Eva Strangert

Swedish Speech Rhythm in a Cross-Language Perspective

Umeå 1985
Almqvist & Wiksell International, Stockholm, Sweden
## ERRATA

<table>
<thead>
<tr>
<th>Page</th>
<th>Line</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>18</td>
<td>Underline: 'patterning'</td>
</tr>
<tr>
<td>4</td>
<td>14</td>
<td>(Olsen, 1971) &gt; (Olsen, 1972)</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>Insert: '(k=2)' after recursively</td>
</tr>
<tr>
<td>8</td>
<td>5-6</td>
<td>Fisher-Jorgensen &gt; Fischer-Jørgensen</td>
</tr>
<tr>
<td>9</td>
<td>4</td>
<td>Lehiste, 1980 &gt; Lehiste, 1980a</td>
</tr>
<tr>
<td>12</td>
<td>18;29</td>
<td>Christovich &gt; Chistovich</td>
</tr>
<tr>
<td>15</td>
<td>31</td>
<td>(1977) &gt; (1972)</td>
</tr>
<tr>
<td>19</td>
<td>8</td>
<td>Kohler, 1983a,b) &gt; (Kohler, 1983; Kohler et al., 1982)</td>
</tr>
<tr>
<td>20</td>
<td>5</td>
<td>&quot;content&quot; &gt; &quot;articulatory content&quot;</td>
</tr>
<tr>
<td>25</td>
<td>35</td>
<td>articulator &gt; articulatory</td>
</tr>
<tr>
<td>29</td>
<td>25</td>
<td>1978 &gt; 1979</td>
</tr>
<tr>
<td>32</td>
<td>Fig 3-1</td>
<td>Change the figures for NUMBER OF SYLLABLES to: 12345 2345 345 45 5</td>
</tr>
<tr>
<td>51</td>
<td>20</td>
<td>Insert: 'segments in monosyllabic intervals could be seen as lengthened, while on the other hand' after For example</td>
</tr>
<tr>
<td>51</td>
<td>33</td>
<td>Enclose:'lengthening' within double quotation marks</td>
</tr>
<tr>
<td>67</td>
<td>1</td>
<td>/e/, the first segment &gt; /er/, the first segments</td>
</tr>
<tr>
<td>75</td>
<td>26</td>
<td>Since U &gt; Since Ü</td>
</tr>
<tr>
<td>78</td>
<td>11</td>
<td>syllable &gt; syllables</td>
</tr>
<tr>
<td>79</td>
<td>Table 6-4</td>
<td>Reverse the order of: FAST and SLOW</td>
</tr>
<tr>
<td>89</td>
<td>4</td>
<td>stress-groups &gt; stress-group</td>
</tr>
<tr>
<td>89</td>
<td>5</td>
<td>positive &gt; negative</td>
</tr>
<tr>
<td>91</td>
<td>14</td>
<td>och &gt; and</td>
</tr>
<tr>
<td>98</td>
<td>5-6</td>
<td>upper &gt; left; lower &gt; right</td>
</tr>
<tr>
<td>102</td>
<td>lower fig</td>
<td>533 &gt; 553</td>
</tr>
<tr>
<td>106</td>
<td>6</td>
<td>Insert: 'the different' after for</td>
</tr>
<tr>
<td>113</td>
<td>12-13</td>
<td>number of syllabes &gt; number of syllable types</td>
</tr>
<tr>
<td>120</td>
<td>11</td>
<td>'a nonsense word' &gt; (a nonsense word)</td>
</tr>
<tr>
<td>121</td>
<td>2</td>
<td>three &gt; four</td>
</tr>
<tr>
<td>126</td>
<td>21</td>
<td>factors &gt; function</td>
</tr>
<tr>
<td>138</td>
<td>16</td>
<td>an &gt; a</td>
</tr>
<tr>
<td>140</td>
<td>10</td>
<td>ratios &gt; durations</td>
</tr>
<tr>
<td>141</td>
<td>10</td>
<td>lower &gt; higher</td>
</tr>
<tr>
<td>142</td>
<td>15</td>
<td>Insert: 'Data. Table 10-3 shows the durations of the vowels and consonants in the test words.' before Table 10-3.</td>
</tr>
<tr>
<td>143</td>
<td>18</td>
<td>word &gt; syllable</td>
</tr>
<tr>
<td>148</td>
<td>28</td>
<td>Insert: 'the' after of</td>
</tr>
</tbody>
</table>
Swedish Speech Rhythm in a Cross-Language Perspective

AKADEMISK AVHANDLING

som för avläggande av filosofie doktorsexamen vid Universitetet i Umeå kommer att offentligt försvaras måndagen den 3 juni 1985 kl 10.15,
i Humanisthuset, hörsal F

Av
Eva Strangert
Fil. kand.
Abstract

The study aims to describe and to explain some temporal aspects of the rhythmic structure of Swedish, and to compare Swedish, Spanish, and Finnish, characterized as rhythmically different languages.

The temporal effects of some manipulations believed to have consequences for rhythm were studied in a series of experiments based on sets of Swedish sentences. These manipulations included the length of stress groups - the number of unstressed syllables between stressed ones - and their structural properties. Also the length of the surrounding stress group was varied, as well as speech rate.

The results indicated the presence of several interacting factors. First, there was an inverse relation between the length of the phrase and the duration of the stress group: phrase-length adjustments. The effects were similar to those obtained when speech rate was independently varied. Secondly, adjacent stress groups were adjusted so that differences of duration were counteracted: stress-group adjustments. In addition, in longer stress groups rhythmic alternation between unstressed syllables occurred. Finally, there was an interplay between rhythm and language structure. This was evident from the stability of the temporal manifestations of the quantity distinction and the distinction between stressed and unstressed syllables irrespective of the experimental manipulations.

The results were interpreted in terms of a process model of speech rhythm, the generality of which was considered in the cross-language comparisons. The presence of similar temporal effects in all three languages was taken to support assumptions of language-independent processing constraints. Deviations from the general pattern were explainable in terms of differences of language structure and adjustments to preserve important structural properties, especially quantity relations.

Keywords: speech rhythm, temporal organization, speech production, language structure, Finnish, Spanish, Swedish.
Swedish Speech Rhythm in a Cross-Language Perspective
Abstract

The study aims to describe and to explain some temporal aspects of the rhythmic structure of Swedish, and to compare Swedish, Spanish, and Finnish, characterized as rhythmically different languages.

The temporal effects of some manipulations believed to have consequences for rhythm were studied in a series of experiments based on sets of Swedish sentences. These manipulations included the length of stress groups - the number of unstressed syllables between stressed ones - and their structural properties. Also the length of the surrounding stress group was varied, as well as speech rate.

The results indicated the presence of several interacting factors. First, there was an inverse relation between the length of the phrase and the duration of the stress group: phrase-length adjustments. The effects were similar to those obtained when speech rate was independently varied. Secondly, adjacent stress groups were adjusted so that differences of duration were counteracted: stress-group adjustments. In addition, in longer stress groups rhythmic alternation between unstressed syllables occurred. Finally, there was an interplay between rhythm and language structure. This was evident from the stability of the temporal manifestations of the quantity distinction and the distinction between stressed and unstressed syllables irrespective of the experimental manipulations.

The results were interpreted in terms of a process model of speech rhythm, the generality of which was considered in the cross-language comparisons. The presence of similar temporal effects in all three languages was taken to support assumptions of language-independent processing constraints. Deviations from the general pattern were explainable in terms of differences of language structure and adjustments to preserve important structural properties, especially quantity relations.

Keywords: speech rhythm, temporal organization, speech production, language structure, Finnish, Spanish, Swedish.
Eva Strangert

Swedish Speech Rhythm in a Cross-Language Perspective
ACKNOWLEDGEMENTS

Many people have contributed to this work.

Claes-Christian Elert has been my teacher and advisor, and a tire¬
less partner in discussions of speech rhythm over the years. He has
read several versions of the manuscript.

Iréne Johansson read most of the final manuscript. Pirkko Forsman¬
Svensson and Per Löfstrand assisted in the preparation of the Finn¬
ish and Spanish material, and read parts of the introduction to the
cross-language study. Bo Strangert read and typed most of the final
manuscript. Margit Tornéus and Mona Wiklund helped by typing some
parts. The word processing system in the Psychology Departement
greatly facilitated this work. Inger Nilsson typed earlier versions
of the manuscript. Jan Robbins revised my English. Minje Zhi and
Claes-Christian Elert used the computer system in the Department of
Phonetics to prepare the figures. I received great help from Roland
Åström and his team at Humanisthusets Printing Center.

Eva Ejerhed, Elisabeth Engdahl, Christina Heldner, and Birgitta Holm
served as speakers in the study based on Swedish. Pirkko Forsman¬
Svensson served as the Finnish and Henrique Jimenez Gaona as the
Spanish speaker in the cross-language study. Helena Veikkola-Nilsson
participated in the preliminary tests of the Finnish material.

Finally, Susanne and Christine, my daughters, helped in various
ways, by compiling the bibliography, and by relieving me from domes¬
tic duties for a long period of time.

To all these people, colleagues, friends, and my family, I express
my warm thanks.

Umeå, April 1985
Eva Strangert
## CONTENTS

### INTRODUCTION

<table>
<thead>
<tr>
<th>1 BACKGROUND</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 Definitions and research strategies</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Observations of temporal regularity</td>
<td>3</td>
</tr>
<tr>
<td>1.2.1 Isochrony and the stress-timing/syllable-timing dichotomy</td>
<td>3</td>
</tr>
<tr>
<td>1.2.2 Adjustments to regularity: Segment duration data</td>
<td>5</td>
</tr>
<tr>
<td>1.2.3 Temporal regularity in non-speech rhythmic behavior</td>
<td>9</td>
</tr>
<tr>
<td>1.3 Some models of temporal organization</td>
<td>10</td>
</tr>
<tr>
<td>1.3.1 Models based on Swedish segment duration data</td>
<td>11</td>
</tr>
<tr>
<td>1.3.2 The rhythm generator</td>
<td>12</td>
</tr>
<tr>
<td>1.3.3 The theory of intrinsic timing</td>
<td>13</td>
</tr>
<tr>
<td>1.3.4 Models emphasizing functional aspects of speech production</td>
<td>14</td>
</tr>
<tr>
<td>1.4 Rhythm and language structure</td>
<td>15</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2 A CONCEPTUAL FRAME FOR STUDYING SPEECH RHYTHM</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 The problem</td>
<td>19</td>
</tr>
<tr>
<td>2.2 Units in speech production</td>
<td>21</td>
</tr>
<tr>
<td>2.3 The structure of information</td>
<td>22</td>
</tr>
<tr>
<td>2.4 Outline of a two-stage process model</td>
<td>23</td>
</tr>
<tr>
<td>2.5 Predictions</td>
<td>26</td>
</tr>
</tbody>
</table>

### THE SWEDISH MAIN STUDY

<table>
<thead>
<tr>
<th>3 INTRODUCTION TO THE SWEDISH MAIN STUDY</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Purpose and outline</td>
<td>28</td>
</tr>
<tr>
<td>3.2 Structural characteristics of Swedish</td>
<td>29</td>
</tr>
<tr>
<td>3.3 General experimental procedure</td>
<td>31</td>
</tr>
<tr>
<td>3.4 Characteristics of the material</td>
<td>31</td>
</tr>
<tr>
<td>3.4.1 Motives for the use of normal sentences as test material</td>
<td>31</td>
</tr>
<tr>
<td>3.4.2 The basic material and the test interval</td>
<td>34</td>
</tr>
<tr>
<td>3.5 Speakers</td>
<td>36</td>
</tr>
<tr>
<td>3.6 Recordings and registrations</td>
<td>36</td>
</tr>
<tr>
<td>3.7 Segmentation, measurements, and statistics</td>
<td>37</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>4 THE BASIC EXPERIMENTAL VARIATION: EFFECTS OF THE NUMBER OF SYLLABLES IN THE TEST INTERVAL</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 Purpose</td>
<td>39</td>
</tr>
<tr>
<td>4.2 Procedure</td>
<td>40</td>
</tr>
<tr>
<td>4.3 The total duration of the test interval as a function of the number of syllables</td>
<td>41</td>
</tr>
<tr>
<td>4.4 Effects on vowel duration of the length of the test interval</td>
<td>44</td>
</tr>
<tr>
<td>4.4.1 Shortening effects</td>
<td>49</td>
</tr>
<tr>
<td>4.4.2 Lengthening effects</td>
<td>52</td>
</tr>
<tr>
<td>4.5 Conclusions</td>
<td>54</td>
</tr>
</tbody>
</table>
5 EFFECTS OF VARIATIONS OF LANGUAGE STRUCTURE: QUANTITY AND BOUNDARIES

