A Linguistic Theory of Timing

by Una M. Cunningham, B.A.

Thesis submitted to the University of Nottingham
for the degree of Doctor of Philosophy, May 1986
ACKNOWLEDGEMENTS

I wish to acknowledge the support of my supervisor Dr. A.S. Crompton, always a source of inspiration and encouragement. I would also like to thank all my informants, but particularly Mrs. E.J. Agrotis and Mr T.A. Lane.

This work was financed by a Major State Scholarship from the Department of Education in Northern Ireland.
ABSTRACT

This dissertation is an account of an investigation of timing in speech. A method of dealing with linguistic timing in the phonetic representation of utterances is developed. One of the major problems in the treatment of speech timing in the phonetic representation has been how to handle segments. The method of phonetic representation proposed here is not based on segments, but is rather a continuous representation, with timing being represented by timing contours.

This contour-based approach is shown to be better than the segment-based approach on both theoretical and practical grounds. The two approaches are compared in a series of experiments on English and Spanish speech material.
TABLE OF CONTENTS

Chapter 1 - Introduction 1

Chapter 2 - Review of previous work 5

Chapter 3 - Methodology 55

Chapter 4 - Segmental timing effects 72

Chapter 5 - Final lengthening 104

Chapter 6 - Stress 127

Chapter 7 - Global rate 135

Chapter 8 - Timing effects in combination 156

Chapter 9 - Conclusion 197

Bibliography 199
CHAPTER 1  INTRODUCTION

This dissertation is the account of an investigation of timing in speech. The ultimate goal of the investigation is to construct a general linguistic theory of timing in speech. A linguistic theory is one which accounts for an aspect of the speaker-hearer's linguistic competence. By timing is meant the relative position of speech events along a time axis.

A method of dealing with timing in the phonetic representation of utterances is developed. The phonetic representation is the exact specification of the pronunciation of an utterance, more abstract than the acoustical representation, but less abstract than the phonological representation.

One of the major problems in the treatment of speech timing in the phonetic representation has been how to handle segments (phones, syllables stress-groups etc.). These segments are necessary in the phonological representation of the utterance, but do not exist in the acoustical representation. The way around these problems is to abandon the idea of segment-based phonetic representations, which in turn calls for a different method of dealing with timing in the phonetic representation. The contour-based approach presented in this dissertation is designed to do precisely this.

Speech timing can be expressed as the rate of utterance (or rate of movement through phonetic space)
at different points in the utterance. It is then possible to characterize the difference between two very similar utterances as a relative timing contour expressing the proportional increase or decrease in rate from an arbitrary norm. This means that a timing effect, such as the well known vowel lengthening before a voiced stop consonant relative to the duration of the vowel before a voiceless stop, can be expressed as a drop in the contour showing the vowel before voiced stop relative to the vowel before the voiceless one.

It may seem that the major advantage in representing this kind of timing effect as a contour instead of as a simple difference in segmental duration is that it avoids the necessity of putting together a segment-based phonetic representation that relates sensibly to the acoustic continuum.

In the process of phonetic investigation it is necessary to be able to analyze speech material, to establish what is happening. If speech timing is the object of the investigation, measurements must be made of the durations of intervals in the acoustic representation of the utterance. This does not affect the issue of whether or not there should be segments in the PR. A segment-based PR, has the advantage of being intuitively more appealing than a non-segmented one. It is usually assumed that we hear in segments (ie that we perceive speech as a series of phones, rather than as a continuum). This is probably a learned feat however (cf Morais,
Cary, Alegria and Bertelson 1979 who found that illiterate adults could neither delete nor add a phone at the beginning of a non-word, although this was easy for people with similar backgrounds who learned to read as adults.)

Serious problems arise however when a segment-based PR is used. If the PR of an utterance is to be made, the utterance must first be segmented into phone-sized units. This is easy in the phonological representation, since that is a rather abstract level of representation. In the PR, all aspects of the pronunciation must be specified. Any parts of the utterance where one segment has coarticulatory influence on another present severe problems for the investigator. In the case of anticipatory lip-rounding for example, is the segment boundary of the segment for which lip-rounding is marked at the onset of lip-rounding or at some later point? If the boundary is at the onset of lip-rounding, are any intervening segments then part of the lip-rounded segment? At times the segment boundaries must be some kind of average position of the boundaries of the various features of the segments.

A non-segmental PR on the other hand has no need of segment boundaries. The representation is continuous, just like the acoustic representation. Where the segment-based PR handles speech timing in terms of segment durations, the continuous representation can refer to variations in the rate of utterance. Other aspects of
the PR, such as vowel quality, can also be expressed in a non-segmental way. Ainsworth (1981) for example, proposes that idealized forms of vowels be represented as points in a multi-dimensional space. The dimensions of this space correlate with the acoustic dimensions of the frequency of the first and second formants, the duration of the vowel, and possibly others such as the bandwidth of the formants and the frequencies of the higher formants.

The fact that segmentation is sometimes very difficult is perhaps not a sufficient argument in itself for adopting a non-segmental PR, but more convincing reasons will become apparent later in the dissertation. Throughout this work, the two kinds of PR will be compared and contrasted.

In the dissertation the timing-contour-based phonetic representation of speech timing is developed fully. It is then applied to four distinct areas where temporal effects have been noted, and compared in each case with the more traditional segment based approach. These areas are: the inverse temporal relationship between a vowel and a following voiced v voiceless stop consonant; the phenomena known as stress-group final lengthening; the temporal effects of stress, and the temporal effects of a change in global rate of utterance. The fifth application is when various of these effects operate together, in combination. In no case is the contour-based representation found to be less satisfactory than the segment based approach, and often it provides a very simple account of what is otherwise a very complicated effect.
CHAPTER 2 REVIEW OF PREVIOUS WORK

1. Theoretical Frameworks

There are durational regularities in speech which must be described by any adequate phonetic theory. These are of many different kinds, ranging from the way in which vowel duration is influenced by the voicing of a following stop consonant, to the way in which changes in global rate of utterance are achieved. Many writers have examined such aspects of speech timing; some have looked at the durational correlates of phonological units such as phonemes or syllables; others have looked at how speech timing is used by listeners as part of perceptual cues; yet others have tried to establish the strategies employed by the articulatory mechanism to achieve certain timing effects.

Chomsky (1965) made the distinction between competence and performance. He defined competence as the speaker-hearer's knowledge of his language and performance as the actual use of language in concrete situations. Chomsky claimed that performance would be a direct reflection of competence given an ideal speaker-hearer in a completely homogeneous speech community. Real performance data fails to be a direct reflection of competence because of noise in the system: false starts, deviations from rules, changes of plan etc. Chomsky sees the linguist's problem being to determine from the data of performance the underlying system of rules that has been mastered by the speaker-hearer and that he puts to use in actual performance. There are therefore theories of performance in speech and theories of competence. This dissertation seeks to provide a competence model of timing in speech.
The speaker-hearer's linguistic competence is then his knowledge of his language; the underlying system of rules that he puts to use in performance. That part of linguistic competence which deals with pronunciation can be thought of as being based on two levels of abstraction: a more abstract phonological representation and a less abstract phonetic representation. The latter specifies everything about the pronunciation of an utterance which is linguistically significant.

There are several possible theoretical frameworks upon which the phonological representation can be built. Nakatini, O'Connor and Aston (1981) claim that two types of temporal structure interact to produce speech rhythm. First a macrostructure which determines the temporal patterning of units the size of a syllable or larger. Second, a temporal microstructure which reflects segmental and articulatory influences which change the temporal macrostructure, making it appear less rhythmic than listeners perceive it to be. Nakatini et al suggest that this is related to rhythmic techniques in poetry. They relate the temporal macrostructures to metrical feet, such as the trochee and dactyl, and the temporal microstructures to poetic devices such as alliteration, vowel harmony, and the choice of consonants, to control rhythmicity by segmental manipulations.

Coates (1980), like Fowler (1980) and Bell-Berti and Harris (1981) (see also section on coarticulation below) rejects the phonological representation of speech as a sequence of segments subjected to allegro and lento rules which are intended to account for supra-segmental timing phenomena. Coates suggests that time should be
considered as an integral dimension of phonological representations, rather than what he describes as a dimension of performance, because: (1) vowels and semi-vowels differ critically in absolute length but not in unit (phonological) length for some speech styles; (2) one of the major cues for distinguishing stops and fricatives is duration; (3) initial voicing contrasts hinge on the amount of delay in the onset of voicing relative to the onset of other features. This seems a reasonable position to adopt.

Fowler (1980) claims that accounts of coarticulation belong to a single class of theory which she calls extrinsic timing theories of speech production, because all assume that the dimension of time is not part of the articulatory plan for an utterance, and propose that the utterance receives coherence in time only by its actualization. Fowler suggests that this is why these theories fail to predict or account for coarticulation. She proposes instead a theory of intrinsic timing which would: (1) define segments as four dimensional entities (i.e., three spatial dimensions and time); (2) rationalize the classification of segments into vowels and consonants (which Fowler sees as necessary if the model of coarticulation is to be compatible with what she terms the 'coarse-grained timing phenomena of rate and stress timing', as well as capturing the intuitions that vowels and consonants are different kinds of entities); (3) merge the plan and its executor by incorporating time into the plan for an utterance, and (4) rationalize coarticulatory effects such that a model derived from the theory would be able to generate them.
Bell-Berti and Harris (1981) agree with Fowler (1980) that the reason that previous models of production and coarticulation have failed to account for observations of real speech is because they have failed to consider time as an integral organizing factor of the speech motor plan rather than as a by-product of articulatory events. They propose a model where the units of speech are taken to be inherently dynamic gestures rather than static vocal tract configurations or invariant commands to the articulators. They assume that segment characteristics have temporal consequences, but unlike Fowler, they do not have temporal specification as part of the segment description. Bell-Berti and Harris propose a set of rules to generate coarticulatory patterns, including a claim that the articulatory period of a segment is longer than its acoustic period (i.e., movements towards and away from the articulatory goals for a segment begin and end before and after the acoustic period of the segment, but are still as much part of the specification of the segment as the goal itself) and that the articulatory period may begin at different times for different articulators.

There are other theoretical frameworks which make use of segments larger than the phone. McCarthy (1982) discusses autosegmental and metrical phonology which include larger phonological structures such as the syllable and the foot. This is motivated by reference to stress assignment, syllabification and other phenomena.

Many abstract phonological segment types have been posited as having durational correlates. The basic unit is the phone-sized segment, but any theory of timing dealing solely with such small segments would miss most of the temporal phenomena in speech. Larger segments are needed to account for the way in which temporal patterns
fit together. To this end the following abstract phonological units have been suggested as being relevant in the phonological representation of timing:

a) phone-sized segment b) syllable - defined as a vowel preceded by and/or followed by a number of consonants which may be zero ( (Cn) V (Cn) ). c) word - no satisfactory definition has been suggested for the word, but it is usually represented as the material between spaces in an orthographic representation. d) foot - defined as the material from and including the beginning of one stressed syllable up to the beginning of the next. e) stress-group - defined as a syllable bearing word (lexical) stress and the unstressed syllables which precede and follow it being more closely syntactically connected to that stressed syllable than to any other neighbouring stress. f) tone-group - this is a unit of intonation. The syntactic structure of the utterance determines where their boundaries may occur. The presence or absence of such a boundary in an utterance can only be determined by listening to the pitch patterns of the utterance. The boundaries are often accompanied by pause. There may be different kinds of tone-group depending on the presence and length of pauses in combination with the syntactic structure of the unit.

Things other than segments should also have a place in the phonological representation, such as the following: g) stress - a distinction must be made here between word (also called lexical) stress, and sentence stress. Word stress can be thought of as marking the stress a word would receive if it was uttered in isolation. In running speech word stress is potential sentence stress, such that if a word is to receive sentence stress it will be the syllable bearing word stress which receives sentence stress. Sentence stress
is usually the realization of word stress (cf Huss 1978), but contrastive emphasis may cause some syllable other than the one marked with word stress to receive sentence stress. h) Global rate of utterance - the rate at which an utterance is executed. i) Stress-timing and syllable timing should also have a place at this level of abstraction of a competence model of timing. Stress-timing is a hypothetical rhythmic arrangement whereby stressed syllables reoccur at approximately regular intervals. Syllable timing is a similar arrangement whereby syllables, regardless of whether they are stressed or not are said to occur at approximately regular intervals. It has been claimed that all the world's languages can be classed as having one or other of these kinds of rhythmic structures.

2. Discussions of temporal phenomena.

The ways in which timing effects operate in speech have been studied by many investigators. These investigations usually imply a segment-based PR, and can be categorized as accounting for the temporal correlates of the phonological units discussed above.

2.1 Phone-sized segments

To begin with the smallest and least controversial elements in the phonological representation, we shall examine past work directed at establishing the temporal correlates of the phone-sized segment.

2.1.1 Inherent duration of segments.

Klatt (1973) makes a study of the interaction between two timing rules: that stressed vowels are shorter
before voiceless Cs than voiced and that stressed vowels are shorter before an unstressed syllable in a bisyllabic word than they are in a monosyllabic word. Klatt interprets his results as showing that vowels become strongly incompressible beyond a certain amount of shortening. He proposes that every vowel has an inherent duration (generally its longest duration) and a minimum duration (which Klatt tentatively assumes to be around 45% of the inherent duration).

Lehiste (1970) also discusses the inherent duration of segments. She claims that high vowels are inherently shorter than low vowels due to the greater extent of articulatory movements in low vowel production.

Klatt (1976) claims that duration systematically varies (and is picked up as a perceptual cue) in the distinctions between inherently long versus short vowels; voiced versus voiceless fricatives; phrase-final versus non-final syllables; voiced versus voiceless post-vocalic consonants, the durational difference being in the preceding vowel in the phrase-final positions; stressed versus unstressed or reduced vowels; the presence or absence of emphasis.

Mitleb (1984b) looks at differences in vowel length in Arabic. He claims that tense vowels are longer than lax vowels since the approximate configuration for tense vowels requires more effort than that for lax vowels. Also the fact that low vowels tend to be longer than high ones is attributed to the degree of jaw lowering required for the configuration of low vowels. Mitleb addresses the question of how long and short features get converted into actual time intervals. The height effect on duration was found to operate in both English
and Arabic, while the tense-lax distinction did not work in the same way for both languages.

Different researchers mean different things when they refer to the inherent duration of segments. Klatt (1973) calls the longest duration that a segment can have its inherent duration, while Lehiste (1970) and Mitleb (1984b) refer to high vowels being inherently shorter than low vowels, as a direct consequence of the way the articulatory process works.

2.1.2 Influence of consonant environment on segment duration.

House and Fairbanks (1953) systematically varied the consonant environment of vowels by forming nonsense syllables. (For a discussion of the validity of experimental results based on nonsense material, see Liberman and Streeter (1978) and chapter 3 on methodology.) They measured the duration, fundamental frequency and the relative intensity of the vowels, and found that these things varied systematically in response to changes in the consonant environment, the most powerful influence being the presence or absence of vocal fold vibration, followed by the manner and place of articulation, in that order.

Lisker (1957) made a similar comparison of voiced and voiceless pairs of intervocalic stops. He examined the /p,b/ contrast in a limited experiment. He found the following acoustic differences to be consistent:

(1) The first, second and third formant transitions following closure begin at lower frequencies, and move less abruptly for /b/ than for /p/.
(2) The closure duration is greater for /p/ than for /b/.

(3) The vowels preceding /p/ tend to be shorter in duration than those preceding /b/. This last difference was less consistent than the others.

Zimmerman and Sapon (1958) also discuss the linguistic implications of the influence of a following consonant on the duration of a tonic vowel in English. They approach the problem cross-linguistically, using English and Spanish material based on word pair lists contrasting voice. They claim to have found that the languages are qualitatively similar but quantitatively different when a vowel precedes a consonant. The durational difference between vowels preceding voiced v voiceless consonants was found to be smaller in Spanish than in English. Since, however, the post-vocalic voiced stops have fricative allophones in Spanish, it is difficult to see what conclusions can be drawn from this.

Peterson and Lehiste (1960) investigated the influence of preceding and following consonants on the duration of stressed vowels and diphthongs in American English. They analysed a set of 1263 CNC words in identical frames, and measured the influences of various classes of consonants on the duration of the nucleus. They found that the initial consonant does not affect the nucleus, but that the final one does. In general they claim that the syllable nucleus is shorter when followed by a voiceless C, and longer when followed by a voiced C, with a ratio of approximately 2:3. Plosives were found to be preceded by the shortest nuclei, while nasals had approximately the same influence as voiced plosives.
Syllable nuclei were longest of all before voiced fricatives.

House (1961) measured the average durations of twelve vowels of American English as they occurred in bisyllabic nonsense words. The consonant environments were symmetrical, and consisted of the voiced and voiceless versions of three stop, one affricate and three fricative consonant articulations. Various vowel lengthenings were found and House separates these into what he calls primary lengthening of vowels (that found in tense vowels and in vowels before voiced Cs) which he claims to be part of the phonology, and secondary lengthening of vowels (found in open vowels and in vowels before fricatives), which he claims to be a function of the articulatory process. It is this secondary lengthening which may be said to belong to the production process. The primary lengthening is part of competence.

