences in reading rate. Consequently the data from both both subjects has been combined for the following presentation of results.

Source of variation	Sum of squares	DF	mean square	F	sig. of F
Main Effects	53883.797	3	17961.266	10.394	0.000
Position	534.766	1	534.766	0.309	0.580
Word class	276.391	1	276.391	0.160	0.691
Cont. stress	53072.641	1	53072.641	30.713	0.000
2-way Interactions	56434.266	3	18811.422	10.886	0.000
Pos. Word c.	41158.266	1	41158.266	23.818	0.000
Pos. Stress	3585.016	1	3585.016	2.075	0.155
Word c. Stress	11691.016	1	11691.016	6.765	0.012
3-way Interactions	6420.000	1	6420.000	3.715	0.059
Pos. Word. Stress	6420.000	1	6420.000	3.715	0.059

Table 3.3 Contrastive Stress, Sentence Position, and Word Class, effects on Word Duration

	Position 1		Position 2	
	Stressed	Unstressed	Stressed	Unstressed
Noun	381	261	403	353
Verb	380	355	341	305

Table 3.4 Mean duration of key words in msec.

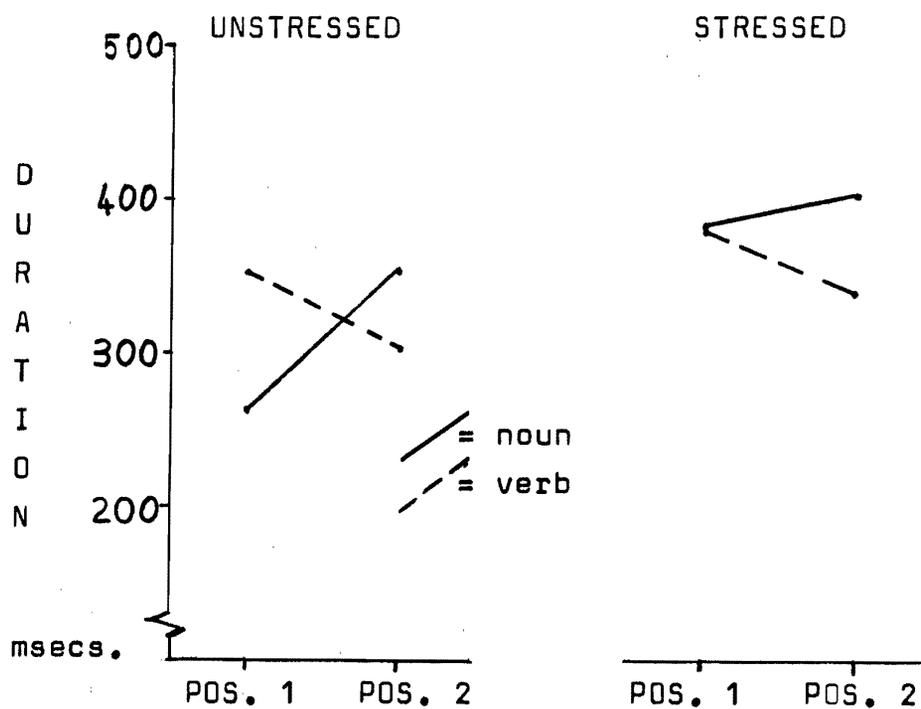
The only significant main effect was for the presence or absence of contrastive stress. There was an average 57 msec. (18%) increase in duration on the stressed word, but this effect varies substantially with Word Class and Position, as the higher order interactions on Table 3.3 indicate.

The absence of a significant main effect for word class, despite the fact that the verb contains an extra phonological segment, the subject-agreement marker, /-s/, which was counted in the total word duration, is attributable to the phonetic (and arguably phonological) fact that the noun [pɜːmt] in Australian English contains a long vowel, whereas the verb [pəmt] contains two short vowels.

Sentence Position affects the duration of the verb form in the same way, regardless of Contrastive Stress. But this appears not to be the case for the noun form, where there is a much stronger positional effect on duration for the unstressed than the stressed condition. Hence the significant 3-way interaction. Closer inspection of the data base however shows that this effect is an artifact of measurement difficulties.

The release of the final stop for the noun [pɜːmt] in subject position was only observable under contrastive stress. This problem did not arise in sentence final position where both subjects employed

INTERACTIVE EFFECTS  
OF STRESS, WORD CLASS, & SENTENCE POSITION  
ON WORD DURATION



The permit        permits him to go.  
He permits the permit.  
      POS. 1        POS. 2

Fig. 3.1 Combined data for subjects RM and JI showing the effects on word duration of the 3 ANOVA factors. See text for interpretation.

a clearly released [t] regardless of contrastive stress. Correcting for the truncation of the silent period and (absent) release of the [t], by comparing the duration of only the first 4 segments (i.e. p3m1t<sup>7</sup> vs. p3m1t ), indicates an actual increase for contrastive stress on the noun in subject position of 14% for JI and 11% for RM; figures in accord with the increase observed for sentence final position.