5.1 Experiment I: Effects of quantity
5.1.1 Purpose
5.1.2 Procedure
5.1.3 Results: The duration of vowels and consonants
5.1.4 The temporal relation between the vowel and the following consonant
5.1.5 The temporal relation between VC:- and V:C-sequences

5.2 Experiment II: Effects of boundary variations
5.2.1 Purpose
5.2.2 Procedure
5.2.3 Data and interpretations

5.3 Conclusions

6 EFFECTS OF VARIATIONS IN SPEECH TEMPO

6.1 Purpose
6.2 Procedure
6.3 Results: The total duration of the test interval
6.4 Syllable durations
6.4.1 The duration of the unstressed syllables
6.4.2 The duration of the stressed syllable
6.5 Speech rate and the distinction between stressed and unstressed syllables

6.6 Conclusions

7 EFFECTS OF CONTEXT

7.1 Purpose
7.2 Experiment I
7.2.1 Procedure
7.2.2 Results: Effects of the size of the surrounding stress groups
7.2.3 Contextual variations and the distinction between stressed and unstressed syllables

7.3 Experiment II
7.3.1 Procedure
7.3.2 Results: Effects of preceding versus following stress groups

7.4 Conclusions

CROSS-LANGUAGE COMPARISONS

8 INTRODUCTION TO THE CROSS-LANGUAGE STUDY

8.1 Purpose and outline
8.2 The selection of languages for comparison
8.3 Main results from earlier research related to rhythm
8.3.1 Spanish
8.3.2 Finnish
8.4 Cross-language comparisons of structural properties
8.4.1 Contrasting word stress, quantity, and syllable characteristics
1 BACKGROUND

1.1 Definitions and research strategies

The concept of rhythm. Rhythm has been defined in various ways depending on the type of activity described. However, the following definition by Martin (1972, p.489-490) is intended to apply to rhythmic action in speech as well as other kinds of human behavior:

"Rhythmic patterns are defined as event sequences in which some events (elements) are marked from others...; call the marked elements "accents". The accents recur with some regularity, regardless of tempo (fast, slow) or tempo changes within the pattern (accelerate, retard). Together an accented and unaccented element constitute the simplest rhythmic pattern; more complex patterns consist of patterns within patterns; that is, they are hierarchically organized."

This definition points to the common properties of rhythm in speech and many other activities. Three basic aspects of rhythm are emphasized: alternation, grouping, and temporal patterning. These concepts are clearly inherent in the elaboration of the term rhythm in speech by Lehiste (1980a, p.8). According to her it refers to:

"...the timing of stressed syllables and the intervals (usually filled with unstressed syllables) that separate them."

Lehiste gives emphasis to the alternation of stressed and unstressed syllables as the basis for grouping, which is natural in the study of languages in which stress is an important structural feature. Dauer (1983, p.8), with a cross-language perspective on rhythm, broadens the basis for grouping to include also other aspects of language structure as well as stress. In doing this Dauer points to a strong dependence of rhythm on language structure:
"In all languages we would expect syllables to be grouped into larger units, even if the basis for grouping is something other than a stress beat. Repetition of particular sounds, syllables, grammatical markers, or pitch patterns, might also be used to group syllables into larger units."

**Rhythmic units.** In studies of rhythm the elementary units are generally assumed to be syllables (Lehiste, 1980a). The rhythm hypothesis (Kozhevnikov & Chistovich, 1965) is based on the open CV syllable as the unit of rhythm. Alternatively, VC sequences are taken to be the relevant units (e.g. Bannert, 1979).

At a higher level syllables are organized into stress groups, referred to variously as rhythmic groups, rhythmic feet, speech measures or interstress intervals (ISI). This type of unit is usually assumed to contain a stressed syllable and a number of adjacent unstressed ones. The exact definition may differ depending on the language studied. Thus, in studies based on English rhythmic groups are generally considered to contain one stressed syllable and all following unstressed ones up to the next stressed syllable (e.g. Dauer, 1983). However, in a language like French, assumed to have a rising rather than a falling rhythm, rhythmic groups may be defined as one stressed and all preceding unstressed syllables (Wenk & Wioland, 1982). Stress groups often span word boundaries. Thus, words are not seen as primary units in studies of rhythm, though influences from word boundaries may be recognized. (For a review of this issue, see Fischer-Jørgensen, 1982.)

At an even higher level stress groups are organized into phrases. A phrase may be identical to a sentence or part of a sentence and is therefore at the same time an articulatory unit and a unit of meaning. Alternative terms defined similarly are "syntagma" (Kozhevnikov & Chistovich, 1965, p.74) and "breath group" (Jones (1918), 1962, p.274).
1.2 Observations of temporal regularity

1.2.1 Isochrony and the stress-timing/syllable-timing dichotomy

The isochrony hypothesis was put forward together with the dichotomy of stress-timing and syllable-timing by Pike (1946, p.35). The underlying idea was that all languages appear to be rhythmic in the sense that some units of speech recur at equal intervals. In some languages, characterized as stress-timed, intervals between stressed syllables were assumed to be approximately equal, while in other languages, characterized as syllable-timed, syllables were assumed to be more or less of equal duration.

Studies based on English. This dichotomy, based mainly on a subjective categorization, seems to account for certain differences between languages. Stress-timed and syllable-timed languages were also associated with different modes of coordination of the breathing muscles (Abercrombie, 1967, p.97). However, right from the start it was clear that isochrony in the strict sense, i.e. equal durations of intervals, did not exist. Even before this proposal was made, Classe (1939) in a study of English, had shown the hypothesis of strict isochrony to be untenable, as the addition of unstressed syllables between two stressed ones always lead to an increase of the interval duration. Classe, however, recognized tendencies to isochrony and even concluded that isochrony, though imperfectly reflected in the speech signal, should be considered the basis of rhythm in English. The conclusions reached in later studies of a similar character seem to depend mainly on the criteria set for isochrony by the specific investigator. Thus, those who take isochrony to imply equal intervals between stressed syllables have rejected it (e.g. Shen & Peterson, 1962), while others willing to accept a weaker definition report tendencies to isochrony (e.g. Uldall, 1971; Lehiste, 1977; Dauer, 1983).

Though Lehiste notifies tendencies to isochrony in the temporal structure of English, she claims that isochrony is mainly a perceptual phenomenon (Lehiste, 1977, 1979, 1980a). This is because listeners seem to be quite insensitive to durational differences like those
appearing most frequently between interstress intervals. As a consequence, Lehiste concludes, intervals between stressed syllables, though not physically the same, may still be perceived as identical.

**Studies based on other languages than English.** Tendencies to stress-timing also have been observed for other languages than English, such as German (Kohler, 1983) and Swedish (Strangert, 1979, 1981, 1984). Major (1981) reports tendencies to stress-timing in Brazilian Portuguese. Moreover, even in languages assumed to be syllable-timed, like Spanish and French, rhythm seems to be adapted to the stressed syllables. This is apparent, for example, from the studies by Manrique & Signorini (1983) on Spanish, and Wenk & Wioland (1982) on French. However, other investigators categorize Spanish as syllable-timed on the basis of durational data (Hoquist, 1983), or as a language with both stress-timing and syllable-timing (Olsen, 1971). Wenk & Wioland (1982) reject the proposal that French is syllable-timed by pointing among other things to the marked differentiation between stressed and unstressed syllables. According to Wenk and Wioland, the rhythmic organization of French, and allegedly stress-timed languages like English, share important rhythmic characteristics. The difference mainly concerns the different position of the stressed syllable within the interstress interval. To capture this difference Wenk and Wioland suggest the terms "leader-timed" and "trailer-timed" for languages with the most prominent syllable at the beginning and the end, respectively, of interstress intervals.

Wenk & Wioland (1982) argue strongly that the reason why French has been characterized as syllable-timed is that studies of rhythm have leant too heavily on English as a model for stress-timing. This bias, in addition to the difficulties of defining stress-timing (and syllable-timing), would seem to be the most likely basis for the divergent conclusions reached in different studies. There is, for example, no answer to how regularly stressed syllables should occur for a language to be labelled stress-timed. (However, see Martin, 1972, for a critique of the criteria for regularity.)
Adjustments between stress groups. Though isochrony implies equal or nearly equal durations between successive syllables or stress groups, most studies have only indirectly examined this contextual aspect of the phenomenon. However, from the studies of Wenk & Wioland (1982), based on French, and Cutler (1980), based on English, it appears that in both languages adjustments may be observed between stressed syllables leading to more equal durations. Wenk and Wioland reported that pauses were inserted in stress groups with few syllables when they were adjacent to longer stress groups. In the study by Cutler, based on speech errors, syllable omissions and changes of the stress position were interpreted as adjustments to temporal regularity between stress groups.

1.2.2 Adjustments to regularity: Segment duration data

Studies based on Swedish. The extensive studies of the dependencies between the duration of a segment and the length of the word and the phrase in which it occurs (Lindblom & Rapp 1973; Lindblom et al., 1976; Lyberg, 1977) have a strong bearing on temporal regularity.

Lindblom & Rapp (1973) studied the effects of word and phrase length on the duration of the stressed vowel in Swedish. ("Word" is here used in the sense "phonological word", that is one or several words of which one contains a syllable with primary stress. See Lehiste, 1980a.) They also examined positional effects within words and phrases. One of their main findings, referred to as the word-length effect, was that the stressed vowel was shortened as the number of syllables in the word was increased. The shortening was strongest at the step between a monosyllabic and a bisyllabic word, and next strongest in trisyllabic as compared to bisyllabic words. However, at a certain duration of the vowel no further shortening occurred. Moreover, the shortening effects of following syllables were found to be stronger than those of preceding syllables.

When words were combined to phrases additional shortening occurred as a function of the number of words in the phrase and, just as for
words, shortening levelled out at great phrase lengths. This was referred to as the phrase-length effect. Also in phrases, in a similar way as for isolated words, more drastic shortening resulted from following than from preceding words. The observations were expressed numerically in a rule assumed to apply recursively first at the word and then at the phrase level:

$$V = D \sum_{i=1}^{k} \frac{1}{(a_i + 1)^{\alpha_i} (b_i + 1)^{\beta_i}}$$

where $V =$ the duration of the vowel

$D =$ the inherent duration of the vowel (word level), or the duration of the vowel at the word level (phrase level)

$a =$ the number of syllables that follow the vowel (word level), or the number of words that follow the vowel (phrase level)

$b =$ the number of syllables that precede the vowel (word level), or the number of words that precede the vowel (phrase level)

$\alpha, \beta =$ constants expressing the degree of anticipatory and backward compensation, respectively.

This rule was later (see Lindblom et al., 1976) interpreted in terms of the power-law model and the short-term memory model.