Chen (1970) also examines the effect of following consonants on vowel length, as a function of +/− voicing. He takes data from French, Russian, Korean and English speech material, and finds that in all cases, a vowel is invariably longer before a voiced consonant. He also finds that the occlusion time before the release of voiceless stops is longer than for voiced stops. Chen concludes that the variation of vowel duration as a function of voicing in the following consonant is a language universal (cf however Mitleb 1984a below). He suggests that this might be explained in articulatory terms by the rate of closure transition. Since voiced stops are made with closed glottis, there is relatively low intra-oral pressure, since pressure builds up only in the mouth cavity, while for voiceless stops, pressure
is also built up in the lungs. If this is so, Chen supports the view that it is the anticipation of the difference in muscular effort in the closed position for Cs that causes the transition from vowel to voiceless stop to be faster than the transition from vowel to voiced stop, which in turn results in different vowel lengths.

Mitleb (1984a) challenges the widely held assumption that the non-discrete physical properties of speech can be derived from the discrete elements of what he calls phonetic transcription by universal rules of performance. As an example of this, Mitleb claims that the assumption that vowels are universally longer before voiced consonants than before voiceless ones is not justified. He presents evidence based on a spectrographic study of Arabic minimal pairs, that there is no such effect in Arabic. Neither could the effect be found when Arabs pronounced English words. Mitleb proposes that non-segmental differences between the temporal structures of languages must be accounted for in the linguistic analysis of each language, and must therefore be learned by second language learners. This is equivalent to saying that they are part of the linguistic competence.

Lehiste (1970 pp6-53) discusses the segmental conditioning of vowel duration. She examines the influence of place of consonant articulation on the lengthening of vowels before voiced Cs, and found that short vowels were: longest before /t/, shorter before /k/ and shortest before /p/; longest before /g/, shorter before /d/ and shortest before /b/; longest before /ʃ/, shorter before /s/ and shortest before /f/; longest before /m/,
shorter before /ŋ/ and shortest before /n/; longer before /z/ than before /v/; and that long vowels were: longest before /t/, shorter before /k/ and shortest before /p/; longest before /d/, shorter before /g/ and shortest before /b/; longest before /j/, shorter before /s/ and shortest before /f/; longest before /ʒ/, shorter before /z/ and shortest before /v/; longest before /ŋ/, shorter before /n/ and shortest before /m/.

Fischer-Jörgensen (1979) examines data on the temporal relations in CV syllables with Danish stops, and discusses whether the inverse relationship which exists between vowel and stop duration when syllables contrasting the voicing feature are compared can be attributed to compensation phenomena or to specific psychological conditions.

Port, Al-Ani and Maeda (1980) performed experiments on Japanese and Arabic speech to explore the extent to which vowels adjacent to longer and shorter apical consonants would compensate for the consonant durations. They found a very large difference between languages. Arabic showed little interdependence between the durations of adjacent segments, while in Japanese the duration of two-syllable (in this case two-mora) words remained almost constant. Port et al claim this as corroboration for the notion that Japanese mora have constant duration, but that it undermines the definition of the mora as a simple CV unit, since the vowels on both sides of the consonant were found to vary inversely with the consonant duration, and the consonant constriction duration was found to vary inversely with what Port et al claim to be the inherent duration of the vowel.
Fitch (1981) discusses the relationship between the temporal information for speaking rate and that for intervocalic stop consonant voicing. It has been established that the temporal difference between voiced and voiceless consonants includes both the duration of the acoustic segment corresponding to vocal tract closure and the duration of the vowel preceding the closure. However, Fitch points out that similar temporal differences are also affected by changes in speaking rate. A voicing change from /b/ to /p/ lengthens the closure and shortens the vowel, but a rate change from fast to slow lengthens both, the difference between the two phenomena being that of an inverse relationship and a direct one.

Kohler (1980) claims to have found evidence in EMG data from German that the articulation of lenis plosives is controlled by a slower raising and faster lowering of the velum than is the case with fortis plosives, within a fairly constant time frame for the complete utterance.

Santerre and Suen (1981) examined a corpus of 720 samples of minimal word pairs containing stop consonants, and measured vowel duration, formants and formant transitions, silent intervals and voice onset time. They claim that one feature cannot account for the distinction between voiced and voiceless stops. They found the variation of vowel formants at the transition to consonant is more obvious for words with /b,d,g/ than /p,t,k/. This was especially true of the first formant.

Fowler (1981) suggests that what she claims to be the compensatory shortening of stressed vowels in the contexts of consonants and unstressed vowels may be related
to the coarticulatory influence of a stressed vowel on consonants and unstressed vowels. The two effects were found to pattern very similarly.

To sum up then, the following temporal segmental effects seem to be well established.

1. The duration of a vowel changes according to the voicing of the following consonant: House and Fairbanks (1953); Zimmerman and Sapon (1958); Peterson and Lehiste (1960); House (1961); Chen (1970); Mitleb (1984a).

2. The duration of a vowel changes according to the place of articulation of the following consonant: House and Fairbanks (1953); Lehiste (1970).

3. The duration of a vowel changes according to the manner of articulation of the following consonant: House and Fairbanks (1953); Peterson and Lehiste (1960); House (1961).

4. The vowel formant transitions are different before a voiced consonant than before a voiceless one: Lisker (1957); Santerre and Suen (1981).

5. There is an inverse durational relationship between vowels and post-vocalic voiced versus voiceless stops which may or may not be attributable to a temporal compensation mechanism: Fischer-Jörgensen (1979); Port, Al-Ani and Maeda (1980); Pich (1981).

Perceptual studies

The following writers have examined the way timing is used by the listener as part of perceptual cueing. Many
have worked on the way in which the voicing of post-vocalic stop consonants is perceived. As early as 1959, Wang claimed the following to be cues for the voicing of post-vocalic consonant voicing:

1. the duration of the preceding vowels (found to have greater duration before voiced Cs);

2. the formant transitions (found to be more abrupt for voiceless stops);

3. the duration of the voicing that may follow the formant transitions.
And if the plosive is released, Wang suggests that the following may also be used as cues:

4. the duration of the closure;

5. the intensity of the release;

6. the spectral properties of the release.

Slis and Cohen (1969), working on Dutch, also found the V-C-V formant transitions to be a useful cue in the perception of the voicing of plosives. They too found rapid transitions to be associated with voiceless plosives, and slow transitions to be associated with voiced plosives, for both the transition from V-C and the transition from C-V. Also, they found that a small range of frequency shift in the first formant causes voiceless judgements, while a wide range leads to voiced plosive judgements.

Raphael (1972), in a study of the use of duration as
a cue for post-vocalic voicing, found a range of V durations above which listeners will predict a following voiced stop, and below which they will predict a following voiceless stop. However, within this range, listener's perceptions were found to vary according to the spectral configurations of the vowel: a falling first formant cues a voiced stop judgement, while a level first formant cues a voiceless stop judgement.

Waajskop and Sweerts (1973) examined French VC syllables for cues to post-vocalic consonant voicing. When the release of the stop was synthetically removed, there were fewer correct stop identifications. The contribution of cues such as the presence or absence of laryngeal vibrations and the difference between first formant transitions were found to be less important for correct perception than the release and the duration of the closure.

Parker (1974) also suggests that the manner of termination of the pre-consonantal vowels in VC sequences is a primary perceptual cue for the voicing of the consonant. He claims to have found that a gradually terminated vowel signals a following voiced stop, and an abruptly terminated vowel signals a voiceless stop. Such a solution, Parker suggests, would account for both the variability of vocal fold vibration during stop closure and vowel lengthening before phonologically voiced stops in English.

O'Kane (1978) investigates the hypothetical perceptual cue for the voicing of post-vocalic stops in English, discussed by Parker (1974). He claims to have tested Parker's hypothesis and found no evidence that
the manner or termination of the preceding vowel is the most important cue to the status of such consonants. O'Kane maintains that durational differences provide the most significant cues. When informants were presented with synthetic syllables constructed by replacing the vowels in cob, cod, coq with those from cop, cot, cock, they no longer gave voiced responses. When the syllable was progressively truncated, and the /k/ from cop and the vowel from cob were combined, there was a 90-100% voiced judgement. Contrary to O'Kanes explanation, this suggests that the manner of termination of the pre-consonantal vowel is an important cue to voicing, especially since that when the vowel length was the same, informants still gave correct judgements (that is, if the vowel originally preceded a voiced stop, it was judged as such and similarly for pre-voiced vowels, even when the vowel duration was unchanged).

Walsh and Parker (1981) dispute O'Kane's interpretation of his results. Since it was found that when vowel duration was constant, those vowels originally preceding voiced stops elicited significantly more voiced responses than those before voiceless stops, and that a truncation of the end of the originally pre-voiced vowel resulted in a rapid drop in the percentage of voiced judgements, Walsh and Parker argue that the manner of termination of the vowel must be a major cue in the voicing of post-vocalic stops.

This argument is followed up in Walsh and Parker (1983) with a corroboration of Raphael's (1972) work on synthetic speech, using what is called real nonsense speech. Since vowel termination is taken to be the operative cue during the intermediate range of vowel
durations, then the only possible function of vowel length as a cue is to either reinforce or override the transition cue at extremely long or short vowel durations. Walsh and Parker claim that vowel length is at best a redundant cue, or at worst misleading.

Port (1978) claims that the perceived voicing of the medial stop in word pairs such as rapid/rabid can be controlled by changing the duration of the stop closure. He found that when the tempo of the carrier sentence for the test words was increased, the boundary between rapid and rabid was shortened along a continuum of silent closure durations, but that this shortening was proportionally much less than the percent decrease in sentence duration. Port found that tempo within the test word had a stronger effect on perception of the boundary between rapid and rabid than did tempo in the carrier, and that for a given tempo of carrier sentence, the ratio of the boundary value of closure duration to the duration or the rap/rab syllable was nearly constant.

Wolf (1978) examined voicing cues in final stop syllables. She found that when stimuli which had been truncated at various points before the end of the VC syllable were presented to listeners, the formant transitions, closure, release bursts and vowel duration were all important. She also found differences in the first 50 ms of the vowel which appeared to affect the pattern of voice judgements.

Fitch (1980) found that the perception of intervocalic /b/ can be changed to /p/ by lengthening the silent closed interval. The slower the speaking rate, the longer the closure interval required, roughly in propor-
tion to the duration of the preceding syllable. It was found, using synthetic speech, that longer syllables with proportionately short steady state sections needed less silence than shorter syllables with proportionately long steady state sections. Thus, the perceptual voicing boundary is sensitive to the dynamic structure of the preceding syllable and not simply to its duration.

Wardrip-Fruin (1982) discusses the status of preceding vowel duration as a cue to voicing in final stop consonants. She suggests that vowel duration is neither necessary nor adequate as a cue in natural speech; that voicing during closure may be necessary to disambiguate final voiced stops. Wardrip-Fruin found syllable length to be a more significant cue to the voicing feature than was vowel duration, but this was not a necessary nor adequate cue either. Voicing during closure was found to determine the voicing judgement over the full range of syllable durations.

It is unsurprising that the temporal effects found to be used for the perception of post-vocalic voicing are the same. It seems to be beyond doubt that the vowel before a voiced stop is longer than the same vowel before a voiceless stop (Wang 1959, Raphael 1972, Wolf 1978, Walsh and Parker 1983); that the voiced stop has shorter closure than the voiceless stop (Wang 1959, Port 1978, Wolf 1978 and Pitch 1984); that the formant transitions following stop closure are more abrupt for voiceless stops (Wang 1959, Slis and Cohen 1969, Raphael 1972, Parker 1974, Wolf 1978, Walsh and Parker 1981 and 1983); that the duration of voicing after the formant transitions to a voiced stop is greater than that to a voiceless stop (Wang 1959 and Wardrip-Fruin 1982).
2.1.3 Influence of position in utterance on segment duration.

Lindblom and Rapp (1973) looked at the factors which influence segment duration in Swedish. They propose an equation by which the duration of a vowel can be predicted. Factors accounted for by the formula are the number of syllables following and preceding the syllable in which the vowel occurs; and the degree to which anticipatory and backward coarticulation take place.

Oller (1973) also considered the influence of position in utterance on segment duration. He used nonsense syllables of the type /bab/, /'babab/ etc to measure the durations of speech segments. Final segments were found to be longer than non-final segments, such that the vowel of the final syllable was incremented by about 100ms, while final consonants were about 20ms longer than non-final ones. Oller also discusses other experiments which found that final syllable lengthening occurs in word and phrase as well as utterance final position, and in all types of syllables. He suggests that lengthening in certain utterance positions is a learned aspect of language which serves to cue listeners to the presence of word-, phrase- or sentence- boundaries.

Umeda (1975) investigated vowel durations in continuous text. She aimed to supply values for the durations of vowels in different contexts. She found the average durations of vowels in continuous text when in pre-pausal, monosyllabic and polysyllabic words, and of all vowels when followed by different consonant types. She
found the influence of the different consonants to be greatest in pre-pausal syllables.

Klatt (1975) seeks to demonstrate that it is the syntactic structure of an utterance that determines the duration of vowels. He measured segmental duration for each kind of segment and for stressed and unstressed environments from spectrograms. He established the median values for each segment and found that significantly longer durations occurred at the boundaries of major syntactic units, including noun and verb phrases. Vowels in phrase-final syllables were found to be an average of 40ms longer than those in other positions. Klatt claims that the effect of the post-vocalic consonant on vowel duration is small except in phrase-final syllables. Word-final syllables were also found to be longer than others.

Both Oller (1973) and Klatt (1975) found the boundaries of major syntactic units to have longer segments. Umeda (1975) found longer segments before pauses. This is another way to look at the same phenomenon, since pauses often occur at tone-group boundaries, which are by definition stress-group boundaries, usually occurring at the boundaries of syntactic units. More about various kinds of final lengthening can be found under the sections devoted to the larger phonological units.

2.1.4 Segmental timing factors in combination

Port (1979) aims to provide rules to determine vowel and consonant durations. He systematically varies vowel tension, number of syllables and consonant voicing and proposes a timing model where the following things are
specified: (1) a total duration for the VC interval as a function of the number of syllables in the word; (2) the relative duration of the vocalic and consonantal portions of the VC unit, depending on both consonant voicing and the number of syllables and (3) the mean underlying vowel duration depending on the tensity of the vowel. Port suggests that the VC sequence is a temporal macro-unit which is subsequently internally structured by more local rules.

Port (1981) looks at ways in which linguistic timing factors combine with each other to produce English speech, and specifically tests and explores aspects of Klatt's (1976) timing model. He claims that although it is well-known that the duration of phonetic intervals in speech change according to changes in the phonological features that distinguish words, no good theory describing the role of linguistic timing has been developed. Port argues that not only may several different phonological features affect the duration of a single phonetic interval, such as vowel duration, but many other factors, such as the length of the word may also influence its duration.

2.1.5 Coarticulation

The concept of coarticulation is an example of the kinds of theories which attempt to account for the difference between timing in competence and performance. Coarticulation is the name given to the position arising when the articulatory correlates of the linguistic units (usually phonemes) overlap in time. It is not usually possible to identify a single point in time where it is clear that one phoneme stops and another begins. This is
clearly shown by Kent and Minifie (1977). They claim that speech is usually considered to be a sequence of discrete units, and that the problems that arise when articulatory events are to have segment boundaries imposed upon them are caused by the fact that the articulatory movements overlap each other in a complex way. This complication gave rise to the concept of coarticulation which the authors describe as a concept of speech behaviour which implies discrete and invariant units serving as input to the speech production system, as well as an eventual obscuration of the boundaries at the articulatory or acoustic levels. Kent and Minifie discuss serial ordering, which they express as taking place according to notional time, while the coordination of submovements in an articulatory sequence must be observed according to physical or clock time.

A similar study was carried out by Macneilage and Declerk (1969). They investigated the way in which discrete phoneme commands are modified according to context to produce speech. They had hoped to find invariant characteristics corresponding to phonemes in the acoustic continuum, or to be able to infer them from 'earlier' stages of the production process. They studied 36 CVC monosyllables, and found that in every case, some aspect of the motor control of a later part of the syllable was influenced by the identity of a previous part. Except in a few cases the reverse was also true, earlier segments being influenced by later ones, and in these cases the influence was found to be greater.

The concept of segment is contained in the concept of coarticulation. The acoustic record is a continuum, and shows no segments. Hammarberg (1976) claims that there
are no invariant correlates of a segment in an acoustic or articulatory record of an utterance, and that it takes a human mind to perceive a segment. He adds however that without both mentalistic phonological concepts, segments and the canonical type (phoneme), there would be no way of dealing with psychological phonetic events. Hammarberg claims that methodological and metaphysical presuppositions rather than theory-free observation are necessary for the categorization and identification of data.

Bell-Berti and Harris (1982) discuss anticipatory coarticulation, specifically of lip-rounding. They challenge theories which attribute anticipatory coarticulation to phones for which a feature is unspecified coming before a phone for which the feature is specified and consequently receiving the feature. They provide EMG evidence that the beginning of activity associated with lip-rounding is independent of measures of the acoustic duration of adjacent lingual consonants. They suggest that the rounding gesture simply co-occurs during predictable intervals with portions of preceding and following lingual consonant articulation. Bell-Berti and Harris define the central problem in the relationship between speech production and perception as being the disparity which exists between the perceptual representation of speech as a series of discrete events, while the acoustic representation is a continuously varying stream, without obvious segment markers.

With a continuous phonetic representation such as is proposed in the next chapter, the complex coarticulatory phenomena can be represented in a straightforward way. This is compatible with Fowler (1980)'s position regar-
ding the incorporation of time into the plan for the utterance, and also regarding the specification of segments as four dimensional entities. (This would happen between the phonological representation and the phonetic representation.) This is also in accordance with Hammarberg's (1976) claim that there are no invariant correlates of segments in the acoustic or articulatory record of an utterance.