It still remains however to account for the opposite effect of Position on Duration in the case of the noun and the verb forms. Unlike the verb form, the noun in Position 2 becomes the object of prepausal lengthening:

*The permit permits him to go.*

*He permits the permit.*

#### PREPAUSAL LENGTHENING

We may attempt to estimate the effect of prepausal lengthening upon word duration by comparing the mean durations for the contrastively stressed nouns in Position 1 and Position 2: 22 msec, or, approximately 6% increase. This result is at variance with other studies (Klatt and Cooper, 1975; Klatt, 1975). The effects of prepausal lengthening are usually much larger. Discussion of this problem is taken up in section 4.1.

The durational effect of Position on the verb form is interesting and unanticipated. The verb with the complex complement (i) is consistently shorter than the one with the simple object (ii), regardless of contrastive stress:

(i) *The permit permits him to go.*

(ii) *He permits the permit.*

Kozhevnikov and Chistovich (1965) noted an inverse relationship between segment duration and phrasal complexity. However, in the

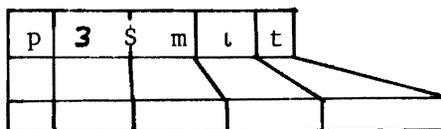
present case, the effect cannot be unambiguously attributed to syntactic factors. There is a difference in the metric structure of the two sentences. The primary stressed syllable in the verb in (i) is separated by two weak syllables from the next primary stressed syllable, but in the case of (ii) only a single weak syllable intervenes. The tendency towards isochronous units of rhythm in English may be responsible, at least in part, for the extra 43 msec observed duration of the verb in (ii)<sup>1</sup>.

#### 4.0 DISCUSSION

##### 4.1 Durational control mechanisms

Different control mechanisms most likely underlie the durational effects of contrastive stress and prepausal lengthening. These mechanisms may be revealed by an examination of the temporal patterning of articulatory events at the syllabic and segmental levels.

A reasonable model for describing prepausal lengthening would appear to be a deceleration curve, in which there is progressive lengthening of segments in the final word before a major constituent boundary, as the articulatory gestures are slowed down in preparation for the pause. This deceleration curve may be diagrammatically represented:



A potential test of the model may be provided by a comparison

---

<sup>1</sup> The author is grateful to H. J. Warkentyne for bringing this point to his attention.

of the relative duration of the first and second syllables of the noun in the target sentences:

*He permits the permit.*

*The permit permits him to go.*

However, it is not possible to make this comparison satisfactorily in the critical case of the non-contrastively stressed noun because of the (forementioned) problem of determining the point of closure for the unreleased [tʰ]. A partial comparison based on the first and truncated second syllable provides weak support for the model in the case of RM but none whatsoever in the case of JI. For RM, the initial syllable of the non-final noun occupies 57% of the total word duration, but only 48% in clause final position. However, for JI, the relative duration of the initial syllable is actually slightly greater in clause final position.

Subject:	RM		JI	
Syllable:	p 3	\$ mt _	p 3	\$ mt _
non-final	133 (57%)	99	148 (53%)	131 msec.
clause final	117 (48%)	125	176 (58%)	127 msec.

Table 4.1 Duration of initial and final syllables  
of unstressed nouns in non-final & clause-final position

The same comparison for the contrastively stressed noun is apparently non-supportive of the model also. Neither subject shows the expected decrease in the relative duration of the initial syllable in clause final position.

Subject:	RM		JI	
Syllable:	p 3	\$ mt_	p 3	\$ mt_
non-final	143 (41%)	205	187 (45%)	226 msec.
clause final	139 (38%)	221	200 (45%)	246 msec.

Table 4.2 Duration of initial and final syllables of contrastively stressed nouns in non-final and clause final position.

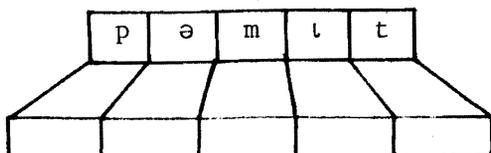
However, the contrastively stressed nouns are a dubious test case for the model of prepausal lengthening, because it is not unlikely that when contrastive stress is required in clause final position, the normal mechanism of prepausal lengthening is overridden. It may be speculatively suggested that English syntax provides the speaker with a way of avoiding this 'conflict of interest' between syntactic boundary marking and semantic highlighting (both of which involve suprasegmental lengthening), by use of the cleft construction, which is a more natural way of focusing the object:

*He permits the permit.* (non preferred)