Lindblom and Rapp based their study on nonsense sequences of the same syllable (dadada ...) modelled on normal words and phrases - so-called reiterant speech. Later, supplementary data were reported on the duration of the stressed vowel /a:/ of real words such as Dag, dagen, idaq and Dagobert uttered in isolation or within systematically varied normal sentence frames (Lindblom et al., 1976). Comparisons of the two types of utterances showed a similar word-length dependence. However, the phrase-length dependence observed in reiterant speech was clearly present only in the final position of the normal sentences. In other positions the stressed vowel was shortened by a constant factor as compared to the duration of the vowel in isolated words.
The study by Lyberg (1977), later elaborated in Lyberg (1981), was based on the data contained in Lindblom et al. (1976). Here Lyberg examined the data for word-length effects within phrases, and reached the conclusion that no such effects seemed to be present, except in the final position of the sentences. This was at least true for the step between the words with one and two syllables. (In the tri-syllabic word Dagobert, on the other hand, the stressed vowel was in effect shorter than in the words with one and two syllables. Also, one of the three speakers in the study actually shortened the vowel gradually between words with one, two, and three syllables.)

Studies based on other languages than Swedish. Observations of dependencies between the length of the word and the duration of its segments date back to the last century and scholars like Rasmus Rask and Henry Sweet, as quoted by Jespersen (1897-1899, p.508; 1914). (For a review of these early findings, see Fischer-Orgensen, 1982).

Especially for words in isolation the word-length effect seems to be of considerable generality. It appears to be present in a great number of languages including languages assumed both to be stress-timed and syllable-timed. Within the first category a word length dependence for isolated words has been observed for English (Lehiste, 1971, 1972; Dutch (Nooteboom, 1972) and German (Rietveld, 1975). Of languages assumed to be syllable-timed word-length effects have been reported for Spanish (Hutchinson, 1973) and French (Roudet, 1910). In addition, word-length effects in isolated words appear to be present in languages like Lappish (Aimä, 1914), Finnish (Iivonen, 1974; Lehtonen, 1974), Hungarian (Tarnóczy, 1965), and Estonian (Eek & Remmel, 1974). Additional references may be found in Lehtonen (1970, p.138-141) Lindblom & Rapp (1973), and Fischer-Jørgensen (1982).

Word-length effects when words (lexical or phonological, see above) appear in phrases seem to be present not only phrase-finally (as would be predicted from Lyberg, 1977) but also in medial positions. Support for this comes from the data of Oiler (1973) using a nonsense material, as well as Klatt (1973), and Klatt & Cooper (1975), using normal English sentences. However, Klatt (1975) claimed word-level
effects to be small and, thus, incompatible with "any strong formulation of an equal-stress timing model for speech production" (ibid, p.139). Earlier Umeda (1972) had rejected the existence of word-length dependencies. Word-length effects within sentences have also been reported for German (Kohler, 1983). According to Fisher-Jorgensen (1982, 1983) word-length effects in Danish are counteracted by tendencies to vowel lengthening in polysyllables as compared to monosyllables. However, only vowels in certain types of syllables were lengthened. Similarly, in Finnish, significant word-length effects were not present in the material studied by Lehtonen (1970, p.138-145).

Effects of phrase-length in English have been reported by Lehiste (1980b). The duration of a test word inserted into a long sentence-frame was shorter than the same word inserted into a short sentence-frame. This finding corresponds to the result obtained in the study by Lindblom & Rapp (1973) for stressed vowels within longer and shorter phrases.

Rhythmic alternation. Durational data for Swedish related to rhythmic alternation between unstressed syllables were presented by Strangert (1981, 1984). Alternation patterns, as revealed by the lengthening of some syllables in a sequence of several unstressed syllables, were quite regular. Either every other or every third syllable was lengthened. Two studies by Bruce (1983, 1984) likewise on Swedish, combine a phonological and phonetic approach to rhythmic alternation. The phonetic parts may be seen primarily as tests on the rhythmic alternation principle suggested by Bruce, according to which every other unstressed syllable from one stressed syllable to the preceding one should be lengthened. The results were interpreted as support for the proposed principle.

Allen (1975), reviewing data for a number of languages, has suggested rhythmic alternation to be a phenomenon of wide generality, present in both stress-timed and syllable-timed languages.
1.2.3 Temporal regularity in non-speech rhythmic behavior

The often very close parallels observed between speech rhythm and rhythmic patterning in other types of activity (e.g. Kozhevnikov & Chistovich, 1965, p. 116; Lehiste, 1980; Dauer, 1983) have been taken to imply the existence of universal rhythmic principles. Such principles according to Allen (1975) are manifested as "rhythmic performance universals" common to speech and non-speech behavior.

Rhythm as a general characteristic of human behavior has been considered from the perspective of production as well as perception. Production studies have been based on spontaneous or induced movements patterns, for example finger-tapping, while studies of perception have been based on reactions to series of simple stimuli such as brief sounds or flashes of light.

Fraisse (1982), reviewing earlier studies (Fraisse, 1946-1947), reports experiments with finger-tapping in which subjects were asked to produce spontaneously regular as well as irregular rhythmic sequences. To estimate the differences between successive taps, the ratios between tapping intervals were calculated. Fraisse found, first, that subjects experienced great difficulties when having to tap irregularly. Secondly, he found obvious regularities in tappings intended both to be rhythmic and arrhythmic. In both types there was a preponderance of nearly-equal ratios and 40 percent were less than 1.2. According to Fraisse subjects tended to produce intervals as equal to the preceding one as possible. It appears as if the first pattern imposed itself on the following interval. Similar observations have been made in perceptual studies by Preusser, Garner & Gottwald, 1970).

Other studies attest to a tendency to adjust both the production and perception of time intervals towards an intermediate or average duration. From the review of Fraisse (1982) it appears that subjects asked to produce groups of taps were found to adjust the intervals between the successive taps so that long groups were tapped at a faster rate than short ones. Therefore, a group of four took only 1.8
times as long as a group of two, and a group of six, 2.2 times as long (MacDougall 1903). Similar findings have been reported by Woodrow (1951), referring to Harrell (1937).

The grouping of a series of elements is largely determined by the specific properties of these elements such as their relative intensities and durations and their temporal relations. In general, regularly recurring elements with greater intensity than the rest will be perceived as leading the group, while an element with a relatively greater duration will be perceived as terminating the group (Woodrow, 1951). However, grouping occurs even if there is no objective basis for it. Listening to sequences of identical and isochronous sounds, they seem to be grouped, primarily by twos and threes. Such grouping seems to reflect a basic characteristic of rhythm, as when longer groups occur they tend to be subdivided into shorter ones. Subgrouping is associated with a strengthening of one of the sounds in the group, primarily the third in a group of four, and either the third or the fourth in a group of five. As a result, subgroups of two or three elements appear (Woodrow 1911, 1951; see also Fraisse, 1982).

1.3 Some models of temporal organization

Models of the temporal organization of speech differ considerably as to their aim and scope. Some lean heavily on data, while others may be characterized as conceptual frames rather than empirically based models. Also, some models aim primarily at data-matching, while others are concerned with the mechanisms behind the observed temporal data. These latter kinds of models are the focus of this section. (For a general review of timing models, see Allen, 1979.)

Another important distinction between the models is whether they conceptualize time as a primary attribute in speech planning or as a secondary consequence of it.
1.3.1 Models based on Swedish segment duration data

The power-law model. The power-law model was the first in a row of models developed on the basis of the segment duration data presented in Lindblom & Rapp (1973) and Lindblom et al. (1976) (see 1.2.2). The model is based on two assumed principles: (1) Word duration is constant, irrespective of the number of syllables of the word. Stressed vowel duration as a consequence is inversely proportional to word length. This principle provides an explanation for shortening. (2) Word length is underestimated by the speaker according to a power-law. As a consequence shortening is reduced. This principle explains why duration appears to level out, when words contain several syllables. The shortening expressed by the power-law model for the duration of stressed vowels within words may also be applied to stressed vowels within phrases. Thus, two hierarchical levels are hypothesized: the word and the phrase level, respectively.

The timing mechanism in this model may be seen as a scanning device which operates on a buffer containing a string of phonetic segments, the length of which corresponds to the phrase. This timing mechanism, sensitive to the hierarchical structure of the phrase, scans the buffer from left to right, thereby triggering the successive segments and assigning proper durations.

The short-term memory model. The further elaboration of this model, the short-term memory model (see Lindblom et al., 1976) aims at an explanation of the principles forming the basis for the power-law model. In this more psychologically oriented model, temporal adjustments are assumed to depend on properties of short-term memory and motor execution. Shortening is attributed to storage economy which explains the crushing together of syllables at great word lengths, or alternatively, words in long phrases. Positional dependencies are explained as consequences of capacity restrictions on memory. Stressed vowels in the final position of words and phrases are shortened less than vowels in other positions due to an asymmetry of space saving. Finally, incompressibility, that is, the tendency for vowel duration to level out in words with many syllables, or in phrases with many words, is attributed to motor constraints.
1.3.2 The rhythm generator

The rhythm generator model was proposed by Kozhevnikov & Chistovich (1965). One of the main problems addressed was whether the timing and articulatory program for an utterance should be considered two separate systems or just one. Two types of observation led Kozhevnikov and Christovich to assume a separate timing mechanism, "a rhythm generator": (1) data based on variations of speech rate and (2) rhythmic patterning as partly independent of segmental characteristics. The uniform increase of the duration of words and syllables in phrases upon an increase or decrease of speech rate was interpreted as support for such a separation. Moreover, the rhythmic figure of a word was assumed to exist as a temporal pattern independent of the segmental composition of the word, though segmental differences could lead to perturbations of the pattern. This was the second kind of evidence for a separate timing mechanism.

The second main problem addressed concerned the units controlled by the timing mechanism. Kozhevnikov and Christovich assumed that the articulatory program operates on units referred to as "syntagmas". A syntagma consists of about seven syllables and is at the same time an articulatory unit, bounded by pauses, and a unit of meaning, being either a sentence or part of a sentence. Its length of seven syllab-
bles is associated with the limitations of operational memory. While the syntagma was considered the largest articulatory unit within the model proposed, the syllable (of the CV type) was seen as the elementary one. (No intermediate units were hypothesized.) The support for the syllable as the basic unit was based on statistical calculations of the correlations between the successive segments of the syllable. Obtained negative correlations were taken to imply compensations in order to maintain an even syllable rhythm. The syllables were therefore assumed to be the units on which the timing mechanism, the rhythm generator, operates. Regular pulses from this generator were assumed to trigger the successive syllables within a syntagma. However, as it was impossible to separate the hypothesized effect from measurement errors and rate variations, this approach was later abandoned.

1.3.3 The theory of intrinsic timing

The model proposed by Fowler (1977) presents a view of speech timing different from that in earlier models, which she referred to as extrinsic. According to Fowler the general characteristic of extrinsic timing models is the separation of "the means of control over the timing of an act of speech from the means of control over the act's configurational properties" (Fowler 1977, p.1). This separation of the timing and articulatory planning mechanisms is claimed to be both uneconomic and untenable.

Fowler bases her arguments on a critical review of two types of timing models: the rhythm generator model (Kozhevnikov & Chistovich, 1965) and scanning models such as, for example, that proposed by Lindblom & Rapp (1973). Her objections concern the assumptions of a separate timing device as well as the underlying assumption of discrete timing units such as syllables, words and phrases.

According to Fowler, the timing of an act is intimately coupled with both the plan and the act itself. The plan is therefore four-dimensional, including time as one parameter. The basis for this assump-
ation is the complex coordination of movements required for the production of speech sounds. Timing has to be tailored very closely to the successive coordinated movements, and is therefore part of the plan rather than external to it. As a consequence independent rhythmic patterns as proposed by Kozhevnikov & Chistovich (1965) are implausible. Scanning devices are also implausible owing to their reliance on discrete production units for which there is no proof. Thus, the absence of clear-cut boundaries of production units removes the necessary markers for the scanning device.

Fowler proposes a planning mechanism based on the concept of "continuous vowel production" by which is meant that speech is produced as one continuous movement from vowel to vowel with consonants as deflections on the ongoing movements. There are therefore in this model no discrete production units. The basic units are assumed to be four-dimensional coordinative structures linked to muscle action. Two types of structure are hypothesized: one basic structure for the movement from one vowel to another and one from one stressed vowel to the next. The latter is associated with the phenomena of stress and stress-timing. Thus, two levels of organization are assumed, one superordinated to the other. Stress-timing, therefore, adds an extra level of constraint, leading to a simplification of the system.