2.2 Syllable

While many have assumed that the syllable has a vital place in any description of linguistic timing, (Gonzalez (1970) claims the syllable to be the acoustic correlate of rhythm), few have looked at how the syllable is made up. Lisker (1978) discusses the possible status of the syllable as the basic unit of articulatory organization. He claims that the durational relationship between vowels and following stops is evidence for this, but ignores the fact that this relationship also holds over syllable boundaries.

Lehiste (1971) discusses the internal organization of mono- and bisyllabic words in English. She looked for correlations in CVC structures between C and V, and between V and C, both with and without intervening morpheme boundaries. The VC structure was found to have closer links in English.

Walsh (1984) presents evidence that consonant cluster abbreviation in English is due in part to a tendency to
produce CVC syllables with a fixed ratio between the CV and the final portions of the syllables, such that CV/C=k.

None of these investigations provide any evidence that the syllable as a unit has any durational correlates.

2.3 Stress

As discussed above, the definition of stress is less than straightforward. However it is generally accepted that duration is one of the acoustic correlates of stress, although Huss (1978) found that word stress that was not realized as sentence stress had no durational correlate. As early as 1955, Fry showed that duration dominates intensity as a cue to stress.

Gonzalez (1970) discusses the acoustic correlates of accent (taken here to mean realized word stress), and hypothesises that accent prolongs syllable duration. He suggests that all syllables should have a duration equal to some empirically discovered duration unit (mora) and that it should be possible to add a predictable fraction of a mora or even a whole mora for accented syllables.

Lea (1977) found that stressed syllables have longer vowels (syllabic nuclei) than unstressed syllables, and also that the intrinsic identity of the vowel has a significant effect on its duration: low vowels have longer duration than high vowels (cf section on inherent durations of segments above), such that unstressed /a/ may sometimes be longer than stressed /i/. Lea claims
that the duration of vowels is also affected by their position in the utterance (cf. section on this topic above), such that a phrase-final unstressed vowel may be longer than the same vowel in a stressed phrase-medial position.

Klatt (1975) claims that consonant durations are a function of the stress assigned to their syllable, such that pre-stressed consonants are longer than pre-unstressed and post-stressed consonants except in positions where pre-pausal lengthening is to be expected.

It is clear then that a syllable bearing sentence stress is longer than one which does not have sentence stress.

2.4 Rhythm - Stress-timing and syllable-timing.

Most phoneticians have assumed that there is rhythm of one kind or another in speech production, but it is chiefly psychologists who have addressed this question. Martin (1972) defines rhythm as the relative timing between adjacent and non-adjacent elements in a behavioural sequence, as opposed to the temporal ordering of elements. Martin claims that natural constraints on movement entail that running speech must be rhythmic, according to his definition, in such a way that accented syllables are generally separated from each other, and will fall at roughly equal intervals. Martin's claim carries the implication that all speech must necessarily be syllable-timed, which is in direct conflict with traditional linguistic opinion.

Allen (1975) agrees with Martin (1972) that the
calls the chest pulses and the stress pulses can be combined. He claims that every language in the world is spoken with one or other of these speech rhythms. In syllable-timed languages (examples given as French, Telugu and Yoroba) Abercrombie claims that the periodic recurrence of movement is supplied by the syllable-producing process (the chest pulses), such that the syllables recur at equal intervals of time - they are isochronous. In stress-timed languages (such as English, Russian and Arabic), it is the stress-producing process, the stress pulses, which provides the recurrence of movement, so that these stress pulses, and hence the stressed syllables occur at regular intervals of time - they are isochronous.

Lightfoot (1970) discusses Pike's (1945) definition of isochronous rhythmic units. Pike suggests that the temporal structure of the utterance is 'a series of simple rhythmic units, each bounded by pauses and containing only one primary contour with one prominent stress', and that there should be approximately isochronous intervals from the beginning of one prominent syllable to the next. Additionally, Pike claims that accents as strong as or stronger than the stresses at the beginning points of the primary contours cannot occur between the isochronous stresses. This seems to be a claim for isochrony of tone-group nuclear stress only. Lightfoot points out that Pike's account does not allow for syncopation where a syllable is strongly accented but is not an isochronous accent. Pike observes that the syllables of larger rhythmic units are hurried to preserve the isochronous spacing of stresses when they are juxtaposed to rhythmic units with few syllables. Lightfoot proposes that in spite of a tenden-
cy in English for syllables bearing sentence stress to be separated by 1, 2 or 3 less accented ones, it is only when the phonetic and grammatical structures and number of syllables are similar that isochrony can be measured in the acoustic record.

Ladefoged (1975) considers the theory that peaks of syllabic stress coincide with peaks of sonority, such that the final /m/ of prism might have more or less sonority than the preceding /z/. Ladefoged defines syllable-timed languages (e.g. French) as having all syllables rather than only stressed syllables tending to recur at regular intervals of time. He defines stress-timed languages (e.g. English and other Germanic languages) as having a tendency (and no more than that) for stresses to recur at regular intervals. Ladefoged claims that Japanese may be analyzed in terms of a unit of timing called the mora, such that each mora takes the same time to say. The most common form for a mora to have is a consonant followed by a vowel, although it can also be just a vowel or even a non-syllabic consonant.

Roach (1982) reviews some previous work on stress-timing and syllable-timing. He points out some of the practical problems involved in assigning a language to one or other of these categories, such as the identification of stresses; where to begin from when measuring inter-stress intervals; how to remove the effects of tempo variations between one tone unit and another; what to do with the beginnings and endings of tone units: they often begin with unstressed syllables, which can only be counted as part of an inter-stress interval if a silent stress is posited at the beginning of the unit. (For more about such problems see the following section
on stress-groups.) Another problem with the measurement is that in English, tone-unit final syllables are often lengthened, such that the material between the last stress and the end of the tone-unit are often ignored. Tone-unit bondaries can also be difficult to identify, although they are sometimes marked by pause. Roach tests for stress-timing by measuring the duration of each tone-unit less any discarded syllables and dividing by the number of inter-stress intervals to give a target duration for isochronous intervals. It was found, however, that so-called stress-timed languages had a higher deviation from this target value than did syllable-timed languages. Roach concludes that a language cannot be assigned to one of the two rhythm categories on the basis of measurements of time intervals in speech. He concludes that a language is syllable-timed if it sounds syllable-timed and suggests that this impression may be due to simple syllable structure, while reduced vowels give an impression of stress-timing.

Shen and Peterson (1962) report an experimental failure to find isochronous stress in English. Uldall (1964) looks at the duration of feet in the RP accent of English. She claims to have found a 'tendency to isochronism' in that more than half of the 'filled' feet fall into a group from 385 to 520ms. However, since she found that feet with more syllables were longer, the evidence seems weak indeed.

Balasubramanian (1980) rejects the traditional assumption that Tamil is either syllable- or stress-timed, since the time taken to utter a foot was experimentally found to increase proportionately with the number of syllables it contains, while the durations of syllables
varied between 30 and 417 ms and are so not isochronous. Balasubramanian found that the duration of a segment depends on the structure of the syllable in which it occurs, such that vowels were longest in syllables consisting of a single vowel, and shortest in CVC syllables. Vowels in VC syllables were always shorter than those in CV syllables, which Balasubramanian attributed to the VC syllables always having a long consonant, while CV syllables have either a non-geminate C or the shorter of two consonants which form a medial consonant group.

Pointon (1980) looks critically at the claim that Spanish is syllable-timed. He proposes a pattern which he calls segment-timing – anti-rhythmic in the conventional sense of the word rhythm. This means that each segment is said to have a standard duration, dependent on its phonetic context (including stress to some extent), with no distortion of this duration to produce either isochronous syllables or stresses.

Major (1981) claims to show that Brazilian Portuguese is in a state of change from being a syllable-timed language to being a stress-timed language. He bases this claim on a comparison of speaking styles ranging from citation through normal to casual, where he found that casual speech showed tendencies associated with stress-timing, while more formal styles tended to be more syllable-timed.

Wenk and Wioland (1982) reject the notion that French may be syllable-timed, on the grounds that listeners tend to rely on rhythm to decode spoken French. They claim to have found that 12 syllables took the same time
to say as 6, and claim that 'to say that French is syllable-timed is to deny the existence of accented syllables in the language'.

Dauer (1983) compares data from continuous texts in Thai, English, Spanish, Italian and Greek. It was found that inter-stress intervals in English are no more isochronous than in Spanish. Dauer suggests that the regular appearance of stress may be universal, and that it is vowel reduction, syllable structure and the phonetic realization of stress and its influence on the linguistic system that constitutes the perceived difference between syllable- and stress-timed languages.

Hoequist (1983a) investigates the reality of impressionistic categories of language (stress-, syllable- and mora-timing). A comparative study of Spanish and Japanese shows no strict isochronous rhythm, but finds evidence for a less rigid hypothesis of rhythm categories. Hoequist claims that discussions of rhythm must be cross-linguistic, since if a particular phenomenon occurs in languages of all three types, it cannot be evidence for a particular type. Hoequist compares known durational phenomena from English with data from Spanish and Japanese and finds that:

1. phonemic length causes durational change only in Japanese;

2. syllable structure and composition affect duration in all three languages;

3. position in the word was found to have more effect on duration in English (final to non-final syllables
than in Spanish (1.17:1) which fits in with the hypothesis of syllable-timing, since Spanish syllables are less variable in duration than English syllables;

4. stress causes increased duration in all the languages, but more in English than in Spanish, which again corroborates the stress-timing versus syllable-timing distinction.

In another paper Hoequist (1983b) compares the duration of real speech syllables to reiterant syllables of the type /mama/, organized to reflect the stress sequence of real speech. He finds that the original morae and stress feet do influence the duration of the nonsense syllables. When English, Spanish and Japanese are compared, Hoequist finds that English speech is organized around stress feet while Japanese is organized around the mora. He finds no evidence that Spanish timing is based on the relative durations of syllables.

Borzone de Manrique and Signorini (1983) agree with this last point. They claim to have shown that in Argentine Spanish, segment duration varies according to several factors, with interstress values clustering around an average value. From this they conclude that Argentine Spanish rhythm has a tendency to stress alteration.

In this section, many claims to have found isochrony in stress-groups/feet/rhythm units, syllables or in other units have been discussed. However, most writers who found the notion of isochrony as a part of linguistic competence to be attractive, offered no experimental evidence to support their theories, while those who performed experiments found only vague tendencies to-
wards isochronous units or even evidence that isochrony did not exist, depending mostly on the interpretation of the data. The most convincing studies were Hoequist (1983a and b) and Dauer (1983) where cross-linguistic evidence is offered challenging the traditional choice between stress- syllable- and (sometimes) mora-timing for any given language.

Other investigations of stress-timing and syllable-timing are based on perceptual material. Allen (1972) investigates the perceived location of rhythmic stress beats in English. Informants were able to tap in time to the perceived rhythm, also to place an audible click where they would have tapped, and judge if the click was correctly placed. Using this method, Allen found the rhythmic beats to be closely associated with the onset of the nuclear vowels of the stressed syllables, but to precede these by an amount positively correlated with the length of the initial consonant(s) of the syllable. Allen interprets this as meaning that the rhythmic beats of English are centered round the release of the initial consonant into the stressed vowel.

Lehiste (1973) hypothesizes that isochrony may be primarily a perceptual phenomenon. She found that rhythmic units were not independent of their syntactic structures. When four different foot types were compared in the same position, they were found to have different durations, although listeners could not consistently tell which foot was longest and which shortest, which led to the impression of isochrony. The last foot in the utterance was always longer, but listeners could not pick up this difference. They performed better with non-speech material, and Lehiste suggests that differen-
ces such as final lengthening are expected and so allowances are made for them.

Lehiste (1977) continues the discussion of isochrony as primarily a perceptual phenomenon. She claims that since the listener expects isochrony, it is possible to manipulate the duration of the inter-stress intervals for linguistic purposes, for example, to signal the presence of a syntactic boundary. Lehiste suggests that isochrony in production occurs only when conditions are favourable.

Morton, Marcus and Frankish (1976) found that words presented such that they are acoustically regular were not perceived as regular. Morton et al investigated the requirements for perceptual isochrony, and defined the P-centre of a word as the psychological moment of its occurrence. Tuller & Fowler (1980) also found that when informants were presented with acoustically regular speech (in this case a sequence of digits), they were not perceived as regular. Only when systematic departures from acoustic isochrony were made did the listeners perceive the sequence as isochronous. These systematic departures from isochrony were the same as those speakers use when asked to produce isochronous speech.

2.5 Word

Many different definitions have been proposed for the word. Orthographically, a word can be defined as the material between two spaces, but there are many problems with this kind of definition.
Lehiste has worked on the temporal correlates of the word and word boundaries. Lehiste (1972a) explores the temporal relationships between the segments in three sets of Estonian words. She hypothesizes that the word is programmed as a whole unit, and that significant relationships exist among all segments that constitute a word, although not all segments participate in segmental quantity oppositions. In Estonian, there are three distinctive quantities for both vowels and consonants. Lehiste claims to have found compensation in the non-contrastive part of the word, such that if the shortest type of consonant occurs the final vowel will be long, or if the longest vowel occurs, the final consonant will be short. From this Lehiste draws the conclusion that words constitute units of programming in Estonian. These words were recorded in isolation, so it is unfortunately impossible to distinguish word-based effects from other possible effects, such as may be based in stress groups or other larger phonological units.

Lehiste (1972b) investigates the effects of morphological and syntactic boundaries on the temporal structure of spoken utterances. She compares sets of words consisting of: (a) a monosyllabic base form; (b) bisyllabic and trisyllabic words derived from the base by the addition of suffixes; and (c) three short sentences in which the base was followed by (1) a stressed syllable, (2) one unstressed syllable and (3) two unstressed syllables, all with a syntactic boundary immediately after the base form. These sentences thus reproduced the syllabic sequence of the words derived from the base form. Measurements of the durations of words and syllables were made and it was found that temporal re-adjustment processes generally ignore morpheme and word boundaries. Shortening of a long vowel before a short
syllable was found. Lehiste reasons that since timing patterns are suprasegmental, there is no reason why they should respect word or morpheme boundaries. It seems here that Lehiste is claiming that there are no durational correlates of the word in English.

Lehiste (1980) discusses Lindblom and Rapp (1972) and Nooteboom (1972) on Swedish and Dutch respectively. They found that the duration of a syllable nucleus decreases as the number of syllables that remains to be pronounced in the word at the beginning of the syllable concerned increases. Lehiste tested this for English using non-sense words in long and short frames, and found the duration of the test words to depend more on the total duration of the utterance than on the position of the test words within the utterance. Again here there is no reason to expect any durational correlates of the word.

2.6 Stress-groups

Jones (1918) claims that if an unstressed syllable occurs between two stressed syllables, it will be shorter if it is grammatically more closely connected with the following stressed syllable than if the closer grammatical relationship is with the preceding stressed syllable. According to the definition of the stress group given above, Jones is claiming that stress-group final syllables are longer than stress-group initial syllables.

Jassem (1949 and 1952 chapter 3, as quoted by Crompton 1980) proposes the stress-group, defined as a stressed syllable with zero or more unstressed syllables preceding and following, as a unit of phonological
structure with durational correlates such that the syllables preceding the stressed syllable are relatively short while the stressed syllable and the syllables following it are relatively long.

Ondrackova (1962) attempts to place stress-group boundaries on Czech material. She defines the rhythmical unit as a section of utterance with a single stress peak. In Czech this prominence occurs most frequently on the first syllable of the word. Ondrackova claims that the length of rhythmical unit, as calculated by the number of syllables in it, is closely related to the tempo of speech, and that they can influence each other. Most units in the material were small, ca 2-3 syllables, which Ondrakova relates to a natural tendency to produce regular rhythm. Ondrackova finds (1) that words pronounced in isolation have accent on the first syllable, but this accent may be lost if the word is other than the first word in the rhythmical unit; (2) that syllables in a single rhythmic unit are very closely joined - as much so as the syllables of a word - although exceptionally the boundary of a rhythmical unit can pass through a word; (3) that boundaries between rhythmical units are not always realized as pause: most frequently the boundary is marked as a weakening of the contact between the syllables.

Ondrackova (1967) considers the status of a monosyllabic unaccented word in Czech. Such a syllable cannot be attached to the previous unit because there is a pause before it. Ondrackova concludes that a monosyllabic word in Czech may be an independent rhythmical unit, a first stressed syllable or other unstressed syllable of a longer rhythmical unit.
Hill, Witten and Jassem (1978) also consider the best way to treat odd syllables when segmenting speech into rhythmic units. They debate the advantages and disadvantages of both Jones' (1918) method of feet, and Jassem's method which ignores proclitic syllables at the end of feet which belong syntactically to what follows, to give rhythm units which are considered to be isochronous.

Crompton (1980) attempts to establish the durational correlates of linguistic units such as segment, syllable, stress-group, phonological phrase etc. in French. He finds that the phonological word (as proposed by Chomsky and Halle (1968)), the phonological phrase (equivalent to the tone-group), clauses and sentences can be associated with particular durational patterns, while no such durational correlate was found for accent.

2.7 Global rate

In Lindblom (1963) an investigation is made of the extent to which formant frequencies in vowels uttered in consonant environments reach their targets as a function of vowel duration. Lindblom compares the effects that an increase in speaking rate and a decrease in stress have on vowel acoustics. The effects are similar - vowels are shorter and reduced in vowel colour relative to vowels in slow stressed speech - and Lindblom proposes that a single production strategy is employed to effect both rate and stress changes.