*It's the permit he permits.* (preferred)

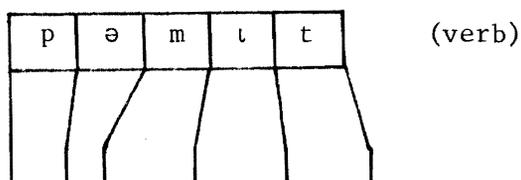
Turning to the question of a durational mechanism for contrastive stress, two competing hypotheses may be offered for highlighting a lexical item in running speech. One model, tentatively labeled 'uniform expansion', posits the whole word or morpheme undergoing contrastive stress as the domain of lengthening. Within limitations imposed by different types of phonetic segment, the

uniform expansion model predicts that all segments in the word will undergo lengthening, as diagrammatically illustrated:



Such a temporal expansion could be simply achieved by retarding the rate of articulation on the contrastively stressed word.

An alternative strategy for contrastive stress, takes as its domain the most prominent syllable in the focused word and selectively expands it, thus changing the internal temporal patterning of gestures within the word, but with less or minimal impact on overall word duration. Thus, diagrammatically:



A test of these alternative models is provided by comparisons among the contrastively stressed and unstressed verbs in the two target sentences:

- (i) *The permit permits him to go.*
- (ii) *He permits the permit.*

The relevant observations are summarized in Table 4.3.

Sentence	JI			RM				
	Stressed	Unstressed		Stressed	Unstressed			
	pə \$ mts	pə \$ mts		pə \$ mts	pə \$ mts			
(i)	78	287 (21)	74	265 (22)	38	261 (18)	70	228 (23)
(ii)	84	341 (20)	75	318 (19)	59	280 (17)	57	265 (17)

Values in brackets ( ) indicate duration of the primary stressed syllable as a percentage of total word duration.

Table 4.3 Relative syllable duration in contrastively stressed and unstressed verbs in msec.

Table 4.3 indicates no proportional increase in the duration of the primary stressed syllable of the word undergoing contrastive stress. Hence, the relatively simple 'uniform expansion' model of contrastive stress is supported by data, over one which would imply some reorganization of the temporal pattern of articulatory gestures within the word.

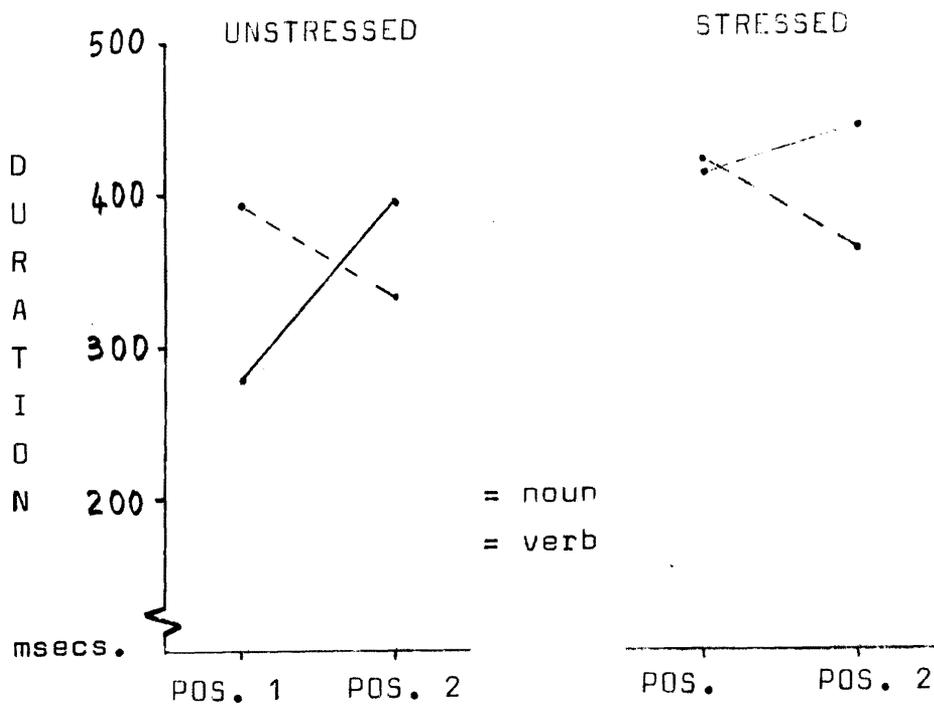
## 5.0 CONCLUSION

On the basis of the very limited data reported in this pilot study it would be quite inappropriate to take too seriously the strong inferences that the author has attempted to draw about mechanisms underlying the suprasegmental control of segment duration in running speech. On the other hand, the systematic study of segment duration by the controlled manipulation of linguis-