1.3.4 Models emphasizing functional aspects of speech production

In the goal-oriented model of Lindblom (1983) the speaker's conception of the communicative demands is the governing factor of speech production. Adjustments to the situation and especially to the listener are made continuously. Such adjustments may include reductions (when communicative demands are low) as well as distinctiveness of articulation (when communicative demands are high). Such a teleological perspective is necessary in order to understand the different manifestations of speech, according to Lindblom.
The gesture theory. The gesture theory (Öhman et al., 1979) sees the temporal structure of utterances in many cases as a consequence of speech planning rather than as a primary goal. The theory places emphasis on the communicative aspects of speech production. Thus, the main object for study is to separate the primary acoustic effects intended by the speaker from the secondary acoustic traces of the efforts (the audible gestures) to produce these effects. The point made is that the primary effects, in general, are not bound to have any specific duration. Therefore, the complexity of the intended effect (or rather, the effort it takes to produce it) determines its time structure. Even durational differences associated with the quantity distinction in Swedish (see 3.1) are seen as secondary effects (of stress) rather than primary temporal effects. Fundamental frequency, on the other hand, is in many cases associated with primary acoustic effects. Changes of fundamental frequency, therefore, especially when complex, should have consequences for duration. This reasoning is reflected in the $F_0$-dependent model of segment duration (Lyberg 1979, 1981; see also Umeda, 1976).

1.4 Rhythm and language structure

Rhythm, if defined as the temporal patterning of stressed and unstressed syllables, is determined to a considerable degree by properties of language structure. For example, Bolinger (1965) measured interstress intervals in two sentences and found, not surprisingly, that in addition to the number of syllables, syllable structure, position in the phrase, and relative semantic importance also influenced the length of the intervals. Classe (1939) had made similar observations.

In studies of rhythm, word boundaries are more or less implicitly assumed not to be important. Stress groups (or feet, interstress intervals) are taken to be the units of rhythm. Lehiste's study (1977) indicates that rhythmic effects span word boundaries (or rather combined word and syntactic boundaries). For example, in speed kills both the stressed vowel and the stem of speed is longer than
in speed increased. Similar observations were made by Strangert (1979). However, other studies point to a certain influence from the word (Kohler et al., 1982; Klatt & Cooper, 1975). Nakatani & Schaffer (1978) even assume that information about word boundaries is always encoded in the speech signal. On the other hand, experiments by de Rooij (1979) point to strong influences from situational factors as to the realization of word boundaries. They may - but need not - be realized.

As is the case with word boundaries, syntactic boundaries appear to be manifested in some, but not in all, studies. The studies of Oller (1973) and Klatt & Cooper (1975) both show boundary effects. However, Klatt & Cooper assume boundary realizations to be optional in short phrases. Lehiste (1980) claims rhythmic structure to be basically independent of syntax. However, when contrast is required boundaries may be temporally reflected. This was shown by Lehiste (1973) in an experiment on the strategies used by speakers to "disambiguate" syntactically ambiguous sentences such as (Steve or Sam) and Bob will come or (Steve) or Sam and Bob will come. As would be expected, the speakers asked to convey one or the other meaning of the sentences signalled the syntactic differences by durational changes. The interstress interval that contained the relevant boundary was substantially increased in length. Thus it appears that when a speaker aims at contrasting two structures, boundaries are clearly manifested. Similarly, Klatt & Cooper (1975) assume phrase-final lengthening within a sentence to be optional in a short phrase, depending on the communicative demands in the specific situation.

**Language structure and rhythmic typology.** Structural differences between languages are reflected in the temporal structure of speech. Such differences, according to Dauer (1983) are the most important source of rhythmic diversity. Dauer compared prose readings of English and Thai, both assumed to be stress-timed, to Spanish with syllable-timing, and Italian and Greek, both rhythmically undefined languages. Measurements of interstress intervals revealed quite similar durations for all five languages (except for the monosyllabic intervals in English and Thai, which were about double the duration
of the monosyllabic intervals in the other languages). The range of interstress intervals in all languages lay between 0.2 s and 1.0 s, clustering around a mean duration of 0.4 s to 0.5 s. These similarities led Dauer to conclude that perceived rhythmic differences between the languages should be attributed to differences within rather than between interstress intervals. She therefore assumed all of these languages to be more or less stress-based, governed by similar rhythmic principles. Dauer singled out three language features thought to be responsible for rhythmic differences between languages: syllable structure, vowel reduction and stress.

Dauer emphasizes the fact that stress has different functions in different languages and that a language with lexically free stress is very often stress-timed. Exceptions are, for example, Spanish and Italian. In stress-timed languages, too, the phonetic correlates of stress are stronger and more varied than in syllable-timed languages. Stress in a stress-timed language may be manifested as longer duration, pitch changes, increased loudness and quality differences, while in a syllable-timed language the realization of stress may be restricted to pitch changes. However, in this respect also, Spanish and Italian appear to be exceptions.

Further, in stress-timed languages syllable structure and the distribution of syllable types over stressed and unstressed syllables reinforce the stressed/unstressed distinction. A greater variety of syllable types, including more complex ones, characterize stress-timed languages. In addition, there is a strong tendency for stressed syllables to be "heavy" (contain many syllables) and unstressed to be "light" (contain few syllables). In syllable-timed languages, on the other hand, simple syllable types predominate and more complex ones, when they occur, are not restricted to stressed syllables.

Dauer also finds different types of vowel reduction in stress-timed and syllable-timed languages: stress-conditioned reduction and reduction conditioned by phonological environment, respectively. The former type of reduction results in reduced, and often centralized, vowels in unstressed position. The second type of reduction often leads
to the complete elimination of syllables which according to Dauer may "reestablish the evenness of successive syllables by eliminating an inherently short syllable (e.g. "chez le garçon" /ʃel.jarson/)."
2 A CONCEPTUAL FRAME FOR STUDYING SPEECH RHYTHM

2.1 The problem

The review in Chapter 1 shows speech rhythm to be a multifaceted phenomenon. The present study is restricted to rhythm as reflected in the temporal structure of the speech signal. However, even if rhythm by definition implies temporal structure it has to be recognized that rhythmic cues may be signalled by other means than duration - for example fundamental frequency patterns (Kohler, 1983a,b).

If rhythmic structure is studied through its temporal manifestations in the speech signal, it is necessary to conceptually isolate the phenomenon of rhythm from other aspects of articulation. Is it, then, possible to single out the problem of rhythmic structure, or more specifically, the temporal dimension of this structure, from the general problem of articulation?

According to Fowler (1977), timing and temporal relations are intrinsic to the configurational properties of a speech act, that is to say, the articulation. So-called models of extrinsic timing, on the other hand, separate the plan for the generation of a sequence of phonetic segments from the timing of these elements (1.3.3). The latter approach presupposes that the phonetic segments have time-invariant representations in the plan, while timing and other prosodic features are superimposed on the phonetic segments extrinsically (ibid, p.2). For example, in scanning models, time is assigned by a device processing an input string of elements in sequential order (1.3.3). These assumptions are considered to be incompatible with well-known facts about contextual variation.

However, from an analytical point of view the problem of rhythmic structure would be simplified if it were possible to separate the sequential processing, or rather some critical aspects of it, from the planning of articulatory configurations. Then, sequential processing, rather than being equated with timing, should be used to denote preservation of temporal order.
A corresponding operationalization would require that an utterance could be articulated in an acceptable way (for the speaker, or listener), even if its temporal structure is systematically varied. Obviously it is very easy to demonstrate empirically a restricted independence between the "content" of an utterance and some of its durational properties. The experiments in the present study are all instances of such a paradigm. (We may note in passing that the use of reiterant speech implies in effect that rhythmic patterns can be separated from articulatory content; see 3.4.1.)

The problem addressed in the present study may be conceptualized in terms of an input-output system with an operational domain over an utterance (a phrase, or a short sentence, bounded by pauses; cf. the "syntagma", Kozhevnikov & Chistovich, 1965, p.74). The input to the system is a structured string of abstract features (phonologic, syntactic, semantic, and pragmatic). This string, the message structure, is transformed by processes of programming and execution to an output string of temporally structured phonetic segments. We assume that under certain experimental conditions the input message may be transmitted without loss of information, i.e a 1-to-1 relation is assumed between the content of the intended and the actually produced utterance, according to the criteria of the speaker or the experimenter.

The problem then is how the sequential order in the abstract string is transformed into the temporally structured output. In other words, how does the temporal variation arise in the system?

The rhythmic programming is assumed to be sensitive to two kinds of input factors, both of which have been shown to affect the rhythmic structure of speech: (1) language structure constraints, i.e. formal and functional properties of the message, including pragmatically based structural features such as contrast and emphasis (1.4), and (2) communicative constraints, which may include criteria for speech rate as well as accuracy and precision of articulation set by the speaker (1.3.4).
Thus, on the basis of empirical facts and certain experimental precautions it may reasonably be concluded that to some extent rhythmic structure can be manipulated and measured independently of the articulatory content. However, the justification for this analytical distinction is not incompatible with the view that time is intimately associated with articulatory programming in human speech processing.

2.2 Units in speech production

Another critical argument of Fowler's (1977) that deserves serious attention concerns the concept of production units. Scanning models generally presuppose some kind of sequential processing of units such as phonetic segments or syllables. These units, being time-invariant, are assumed to be deducible from the input structure of the utterance. Moreover, these units are assumed to be recoverable from the speech signal, which in effect presupposes that the units are discrete and have distinct boundaries. As is well known, this is not the case. The parallel transmission of segments leads to considerable variation of the assumed production units in the speech signal.

The ways of handling this problem in studies of the temporal organization of speech is to try to restrict the variation through a careful choice of the test material and the criteria for segmentation. With an easily segmented material, for example nonsense syllables such as /ta/, a certain invariance is created. (In spite of such precautions, however, errors of measurement may introduce temporal artifacts such as negative correlations between adjacent elements, which may be difficult to separate from the temporal structure under observation (Ohala, 1975; Ohala & Lyberg, 1976)).

Though the use of reliable criteria for segmentation is a necessary condition for the measurement of temporal structure, it does not solve the general problem of temporal variation with context. It is a fact that temporal effects are not exclusively associated with any particular kind of unit or boundary. We know, for example, that the properties of the preceding as well as the following consonant
affect the duration of a vowel (Lindblom & Rapp, 1973). A single intervocalic consonant, then, "belongs to" the preceding as well as the following vowel. Under these conditions how should syllables be defined and segmented in the speech signal, when whatever criteria are chosen, artifacts would be introduced?

The theoretical consequences of this problem led Fowler (1977) to propose the theory of intrinsic timing in which the concept of production units is ruled out. The theory leans, rather, on the assumption of "continuous vowel production". Vowels, and especially vowel onsets, are seen as line-up points for the planning mechanism, which operates on four-dimensional coordinative structures associated with the vowels. Neither syllables, nor larger units, except for the phrase, are hypothesized (1.3.3).

Thus, while a linguistic analysis might lead us to assume invariant units at various levels (phonetic segments, syllables, words, and phrases) in the input string, an analysis of the speech output indicates variance rather than invariance. There is a strongly context-dependent temporal flow of information without distinct boundaries or units, except under certain standardized experimental conditions.

2.3 The structure of information

It is reasonable that a model of rhythmic structure, which presupposes some kind of sequential processing of production units, should provide a hint on how to handle the context problem. In the present study the concept of information network system has been chosen to represent the structural relations between the possible production units.

Formally, the network consists of nodes and their interconnections. Nodes correspond to different units, and several levels of unit exist (e.g. articulatory features, syllables, words). The interconnections correspond to linguistic, learned, and built-in relations or constraints, according to common views. For example, a syllable A is connected to (i.e. represented by) a set of articulatory fea-
tures \((a, b, c...\)), some of which it shares with another syllable \(B\). \(A\) and \(B\) may be joined in a word or a phrase (a collection of syllables or words). A multitude of such interconnections are obviously typical for the human speaker. The interconnections are assumed to have dynamic properties. That is, activation of a unit may influence the states of other connected units, the net effect of which is a possible contextual adjustment. Activation of a set of nodes is equivalent to gaining access to the information represented by the units. It is further assumed that this activation may last only for a brief period of time. (This sketch of an information network is related to the basic outline of recent processing models of memory structure, e.g. Anderson, 1976.)