Kozhevnikov and Chistovich (1965) performed experiments to determine what happens to vowels and consonants
when the rate of utterance changes. They found a
distinct relationship between the relative duration of
the vowel and the consonant in a syllable and rate of
speech. On increasing the duration of a phrase they
found a clear lessening of the portion of time in it
occupied by the consonants. They suggest that this
change in relative duration may be due to pre-defined
maximum and minimum values for consonants. The suggest
that vowels can probably increase in duration without
limit, just as they can disappear (ie become reduced to
zero duration). Kozhevnikov and Chistovich suggest that
vowel deletion in rapid speech may occur because at a
rapid rate of speech, if the programmed timing of
syllables is to be preserved, there is not enough time
for the vowel to be articulated (see also section on
global rate in production).

Gay (1968) studied the effect of changes in speaking
rate on diphthong formant movements. His results showed
that onset target position and second formant rate of
change are fixed features of the diphthong formant
movement, while the target positions for offset were
found to change when the speaking rate changed. Gay also
found that diphthongs were longer when they were follow-
ed by a consonant, and that this lengthening was
achieved by longer steady state phases for some diph-
thongs (/ɔɪ, ʌɪ, əʊ/) while others (/ɛɪ, ɔɪ/) used a
longer glide phase to accomplish the lengthening.

Allen (1973) describes the accuracy with which the
durations of individual segments are controlled over
changes in rate. He claims that speech rate and segmen-
tal duration are separable.
Gay, Ushijima, Hirose and Cooper (1974) made a study of the effect of increased rate of utterance on the articulation of the labial consonants /p, w/ in combination with the vowels /i, a, u/. It was found that target undershoot occurred at the faster rate, but not at the moderate rate.

Bolozky (1977) discusses the differences between fast speech and normal speech. He claims that there are processes unique to fast speech, and also that there are normal speech processes which do not operate in fast speech. Like Port (1977) he discusses the absence of vowel lengthening before voiced stops in fast speech. He suggests that constraints on normal speech are very often relaxed in fast speech, and that different rules apply, such that a distinction should be made between normal, slow and fast speech. Bolozky provides the alternative view that fast speech be regarded as a tempo component in variable rules. He hypothesizes that (1) the faster the speech tempo, the greater the likelihood of variable assimilation processes and of restricting clusters to those which are assimilated in some way; (2) the faster the speech tempo, the more likely are reduction and/or deletion processes and restrictions on the occurrence of weak syllabic elements; (3) normal speech processes, primarily assimilations and reduction processes and constraints enforcing assimilation or reduction, should be restricted with reference to what the speaker perceives as the limit of recoverability of semantic information (i.e., weaker segments may be assimilated, deleted or excluded from fast speech, provided that the speaker believes that their loss would not render the word unrecognizable.)

Port (1977) ran experiments to establish how speech
tempo influences the durations of stressed vowels and medial stops. Using three tempos (fast, neutral and slow) he found that slowing down lengthened all measured intervals by the same proportions, while speeding up resulted in (1) less shortening in the stressed syllable than the sentence as a whole; (2) reduced durational voicing contrasts for non-apical voiced-voiceless pairs.

Isenberg (1978) recorded words distinguished by a fricative-affricate contrast (dish/ditch) at five increasing speech rates. The durations of closure interval in ditch and fricative noise in both words were measured. When linear functions were plotted relating the duration of the acoustically defined intervals to the total sentence length, and fitted to the data, it was found that the slopes were smaller for closure intervals than for fricative noise, which indicates that closure intervals are less flexible than fricative noise duration as speaking rate changes.

Gay (1978) studied the effects of changes in speaking rate on both the attainment of acoustic targets for vowels, and the relative time and speed of movement towards these presumed targets. It was found that mid-point formant frequencies did not vary substantially as a function of rate. However, in fast speech, onset frequencies of second formant transitions were closer to their target frequencies, while CV transition rates were not significantly changed, which indicates that movement towards the vowel simply begins earlier for fast speech. When both speaking rate and lexical stress were varied, the results showed that for stressed vowels, an increase in speaking rate was accomplished primarily by a decrease in duration, while destressed vowels, even if
they were of the same duration as quickly produced stressed vowels, were reduced in overall amplitude, fundamental frequency and to some extent vowel colour. From these results Gay concludes that speaking rate and lexical stress are controlled by two different mechanisms. This contradicts Lindblom's (1963) claim.

It is interesting to note that within the framework used by Gay, the effects he describes appear to be of the same kind, and are controlled directly by a mechanism controlling the rate of utterance. These effects would be quite different within the contour-based framework described in this dissertation. Our theory would recognize some of Gay's effects as changes in the utterance's track through the phonetic space, (the tendency towards vowel centralization, for example), while other effects (such as the greater elasticity of vowels compared to consonants) do not affect the track at all, with the difference showing up in the timing contour.

Tuller, Harris and Kelso (1981) also discuss the similarity of the effects of changes in rate and stress on segmental timing. They compare two types of explanation for the temporal changes: first that the 'commands' for the production of syllables uttered at high rate or without stress show greater temporal overlap than for the same syllables uttered more slowly or with a greater degree of stress; or alternatively that the temporal relations between the articulations remain constant, but the individual gestures themselves are reduced. In an experiment based on nonsense syllables, Tuller et al found that temporal variations accompanying a change in speaking rate were not the same as the variations
accompanying a change in stress.

Gay (1981) demonstrates that speaking rate changes can result in changes in (1) segment duration, (2) articulatory displacement, (3) articulatory velocity and (4) intra-syllabic coarticulation. He claims that these transformations are non-linear, and also that they are achieved by a reorganization of speech motor strategies. Gay documents the following: (1) If speech is faster segments must be shorter, but when segments are shortened by an increase in speaking rate, this shortening is not proportionate - vowels are reduced more than consonants. (2) Increases in rate are accompanied by a tendency to vowel neutralization. Gay finds a positive correlation between vowel shortening and target under-shoot or vowel reduction. This is explained by motor commands at fast speaking rate being issued too quickly to permit the articulators to reach their targets before the next set of commands arrives. (3) An investigation using x-rays showed that although vowels were reduced during fast speech, the velocity of articulatory movements to the reduced vowel target positions remained constant. (4) Gay discusses his hypothesis (outlined in Gay (1978) that speaking rate affects the relative timing of, or coarticulation between, adjacent segmental gestures. Acoustic measures showed the onset frequencies of the second formant of the vowel to be closer to mid-point frequencies during fast speech. Gay claims that these facts show that speakers do not control their rate of speech along a single dimension or by a constant mechanism.

The area of the effect of changes in rate on the internal timing of utterances appears from the above discussed work to be very complex and difficult. An
alternative treatment of changes in the global rate of utterance is proposed in the chapter dealing with global rate.

2.8 Utterance

Cooper and Danly (1981) examined utterance final lengthening in words containing various consonants and vowels. Firstly, words containing the fricatives /f, v, s, z/ in word-initial, -medial and -final positions in both phrase-final and utterance final positions were measured for final lengthening. Some utterance-final lengthening was found when fricatives were in word-initial position, but much more when fricatives were word-final. Secondly, utterance-final lengthening was measured for words with long versus short medial vowels followed by a word final voiced versus voiceless stop consonant. A greater amount of utterance-final lengthening was found for words with a terminal voiced stop.

Borzone de Manrique (1982) investigated lengthening in the final word of the utterance and found that not only final vowels, but also many other final word segments were elongated. She attempted to quantify this effect and found that in open syllables in penultimate stressed words, the greater increment was in the final syllable's vowel. The penultimate syllable was also longer, but it is not clear whether this is due to stress or position or both. Borzone de Manrique found that pre-boundary words were longer than non-final words, but shorter than final words.

Berkovits (1984) compares phonetically matched fi-
nished and unfinished sentences in English and Hebrew. Segments were found to be longer in sentence-final and phrase-final position in English, but there was no difference between finished and unfinished sentences in Hebrew.

It appears then that there is lengthening at the end of a number of units: the stress-group (Jassem 1949 and 1952,); the phrase (approximately equivalent to the tone group or phonological phrase); and the utterance or sentence (Crompton 1980 on French, Berkovits 1984 on English, and Cooper and Danly on English 1981).

3 Phonological representation

As has been shown there are many units which seem to have durational correlates and therefore a place in the phonological representation of timing in speech. Most of these units can be placed in a hierarchy. This means that the units must have boundaries such that the boundary of a large unit must necessarily be a boundary of all smaller units. This means that of the units discussed and defined above, segments, syllables, stress groups, tone-groups and utterances can have a place in the hierarchy in the following way:

Fig 2.1

```
.s.s.s.s.s.s.s.s.s.s.s.s.s.s.s.s.s.s.s.s.s.s.s.s.
 .sy .sy.sy . sy .sy . sy .sy . sy . sy .
 .stress-group. stress-group .stress-gp.
 .tone-group . tone-group .
 . utterance .
```
This arrangement can also be expressed as a tree diagram:

\textbf{Fig 2.2}

Clearly units such as the foot and the word do not fit into a hierarchical system such as the one described above. The boundaries of larger units may come in the middle of these, so that a stress-group (or foot) boundary may come in the middle of a word, (for example in words with more than one word stress such as indiscriminate). Lehiste's (1972a, 1980) work on the durational correlates of the word and word boundaries could equally well apply to the stress-group. Any treatment of isochronous stresses (stress-timing such as Abercrombie (1967), Jones (1918), Pike (1945), Ladefoged (1975), Roach (1982), Shen and Peterson (1962) Balasubramanian (1980), Dauer (1983), Hoequist (1983b), Borzone de Manrique and Signorini (1983) discuss it) entails the notion of the foot. A major problem of the foot, as discussed by Hill, Witten and Jassem (1978) and Roach (1982), is that it is impossible to divide an utterance
cleanly into feet - there are often odd unaccounted for unstressed syllables at the beginning of tone-groups which will not fit into any foot unless an imaginary 'silent stress' is posited at the beginning of the tone-group. The stress-group, being closely related to the syntactic structure of the utterance, can have its boundaries imposed on the utterance in a much simpler way.

Things other than units have also been considered as having durational correlates: final lengthening, stress, changes in global rate of utterance, and the rhythm categories of stress-timing and syllable timing, all of which must have a place in the phonological representation. These can be described in terms of the above units, and are often limited by the boundaries of these units, such that stress-group final lengthening might be supposed to be a lengthening beginning at some time before a stress-group boundary and continuing until the end of the stress-group. The same goes for stress-associated lengthening as discussed by Fry (1955), Gonzales (1970), Lea (1977), and Klatt (1975) who take stress to be a property associated with syllables, and limited by the syllable boundaries. Syllable timing refers to the syllable as the basic unit of rhythm, while stress-timing, as discussed above, refers to the foot, although Hill, Witten and Jassem (1978) suggest that there might be some other kind of isochronous unit, something between feet and stress-groups. In any case no evidence has been found for the isochronous foot. This means that the above hierarchy contains all that is necessary for the phonological representation of timing in speech.

These phenomena are all part of competence - that is,
they are part of the speaker-hearer's knowledge of his language. The phonological model expressed in the hierarchical tree-diagram in figure 2.2 is the basis upon which the description of all the phenomena described earlier in the chapter can be built.
CHAPTER 3  METHODOLOGY

In the previous chapter it was established that the following units have a place in the phonological representation of timing in speech: utterance, tone-group, stress-group, syllable, segment. One way of looking at the phonological representation of an utterance is to say that it is the output from a generative grammar. This representation is still rather abstract. In order for speech to be generated, a much more precise specification of everything about the pronunciation of an utterance that distinguishes it from other utterances must be provided. This specification is arrived at by a set of linguistic rules, such that the final output of the generative grammar is the phonetic representation of the utterance. It is clear that the above segments are present in the phonological representation of timing. It is less clear that there is a need for segments in the phonetic representation. The phonetic representation (PR) is much less abstract than the phonological representation. There are no segments in the acoustical, auditory and articulatory representation of the utterance; they are continua. This means that a non-segmented representation is easier to relate to these representations than a segmented representation would be.

1 Phonetic Representation

Ainsworth's (1981) treatment of vowels included duration as one of the parameters of the vowel. He was still dealing with segments in the treatment of timing, although he could handle vowel quality by reference to a multi-dimensional space. It is not difficult to imagine Ainsworth's phonetic quality space to be expanded to
include those dimensions necessary for the phonetic representation of other speech sounds and whole utterances. A large number of quality dimensions would be necessary. The phonetic representation of an utterance in this form can be pictured as a track through this multi-dimensional quality space. Points along the track correspond then to the coordinates of the sounds of the utterance. Compare this to a path marked out on a road map which a car must follow. The map shows the path that the car must take but contains no information about the speed at which the car should travel. Clearly, even if the car follows the same track many times, the speed can vary at different points along the track. Also, the overall speed of the journey can vary if the path is followed by a faster driver or a faster car, or even the same car and driver in a hurry. Imagine then the position shown in figure 3.1, where a contour represents the speed of the car at all points along the marked path, where ——— = path of car, and …… = speed of car.

**Figure 3.1**

![Speed Map](image)

Leaving this map analogy to return to the PR of
speech, we can see that many local changes in the rate of utterance may occur along the path that an utterance takes through the multi-dimensional phonetic quality space. The rate of utterance can then be expressed as a continuous contour, and becomes the rate at which phonetic space is traversed. There exist no segments in this representation, although there may be rate changes which are associated with the boundaries of phonological segments of all sizes. Let us then take a simple example. The two sequences [iɛ] and [ɪɛ] as occur for example in ear and year, can be said to follow the same track through phonetic space. For ease of exposition, we can express all the qualitative differences between [i] and [ɛ] as differences along a single continuum. The difference between them is then one of relative rate of utterance: the early part of year has higher rate than the early part of ear, while the later part of ear has higher rate than the later part of year. This relationship is shown in figure 3.2 below where the solid line represents ear and the dotted line represents year:

Figure 3.2a plots the position on the quality axis (distance in the phonetic space) against time. Figure 3.2b shows the rate of utterance against the position in
phonetic space (distance travelled) of each of the sequences, such that they can easily be compared.

2 Timing contours

Graphs such as those shown in figure 3.2 cannot be drawn for real utterances, since we cannot measure phonetic distance in a straightforward way. It is however possible to approximate to actual timing contours by measuring the amount of spectral change between one part of an utterance and the next. This gives however an acoustic timing contour rather than a linguistic one. This is, however, not an important issue at the present time. In practice, when we look at a representation such as spectrograms, we make use of easily identifiable points in the utterance. These points are usually easy to identify because they coincide with sudden changes in the spectral character of the utterance. These points correspond then to the series \( q = q_0, q_1, \ldots, q_n \), which can also be thought of as points on the track that an utterance follows through the multidimensional phonetic quality space. The time intervals between these sample points are the series \( t = t_1, t_2, \ldots, t_n \).

Now, the rate of utterance contour:

\[
q'_i = q'_0, q'_1, \ldots, q'_n
\]

is by definition the rate of change of \( q \), or

\[
q'_i = \frac{q_i - q_{i-1}}{t_i}
\]

so that

\[
t'_i = \frac{q_i - q_{i-1}}{q'_i}
\]

The \( q' \)-contour represents then the rate of utterance
at various points in time. From this point onwards, the term timing contour will be used to refer to the discrete contours obtained in this way. As was mentioned previously, we cannot actually draw q'-contours since we have no way of coming at the measurement of distance travelled through phonetic space. This is not important for their theoretical significance. Nothing of theoretical importance is determined by the reality of q'-contours or by the selection of the points $q_1^1, \ldots, q_n^n$. These points are chosen on grounds of convenience, in order to provide a reasonable approximation to the continuous q-contour.

Suppose then that there are two q-contours, $a$ and $b$ made up of the points:

\[
a_0, a_1, a_2, a_3, \ldots, a_n
\]

\[
b_0, b_1, b_2, b_3, \ldots, b_n
\]

where $a_i = (ta_i, q_i)$

$b_i = (tb_i, q_i)$

Note that the $q_i$s are the same in both cases, which means that the two contours represent identical paths through the phonetic quality space, such that it is possible to compare the elapsed time at fixed points along the quality dimension for each of the utterances. Consider first the simple case where the q-contour consists of a straight line (compare with eg /aja/ where the direction changes in mid-utterance). This track is followed at two different constant rates of utterance as is shown here in figure 3.3, where $a$ is shown as a dotted line and $b$ as solid.
Figure 3.3

Example:

In order to make actual calculations, we can calibrate the axes to give the following values for $q_i$, $t_{a_i}$ and $t_{b_i}$:

<table>
<thead>
<tr>
<th>$q$</th>
<th>0</th>
<th>2</th>
<th>4</th>
<th>7</th>
<th>8</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_{a_i}$</td>
<td>0</td>
<td>5</td>
<td>10</td>
<td>17.5</td>
<td>20</td>
<td>25</td>
</tr>
<tr>
<td>$t_{b_i}$</td>
<td>0</td>
<td>15</td>
<td>30</td>
<td>52.5</td>
<td>60</td>
<td>75</td>
</tr>
</tbody>
</table>

Remember that the values shown here for $q$ have no correlates in reality: the purpose of these calculations is to arrive at a practical method of representing speech timing in the PR. However, given these values it is a simple matter to calculate the $q'$-contours $a'$ and $b'$ in the following way:
\[ a' = a'_0, a'_1, a'_2, a'_3, \ldots, a'_n \]
\[ b' = b'_0, b'_1, b'_2, b'_3, \ldots, b'_n \]

where
\[ a'_i = \frac{q_i - q_{i-1}}{t_{a_i} - t_{a_{i-1}}} \]
\[ b'_i = \frac{q_i - q_{i-1}}{t_{b_i} - t_{b_{i-1}}} \]

**Figure 3.4**

As would be expected, the \( q' \)-contours \( a' \) and \( b' \) are completely flat since the gradient of the \( q \)-contours was constant. In speech terms this shows that the rate of utterance is constant throughout for both utterances. The graph shown in figure 3.4 can be called an absolute timing contour. It shows absolute rate of utterance.