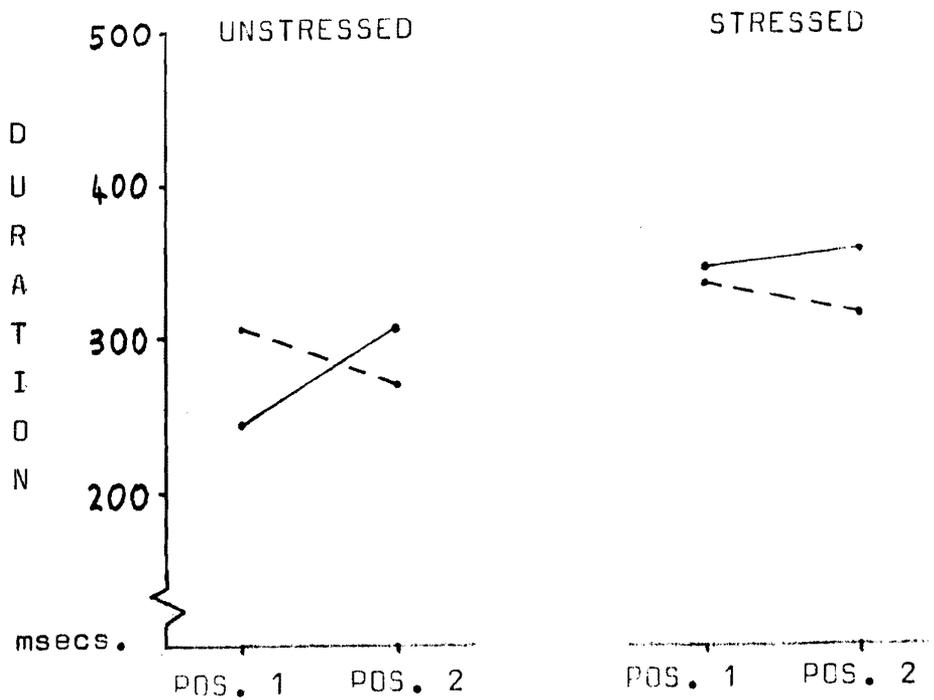
tically contrastive suprasegmental features appears to be a promising avenue of research. It does seem possible to ask and gain answers to important questions regarding the process of speech production and strategies for encoding linguistic information that are expressed, at least in part, by the temporal patterning of segments in speech. In particular, the following areas would appear to warrant further investigation based on the initial findings of this study:

1. Further clarification of the mechanism of contrastive stress and its interaction with lexical stress.
2. Further study of how the durational effects that serve the function of semantic highlighting (such as contrastive stress) interact with those that serve to carry information on (superficial) grammatical constituent structure (such as prepausal lengthening).

INTERACTIVE EFFECTS  
OF STRESS, WORD CLASS, & SENTENCE POSITION  
ON WORD DURATION  
SUBJECT: JI



SUBJECT: RM



## REFERENCES

- Bolinger, D. L. 1961. Contrastive accent and contrastive stress. *Language*. 37. 83-96.
- Boomer, D. S., and Dittman, A. T. 1962. Hesitation pauses and juncture pauses in speech. *Language and Speech*. 5. 215-220.
- Coker, C. H., Umeda, N., and Browman, C. P. 1973. Automatic synthesis from text. *IEEE Trans. Audio Electroacoust.* AU-21. 293-297.
- Cooper, W. E. 1976. Syntactic control of timing in speech production: a study of complement clauses. *Journal of Phonetics*. 4. 151-171.
- Crystal, D. 1969. *Prosodic systems and intonation in English*. Cambridge: Cambridge U.P.
- Ingram, J. C. 1980. Perceptual dimensions of phonemic recognition. In *Experimental Linguistics*. G. D. Prideaux, B. L. Derwing, and W. J. Baker, eds. Ghent, Belgium: E. Story-Scientia.
- Klatt, D. H. 1975. Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics*. 3. 129-140.
- Klatt, D. H. 1976. Linguistic uses of segmental duration in English. *Journal of Acoustical Society of America*. 59. 1208-1221.
- Klatt, D. H., and Cooper, W. E. 1975. Perception of segment duration in sentence contexts. In *Structure and process in speech perception*. A. Cohen and S. Noteboom, eds. Heidelberg: Springer Verlag.
- Kozhevnikov, N. and Chistovich, L. 1965. *Speech: articulation and language*. Translation by Clearinghouse for Federal Technical and Scientific Information. Washington, D.C.
- Lieberman, P. 1967. *Intonation, perception, and language*. Cambridge, Mass.: MIT Press.

Lyberg, B. 1979. Final lengthening - partly a consequence of restrictions on the speed of fundamental frequency change? *Journal of Phonetics*. 7. 187-196.