### 2.4 Outline of a two-stage process model

The concept of a flexible information network is the basis for some speculative assumptions about the origin of rhythmic structure in speech. The following account distinguishes articulatory planning from motor execution. Two sequential stages of planning are assumed (Figure 2-1).

![Figure 2-1. Input-output model of rhythmic processing.](image-url)
At the first step of planning, the preprogramming stage, the system scans an input string about the size of a phrase, or a part of a phrase, which transforms the input information into articulatory coded units. The scanning mechanism is sensitive to prosodic and syntactic features such as stress and boundary specifications. These features form the basis for the partitioning of the phrase into rhythmic units. The stress group, delimited to include one stressed and an unbroken string of unstressed syllables following (or preceding) the stressed one, is considered to be the primary rhythmic unit (1.1).

The stress group is input to the next processing step, the articulatory motor programming, which generates commands to the motor executor. The operational domain is assumed to be restricted to a few syllables, owing mainly to a limited channel capacity of the motor execution system, that is to say a relatively small number of possible states (muscle actions). This capacity restriction corresponds to an estimate of the maximal storage of about three syllables in the short-term memory model by Lindblom et al. (1976).

A consequence of the limited capacity would be that stress groups with many syllables, as delimited from the input string, would have to be subgrouped at the initial step of processing. Such subgrouping is assumed to be the basis for the phenomenon of rhythmic alternation in long sequences of unstressed syllables (cf. subgrouping in non-speech rhythmic behavior, e.g. Woodrow, 1951; Fraisse, 1982).

The exact estimate of processing or storage capacity may depend not only on the channel capacity of the execution system, but also to some extent on the connections between the phonetic segments forming the stress group. Well-connected units imply redundancy, which will facilitate the transmission and storage of information.

It will be assumed that in the experiments of the present study the durational variance of the final stage of motor execution is uncorrelated with the variance introduced by the experimental manipulations, at least for the chosen level of analysis.
The operation of the system. At the preprogramming stage the input string is processed by activating all nodes and interconnections associated with the input, possibly in sequential order. Activation of a node dissipates with time, but for a brief period all elements of the phrase are simultaneously active. During this period grouping or chunking occurs, as well as contextual adjustments. At the articulatory programming stage these chunks (stress groups) will be transformed in turn into sets of motor commands. Even at this stage contextual adjustments may be possible but within a smaller domain. The transmission of information (or control) is assumed to occur cyclically. A phrase may require a few cycles of articulatory motor programming, depending on the number of stressed syllables in the phrase. This would be the principal basis for temporal regularity in speech.

The system probably has a limited capacity in at least two correlated ways. First, it is time-limited, since the activation of the information nodes dissipates with time. The processing would have to be faster for long input strings than for short ones, if the demands on processing are constant. Secondly, the "processing energy" may be limited. Only a small number of nodes may be active at one time (e.g. about seven syllables). If a constant amount of energy is assumed, then the available energy will have to be distributed among the elements. In longer phrases the elements would be less activated, or more superficially processed, given a constant rate of processing. Thus, there will be a trade-off relation between speed and accuracy of processing.

These general predictions must, however, be modified with respect to the structure of the information network, that is, by various constraints on language structure, speaker skills, and characteristics of the test material.

Finally, the first processing step may also be affected by the speaker's conception of the communicative demands in the specific speech situation. The speaker's criteria for speech rate and articulatory precision may vary, and thereby lead to variations of processing rate and articulator target values (Lindblom, 1983).
2.5 Predictions

The purpose of the foregoing speculative account of some preliminaries to a model of rhythmic structure was to facilitate the integration of various empirical findings with some general predictions. Note that no assumptions of timing in terms of durational rules or neural clock mechanisms were made. Nevertheless, it is possible to make some general predictions for temporal relations in speech on the basis of the assumed processing constraints.

First it will be recalled that phrase length, defined by the number of syllables, determines the rate of processing, and thus, the duration of the utterance. This prediction presupposes that the phrase will be within the planning domain of the first processing step. The phrase-length effect would tend to make elements (syllables, or stress groups) in shorter phrases longer in duration than the same elements in longer phrases. The prediction is supported by empirical findings in studies of speech (Lindblom & Rapp, 1973; Lehiste, 1980b). Observations on non-speech behavior also support this type of length-adjustment. In experiments with finger-tapping, elements in a longer unit tend to be produced at a higher rate than elements in a shorter unit (see 1.2.3). This type of temporal effect would be equivalent to that resulting from an intentional variation of speech rate.

The assumption of cyclic processing as an explanation for temporal regularity of speech implies that there exists a variable temporal adjustment of stress groups. This stress-group adjustment would tend to make adjacent stress groups more similar in duration. This type of adjustment is also observed in non-speech rhythm. Fraisse (1946-1947, 1982), experimenting with patterns of finger-tapping, showed that a following interval was adjusted to the preceding one so that the intervals would be more similar.

According to this conceptual frame, rhythmic structure would mainly be created by the dual processes of the partitioning of phrases into stress groups (at the preprogramming stage) and the cyclic generation of motor commands (at the stage of articulatory motor program-
The distinction between stressed and unstressed syllables is assumed to be the primary basis for the partitioning of the phrase into stress groups. Stressed and unstressed syllables would have to be clearly differentiated.

It would be reasonable to expect that the assumed processing constraints would not vary between different languages. Differences of rhythmic structure, then, would be determined by language-specific structural constraints. Structural properties such as quantity and syllable structure should be important determinants of rhythmic structure (cf. Dauer, 1983). In general, then, the processing constraints are reasonably language-independent, but since language structure plays a major role in the shaping of rhythm (e.g. through boundaries, stressmarkers, syllable structure etc.) language-specific characteristics will color the manifestations of rhythm.
3 INTRODUCTION TO THE SWEDISH MAIN STUDY

3.1 Purpose and outline

The purpose of the study is first to enlarge upon earlier observations of longer units of speech and their internal structure (see Strangert, 1979, 1981, 1984-). This gives it to some extent a different profile from that of other Swedish studies relating to speech rhythm, which have mainly focussed on the duration of segments, especially vowels (see 1.2.2).

The Swedish data will be examined for the three main influences on rhythmic structure assumed in Chapter 2: (1) Adjustments of processing rate (uniform and variable) to the length of stress groups and phrases, (2) Adjustments to maintain grouping characteristics, i.e. the distinction between stressed and unstressed syllables, and (3) Adaptations of rhythmic constraints to properties of language structure.

In Chapter 4 the quantitative relation between the number of syllables in the stress group and its duration is analyzed as a base-line for the following experiments. In addition, the influence of language structure on this relation is examined. The distinction between stressed and unstressed syllables as a basis for rhythmic grouping is the basis for the first part of Chapter 5. The temporal characteristics of stressed syllables are analyzed in terms of the relations between their vocalic and consonantal parts, when their phonological lengths are varied. In the second part the effects of syntactic boundaries on the durational patterns of stress groups are studied. The main problem addressed in Chapter 6 concerns the temporal relation between stressed and unstressed syllables when there are experimental variations of speech tempo. Finally, in Chapter 7 the two types of temporal adjustment assumed in Chapter 2 (uniform adjustment to phrase length and variable adjustment between stress groups) are contrasted by experimental manipulations of the contextual frame.
Throughout the study special attention has been directed to inter-speaker variation.

The present chapter includes a short sketch of some structural characteristics of Swedish and an account of the general experimental design.

3.2 Structural characteristics of Swedish

A short summary will be given of some basic aspects of the phonology of Swedish assumed to have relevance to rhythm. It includes word stress, quantity, and syllable structure.

**Word stress.** The lexically determined position of stress in words is free in the sense that it is not bound to occur in any specific position of the word. A shift of stress from one syllable to another may lead to a shift of meaning. However, even if stress may occur in different positions, initially stressed words predominate.

In a traditional analysis of Swedish, for example that used in the Dictionary of the Swedish Academy (SAOB 1898-), several levels of stress are distinguished: primary stress, secondary stress, and two levels of weak stress. In a structural phonological interpretation, based on minimal contrast, only two distinct stress levels may be recognized: stressed and unstressed. Other stress levels in the traditional analysis are either predictable from the occurrence of other phonological elements (such as tonal accent and vowel quantity) or may be attributed to rhythmic alternation. This alternation of two degrees of weak stress has been described by Elert (1957, p.48; 1964, p.22-24; 1970, p.37; 1978, p.191). Lindblom et al. (1976, p.24) suggested an explanation for this strengthening of unstressed syllables in terms of syllable structure, that is, the occurrence of strong clusters, or the derivational structure of the word.

Content words and grammatical words have different stress patterns. In connected speech primary stress is assigned to at least one
syllable in content words, while in grammatical words all syllables are weakly stressed (Elert, 1964, p.24-25).

An increased degree of stress, greater than primary stress, is given to the stressed syllable of words which are emphasized (emphatic, or contrastive stress) or represent something new introduced in the discourse (rhematic stress, or sentence stress) (Elert, 1964, p. 23; Gårding, 1967, p.64).

Stress occupies a strong position in Swedish as a determinant of the distribution of other structural features. Stressed and unstressed syllables are different in several respects. The quantity distinction, syllable structure and vowel reduction, all seem to be conditioned by stress. Phonetically stress is manifested in several domains. Stressed syllables are characterized by longer duration, greater intensity and characteristic peaks of the fundamental frequency contour.

**Quantity.** In stressed syllables there are two distinctive degrees of vowel length: long and short. Also, there is complementarity between the vowel and the following consonant so that a phonologically long vowel is followed by a short consonant (V:C), and a phonologically short vowel by a long consonant (VC:). Consonant length, being thus predictable from vowel quantity and stress, is not phonologically distinctive.

The complementarity between the duration of the stressed vowel and that of the following consonant makes stressed syllables almost equal in duration irrespective of the length of the vowel. However, small differences exist and, in general, a sequence of a long vowel followed by a short consonant is longer than a sequence of a short vowel followed by a long consonant (Elert, 1964, p.158).

**Syllable structure.** Swedish allows for complex consonant combinations in all positions in words. As a consequence, both closed and open syllables occur. However, in general, stressed syllables are closed while unstressed syllables are open (Sigurd, 1965, p.163-165). There
is also a difference between the distribution of vowels in stressed and unstressed syllables. Stressed syllables can contain any vowel while in unstressed syllables there is a more restricted range of vowels (Sigurd, 1965, p.164).

In natural speech unstressed syllables may be reduced, especially in casual speech and when the speech rate is high. The reduction means that unstressed vowels are shortened and centralized (Lindblom, 1963).

Occasionally vowels may even be eliminated, and consonants also may be lost, especially in medial and final clusters. Examples (from Garlén, 1984, p.81) are råspif for rostbiff 'roast-beef', amrikan for amerikan 'American', and aså for alltså 'thus'.

3.3 General experimental procedure

The specific experiments in the present study are, to a certain extent, based on different test material. However, the various parts of the material were constructed according to similar principles. These similarities, as well as the common characteristics of the experimental procedure, will be described in the following sections. Experimental conditions specific to the problem addressed will be described separately for each part of the study.

3.4 Characteristics of the material

3.4.1 Motives for the use of normal sentences as test material

All the test material consists of semantically acceptable, or nearly acceptable, sentences. Reiterant speech would have been preferable for several reasons (control of the segmental composition of syllables, facilitation of measurements). However, as it is questionable to what extent reiterant speech actually copies normal speech, it was not used in the present study.
Reiteration means that the speaker first reads a semantically normal word (or sentence), stores its rhythmic (temporal) pattern in short-term memory, and then utilizes the pattern in the planning and execution of the reiterant version. Thus, the use of reiterant speech in experimental work actually presupposes a certain independence of the rhythmic figure and the segmental composition of the word (or sentence). Therefore, reiterant speech implies that rhythmic patterns can be mentally stored and utilized (see 2.1). However, the utilization of such a plan does not necessarily guarantee a correct mapping of the normal features onto the reiterant version.