Instead of drawing \( q' \)-contours showing the absolute rate of utterance (or rate of movement through phonetic space), we can look at relative rate of utterance. Even if we are unable to say what the rate of utterance is, either globally (the overall rate of utterance) or locally (the utterance rate at particular points in the utterance), we can easily compare two utterances and say
that one is uttered at a higher rate than the other, or as in the case of [i̯] and [ɪ̯] above (in fig 3.2) that the rate at a point in the utterance is higher in one case than in the other. Problems arise however when we want to quantify these rate differences. Perhaps the obvious way is to speak of increases or decreases in rate relative to and proportional to a norm. In order to compare two utterances in this way, it is essential that they both follow the same track through phonetic space; i.e., like the [i̯] and [ɪ̯] example both utterances are made up of the same phones. (This is the basic rule; minor deviations arise when two similar utterances are being compared for temporal effects.) When the track through phonetic space can be taken to be the same in both cases, it is possible to 'cancel out' the qs as is be shown below.

A different kind of timing contour can therefore be drawn by comparing two utterances that have identical (or at least very nearly identical) tracks through the phonetic quality space. Line-up points are selected in the utterances (the procedure for this is described in detail below), and the intervals between the line-up points are measured. One of the utterances is taken to be the norm, and the other the test utterance (it does not matter which is which). Next, the differences in the duration of these intervals as they occur in the norm and the test utterance are calculated and expressed as a percentage increase or decrease. This yields a series of points which define a relative timing contour for the test utterance as compared with the norm.

This process may be clarified by an analogy. Imagine that the test utterance has its spectrogram printed on a
sheet of rubber, such that it can be stretched. If points in the spectrograms are chosen so that they can be identified in both utterances (for example noise bursts associated with plosive release or changes in direction of formant movement) these can be used as line-up points for the comparison of the utterances. Imagine further then that a pin is inserted through each pair of line-up points, joining the norm utterance with the elastic utterance. In order to make the line-up points in each case meet it will be necessary to stretch portions of the elastic spectrogram. Different portions of the spectrogram may be stretched by different amounts and sometimes it may need to be compressed. It is then possible to extract a figure corresponding to the amount of temporal warping ($w$). Since however rubber spectrogram paper does not exist, the same figure can be extracted more simply by measuring the time between line-up points for both the utterance chosen as norm and the one which is to be compared to it. The difference between these times can be expressed as a percentage increase or decrease in inter-line-up-point duration or inverted, as percentage increase or decrease in rate of utterance from the norm. These relative timing contours can be seen as a representation of the sort of temporal warping that is needed to match the test utterance to the norm. They can be called $w$-contours.

This process gives timing contours as follows for the numerical example given above. Because of the constant rate in both cases, the amount of temporal warping is also constant:
It is not immediately clear what label may be given to the horizontal axis. It does not represent absolute time, but rather it represents two different times, one from each utterance. The regularity of intervals on the horizontal axis is spurious, since the sampling frequency is irregular (a sample is taken at every line-up point, and no effort is made, or could be made, to have regular line-up points).

If the time taken for each of the utterances was the same, the area under the w-contour should ideally be equal to zero, since w is a measure of the amount of time warping which is needed to match the utterances to each other, and if they take the same total time, the warpings in each direction should balance each other. For the current example, this is not the case since the two utterances did not take the same total time to utter. In order for this to happen the horizontal axis must represent time, that is, instead of having regular
samples, the samples will be placed on a time axis at the points where they occurred for the utterance taken as norm, as follows. Imagine the case where the duration of a vowel plus the duration of the following plosive is found to be constant regardless of whether the plosive is voiced or not. This in spite of the fact that the duration of the vowel is much less when followed by a voiceless plosive, and the duration of the stop closure is greater when the stop is voiceless. Whether or not this is deliberate temporal compensation (cf Port, Al-Ani and Maeda 1980), it can be represented very effectively using w-contours, especially if these are normalized in time. The contours below show the picture as if several line-up points had been taken in each the vowel and the consonant.

**Figure 3.6a**

![Graph showing line-up points on a time axis](image)

No label on x-axis; regular line-up points

**Figure 3.6b**

![Graph showing line-up points on a time axis](image)

X-axis represents time; irregular line-up points
It is the kind of normalized w-contour shown in figure 3.6b which is meant when the term timing contour is used. It may seem that these timing contours are much more abstract than the q'-contours discussed above. The relationship between q'-contours and timing contours is not immediately obvious, but is quite straightforward:

Let us return to the series representing the two q-contours, a and b:

\[ a = a_0', a_1, a_2', a_3, \ldots, a_n \]
\[ b = b_0', b_1, b_2', b_3, \ldots, b_n \]

where \[ a_i = (ta_i, q_i) \]
\[ b_i = (tb_i, q_i) \]

The normalized w-contour is obtained in the following way:

\[ w = w_0', w_1, w_2', w_3, \ldots, w_n \]

\[ \frac{(ta_i - ta_{i-1}) - (tb_i - tb_{i-1})}{(ta_i - ta_{i-1})} \]

\[ w_i = 1 - \frac{(tb_i - tb_{i-1})}{(ta_i - ta_{i-1})} \]

The q'-contours \( a' \) and \( b' \) are obtained as follows:

\[ a' = a'_0', a'_1, a'_2', a'_3, \ldots, a'_n \]
\[ b' = b'_0', b'_1, b'_2', b'_3, \ldots, b'_n \]

where \[ a'_i = \frac{q_i - q_{i-1}}{ta_i - ta_{i-1}} \]
and 
\[
    b'_i = \frac{q_i - q_{i-1}}{tb_i - tb_{i-1}}
\]

so that
\[
    a'_i = \frac{q_i - q_{i-1}}{ta_i - ta_{i-1}} * \frac{tb_i - tb_{i-1}}{q_i - q_{i-1}}
\]
\[
    b'_i = \frac{tb_i - tb_{i-1}}{ta_i - ta_{i-1}} = 1 - w_i
\]

This relationship is very important. We cannot make \(q'\)-contours for real utterances. If we could, we would have a measure of absolute rate of utterance. We do not even know what the units of rate of utterance might be. If however we divide one \(q'\)-contour by another, the \(q\)s cancel out, leaving us with a figure we can obtain without reference to a quality (or phonetic) space – the \(w\)-contours (relative rate contours).

The relationship between the \(w\)-contours and the \(q'\)-contours is simple. The \(w\)-contours have a maximum value for \(w\) of 1, although there is no minimum value. In other words this means that the relative rate can be reduced infinitely; if the warp factor \(w = -\infty\), the duration of the compared sample becomes infinitely long. If however the value of \(w\) approaches 1, the duration of the compared sample approaches zero, since when \(w = 1\) the duration of the compared sample has been decreased from the norm by 100%, and so is zero. This is one way of
looking at what happens when a segment is deleted - a
destressed vowel or a consonant in a cluster.

Timing contours (w-contours) have then a close and
straightforward relationship to g'-contours. This means
that even though a quality dimension is not accessible,
calculations of relative rate may be made using the
above described method.

3.2 How to make a timing contour

Timing contours (w-contours) are a representation of
the fluctuations in rate of utterance in one utterance
relative to another almost identical utterance. In order
to draw w-contours, it must be possible to identify a
common set of events in the acoustic record of both
utterances such that measurements of the time elapsed
between these line-up points can be measured and com-
pared. The technique used to obtain the acoustic record is
described in chapter four.

3.21 Line-up points

The line-up points are simply chosen as easily
identified points in the acoustic record. Let us suppose
that we wish to make a timing contour of an utterance
pronounced at both high and low global rates. The
utterance is the same in both cases, for example I say
price every day as is shown in the following figure 3.7.

Suitable line-up points in such an utterance might be
the beginning of low frequency activity for the initial
vowel, the beginning and end of the friction of the /s/,
the closure of the /p/, the release of the /p/, the
beginning of the vowel, the beginning of friction for
the /s/, the end of the friction for /s/, the beginning and end of friction for the /v/, the closure and release for the /d/, and the end of low frequency activity at the end of the utterance.

Figure 3.7

These are not the only line-up points that could be taken. Depending on the acoustic information available, quite different line-up points might have been taken, perhaps making reference to the direction of formant movements, or to formant frequency peaks and troughs. Provided that the informants make no pauses in their pronunciation of the utterances, the above method of line-up point definition, (which covers the entire utterance) should work satisfactorily. Otherwise the process might be considerably more complicated.

3.22 Measurement of inter-line-up point durations

The time taken to utter the material coming between
each pair of line-up points can be measured quite straightforwardly from the time base of the acoustic record.

3.23 Calculation of $w$ values - vertical axis of contour

One of the two rates must be chosen to represent the norm. In this case the choice is purely arbitrary, so we can choose the high rate as norm. The relationship between the intervals between a given pair of line-up points at each of the two rates can be expressed as follows:

$$w = \frac{d_{\text{fast}} - d_{\text{slow}}}{d_{\text{fast}}}$$

where $d_{\text{fast}}$ is the time taken to utter the material between any given pair of line-up points at fast rate and $d_{\text{slow}}$ is the time taken between the same pair of line-up points at slow rate.

If $w$ is multiplied by 100 it represents the percentage of the norm (fast) utterance, but this is simply a scaling factor, and so unnecessary.

3.24 Calculation of the $t$ values - horizontal axis of timing contour:

The horizontal time axis of the timing contour can be measured in milliseconds. Points of the timing will be plotted at those values of $t$ when the norm utterance reached each line-up point. The $w$ values can then be plotted against the $t$ values to give a timing contour. For ease of drawing, however, it is only in the contours shown in chapter four on segmental timing effects that the horizontal axis is calibrated in the above described way. Since in most cases the integral of the timing contour is not crucial, the horizontal axis is uncalli-
brated. It cannot then be said to represent real time, but does in some sense represent notional or phonological time, or the sequential ordering of the line-up points.
CHAPTER 4 SEGMENTAL TIMING EFFECTS

In the previous chapter, the notion of continuous contours in the phonetic representation of speech timing was introduced. In this chapter, these contours will be used to describe some of the temporal relationships operating at the level of the phoneme-sized phonological segment. Timing contour techniques can then be compared with a more traditional segmented approach.

1 A temporal effect.

One of the most discussed aspects of segmental timing is perhaps the durational differences between post-vocalic voiced versus voiceless stop closure durations - that voiceless stops are longer than voiced stops (Wajskop and Sweerts(1973), Port (1978), Fitch (1980)), and between the vowels preceding them - that the vowel before a voiced stop is longer than the vowel before a voiceless stop (Wang (1959), Raphael (1972), O'Kane (1978), Walsh and Parker (1981), Wardrip-Fruin (1974), and Mitleb (1984a)).

Some writers (eg Port (1979) and Fischer-Jörgensen (1979)) have gone further and suggest that the durational differences of the vowel and the consonant may be related. Since if the consonant is voiced the vowel is longer and the closure shorter than if the stop is voiceless, there is an inevitable tendency for the VC total duration to be constant. It has been suggested that this inverse relationship is part of the linguistic competence of the speaker, that the vowel duration is used to compensate for the difference in consonant
duration, or vice-versa. (Port, Al-Ani and Maeda (1980) examined a similar effect where different consonant lengths were compensated for by differences in the vowel durations in Japanese, but not in Arabic.)

2 An experimental investigation of the temporal effect.

Since this is the first experiment to be reported in the dissertation, the experimental method used will be fully related.

2.1 Experimental methods: The hypothesis

The first stage in the design of an experiment is the statement of a hypothesis in a testable form. It is often possible to test more than one hypothesis with a single set of material, though it is important then to treat the hypotheses separately. Since this dissertation investigates timing in linguistic competence, which is by definition knowledge of a specific language, hypotheses can be made concerning only one language at a time. A finding about the temporal structure of French cannot be assumed to work also for Spanish.

2.2 Choice of material

The material for an experiment must of course be carefully chosen. Most important in the criteria for the selection of material is that the experiment must provide a satisfactory test of the hypothesis. The material must contain contrasts of the sort required to test the claims of the hypothesis. Many experimenters use nonsense words in experiments. Often they use reiterative syllables of the type /bababa/ in an attempt
to eliminate the effects of varying phonetic content on the experiment. It has never been satisfactorily shown that informants treat this kind of material in the same way that they treat genuine speech material, especially as in some languages (including English) a sequence of the form /bababa/ would not be a permissible word. Although this kind of experiment undoubtedly makes life easier for the experimenter, it may mean that a false picture is formed. Liberman and Streeter (1978), however, claim to have found that the temporal patterns produced by this method are stable and reproducible both within and across speakers. An alternative method, the one adopted in this dissertation, is to use only genuine speech material, that is words of the language under investigation, or possible words that happen not to be words (e.g. mabbing).

The material chosen to test the hypothesis must be carefully chosen in other ways. If timing contours are to be drawn from the material it is necessary to have material in which line-up points can be defined relatively easily and in sufficient quantity to give a reasonable resolution in the timing contour.

For the purposes of statistical analysis many samples are necessary. This involves the informants repeating the test utterances a large number of times, at least ten times and ideally thirty times. The confidence levels of the statistical tests used make allowance for the number of samples used.

Several reading lists were compiled for each experiment, each containing between three and ten similar utterances repeated a number of times, to give the
required number of repetitions of each utterance. In those experiments where only a single word was under investigation, this was contained in a suitable frame sentence. In all experiments, each list began and ended with a number of dummy sentences to eliminate any effects (specifically timing effects) associated with utterances at the beginning and end of lists. The lists were compiled to make reading as easy as possible, i.e. to avoid tongue-twisting sequences of utterances.

2.3 Choice of informants

Ideally more than one informant should be used to test each hypothesis, to reduce the possibility of error, and also the possibility of an informant who uses speech timing in a way very different from most people. If two informants show very different results, a third and fourth informant can be used. All informants should in this case speak with similar accents, and without obvious speech defects or irregularities. This implies the expectation that all speakers of a given accent will have similar speech timing.

2.4 Recording

The recordings were made in a sound-proof recording booth using studio quality magnetic tape and tape recorders. A recording speed of 7.5in/s was used. Each informant recorded all utterances for any single experiment on a single day, although with larger sets of material it was necessary to break several times to allow the informant to rest.
2.5 Acoustic analysis of material

The recorded material was passed through an intensity meter to give two traces on ultra-violet light sensitive recording paper. The first trace was a high pass filtered intensity trace with an integration time of twenty, and the second a low pass filtered intensity trace with a low integration time (five) so that voicing information was also present. A microphone trace was included for additional information. The paper speed was 50 or 100 mm/s depending on the type of measurements to be made. A sample of the traces was shown in the previous chapter.

2.6 Statistical Analysis

As is described above under choice of material, a number of examples of each utterance were recorded. This was for the purposes of statistical testing. While a durational difference can be found simply by subtracting one difference from another or by drawing a timing contour, it is not easy to know if this difference is genuine and systematic, or if it is a chance fluctuation in rate of utterance. T-tests can tell if there is a difference between two sets of measurements, and also how great the probability is that this difference is genuine (ie not due to chance). In deciding whether there is a difference or not, the test uses the average value for each set of data as well as the square of each individual measurement in the data. This means that the effect of a single very large or very small measurement is considerably lessened. In deciding how much confidence the experimenter can place in the result the test
uses a number related to the size of this difference, (t), and a number related to the number of measurements in each data set, (df (degrees of freedom) = Na+Nb-2).

The value of t is obtained from the following formula:

\[
t = \frac{M_a - M_b}{\sqrt{\frac{\sum_{i=1}^{a} (X_{a,i} - M_a)^2}{a} + \frac{\sum_{i=1}^{b} (X_{b,i} - M_b)^2}{b}}} \frac{1}{\sqrt{(N_a + N_b - 2) \times \frac{1}{a} + \frac{1}{b}}}
\]

In the above formula, \(M_a\) and \(M_b\) are the average values for each of the data sets, \(N_a\) and \(N_b\) are the number of samples in each data set; \(X_{a,i}\) and \(X_{b,i}\) are the sums of the squares of each member of the data sets.

Another statistical test was used to measure the correlation between two sets of data under varying conditions. This was the Pearson product movement correlation, which gives a statistic "r" which can have any value between +1 and -1. A positive correlation means that as one of the data sets increases, the other increases proportionately. A negative correlation means that if one data set increases, the other decreases in proportion. A zero correlation means that there is no simple linear relationship between the two sets of variables.

R is calculated from the following formula:

\[
r = \frac{N \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{N \sum_{i=1}^{n} (X_i - \bar{X})^2 \sum_{i=1}^{n} (Y_i - \bar{Y})^2}}
\]
Where $X$ is the sum of the measurements under the first set of conditions; $Y$ is the sum of the measurements under the second set of conditions; $XY$ is the sum of the products of the measurements under each set of conditions; $N$ is the number of measurements in each set.

3 Experiments employing the segment-based phonetic representation.

The first three experiments to be described here investigate the temporal relations between segments. This involves the identification of segment boundaries and measurements of segment durations, insofar as this is possible. Later on in this chapter experiments using the continuous, non-segmental, contour-based phonetic representation will be described and the two representations can then be compared.