There is evidence that reiteration introduces artifacts of its own. At least two types of discrepancy between normal and reiterant speech have to be considered:

(1) Syllable structure and segmental composition differ for stressed and unstressed syllables in Swedish (3.2). This seems to be a property of Swedish with important implications for speech rhythm. By reiterating all syllables as, for example, ba, and thereby reducing the differences between stressed and unstressed syllables, one would introduce distortions into the rhythmic patterning of syllables. The consequences would be especially serious with regard to unstressed syllables for which the difference between nonsense and normal speech would be expected to be more marked. Such differences between unstressed syllables in hummed and normal speech have been observed by Svensson (1974, p.15-17). Similar arguments against the use of reiterant speech have been advocated by Dauer (1983).

(2) Reiteration appears to be a difficult task for the speaker, even after considerable training. One would expect such difficulties to be reflected in the temporal domain and it would therefore be impossible to separate them from the induced experimental variation. Artifacts would be especially noticeable in renderings of longer utterances which tax the capacity of the speaker.

The view that reiterant speech may introduce this second type of artifact is supported by comparative data on normal (model) and reit-
erant versions of some test sentences. Figure 3-1 shows the duration of stressed and unstressed vowels, when the number of syllables between two primary stressed ones was increased from one to five. The data pertain to the second stress group of semantically normal sentences like Bönderna böt bilen i Bergamo, 'The peasants changed the car in Bergamo', and enlargements of this sentence through the insertion of extra unstressed syllables after böt. In the reiterant version the syllables in the second stress group were exchanged for ba.

![Figure 3-1](image-url)
Apart from the overall increase of duration in the réitérant as compared to the normal speech, the patterning of réitérant speech deviates from that of the normal version. The successive shortening of vowel duration as a function of the number of following unstressed syllables evidenced in the normal version is only partially copied in the réitérant speech. The deviances seem to arise when stress groups contain more than three or four syllables. The longer durations in réitérant speech, therefore, can be considered to be an effect of the difficulties experienced by the speaker having to reiterate syllables to form long stress groups.

Few studies have been devoted to direct comparisons of réitérant and normal speech. In addition to the data of Svensson (1974-), a study of Liberman and Streeter (1978) included observations of both normal and réitérant versions of sentences. However, the main purpose of that study was to examine the stability and reliability of réitérant speech as modelled on different normal sentences. The crucial point of the correspondence between normal and réitérant speech was not considered.

3.4.2 The basic material and the test interval

The study is based on sets of systematically varied semantically acceptable or nearly acceptable sentences. Some semantically odd (though phonologically and syntactically regular) sentences had to be included to fill gaps in the word inventory required by the experimental variation.

The basic material. Even if the different problems addressed require different sets of sentences, efforts were made to reduce all but the necessary variation. Therefore, a specific set of sentences, the basic material, was constructed so that it could be used throughout the study with certain modifications required by the specific experimental conditions.
The basic material consists of five sentences with four stress groups in each. In the second stress group the syllable number is varied from one to five with the stressed syllable leading the group:

böt \[b\ddot{\text{o}}:t\]  
byter \[b\ddot{\text{y}}:\text{ter}\]  
byter väl \[b\ddot{\text{y}}:\text{terve}l\]  
byter välan \[b\ddot{\text{y}}:\text{terve}lan\]  
byter välan då \[b\ddot{\text{y}}:\text{terve}l\text{and}o\]  

The varied stress group occurred within the following constant frame:

Bönderna ---- bilen i Bergamo.  
'The peasants ---- the car in Bergamo.'

The choice of b as the initial sound of all stressed syllables was made partly to facilitate segmentation and partly because stress beats associated with vowels seem to vary according to the preceding consonant (Rapp, 1971; Morton et al., 1976). The choice of b throughout should therefore reduce undesired variation in the material.

The second stress group contains a verb in the past or present tense, böt /b\ddot{\text{o}}:t/ 'changed', and byter /by:ter/ 'changes', respectively. In the sentences with three, four or five syllables in the second stress group, words with low semantic weight, particles which express an attitude of undecidedness, have been added. The word då /do:/ 'then', which can be both a temporal adverb (with stress) and a particle (without stress), was read in the test material as a particle.

In the construction of the material it was required that the addition of syllables should not lead to modifications of the preceding syllables. This general requirement, however, could not be met fully if all sentences were to be semantically acceptable. The change of vowel quality in sentences with the syllables böt and byt-, respectively, introduced a difference in inherent vowel duration. This difference was compensated for throughout the duration data by a subtraction of 28 ms from the duration of böt. (28 ms corresponds to
the mean durational difference obtained by Elert (1964, p.95-96) for the vowels /ø:/ and /y:/ in monosyllabic words contained within sentences. Checks on the speakers in the present study confirmed this estimation.

The test interval. The second interstress interval in the test sentences with its variation of the number of syllables, is the focus of the experimental study. With minor exceptions all measurements were made on this part of the sentences. This interval, therefore, will be referred to as the test interval.

3.5 Speakers

Five speakers (labelled as CH, EEJ, ES, EEN and BH) served as subjects in the study. They were all university teachers or researchers, female, and 35-45 years of age. Having grown up in Stockholm they were all representative of that variety of Swedish. CH, EEJ and ES served as subjects in all experiments, while EEN and BH participated only in a part of them.

3.6 Recordings and registrations

The test sentences occurred in random order on lists. The speakers read the sentences in a neutral way without any kind of emphasis or contrast. To avoid hesitations within sentences the speakers were instructed to pause between each sentence to plan the next sentence as a whole before reading it out.

The sentences were recorded on tape, processed by a Fonema speech analysis unit, and registered on a Siemens mingograph with a paper speed of 100 mm/s. Durations were measured from the duplex and two intensity curves (full range and low-pass filtered) to the nearest 5 ms.
3.7 Segmentation, measurements, and statistics

Segmentation and measurements. There were three types of measurements: (1) the total duration of the test interval, (2) the duration of syllables, and (3) the duration of vowels and consonants. The following criteria were used for their delimitation:

(1) The total duration of the test interval was measured from the onset of the stressed vowel to the onset of the next stressed one. These segmentation points were chosen because previous research (Rapp, 1971; Huggins, 1972; Morton et al., 1976, Marcus, 1979; Carlson et al., 1979) has shown that the center of rhythm in production as well as in perception appears to be tied to the onsets of stressed vowels.

(2) The duration of syllables was measured from the onset of one vowel to the onset of the next one. Thus, syllables in the present study correspond to VC(C) sequences. Support for the use of this type of temporal unit stems from e.g. Huggins (1972), Bannert (1979), Kohler et al. (1982), Strangert (1983), Niemi (1984).

(3) Boundaries between vowels (V) and the consonants (b, d, t, v, l, n, r) occurring in the test interval were determined according to the following criteria. (Consonants occurring in sequence were not separated from each other.)

The boundary between the vowel and the following consonant was set at

\[
V \rightarrow /b,d,t/ \text{ the point where intensity started to fall rapidly} \\
/b,d/ \rightarrow V \text{ the explosion of the consonant} \\
/t/ \rightarrow V \text{ the point where there was a rise of intensity associated with the vowel}
\]

There are thus different definitions for the respective endings of voiced and unvoiced plosives. The interval after the release of the
consonant is included in the unvoiced plosive interval, while it is referred to the vowel when the plosive is voiced.

\[ /v,l,n,r/ - V \] and \[ V - /v,l,n,r/ \] The boundaries between the vowel and the consonant were set at the rapid increase or decrease of intensity associated with the beginning or ending of the vowel, respectively.

These boundaries are fairly easy to determine from discontinuities in the duplex curve.

**Statistics.** All calculations were based on ten readings of each sentence. For every type of measured duration the basic statistics included the arithmetic mean and standard deviation. The level of confidence indicated in the figures is .95. For details of additional statistics, see the specific experimental parts.
4 THE BASIC EXPERIMENTAL VARIATION: EFFECTS OF THE NUMBER OF SYLLABLES IN THE TEST INTERVAL

4.1 Purpose

The basic variation throughout the Swedish main study concerns the number of syllables contained within the test interval. The first experiment is therefore partly preparatory, aiming at a descriptive baseline for the following experiments. The purpose is to give an adequate description of the test interval and its component parts in the basic material (3.4.2). In addition, some assumptions related to the model outlined in Chapter 2 will be examined.

First, the temporal relations between the total duration of the test interval, when it contains from one to five syllables, will be explored.

The data should also reveal whether there are dependencies between the segment durations and the number of syllables in the test interval in the medial position of phrases. Such dependencies for Swedish have been refuted by Lyberg (1977, 1981) while on the other hand observed by Strangert (1979). The presence of such dependencies may be interpreted as evidence for processing constraints manifested as temporal adjustments (2.5).

Another type of temporal patterning to be examined is rhythmic alternation between unstressed syllables. It is assumed that such alternation would be reflected within the long stretches of unstressed syllables (up to four) contained within the test material (cf. Strangert, 1981; Bruce, 1983, 1984). In the model outlined in section 2.4 rhythmic alternation is seen as a consequence of restrictions on the capacity for articulatory programming.

The following aspects of the temporal structure of the test intervals were examined:
(1) The duration of the test interval as a function of the number of syllables it contains.
(2) Effects on stressed and unstressed vowel duration of the number of syllables in the test interval.
(3) The relation between rhythmic patterning and language structure.

Some of the data in this chapter have been reported in a previous study (Strangert, 1984).

4.2 Procedure

All five speakers participated in the test.

The measurements were based on readings of the following sentences, identical to the basic material described in detail in 3.4.2:

Bönderna bót bilen i Bergamo.
Bönderna byter bilen i Bergamo.
Bönderna byter väl bilen i Bergamo.
Bönderna byter välan bilen i Bergamo.
Bönderna byter välan då bilen i Bergamo.

Two of the speakers (CH and ES) also read another version of the sentences in which the order of välan and då was reversed. Thus, the following sentences were included, too:

Bönderna byter då bilen i Bergamo.
Bönderna byter då väl bilen i Bergamo.
Bönderna byter då välan bilen i Bergamo.

Measurements of the test interval, underlined above, included the total duration and the durations of the stressed and unstressed vowels. For segmentation and measurements, see 3.7.

The basic material was read together with two different sets of sentences including one and five syllables respectively in the stress
4.3 The total duration of the test interval as a function of the number of syllables

Figure 4-1 presents the durations of the test interval as a function of the number of syllables it contains for all five speakers. The data refer to the readings of the basic version of the test sentences (see 4.2).

**Figure 4-1.** Duration of the test interval as a function of the number of syllables (basic version of the test sentences).
The relation between the duration of monosyllabic and polysyllabic intervals. The typical curve for all speakers starts at a relatively high value, indicating a long monosyllabic duration. The increase associated with each additional syllable is small in relation to this interval, with duration ratios ranging between extreme values of 1.08 and 2.44 for the duration of intervals with two and five syllables respectively, to monosyllabic intervals (Table 4-1).

Table 4-1. Duration ratios for interstress intervals with 1-5 syllables. Ratios calculated as the duration of intervals with 2-5 syllables in relation to the duration of the monosyllabic intervals.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Number of syllables</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>CH</td>
<td>1.00</td>
</tr>
<tr>
<td>EES</td>
<td>1.00</td>
</tr>
<tr>
<td>ES</td>
<td>1.00</td>
</tr>
<tr>
<td>EEN</td>
<td>1.00</td>
</tr>
<tr>
<td>BH</td>
<td>1.00</td>
</tr>
</tbody>
</table>

It is apparent that all speakers uttered the five-syllable interval within little more than double the time for the monosyllabic interval. Intervals with different numbers of syllables thus cover a relatively restricted range of durations.

These findings are in agreement with observations based on a Swedish prose text (Strangert, 1981). They may therefore be considered representative for Swedish. Similar relations between monosyllabic and longer intervals have been observed for English, too, for example by Bolinger (1965), Uldall (1971), and Dauer (1979, 1983).

Influences of language structure. The long monosyllabic duration may be seen, at least partly, as a consequence of the quantity distinction of stressed vowels in Swedish. Phonologically long vowels (as in the present study) must have long durations in order to be separated from phonologically short ones. Long vowel durations, therefore,
should be associated with long monosyllabic interstress intervals. Furthermore, because of the complementarity between the vowel and the following consonant, short stressed vowels followed by long consonants should be associated with long monosyllabic interstress intervals, too. Assuming that the duration of interstress intervals reflects the rhythm of a language, one might expect to find that structural constraints such as, for instance, the distinction between phonologically long and short vowels, could influence rhythm to a high degree.

The functions in Figure 4-1 have a weak tendency to positive acceleration, which is primarily a consequence of the relatively small increase in duration at the step between a monosyllabic and a polysyllabic interval. However, the specific form of the curves is to a certain extent determined by the segmental composition of the syllables contained in the intervals. The weak, almost neutral character of -er, added to the monosyllabic interval might very well be one such source. Such weak unstressed syllables very often occur after stressed syllables and should be considered a characteristic feature of Swedish. Other moderations of the curves over the different intervals should in a similar way be attributed to the different segmental composition of the added syllables.

Inter- and intraspeaker variation. The intraspeaker variation expressed as the standard error for each separate interval was small and in most cases within the limits of measurement error. The range over the different intervals for each speaker was: 3-7 ms (CH), 5-8 ms (EEJ), 3-7 ms (ES), 5-19 ms (EEN), and 3-12 ms (BH). There was no systematic relation between the standard errors and the length of the test interval.

The absolute durations as shown in Figure 4-1 differed clearly between the speakers within a range of less than 200 ms for each length of the interstress interval. However, in spite of this variation associated with different speech rates, interstress intervals varied systematically in a similar way for all speakers as a function of the number of syllables in the interval. Thus, the data reveal a considerable degree of stability both within and between speakers.
4.4 Effects on vowel duration of the length of the test interval

The durations of the vowels occurring in the first (stressed) through fifth position of the test interval, when it contains from one to five syllables, are given in Table 4-2. The data (means and standard deviations) refer to the basic versions (see 4.2) as read by all five speakers. The same type of data for CH and ES are given in Table 4-3 for the alternative version of the sentences, with välan and då appearing in the opposite order to that in the basic version. Figure 4-2 (a-e) and 4-3 (a-b) depict the same data.

Table 4-2. Means and standard deviations for the duration of vowels as a function of the position and the number of syllables (N) in the test interval (basic version).

<table>
<thead>
<tr>
<th>Position: Speaker N</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>EEJ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>164</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>149</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>146</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>142</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>142</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EEN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>162</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>137</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>128</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>128</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>158</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>121</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>103</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>104</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>109</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>109</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>183</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>183</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>129</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>126</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>122</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ES</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>183</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>138</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>129</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>126</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>122</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4-3. Means and standard deviations for the duration of vowels as a function of the position and the number of syllables (N) in the test interval (alternative version).

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Position</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>CH</td>
<td></td>
<td>119</td>
<td>33</td>
<td>5</td>
<td>78</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>116</td>
<td>36</td>
<td>9</td>
<td>92</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td></td>
<td>120</td>
<td>33</td>
<td>8</td>
<td>97</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>120</td>
<td>43</td>
<td>4</td>
<td>93</td>
<td>7</td>
</tr>
<tr>
<td>ES</td>
<td></td>
<td>120</td>
<td>44</td>
<td>3</td>
<td>102</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>120</td>
<td>47</td>
<td>3</td>
<td>108</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>120</td>
<td>8</td>
<td>7</td>
<td>83</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>120</td>
<td>5</td>
<td>4</td>
<td>102</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>120</td>
<td>6</td>
<td>5</td>
<td>108</td>
<td>5</td>
</tr>
</tbody>
</table>

Figure 4-2 (a-e). Mean duration of the stressed and following unstressed vowels as a function of position and number of syllables in the test interval (basic version).
Figure 4-3 (a-b). Mean duration of the stressed and following unstressed vowels as a function of position and number of syllables in the test interval (alternative version)
Two different tendencies appear from the data. Some vowels were shortened when new syllables were added to the interval, while others, on the other hand, were lengthened. ("Shortening" and "lengthening", respectively, indicate a decrease and increase of the duration for a certain position of the vowel.) Shortening was most marked in the first three positions of the interval. However, when the test interval contained more than three syllables, the vowel in the third position was sometimes lengthened. In longer intervals, there was also an increase in the duration of the fourth or fifth syllable.

4.4.1 Shortening effects

As mentioned above, shortening was most marked in the first three positions of the test interval, specifically in the stressed syllable and the following unstressed one. Generally, the most drastic effects in the stressed vowel were found at the step between one and two syllables, and in the first unstressed syllable at the step between two and three. This was not true for EEJ to the same extent as for the other speakers. For EEJ the first unstressed vowel was shortened on the addition of a new syllable, but the stressed one was relatively unaffected by the length of the test interval. However, both CH and EEJ shortened the second unstressed vowel (väl). For BH the shortening was more gradual than for the other speakers. For each syllable added to the interval, the duration of the stressed and the first unstressed vowel decreased successively.

Incompressibility and language structure. For all speakers except BH, there was a strong tendency for vowel duration to level out - that is to reach an asymptote after an initial drastic fall. This tendency was obvious in the stressed as well as in the following unstressed syllable, and for CH and EEJ also in the second unstressed one (see above). However, the asymptote values were clearly differentiated for the stressed and unstressed syllables. The duration of the stressed vowel leveled out at considerably higher values (about 50-100 ms) than the unstressed vowels. The asymptote values differed also be-
tween the different unstressed vowels.

This tendency for vowels to reach a point beyond which there is no further shortening has been treated by Klatt (1973), and later by Lindblom et al. (1976) under the heading of "incompressibility". Incompressibility is assumed to reflect constraints on articulatory movement. The asymptote value is considered to reflect a minimum execution duration for the vowel (1.3.1).

However, the minimum duration apparently differs for stressed and unstressed vowels and for different unstressed vowels as observed in the present study and previously by Klatt (1973). This makes incompressibility a dubious phenomenon, especially, as in certain conditions, (unstressed vowels, high speech rate, casual speech) vowels may be completely deleted. Therefore, the different asymptotic values of the stressed and unstressed vowels mainly reflect their different stress characteristics. Vowels to be perceived as stressed should thus level out at a higher value than unstressed ones.

Comparisons with other studies. The shortening effects observed in the present study support earlier findings for Swedish (Strangert, 1979) as well as for languages like English and German (1.2.2). They are, however, somewhat unexpected when seen in the light of the conclusions found in Lyberg (1977, 1981) that temporal adjustments are restricted to phrase-final position (1.2.2; 1.3.1).

It should be remembered (see 1.2.2) that in Lyberg's data also, durational changes occurred in other positions than phrase-finally. A decrease of duration occurred systematically at the step between words with two and three syllables. Examination of the data reveals moreover that one of the three speakers in the study, made in effect, adjustments similar to those in the present study. In conclusion then, it appears that in Lyberg's data also there was some evidence for a decrease of stressed vowel duration, when the number of unstressed syllables was increased.
The more restricted effects in Lyberg's study, as compared to the present one, might be due to the different types of material used as well as the different units studied (words and stress groups respectively). Boundaries of words and constituents in Lyberg's study occurred within stress groups as defined in the present work. As it appears that linguistic boundaries might be reflected in the temporal pattern of speech, though to a varying degree, depending on the situation and the speaker (1.4 and 5.2.3), this might well have influenced the speakers in Lyberg's study.

Alternative interpretations of the observations. The data are compatible with the assumption of temporal adjustments related to rhythmic constraints on the planning mechanism. However, they do not allow any decisions as to the type of adjustment involved, since the observations may be interpreted as adjustments to the regularity of phrases as well as stress groups (cf. 2.5). Contrasting these two types of adjustment will be postponed until Chapter 7.

Viewed as adjustments (to phrases, or stress groups, or both) the temporal changes hitherto referred to technically as shortenings may well be seen as either shortening or lengthening effects. For example, segments in polysyllabic intervals could be seen as shortened (to the length of the phrase or the stress group). This more neutral approach to the shortening/lengthening dichotomy will be adopted henceforth in relation to the observed adjustment effects.

Now, are there alternative interpretations of these data? Are the durational changes really adjustments to the regularity of stressed syllables or phrases? Might they, for example, be considered as being due to preboundary lengthening? This seems not unreasonable as the durational changes are considerable when a vowel is moved forward from the final position upon the addition of a new syllable. The presence of a major syntactic boundary at the end of the test interval also supports such an assumption (cf. 1.4). However, even if the data are described in terms of lengthening, this cannot explain all of the changes. First, lengthening in the present data seems to be restricted to the stressed and the following one or two unstressed
vowels. But one would expect preboundary lengthening to occur for all syllables, when in final position. Furthermore, durational changes occurred even between different non-final positions. This should not be expected if preboundary lengthening is assumed to be the main source of the durational changes. We therefore assume that the changes result from adjustments to the regularity of phrases or stress groups (or both).

4.4.2 Lengthening effects

As has been pointed out, lengthening affected the vowel of -an, and in some cases also väl-. (Henceforth, vowel duration is intended, when referring to the duration of väl-, -an, and då.) In Figure 4-2 and 4-3 lengthening may be seen as upward steps in the curves for a certain position, when a new syllable is added to the test interval. Let us first examine Figure 4-2.

Alternation pattern and interspeaker variability. Obviously this type of lengthening, observed primarily in the third and fourth position, occurred when the test interval contained four or five syllables. Most probably it reflects a tendency to differentiate between unstressed syllables, giving extra weight to some in a sequence of several that are unstressed. Thus, unstressed syllables will alternate with stressed ones, producing a subgrouping within stress groups, or alternatively, the splitting of a stress group into two new ones.

Though all five subjects exhibited some kind of lengthening, each had a different pattern. For CH, EEJ, ES, and BH there was no significant lengthening in the test interval with four as compared to three syllables. At the same time -an was long as compared to väl-, probably indicating that -an had been given a certain degree of prominence. For EEN, on the other hand, -an was not longer than väl-, probably because väl- had been given some prominence. In the interval with five syllables, as compared to that with four, lengthening was most marked for -an (CH, EEJ, EEN). BH and ES had weaker and more
diffuse lengthening patterns. Thus, though there was a clear tendency to lengthening of some syllables in a sequence of unstressed ones, the pattern was not very rigid, allowing for variability between different speakers.

In spite of this variability, however, the syllable given prominence was in all cases separated from the stressed one by at most two unstressed syllables. The most frequent patterns, with prominence indicated by +, may be diagrammed schematically as:

- \text{-} \text{-} \text{-} \text{-} \text{+} \quad \text{Four-syllable intervals}
- \text{-} \text{-} \text{-} \text{+} \quad \text{+} \quad \text{Five-syllable intervals}

These patterns resemble those in traditional descriptions of Swedish stress, according to which either one or two weak syllables intervene between two stressed ones (3.1).

\textbf{Influences of language structure.} On the basis of the present data it seems reasonable to conclude that the alternation pattern of a stress group is to a considerable degree influenced by the structural properties of the elements it contains. This seems evident, when the word välan is considered. When spoken in isolation it may be stressed on either the first or the second syllable. This peculiarity seems to have been reflected in the present data, as some of the speakers tended to favor an alternation pattern with prominence primarily on väl-, while the others (the majority) seem to have treated -an as the element to stress. Thus, given that maximally two unstressed syllables intervene between two stressed ones, there seems to be a certain flexibility of the rhythmic alternation, allowing both for structural influences and variation between different speakers.