3.1 Experiment SE1

An experiment was designed to systematically investigate the temporal relationship between vowels and postvocalic consonants. This experiment applies the traditional segment-based phonetic representation. The following hypotheses were tested:

1. that the vowel before a voiced stop should be longer than the vowel before a voiceless stop;

2. that the closure of a voiceless stop should be longer than the closure of a voiceless stop;
3. that there will be a significant negative correlation between the vowel and stop closure duration in word pairs where the post-vocalic voicing changes;

4. that the above will be true with different place of consonant articulation and different vowel height and frontness and length. (Cf House and Fairbanks (1953) on the effect of different place of consonant articulation on vowel duration);

5. that the above described relationships will hold even if a syllable boundary comes between the vowel and the consonant.

Experiment design

Thirty-six test words with the form CVC were chosen such that the consonants /p,b,t,d,k,g/ and the vowels /i:,i,ə:,ʌ,ə:/ occurred in all possible combinations in the VC positions. In order to have a complete set of all the combinations of V and C, some of the test words were not actual words of English, although all were possible English words. The test words were as follows:

beep, beeh, beat, bead, beak, beag
dip, dib, bit, bid, dick, dig
burp, herb, Bert, bird, berk, burg
pup, pub, but, bud, buck, bug
tap, tab, bat, bad, back, bag
barp, barb, bart, bard, bark, barg

This choice of words means that the fourth part of the hypothesis can be tested. The words were set into a
frame sentence: It's ........ I'm saying now. The use of a frame helps the informants to maintain a more constant rate of utterance than would otherwise be possible. The frame sentence was designed to ensure release of the final stops of the test words. Since all the test words began with stop consonants, the preceding /s/ of the frame sentence could be expected to have no special influence on any test word. The sentences were recorded in random order in lists of 24 (that is, each list contained 12 pairs of test words). Each list began with 3 dummy sentences and ended with 4 to overcome any temporal effects associated with the beginning and end of lists.

Two informants, F.S. and A.R. recorded four examples of each sentence. The sound and acoustic recordings were made as described above. The vowel and final consonant closure durations were measured for each example of each word. Most of the test words began with /b/ and one pair with /d/. In these words segmentation was no problem. One pair began with /t/ and one with /p/; in these cases the vowel was measured from the onset of voicing. This was perhaps not a satisfactory solution since these words did not have the same timing patterns as those which began with voiced stops.

This whole procedure was then repeated for a second set of words of the structure CVCV, with an unstressed vowel (/ə/ or /œ/) for example beater/beader. These words were included to investigate the effect of a syllable boundary on the VC relationship. They were embedded in the frame: This word's ........ . The frame was changed because these bisyllabic test words need no following vowel to ensure release of the post-vocalic
stop, and such a following vowel in the frame would perhaps make the informants pronounce an /r/ in those test words which end in an orthographic <r>. It is however unclear exactly where the syllable boundary should come in words such as beater. The /t/ can be taken as belonging to the first or second syllable, or even as belonging to both syllables simultaneously.

Results

To test points one and two of the hypothesis, t-tests were taken to compare the duration of the vowels before voiceless consonants with the same vowels as they occurred before voiced consonants, and the durations of the voiceless consonants with the voiced consonants. The following results were obtained. The vowel and consonant tests are one-tailed tests, while the V+C tests are two-tailed.

Table 4.1

<table>
<thead>
<tr>
<th>Informant F.S. ;VC</th>
<th>T-tests</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Word</td>
</tr>
<tr>
<td>bee b v beeb</td>
<td>-2.3364</td>
</tr>
<tr>
<td>bee b v b eed</td>
<td>-10.145</td>
</tr>
<tr>
<td>bee k v beeg</td>
<td>-12.465</td>
</tr>
<tr>
<td></td>
<td>dip v dib</td>
</tr>
<tr>
<td></td>
<td>hit v hid</td>
</tr>
<tr>
<td></td>
<td>dick v dig</td>
</tr>
<tr>
<td></td>
<td>burp v  b eeb</td>
</tr>
<tr>
<td></td>
<td>bar t v  bixi</td>
</tr>
<tr>
<td></td>
<td>beak v burg</td>
</tr>
<tr>
<td></td>
<td>pup v pub</td>
</tr>
<tr>
<td></td>
<td>hit v hul</td>
</tr>
<tr>
<td></td>
<td>buck v hug</td>
</tr>
<tr>
<td>Task</td>
<td>Word 1</td>
</tr>
<tr>
<td>------------</td>
<td>--------</td>
</tr>
<tr>
<td>tap v tab</td>
<td>-7.2898</td>
</tr>
<tr>
<td>hat v bead</td>
<td>-5.826</td>
</tr>
<tr>
<td>back v leg</td>
<td>-10.846</td>
</tr>
<tr>
<td>harr v hab</td>
<td>-9.9420</td>
</tr>
<tr>
<td>bart v hand</td>
<td>-4.4653</td>
</tr>
<tr>
<td>back v barg</td>
<td>-4.2079</td>
</tr>
<tr>
<td>all</td>
<td>-6.7344</td>
</tr>
</tbody>
</table>

**Informant A.R.**; VC T-tests

<table>
<thead>
<tr>
<th>Task</th>
<th>Word 1</th>
<th>Word 2</th>
<th>p(t)</th>
<th>NS</th>
<th>p(t)</th>
</tr>
</thead>
<tbody>
<tr>
<td>harr v hab</td>
<td>-3.8332</td>
<td>7.3054</td>
<td>p(t)&lt;0.0005</td>
<td>-1.543</td>
<td>p(t) NS</td>
</tr>
<tr>
<td>bart v hand</td>
<td>-8.5205</td>
<td>3.1622</td>
<td>p(t)&lt;0.01</td>
<td>-1.797</td>
<td>p(t) NS</td>
</tr>
<tr>
<td>back v leg</td>
<td>-1.1522</td>
<td>2.1636</td>
<td>p(t)&lt;0.05</td>
<td>-0.0842</td>
<td>p(t) NS</td>
</tr>
<tr>
<td>dip v dib</td>
<td>-1.4828</td>
<td>1.6712</td>
<td>p(t) NS</td>
<td>0.1753</td>
<td>p(t) NS</td>
</tr>
<tr>
<td>bit v bid</td>
<td>-1.8820</td>
<td>3.4292</td>
<td>p(t)&lt;0.01</td>
<td>2.5011</td>
<td>p(t)&lt;0.05</td>
</tr>
<tr>
<td>click v dig</td>
<td>-0.2401</td>
<td>2.7457</td>
<td>p(t)&lt;0.0025</td>
<td>0.6535</td>
<td>p(t) NS</td>
</tr>
<tr>
<td>harr v hab</td>
<td>-2.1371</td>
<td>3.6055</td>
<td>p(t)&lt;0.01</td>
<td>0</td>
<td>p(t) NS</td>
</tr>
<tr>
<td>bart v hand</td>
<td>-1.0141</td>
<td>2.8465</td>
<td>p(t)&lt;0.0025</td>
<td>0</td>
<td>p(t) NS</td>
</tr>
<tr>
<td>back v leg</td>
<td>-4.2066</td>
<td>3.1304</td>
<td>p(t)&lt;0.0025</td>
<td>0.9496</td>
<td>p(t) NS</td>
</tr>
<tr>
<td>pup v pub</td>
<td>0.1873</td>
<td>2.1495</td>
<td>p(t)&lt;0.05</td>
<td>1.6829</td>
<td>p(t) NS</td>
</tr>
<tr>
<td>bit v bid</td>
<td>-1.4111</td>
<td>1.5126</td>
<td>p(t) NS</td>
<td>0</td>
<td>p(t) NS</td>
</tr>
<tr>
<td>back v leg</td>
<td>-6.5996</td>
<td>1.7822</td>
<td>p(t)&lt;0.005</td>
<td>-2.1259</td>
<td>p(t) NS</td>
</tr>
<tr>
<td>tap v tab</td>
<td>-2.6945</td>
<td>2.0604</td>
<td>p(t)&lt;0.0025</td>
<td>-2.079</td>
<td>p(t) NS</td>
</tr>
<tr>
<td>hat v bead</td>
<td>-2.6396</td>
<td>2.7331</td>
<td>p(t)&lt;0.0025</td>
<td>-0.5301</td>
<td>p(t) NS</td>
</tr>
<tr>
<td>back v leg</td>
<td>-2.1669</td>
<td>3.7936</td>
<td>p(t)&lt;0.005</td>
<td>0.6790</td>
<td>p(t) NS</td>
</tr>
<tr>
<td>harr v hab</td>
<td>-3.5826</td>
<td>2.2000</td>
<td>p(t)&lt;0.05</td>
<td>-2.1143</td>
<td>p(t) NS</td>
</tr>
<tr>
<td>bart v hand</td>
<td>-2.6349</td>
<td>3.3946</td>
<td>p(t)&lt;0.001</td>
<td>-0.4649</td>
<td>p(t) NS</td>
</tr>
<tr>
<td>back v barg</td>
<td>-6.4001</td>
<td>3.0644</td>
<td>p(t)&lt;0.0025</td>
<td>-1.3206</td>
<td>p(t) NS</td>
</tr>
<tr>
<td>all</td>
<td>-3.8260</td>
<td>10.608</td>
<td>p(t)&lt;0.0005</td>
<td>-0.3976</td>
<td>p(t) NS</td>
</tr>
</tbody>
</table>

**Informant P.S.**; CV T-tests

<table>
<thead>
<tr>
<th>Task</th>
<th>Word 1</th>
<th>Word 2</th>
<th>p(t)</th>
<th>NS</th>
<th>p(t)</th>
</tr>
</thead>
<tbody>
<tr>
<td>harr v hab</td>
<td>-4.6358</td>
<td>3.6288</td>
<td>p(t)&lt;0.005</td>
<td>0.2175</td>
<td>p(t) NS</td>
</tr>
<tr>
<td>bart v hand</td>
<td>-13.498</td>
<td>3.6055</td>
<td>p(t)&lt;0.01</td>
<td>1.1344</td>
<td>p(t) NS</td>
</tr>
<tr>
<td>back v leg</td>
<td>-6.8355</td>
<td>3.1305</td>
<td>p(t)&lt;0.0025</td>
<td>0.7333</td>
<td>p(t) NS</td>
</tr>
<tr>
<td>dip v dib</td>
<td>-3.4167</td>
<td>2.7791</td>
<td>p(t)&lt;0.005</td>
<td>0.2667</td>
<td>p(t) NS</td>
</tr>
<tr>
<td>bit v bid</td>
<td>-3.0000</td>
<td>2.3763</td>
<td>p(t)&lt;0.005</td>
<td>0.1525</td>
<td>p(t) NS</td>
</tr>
<tr>
<td>click v dig</td>
<td>-3.0833</td>
<td>4.3994</td>
<td>p(t)&lt;0.0005</td>
<td>-0.1705</td>
<td>p(t) NS</td>
</tr>
<tr>
<td>harr v hab</td>
<td>-3.2002</td>
<td>5.7446</td>
<td>p(t)&lt;0.005</td>
<td>0.1877</td>
<td>p(t) NS</td>
</tr>
<tr>
<td>bart v hand</td>
<td>-1.7963</td>
<td>1.0000</td>
<td>p(t) NS</td>
<td>0.1656</td>
<td>p(t) NS</td>
</tr>
<tr>
<td>back v leg</td>
<td>-3.2857</td>
<td>3.1529</td>
<td>p(t)&lt;0.01</td>
<td>0.1641</td>
<td>p(t) NS</td>
</tr>
<tr>
<td>pup v pub</td>
<td>-2.0997</td>
<td>2.5923</td>
<td>p(t)&lt;0.005</td>
<td>-0.3659</td>
<td>p(t) NS</td>
</tr>
<tr>
<td>bit v bid</td>
<td>-0.1706</td>
<td>0.2425</td>
<td>p(t) NS</td>
<td>0</td>
<td>p(t) NS</td>
</tr>
<tr>
<td>back v leg</td>
<td>-1.7597</td>
<td>0</td>
<td>p(t) NS</td>
<td>0.2356</td>
<td>p(t) NS</td>
</tr>
<tr>
<td>Word</td>
<td>t-value</td>
<td>p-value</td>
<td>NS</td>
<td>t-value</td>
<td>p-value</td>
</tr>
<tr>
<td>---------------------</td>
<td>---------</td>
<td>---------</td>
<td>----</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>tapper v tabber</td>
<td>-1.7955</td>
<td>1.3887</td>
<td>0.2942</td>
<td>p(t) NS</td>
<td>0.2942</td>
</tr>
<tr>
<td>hatter v hadder</td>
<td>-1.8773</td>
<td>2.0702</td>
<td>0.0200</td>
<td>p(t)&lt;0.05</td>
<td>0.0200</td>
</tr>
<tr>
<td>backer v bagger</td>
<td>-2.4065</td>
<td>1.8999</td>
<td>0.0889</td>
<td>p(t) NS</td>
<td>0.0889</td>
</tr>
<tr>
<td>harper v hadder</td>
<td>-2.3816</td>
<td>2.2360</td>
<td>0.0719</td>
<td>p(t)&lt;0.05</td>
<td>0.0719</td>
</tr>
<tr>
<td>hatter v hadder</td>
<td>-5.1418</td>
<td>5.5549</td>
<td>0.1461</td>
<td>p(t)&lt;0.005</td>
<td>0.1461</td>
</tr>
<tr>
<td>backer v hagger</td>
<td>-3.9302</td>
<td>2.0000</td>
<td>0.0884</td>
<td>p(t)&lt;0.005</td>
<td>0.0884</td>
</tr>
<tr>
<td>all words</td>
<td>-3.9264</td>
<td>9.3704</td>
<td>0.3805</td>
<td>p(t)&lt;0.0005</td>
<td>0.3805</td>
</tr>
</tbody>
</table>

**In all cases vowels before a voiceless stop are significantly shorter than those before a voiced stop, while voiceless stops are seen to have a significantly longer closure duration than voiced stops. There are more individual non-significant results for CVCV words than for CVC words. The syllable boundary between the V and the C seems to block the relationship that existed between them in CVC words to some extent. For F.S. both the V and the C are less likely to have significantly different durations when the voicing of the C is changed in CVCV words, while for A.R. the pre-boundary vowels**
are less likely to vary their duration than are the following consonants.

In order to test point three of the hypothesis, the correlation of the vowel duration to the stop closure duration was measured for all the vowels and consonants in each word pair, the following results were obtained. The 5% significance level was used.

Table 4.2

Informant F.S.; VC correlations

\[
\begin{array}{cccc}
\text{beep} & -0.669 & \text{beat} & -0.931 \\
\text{dip} & -0.416 & \text{bit} & -0.957 \\
\text{larp} & -0.759 & \text{bart} & -0.523 \\
\text{pup} & -0.749 & \text{burt} & -0.822 \\
\text{tap} & -0.932 & \text{bat} & -0.606 \\
\text{larp} & -0.814 & \text{bart} & -0.867 \\
\text{all words: } r & -0.425 & p(r)<0.0005
\end{array}
\]

Informant A.R.; VC correlations

\[
\begin{array}{cccc}
\text{beeb} & -0.756 & \text{bead} & -0.856 \\
\text{dib} & -0.005 & \text{bid} & -0.840 \\
\text{lurb} & -0.512 & \text{birt} & -0.087 \\
\text{pob} & 0.050 & \text{bud} & 0.871 \\
\text{tab} & -0.061 & \text{bat} & -0.270 \\
\text{lurb} & -0.710 & \text{bard} & -0.688 \\
\text{all words: } r & -0.201 & p(r)<0.025
\end{array}
\]

Informant F.S. VC correlations (CVV words)

\[
\begin{array}{cccc}
\text{beeder} & -0.764 & \text{beeder} & -0.888 \\
\text{ditteber} & -0.924 & \text{bittle} & -0.815 \\
\text{berber} & -0.833 & \text{britle} & -0.112 \\
\text{pubby} & -0.900 & \text{hitter} & 0.363 \\
\text{tabber} & -0.381 & \text{beeder} & -0.785 \\
\text{berber} & -0.833 & \text{banger} & -0.916 \\
\text{all words: } r & -0.318 & p(r)<0.005
\end{array}
\]
Informer A.R. VC correlations (CVC words)

<table>
<thead>
<tr>
<th></th>
<th>r</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>heeber</td>
<td>-0.523</td>
<td>p(r) NS</td>
<td></td>
</tr>
<tr>
<td>header</td>
<td>-0.569</td>
<td>p(r) NS</td>
<td></td>
</tr>
<tr>
<td>digger</td>
<td>-0.799</td>
<td>p(r)&lt;0.025</td>
<td></td>
</tr>
<tr>
<td>Buffer</td>
<td>-0.849</td>
<td>p(r)&lt;0.005</td>
<td></td>
</tr>
<tr>
<td>bagger</td>
<td>0.361</td>
<td>p(r) NS</td>
<td></td>
</tr>
<tr>
<td>phoney</td>
<td>0.029</td>
<td>p(r) NS</td>
<td></td>
</tr>
<tr>
<td>hobber</td>
<td>-0.209</td>
<td>p(r) NS</td>
<td></td>
</tr>
<tr>
<td>tacker</td>
<td>-0.756</td>
<td>p(r)&lt;0.025</td>
<td></td>
</tr>
<tr>
<td>haller</td>
<td>-0.548</td>
<td>p(r) NS</td>
<td></td>
</tr>
<tr>
<td>beller</td>
<td>-0.534</td>
<td>p(r) NS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>all words: r=-0.086 p(r) NS</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Except for A.R.'s CVCV words there is a significant negative correlation between the vowels and consonants. This means that as the vowel gets longer, as it does before a voiced consonant, the consonant associated with it gets shorter.

To test point five of the hypothesis, and investigate the effects of the syllable boundary between the V and C in CVCV words, t-tests were taken to establish whether there is a significant difference between the V durations and the C durations as they occur in each type of word. The following results were obtained:

<table>
<thead>
<tr>
<th></th>
<th>VC correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Informant F.S.</td>
<td>Informant A.R.</td>
</tr>
<tr>
<td>VC vowel v VC vowel: t=7.1222 p(t)&lt;0.0005</td>
<td>t=3.9703 p(t)&lt;0.0005</td>
</tr>
<tr>
<td>VC stop v VC stop: t=3.2182 p(t)&lt;0.005</td>
<td>t=-0.319 p(t) NS</td>
</tr>
</tbody>
</table>

This means that for both informants the vowel is shorter in CVCV words, but the consonant is shorter only for F.S. This could explain the lack of a significant negative correlation between CVCV consonants and vowels for A.R.

3.2 Experiment SE2

Another experiment was designed to investigate the durational properties of unstressed vowels followed by voiced and voiceless stop consonants or fricatives. The intention is to establish whether differences are main-
tained between the vowel and or consonant durations in unstressed syllables, as is the case for stressed syllables (cf experiment SE1). Port, Al-Ani and Maeda (1980) claim that compensation in English is restricted to stressed syllables. This claim will be tested here.

Hypotheses: 1. that the vowel of an unstressed syllable is longer if followed by a voiced consonant than if followed by a voiceless consonant;

2. that voiced consonants are longer than voiceless consonants;

3. that there is a negative correlation between vowels and consonants when both voiceless and voiced syllables are taken.

Experiment design

Two informants, T.L. and L.K. each recorded 10 examples of each of 9 pairs of utterances, composed of a frame sentence and a test sequence. The recordings were made as described above. The frame sentence had the form: I say ........ every day. The test sequences had a stressed syllable followed by an unstressed syllable. In three cases, the words ended in a voiceless v voiced apical fricative, and the other forms were of the form CV,CVC v CVC.VC where ',' is a syllable boundary and '.' is a word boundary, and the final C was a voiceless v voiced alveolar stop. There are not many true minimal pairs in English where a voicing contrast operates in an unstressed syllable, but this should not be important. In experiment SE1 it was shown that the inverse relationship between V and C durations holds fairly well
over syllable boundaries, while Lehiste (1972) established that the same is true for morpheme and word boundaries. The following pairs of test sequences were used:

shyness, shiners; coyness, coiners; handless, handlers; head it, beaded; pad it, padded; boot it, booted; part it, parted; bed it, bedded; dot it, dotted.

The sentences were recorded as described above, in random order in lists of 18, whereof the first three and the last three were dummy sentences included to overcome any effects associated with the beginning and end of lists. Three pairs of test sentences each occurred twice in each list, which the informants read five times, to give the ten examples of each sentence.

The durations of the unstressed vowels, the consonants that followed them and the CVCVC sequence in which they occurred were measured. The vowels were measured from the release of the preceding stop or the change in intensity readings when the vowel was preceded by something other than a stop. The final stops were measured from closure to release, and for final fricatives, the duration of friction was measured.

Results

In order to test the hypothesis, the vowels were examined, comparing the durations of those followed by a voiceless consonant with those followed by a voiced consonant. The hypothesis claims that the former group will be longer. T-tests gave the following results:
Table 4.4

UNSTRESSED VOWELS in:

<table>
<thead>
<tr>
<th>Informant</th>
<th>T.L.</th>
<th>L.K.</th>
</tr>
</thead>
<tbody>
<tr>
<td>shyness v shiners</td>
<td>t=-0.7093 p(t) NS</td>
<td>t=0.1588 p(t) NS</td>
</tr>
<tr>
<td>bead it v headed</td>
<td>t=0</td>
<td>p(t) NS</td>
</tr>
<tr>
<td>pad it v patted</td>
<td>t=-1.6330 p(t) NS</td>
<td>t=-2.0604 p(t) &lt;0.05</td>
</tr>
<tr>
<td>coyness v coiners</td>
<td>t=1.03237 p(t) NS</td>
<td>t=0.5896 p(t) NS</td>
</tr>
<tr>
<td>look it v looked</td>
<td>t=0.81818 p(t) NS</td>
<td>t=-2.769 p(t) &lt;0.01</td>
</tr>
<tr>
<td>part it v parted</td>
<td>t=1.32690 p(t) NS</td>
<td>t=2.7330 p(t) &lt;0.01</td>
</tr>
<tr>
<td>handless v handlers</td>
<td>t=-0.8050 p(t) NS</td>
<td>t=-1.1404 p(t) NS</td>
</tr>
<tr>
<td>head it v headed</td>
<td>t=0.51449 p(t) NS</td>
<td>t=0.2835 p(t) NS</td>
</tr>
<tr>
<td>dot it v dotted</td>
<td>t=-0.8955 p(t) NS</td>
<td>t=-0.7093 p(t) NS</td>
</tr>
</tbody>
</table>

All unstressed vowels, preceding voiceless v voiced consonants:

<table>
<thead>
<tr>
<th>Informant</th>
<th>T.L.</th>
<th>L.K.</th>
</tr>
</thead>
<tbody>
<tr>
<td>t=0</td>
<td>p(t) NS</td>
<td>t=0.1718 p(t) NS</td>
</tr>
</tbody>
</table>

Clearly the hypothesis is not corroborated on this point. There does not seem to be any kind of systematic significant vowel lengthening before a voiced consonant.

Given this situation for unstressed vowels, compensation for the vowel duration cannot be a reason to expect the voiced consonants to be shorter than the voiceless consonants. However, the following results were obtained when t-tests were calculated:

Table 4.5

FINAL CONSONANTS in:

<table>
<thead>
<tr>
<th>Informant</th>
<th>T.L.</th>
<th>L.K.</th>
</tr>
</thead>
<tbody>
<tr>
<td>shyness v shiners</td>
<td>t=4.2302 p(t)&lt;0.005</td>
<td>t=2.1548 p(t)&lt;0.025</td>
</tr>
<tr>
<td>bead it v headed</td>
<td>t=2.8461 p(t)&lt;0.01</td>
<td>t=1.6330 p(t) NS</td>
</tr>
<tr>
<td>pad it v patted</td>
<td>t=0.6124 p(t) NS</td>
<td>t=2.6886 p(t)&lt;0.01</td>
</tr>
<tr>
<td>coyness v coiners</td>
<td>t=2.6349 p(t)&lt;0.01</td>
<td>t=8.3206 p(t)&lt;0.0005</td>
</tr>
<tr>
<td>look it v looked</td>
<td>t=1.3416 p(t) NS</td>
<td>t=0</td>
</tr>
<tr>
<td>part it v parted</td>
<td>t=1.524 p(t) NS</td>
<td>t=2.948 p(t)&lt;0.005</td>
</tr>
<tr>
<td>handless v handlers</td>
<td>t=3.8376 p(t)&lt;0.005</td>
<td>t=4.5646 p(t)&lt;0.0005</td>
</tr>
<tr>
<td>head it v headed</td>
<td>t=2.9104 p(t)&lt;0.005</td>
<td>t=1.3416 p(t) NS</td>
</tr>
<tr>
<td>dot it v dotted</td>
<td>t=3.2797 p(t)&lt;0.005</td>
<td>t=0.5721 p(t) NS</td>
</tr>
</tbody>
</table>

All voiceless v voiced consonants:

<table>
<thead>
<tr>
<th>Informant</th>
<th>T.L.</th>
<th>L.K.</th>
</tr>
</thead>
<tbody>
<tr>
<td>t=5.1514 p(t)&lt;0.0005</td>
<td>t=3.8641 p(t)&lt;0.0005</td>
<td></td>
</tr>
</tbody>
</table>
The overall result shows that for both informants the voiceless consonant is longer than the voiced one.

When the correlations of the vowel to the consonant duration were measured, the following results were obtained:

Table 4.6

<table>
<thead>
<tr>
<th>UNSTRESSED V v C in:</th>
<th>INFORMANT T.L.</th>
<th>INFORMANT L.K.</th>
</tr>
</thead>
<tbody>
<tr>
<td>shyness &amp; shiners:</td>
<td>r=0.0806 p(r) NS</td>
<td>r=0.3831 p(r) NS</td>
</tr>
<tr>
<td>head it &amp; heeded:</td>
<td>r=0.1592 p(r) NS</td>
<td>r=-0.3068 p(r) NS</td>
</tr>
<tr>
<td>pad it &amp; pedaled:</td>
<td>r=-0.0769 p(r) NS</td>
<td>r=-0.5412 p(r) NS</td>
</tr>
<tr>
<td>coyness &amp; coiners:</td>
<td>r=0.0416 p(r) NS</td>
<td>r=-0.1226 p(r) NS</td>
</tr>
<tr>
<td>boot it &amp; booted:</td>
<td>r=0.3742 p(r) NS</td>
<td>r=0.3299 p(r) NS</td>
</tr>
<tr>
<td>part it &amp; parted:</td>
<td>r=0.6727 p(r) NS</td>
<td>r=0.3245 p(r) NS</td>
</tr>
<tr>
<td>hardless &amp; handled:</td>
<td>r=-0.0694 p(r) NS</td>
<td>r=-0.0146 p(r) NS</td>
</tr>
<tr>
<td>bed it &amp; bedded:</td>
<td>r=0.1702 p(r) NS</td>
<td>r=-0.3464 p(r) NS</td>
</tr>
<tr>
<td>dot it &amp; dotted:</td>
<td>r=-0.4752 p(r)0.025</td>
<td>r=-0.3378 p(r) NS</td>
</tr>
<tr>
<td>All pairs:</td>
<td>r=0.1769 p(r) NS</td>
<td>r=-0.0467 p(r) NS</td>
</tr>
</tbody>
</table>

There is no significant negative correlation as was predicted by the hypothesis. This is not unexpected given the lack of a difference between the vowel durations. It seems that the timing rules of stressed syllables do not apply to unstressed syllables.

3.3 Experiment SE3

Another experiment was designed to investigate the temporal relationships between vowels and post-vocalic voiced v voiceless fricatives, to see if the same relationship holds as for post-vocalic stops. The following hypotheses were tested:
1. that the vowels preceding voiced fricatives are longer than those preceding voiceless fricatives;

2. that voiced fricatives have shorter duration than voiceless fricatives;

3. that the above two durational differences are related in such a way that there will be a negative correlation between vowel and fricative durations;

Experiment design

Material: Two informants, T.L. and L.K. each recorded 10 examples of each of 9 pairs of words which contrasted the voicing feature in post-vocalic fricatives. The words were embedded in frame sentences of the form: I say ..... every day. Different fricatives and vowels were tested, and all words were of the form: stop or /s/, vowel, fricative, except for one pair: mesher/measure. The word pairs tested were as follows:

Piece, peas; surf, serve; teeth, teethe; dose, doze; mesher, measure; duff, dove; bus, buzz; safe, save; sheath, sheathe.

The durations of the vowels were measured from the onset of voicing for words beginning with a voiceless stop or /s/ and from the release of the stop for words beginning with a voiced stop. Fricatives were measured from the onset of friction.
Results

Firstly the vowels were examined, comparing the durations of those vowels followed by a voiceless fricative with those followed by a voiced fricative. The hypothesis predicts that the latter category will be the longer. T-tests were performed on the vowels as they occurred individually by pairs and for all tokens of all words. The results of the tests are shown in the following table:

Table 4.7

<table>
<thead>
<tr>
<th>VOWELS in:</th>
<th>INFORMANT</th>
<th>T.L.</th>
<th>INFORMANT</th>
<th>T.L.</th>
</tr>
</thead>
<tbody>
<tr>
<td>piece v peas:</td>
<td>t=-5.0952</td>
<td>p(t)&lt;0.0005</td>
<td>t=-5.9999</td>
<td>p(t)&lt;0.0005</td>
</tr>
<tr>
<td>surf v serve:</td>
<td>t=-2.3985</td>
<td>p(t)&lt;0.025</td>
<td>t=-4.2960</td>
<td>p(t)&lt;0.0005</td>
</tr>
<tr>
<td>teeth v teeth:</td>
<td>t=-5.9764</td>
<td>p(t)&lt;0.0005</td>
<td>t=-8.4604</td>
<td>p(t)&lt;0.0005</td>
</tr>
<tr>
<td>close v close:</td>
<td>t=-4.4471</td>
<td>p(t)&lt;0.0005</td>
<td>t=-9.3244</td>
<td>p(t)&lt;0.0005</td>
</tr>
<tr>
<td>mesh er v measure:</td>
<td>t=0</td>
<td>p(t) NS</td>
<td>t=3.8820</td>
<td>p(t)&lt;0.0005</td>
</tr>
<tr>
<td>cluff v close:</td>
<td>t=0</td>
<td>p(t) NS</td>
<td>t=6.1482</td>
<td>p(t)&lt;0.0005</td>
</tr>
<tr>
<td>bus v buzz:</td>
<td>t=3.8341</td>
<td>p(t)&lt;0.0005</td>
<td>t=11.001</td>
<td>p(t)&lt;0.0005</td>
</tr>
<tr>
<td>safe v save:</td>
<td>t=8.7198</td>
<td>p(t)&lt;0.0005</td>
<td>t=8.1744</td>
<td>p(t)&lt;0.0005</td>
</tr>
<tr>
<td>sheath v sheath:</td>
<td>t=-11.005</td>
<td>p(t)&lt;0.0005</td>
<td>t=-5.8936</td>
<td>p(t)&lt;0.0005</td>
</tr>
</tbody>
</table>

All words, preceding voiced v voiceless fricative:

<table>
<thead>
<tr>
<th></th>
<th>INFORMANT</th>
<th>T.L.</th>
<th>INFORMANT</th>
<th>T.L.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>t=6.1462</td>
<td>p(t)&lt;0.0005</td>
<td>t=10.339</td>
<td>p(t)&lt;0.0005</td>
</tr>
</tbody>
</table>

This overall result shows the vowels preceding voiced fricatives to be significantly longer than those preceding voiceless fricatives. The same tests were performed on the fricatives and the results are shown in the following table:

Table 4.8

<table>
<thead>
<tr>
<th>FRICATIVES in:</th>
<th>INFORMANT</th>
<th>T.L.</th>
<th>INFORMANT</th>
<th>T.L.</th>
</tr>
</thead>
<tbody>
<tr>
<td>place v peas:</td>
<td>t=1.4816</td>
<td>p(t) NS</td>
<td>t=13.1663</td>
<td>p(t)&lt;0.0005</td>
</tr>
<tr>
<td>surf v serve:</td>
<td>t=1.06274</td>
<td>p(t) NS</td>
<td>t=8.81964</td>
<td>p(t)&lt;0.0005</td>
</tr>
<tr>
<td>teeth v teeth:</td>
<td>t=0.1506</td>
<td>p(t) NS</td>
<td>t=8.3307</td>
<td>p(t)&lt;0.0005</td>
</tr>
<tr>
<td>close v close:</td>
<td>t=1.68676</td>
<td>p(t) NS</td>
<td>t=7.51955</td>
<td>p(t)&lt;0.0005</td>
</tr>
<tr>
<td>mesh er v measure</td>
<td>t=1.5492</td>
<td>p(t) NS</td>
<td>t=6.84111</td>
<td>p(t)&lt;0.0005</td>
</tr>
<tr>
<td>cluff v close:</td>
<td>t=4.0379</td>
<td>p(t)&lt;0.0005</td>
<td>t=5.52346</td>
<td>p(t)&lt;0.0005</td>
</tr>
<tr>
<td>bus v buzz:</td>
<td>t=12.7281</td>
<td>p(t)&lt;0.0005</td>
<td>t=10.2862</td>
<td>p(t)&lt;0.0005</td>
</tr>
<tr>
<td>safe v save:</td>
<td>t=4.84616</td>
<td>p(t)&lt;0.0005</td>
<td>t=8.7207</td>
<td>p(t)&lt;0.0005</td>
</tr>
<tr>
<td>sheath v sheath</td>
<td>t=3.58395</td>
<td>p(t)&lt;0.0005</td>
<td>t=21.5692</td>
<td>p(t)&lt;0.0005</td>
</tr>
</tbody>
</table>

All words, voiced v voiceless fricative:

<table>
<thead>
<tr>
<th></th>
<th>INFORMANT</th>
<th>T.L.</th>
<th>INFORMANT</th>
<th>T.L.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>t=6.50859</td>
<td>p(t)&lt;0.0005</td>
<td>t=21.5692</td>
<td>p(t)&lt;0.0005</td>
</tr>
</tbody>
</table>
The hypothesis is clearly corroborated for both informants in the overall results. The voiceless fricatives are significantly longer than the voiced ones. For L.K. this is always true, while for T.L. the fricatives in piece v peas, surf v serve, teeth v teethe, dose v doze and mesher v measure are not significantly different in duration. This does not seem to be anything to do with the particular fricatives involved, since except for mesher/measure the fricatives all show the expected longer duration when voiced in other words. Similarly in the vowels, the vowel in duff v dove which was not significantly shorter for T.L. was significantly longer in buzz than in bus, although it may be relevant that the voiceless fricatives do not seem to become significantly shorter in five out of the seven cases with long vowels.

So far then, it seems that the hypothesis predicts durational differences between the pairs of test words accurately for L.K., but less accurately for T.L.. The next part of the hypothesis predicts that there will be a negative correlation between the vowels and the fricatives when all vowels are compared with all fricatives. The following results were obtained when the correlations were calculated:

Table 4.9

<table>
<thead>
<tr>
<th>CORRELATION OF VOWELS TO FRICATIVES in:</th>
<th>INCUMENT T.L.</th>
<th>INCUMENT L.K.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Piece &amp; peas: r=0.2946 p(r) NS</td>
<td>r=0.7972 p(r)&lt;0.0005</td>
<td></td>
</tr>
<tr>
<td>Surf &amp; serve: r=0.2730 p(r) NS</td>
<td>r=0.8419 p(r)&lt;0.0005</td>
<td></td>
</tr>
<tr>
<td>Teeth &amp; teethe: r=0.0206 p(r) NS</td>
<td>r=0.8200 p(r)&lt;0.0005</td>
<td></td>
</tr>
<tr>
<td>Dose &amp; doze: r=0.6897 p(r)&lt;0.0005</td>
<td>r=0.8219 p(r)&lt;0.0005</td>
<td></td>
</tr>
<tr>
<td>Mesher &amp; measure: r=0.0192 p(r) NS</td>
<td>r=0.3971 p(r)&lt;0.05</td>
<td></td>
</tr>
<tr>
<td>Diff &amp; dove: r=0.0418 p(r) NS</td>
<td>r=0.6434 p(r)&lt;0.005</td>
<td></td>
</tr>
<tr>
<td>Bus &amp; buzz: r=0.7319 p(r)&lt;0.0005</td>
<td>r=0.9146 p(r)&lt;0.0005</td>
<td></td>
</tr>
<tr>
<td>Safe &amp; save: r=0.8515 p(r)&lt;0.0005</td>
<td>r=0.8205 p(r)&lt;0.0005</td>
<td></td>
</tr>
<tr>
<td>Sheeth &amp; sheather: r=0.7435 p(r)&lt;0.0005</td>
<td>r=0.7225 p(r)&lt;0.0005</td>
<td></td>
</tr>
<tr>
<td>All words, with voiced v voiceless fricative: r=0.2899 p(r)&lt;0.005</td>
<td>r=0.5096 p(r)&lt;0.0005</td>
<td></td>
</tr>
</tbody>
</table>
Clearly then, there is a significant overall negative correlation between the vowels and the fricatives for both informants. The non-significant results in some cases for T.L. are due to some fricatives failing to shorten when voiced as was shown in table 2.8.

Data from L.K.'s speech corroborated the hypothesis on every point. The overall results for T.L. were also corroborative, but there was also something else happening at the level of individual word pairs; the voiced fricative sometimes does not shorten very much after a long vowel, even though the vowel itself is usually longer before a voiced fricative.

4 Experiments employing the contour-based representation.

Experiments SE1, SE2 and SE3 give a fairly clear picture of the temporal relationships between vowels and the consonants following them. All the above described experiments were conducted using a segment-based phonetic representation. This is not the only way of looking at the temporal relationships between adjacent phonesized phonological segments. The question must be raised as to whether this approach is fully explanatory. It may be that the contour-based phonetic representation described in chapter three can offer improved insights into the way in which the timing relationships between phoneme-sized phonological segments work. The following experiments apply the contour-based representation.
4.1 Experiment SE4

The following utterances were recorded, and spectrograms were made:

1a  I say price every day
1b  I say prize every day
2a  I say beat every day
2b  I say bead every day

Timing contours were made of the b utterances (with voiced post-vocalic consonant) relative to the a utterances in the way described in chapter 3. The line-up points chosen were as follows:

Label  line-up point
I      beginning of utterance
s      end of first formant for diphthong
ay     end of friction for s
       end of second formant for diphthong

(For utterances 1a and 1b:
P      release of plosive
r      end of friction for r
i      end of second formant for diphthong
ce/ze  end of friction)

(For utterances 2a and 2b:
b      release of stop
ea     end of second formant for vowel
t/d    release of stop)
ev     beginning of formants after v
d      end of second formant
ay     release of stop
      end.

When the calculations were made the following contours were obtained:
Fig 4.1 Contour of I say prize every day relative to I say price every day.

Fig 4.2 Contour of I say head every day relative to I say beat every day.
The inverse relationship between vowel and post-vocalic consonant can be seen very clearly from the contours. Other peaks and troughs in the contour must be attributed to noise caused by the low number of examples of each utterance (2 examples). This method is also very useful in discussions of an isochronous VC unit (cf Port (1979)), since the area under the timing contour represents the time taken to utter the material, such that if the VC unit was isochronous, the trough representing the lower rate of the vowel and the peak for the higher rate of the consonant in utterances with a voiced post-vocalic consonant would be seen to be equal in area.

It is important to note that the intervals between line-up points are not segments. Often the line-up point may be chosen at a segment boundary, but this is simply because easily observable acoustic events happen mainly at segment boundaries. (For a further discussion of events at and around segment boundaries see chapter 6 on global rate.) It has been suggested that the inverse durational relationship between vowels and post-vocalic consonants has to do with the production of the segments, that this is a property of the VC unit. It has been suggested that this inverse relationship is an automatic effect of vocal tract acoustics, (cf Chen (1970), Kohler (1980), Parker (1974)). When a voiceless fricative is articulated, the vocal folds are abducted; this means that more air passes through than is the case for voiced fricatives, where the vocal folds are vibrating. For voiceless fricatives this means that the offset of voicing at the beginning of the voiceless fricative, is accompanied by an increase in airflow,
which causes an earlier onset of friction. If then the VF boundary is taken to be the onset of friction, the vowel before the voiceless fricative will end earlier than that before a voiced fricative, and so the fricative will begin earlier in the voiceless case. The question then arises as to what controls the end of the fricative. If the end is linked to the beginning of the fricative no inverse relationship between the vowel duration and the fricative duration would be expected. On the other hand, if the end of the fricative is linked to an earlier point, such as the onset of the previous vowel, the inverse relationship between the durations would be expected to hold, and the VF sequence would be isochronous through changes in voicing.

It is not difficult to imagine that something similar might happen with stop consonants too, the increased muscular tenseness of voiceless stops causing some stop closure command issued by the production process to be executed more quickly, thus making the vowel shorter and the stop closure longer than is the case for voiced stops. That this is indeed what happens is shown by the difference of rate of formant transitions between the voiceless and voiced stops; the transitions are more abrupt for voiceless stops (cf Lisker (1957), Santerre and Suen (1981)).

In both the fricative and stop case, it is the choice of segment boundary which causes the effect. A segment boundary is usually a collection of various acoustic events, such as offset of voicing, onset of friction, stop closure etc. But it is just this fact that makes it so difficult to segment speech (cf also discussion of coarticulation in chapter 2). When these acoustic events do not happen simultaneously, a choice must be made as
to which of the events will be taken as the segment boundary. This choice is apparently quite arbitrary. Using the contour-based phonetic representation the position is not so arbitrary. Inherent in the concept of timing contours is the assumption that if the line-up points were taken more closely together, the shape of the contour would be more or less the same. No claim is made about the significance of the line-up points of the material that happens to be cut off between them, in contrast to the segment-based approach which attaches great importance to the segment boundaries and expects temporal effects to be bounded by them. In addition, the contour-based representation shows that the compensatory effect of the adjacent rate changes in a way that a segmental approach cannot. The segment-based approach can only come up with findings relating to smaller or larger segments. Segment durations can be compared with each other; they can be averaged and statistical tests can be performed to test for differences and correlations between pairs of sets of measurements. This is quite informative in its way, but fails to give any insight into the way in which the timing of entire utterances or adjacent sections of speech works. The temporal influences that some parts of the utterance have on others can be seen, and new relationships discovered. It can be the case perhaps that a small change in the phonological structure of an utterance has much more far-reaching temporal consequences than would be expected. It is extremely unlikely that such effects would be discovered using the segment-based representation, while the contour-based representation exposes the whole picture neatly for inspection.

One way to test whether or not the inverse relation-
ship between the duration of vowels and post-vocalic consonants is part of the production process and therefore only due to the articulatory mechanism, or whether it is also part of the speaker's competence, is to see what happens when speakers of a language without a tense-lax contrast in post-vocalic consonants pronounce words such as those used in experiments SE1 and SE3.

4.2 Experiment SE5

This experiment was designed to test whether the inverse relationship between the durations of a vowel and a post vocalic +/-voice stop consonant (as established in experiment SE1) is a necessary consequence of vocal-tract aerodynamics, or if it is a learned feature, part of the speakers competence. In the latter case, it seems reasonable to suppose that non-native speakers would not have learned to use this cue consistently, and so logically, if the relationship does not occur in their native language, it should not be present in their English. If however, the relationship is a consequence of the way in which the vocal tract works, it should hold even for non-native speakers.

Hypothesis: that there will be an inverse relationship between the duration of vowels and post-vocalic voiced and voiceless stop consonants when native Spanish speakers read English sentences of the type:

I say beat every day

I say bead every day.

If the hypothesis is corroborated by the experiment,
this will indicate that the VC relationship is in some way an inherent part of the voicing distinction in post vocalic stops. In Spanish there is no voicing contrast in non-syllable initial post-vocalic stops. The voiceless stops /p/, /t/, /k/ occur in this position (eg captar, futbol, técnico), but /b/, /d/, /g/ have fricative allophones in this position.

Two native speakers of standard Castillan Spanish, M.P. and M.G.P., each recorded 10 tokens of twelve word pairs which had post-vocalic stop consonants contrasting a) voicing, (as in 

| beep | beeb | r = -0.8686 p(r) < 0.0005 | r = -0.2091 p(r) < NS  
|------|------|--------------------------|------------------------
| beat | bead | r = -0.7894 p(r) < 0.0005 | r = -0.5136 p(r) < 0.025  
| book | beak | r = -0.6779 p(r) < 0.005  | r = -0.3884 p(r) < 0.05  
| dip   | dib  | r = -0.6543 p(r) < 0.005  | r = -0.6847 p(r) < 0.0005  
| bit   | bid  | r = -0.9470 p(r) < 0.0005 | r = -0.5198 p(r) < 0.01  
| dick  | dig  | r = -0.6881 p(r) < 0.0005 | r = -0.2833 p(r) < NS  
| hump  | hobo | r = -0.8305 p(r) < 0.0005 | r = -0.4426 p(r) < 0.05  
| Bert  | birt | r = -0.6969 p(r) < 0.0005 | r = 0.0343 p(r) < NS  
| berk  | beg  | r = -0.5861 p(r) < 0.005  | r = -0.2253 p(r) < NS  
| pub   | pub  | r = -0.7587 p(r) < 0.0005 | r = -0.6939 p(r) < 0.0005  
| but   | bud  | r = -0.8315 p(r) < 0.0005 | r = -0.4055 p(r) < 0.025  
| buck  | bug  | r = -0.7316 p(r) < 0.0005 | r = -0.3824 p(r) < 0.05  

The test words were embedded in frame sentences of the type: I say ...... every day and arranged into three lists, each of which began and ended with three dummy sentences as described above. The durations of the vowel and post-vocalic consonant in each case were measured. The correlation between the vowel and consonant duration for each voiced v voiceless pair was measured. If there is an inverse relationship between the forms there should be a negative correlation between the vowel and consonant durations. The results of this test are shown in the following table:

Table 4.10

<table>
<thead>
<tr>
<th>Informant M.G.P.</th>
<th>Informant M.P.</th>
</tr>
</thead>
<tbody>
<tr>
<td>beep beeb</td>
<td>r = -0.8686 p(r) &lt; 0.0005</td>
</tr>
<tr>
<td>beat bead</td>
<td>r = -0.7894 p(r) &lt; 0.0005</td>
</tr>
<tr>
<td>book beak</td>
<td>r = -0.6779 p(r) &lt; 0.005</td>
</tr>
<tr>
<td>dip dib</td>
<td>r = -0.6543 p(r) &lt; 0.005</td>
</tr>
<tr>
<td>bit bid</td>
<td>r = -0.9470 p(r) &lt; 0.0005</td>
</tr>
<tr>
<td>dick dig</td>
<td>r = -0.6881 p(r) &lt; 0.0005</td>
</tr>
<tr>
<td>hump hobo</td>
<td>r = -0.8305 p(r) &lt; 0.0005</td>
</tr>
<tr>
<td>Bert birt</td>
<td>r = -0.6969 p(r) &lt; 0.0005</td>
</tr>
<tr>
<td>berk beg</td>
<td>r = -0.5861 p(r) &lt; 0.005</td>
</tr>
<tr>
<td>pub pub</td>
<td>r = -0.7587 p(r) &lt; 0.0005</td>
</tr>
<tr>
<td>but bud</td>
<td>r = -0.8315 p(r) &lt; 0.0005</td>
</tr>
<tr>
<td>buck bug</td>
<td>r = -0.7316 p(r) &lt; 0.0005</td>
</tr>
</tbody>
</table>
Informant M.G.P. has a significant negative correlation in all cases, while in 4 out of the 12, M.P. shows no significant correlation, but in three out of these four cases there is a non-significant negative correlation.

To clarify the position, t-tests were taken to compare the durations of first the vowels when followed by voiced v voiceless stops, and then the voiced v voiceless post-vocalic stops themselves. The results of these tests are shown in tables 4.11 and 4.12:

**Table 4.11**

<table>
<thead>
<tr>
<th>T-tests comparing vowels in:</th>
<th>Informant M.G.P.</th>
<th>Informant M.P.</th>
</tr>
</thead>
<tbody>
<tr>
<td>beep beeb</td>
<td>t = -4.9667 p(t) &lt; 0.0005 t = -0.8767 p(t) &lt; NS</td>
<td></td>
</tr>
<tr>
<td>beat bead</td>
<td>t = -4.2301 p(t) &lt; 0.0005 t = -1.0812 p(t) &lt; NS</td>
<td></td>
</tr>
<tr>
<td>beep beeg</td>
<td>t = -2.8846 p(t) &lt; 0.005 t = -1.7717 p(t) &lt; 0.05</td>
<td></td>
</tr>
<tr>
<td>dip dib</td>
<td>t = -4.1295 p(t) &lt; 0.0005 t = 0.6042 p(t) &lt; NS</td>
<td></td>
</tr>
<tr>
<td>bit bid</td>
<td>t = -7.0710 p(t) &lt; 0.0005 t = -1.4536 p(t) &lt; NS</td>
<td></td>
</tr>
<tr>
<td>dick dig</td>
<td>t = -8.7781 p(t) &lt; 0.0005 t = 0.8763 p(t) &lt; NS</td>
<td></td>
</tr>
<tr>
<td>lump herb</td>
<td>t = -5.6184 p(t) &lt; 0.0005 t = -0.5817 p(t) &lt; NS</td>
<td></td>
</tr>
<tr>
<td>Bert bird</td>
<td>t = -4.5491 p(t) &lt; 0.0005 t = -0.2689 p(t) &lt; NS</td>
<td></td>
</tr>
<tr>
<td>beep beeg</td>
<td>t = -5.3034 p(t) &lt; 0.0005 t = -3.6905 p(t) &lt; 0.005</td>
<td></td>
</tr>
<tr>
<td>pup pub</td>
<td>t = -5.0185 p(t) &lt; 0.0005 t = -1.6552 p(t) &lt; NS</td>
<td></td>
</tr>
<tr>
<td>hut bud</td>
<td>t = -9.3709 p(t) &lt; 0.0005 t = -3.7101 p(t) &lt; 0.005</td>
<td></td>
</tr>
<tr>
<td>buck bug</td>
<td>t = -5.6695 p(t) &lt; 0.0005 t = -1.2256 p(t) &lt; NS</td>
<td></td>
</tr>
</tbody>
</table>

**Table 4.12**

<table>
<thead>
<tr>
<th>T-tests comparing final consonants in:</th>
<th>Informant M.G.P.</th>
<th>Informant M.P.</th>
</tr>
</thead>
<tbody>
<tr>
<td>beep beeb</td>
<td>t = 9.2851 p(t) &lt; 0.0005 t = 1.1711 p(t) &lt; NS</td>
<td></td>
</tr>
<tr>
<td>beat bead</td>
<td>t = 12.05 p(t) &lt; 0.0005 t = 4.2427 p(t) &lt; 0.0005</td>
<td></td>
</tr>
<tr>
<td>beep beeg</td>
<td>t = 5.5628 p(t) &lt; 0.0005 t = 1.6348 p(t) &lt; NS</td>
<td></td>
</tr>
<tr>
<td>dip dib</td>
<td>t = 13.6259 p(t) &lt; 0.0005 t = 1.3760 p(t) &lt; NS</td>
<td></td>
</tr>
<tr>
<td>bit bid</td>
<td>t = 9.7164 p(t) &lt; 0.0005 t = 0.4932 p(t) &lt; NS</td>
<td></td>
</tr>
<tr>
<td>dick dig</td>
<td>t = 4.0611 p(t) &lt; 0.0005 t = 1.6164 p(t) &lt; NS</td>
<td></td>
</tr>
<tr>
<td>lump herb</td>
<td>t = 9.0000 p(t) &lt; 0.0005 t = 2.5456 p(t) &lt; 0.025</td>
<td></td>
</tr>
<tr>
<td>Bert bird</td>
<td>t = 10.4355 p(t) &lt; 0.0005 t = -0.2308 p(t) &lt; NS</td>
<td></td>
</tr>
<tr>
<td>beep beeg</td>
<td>t = 4.0762 p(t) &lt; 0.0005 t = 0.8783 p(t) &lt; NS</td>
<td></td>
</tr>
<tr>
<td>pup pub</td>
<td>t = 5.8012 p(t) &lt; 0.0005 t = 0.2000 p(t) &lt; NS</td>
<td></td>
</tr>
<tr>
<td>hut bud</td>
<td>t = 10.5847 p(t) &lt; 0.0005 t = 3.0769 p(t) &lt; 0.005</td>
<td></td>
</tr>
<tr>
<td>buck bug</td>
<td>t = 3.8007 p(t) &lt; 0.0005 t = 1.4230 p(t) &lt; NS</td>
<td></td>
</tr>
</tbody>
</table>