The emerging pattern is thus much more complex than that in the studies of Bruce (1983, 1984). However, Bruce claims to have considered only the simplest cases of alternation, which explains the difference between his studies and the present one.
**Supplementary test.** To test the generality of the observed patterns the data in Figure 4-3 may be examined. The observations refer to the alternative readings of the test sentences with då and välan appearing in opposite order to that depicted in Figure 4-2. Thus då occurred in the third position of the test interval as compared to the fifth in the basic material.

It appears that the same type of lengthening occurred when word order was reversed. However, now the lengthening was most marked in the third position (då) as compared to the fourth position (-an) in the previous data. Also -an had about the same duration in the two sentence version, which seems to indicate that it was stressed to about the same degree irrespective of its position. Thus, the five-syllable interval contained two prominence peaks (on the third and fifth syllable) as reflected in the lengthening patterns. As this realization of the alternation pattern undoubtedly depends on the characteristics of the word välan, it demonstrates a strong interaction between language structure and rhythmic alternation.

### 4.5 Conclusions

In test intervals with several syllables at least one of the unstressed vowels were lengthened. This lengthening was associated with the phenomenon of rhythmic alternation. Test intervals with many syllables were subgrouped into smaller units containing one stressed and one or two unstressed syllables. Such a subdivision would be expected on the assumption that articulatory programming is adapted to a motor execution system with limited capacity (2.4). The various types of subgrouping indicated an interaction with language structure constraints.

In stress groups with less than four syllables there was a decrease of duration of stressed and unstressed vowels, when the number of syllables increased. This indicated temporal adjustments either to the length of the phrase or the length of adjacent stress groups.
5 EFFECTS OF VARIATIONS OF LANGUAGE STRUCTURE: QUANTITY AND BOUNDARIES

5.1 Experiment I: Effects of quantity

5.1.1 Purpose

The study is supplementary to the previous one which focussed on the duration of vowels in stressed and unstressed syllables. As a basis for the analysis of the distinction between stressed and unstressed syllables in Chapter 6, the properties of the stressed syllable will now be considered in more detail. The durations of phonologically long and short stressed vowels and the following consonants will be examined under the same conditions as in the previous study, that is, as a function of the number of syllables in the test interval. Owing to the complementarity between the vowel and the following consonant the sequences to be observed consists either of a short vowel followed by a long consonant (VC:), or a long vowel followed by a short consonant (V:C). Thus, "syllables" in this study are equivalent to VC-sequences.

The first problem addressed is whether vowels and consonants (long and short) are affected in a similar way, when the test interval increases in length. On the assumption that the complementarity is an important aspect of the quantity distinction we shall expect durational changes to occur in such a way that the relation between the vowel and the following consonant is not disrupted. This should be reflected in an invariant ratio between the vowel and the consonant.

Further, though the duration of a sequence of a stressed vowel and a following consonant (VC: or V:C) should change with a variation of the number of syllables in the test interval, a stable relation between VC: and V:C is assumed. This relation should approach perfect complementarity (a ratio of 1:1), as according to Elert (1964, p.158) sequences of VC: and V:C have about the same duration. (The V:C-se-
quence, in general, appears to be somewhat longer than the VC:-sequence.)

The following aspects of the manifestation of the quantity distinction in differently-sized test intervals will thus be examined:

(1) The duration of the stressed vowel V(:) and the following consonant C(:) and their temporal relation in sequences of VC: and V:C.

(2) The duration of VC:- and V:C-sequences and their temporal relation.

5.1.2 Procedure

Three of the speakers, CH, EEJ, and ES, participated in the study.

The material differed on two points from the basic material. First the word bilen 'the car' in the sentence frame was exchanged for enen 'the juniper' to facilitate segmentation of the consonant following the stressed vowel in monosyllabic test intervals. Secondly, the verb forms in the test interval were not identical to those in the basic material. They now included long and short varieties of the vowel /y/ followed by /t/. Due to the complementarity between the vowel and the following consonant the test sentences were:

VC: *Bönderna být enen i Bergamo.
    *Bönderna býtter enen i Bergamo.
    *Bönderna býtter väl enen i Bergamo.
    *Bönderna býtter välan enen i Bergamo.
    *Bönderna býtter välan då enen i Bergamo.

V:C *Bönderna být enen i Bergamo.
    Bönderna býter enen i Bergamo.
    Bönderna býter väl enen i Bergamo.
    Bönderna býter välan enen i Bergamo.
    Bönderna býter välan då enen i Bergamo.

*In sentences marked by asterisks the verbs are not real Swedish words.
Only the stressed vowel and the following consonant were measured. For segmentation and measurements, see 3.7.

5.1.3 Results: The duration of vowels and consonants

The duration of phonologically long and short /y/ and the following complementary short and long /t/, has been diagrammed in Figure 5-1 (a-c) for the three speakers.

The durational changes of the vowels appeared to be somewhat more gradual than in the previous study (4.4.1). At least this was true for CH and ES. Though the absolute changes were also much greater for CH and ES as compared to EEJ, they were nevertheless quite similar for phonologically long and short vowels for all three speakers. Thus it seems that the length of the vowel had no influence on the degree to which it was affected by the addition of unstressed syllables to the test interval.

The consonants decreased in duration to a greater degree than the vowels. This was true for all of the speakers, though, as was the case for the vowels, CH and ES exhibited greater effects than EEJ. The long variant of the consonant was affected to about the same degree as the short one. Also, for CH and ES both the long and short consonant shortened gradually as a function of the number of syllables in the test interval. This gradual decrease supports the previous assumption (4.4.1) of the presence of durational adjustments to the length of the stress group (or the phrase).

The greatest durational change of the consonant occurred for all three speakers at the step between a monosyllabic and a bisyllabic interval. For CH and ES the duration of the consonant appeared to be associated with the very long interval between the explosion of the consonant and the following vowel onset. This may be observed for CH in Figure 5-2, which also shows a (probably compensatory) short duration of the occlusive interval of the long consonant, representative also for ES. For EEJ the long duration of the monosyllabic interval
Figure 5-1 (a-c). Mean durations of the short and long stressed vowel (upper part) and the following long and short consonant (lower part) as a function of the number of syllables in the test interval.
was brought about by a long occlusive interval without any contribution from the interval between the explosion and the vowel onset. Thus, for all three speakers the consonant interval was longest in the monosyllabic interstress interval, but the strategy used by CH and ES apparently resulted in the longest duration.

![Graph](image)

**Figure 5-2.** Mean duration of the occlusion of the consonant and the interval between the explosion and the following vowel onset as a function of the number of syllables in the test interval. Data for long and short consonants following the stressed vowel.

**Shortening versus lengthening.** One possible explanation for the very long consonant interval rests on the assumption that speakers, in effect, attempt to make stressed vowels (and especially their onsets) recur regularly. While for longer interstress intervals this should imply shortening of the vowels and consonants contained in it, for monosyllabic intervals lengthening might be the solution to regular recurrence.
There is some evidence for such lengthening in the present data, especially for CH and ES. The very long interval between the explosion and the following vowel contained, in addition to the fricative noise phase, either a silent interval, or what should be considered as "false starts" of the following vowel.

5.1.4 The temporal relation between the vowel and the following consonant

The ratios between the long and short vowels and the following consonants for test intervals with one to five syllables are shown in Table 5-1 for the three speakers.

Except for the monosyllabic interval, the ratios of V:C: and V:/C are relatively stable over the different lengths of the test interval. However, for both CH and ES the ratio of V:/C in the monosyllabic interval is lower than the ratios for polysyllabic intervals containing the long vowel. For CH also the ratio of V:C: is lower in the monosyllabic as compared to longer intervals.

Table 5-1. Ratios between the vowel and the following consonant (V:C: and V:/C) in test intervals with one to five syllables.

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Ratio</th>
<th>Number of syllables in the test interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>CH</td>
<td>V:C:</td>
<td>.48</td>
</tr>
<tr>
<td></td>
<td>V:/C</td>
<td>.81</td>
</tr>
<tr>
<td>EEJ</td>
<td>V:C:</td>
<td>.52</td>
</tr>
<tr>
<td></td>
<td>V:/C</td>
<td>1.07</td>
</tr>
<tr>
<td>ES</td>
<td>V:C:</td>
<td>.55</td>
</tr>
<tr>
<td></td>
<td>V:/C</td>
<td>.87</td>
</tr>
</tbody>
</table>
This exceptional patterning is a consequence of the very long monosyllabic consonant duration of CH and ES. EE3, with more stable consonant durations, also exhibits more similar monosyllabic and polysyllabic VC-ratios. For ES a consequence of the long consonant duration is a weakening of the distinction between V/C: and V:/C, while for CH it is unchanged, owing to the similar adjustments of the V/C:- and V:/C-ratios.

However, in spite of the differences observed in differently-sized intervals, VC:-sequences were always clearly separated from V:C-sequences. (Theoretically, a strengthening of the distinction between the two types of VC-sequences could have been expected to occur in the monosyllabic interval rather than the stability or weakening observed. According to Bannert (1979), strengthening seems to be required for perceptual reasons when the VC-sequences increase in duration. Such a strengthening seems to have been overruled in the present study by the long monosyllabic consonant duration.)

5.1.5 The temporal relation between VC:- and V:C-sequences

Figure 5-3 shows the durations of the VC:-sequences plotted against the durations of the V:C-sequences for the three speakers. Only the slopes for CH and ES have been indicated, as for EEJ the data cluster around a small area rather than exhibiting any constant relationship. Within this area the ratios of VC:-sequences to V:C-sequences cluster around 1.0 with the following values for intervals with one to five syllables: 1.02, .98, .95, 1.00, .99.

The slopes approach one for CH and ES, which confirms the complementarity between the vowel and the consonant. This, in addition to the almost perfect complementarity as revealed by the data for EEJ, points to the stability of the VC-unit across the variation of the test interval. The relevance of this unit has earlier been claimed by e.g. Bannert (1979), Kohler et al. (1982), and Strangert (1983).
5.2 Experiment II: Effects of boundary variation

5.2.1 Purpose

The evidence for temporal correlates of boundaries between words and constituents is conflicting (1.5.2 and the review in Fischer-Jørgensen, 1982). It would be most reasonable for the different results to be seen as an indication that boundaries may, but need not be manifested. It seems that whether they are attended to or not depends on pragmatic factors related to the entire speech situation (de Rooij, 1979). For example, a desire on the part of the speaker to give extra prominence to a word or a phrase, or to contrast two words or phrases will necessarily be reflected in the speech signal. In the present study, therefore, the purpose was to find out whether the
insertion of a syntactic boundary within a stress group will affect its temporal structure in relatively neutral renderings of sentences.

5.2.2 Procedure

Three of the speakers, CH, EEJ, and ES, participated in the study.

The observations were based on a material in which the test interval contained two versions of a two-syllable sequence. In one, boundaries between constituents and words coincided with the boundaries of the stress group (as defined in the present study), while in the other a combined word and syntactic boundary occurred within the stress group. The two alternatives were:

```
----- # böter # -----  
----- # bôt # er-----
```

The two test sequences were inserted within ten sentence frames, all with the same prosodic structure, but with a variation of the initial and final words. Two examples will be given:

```
Barabbas ----- bilen med bönderna.
'Barabbas ----- the car with the peasants.'
Barbara ----- bilen i Brandenburg.
'Barbara ----- the car in Brandenburg.'
```

The boundary position in the first test sequence made the unstressed er the last part of *böter. Though not a real Swedish word, the speakers experienced no difficulty when asked to produce this word as a normal Swedish verb. The boundary position in the second test sequence made er the first part of the word *erbilen. Neither was this a real Swedish word. However, the speakers treated it without difficulty as a medially stressed noun.

The different sentence frames were chosen so as to attract the speaker's attention. This measure was taken in order to avoid contrastive
renderings of the different versions of the test sequence.

The duration of the vowels and the consonants within the test sequences as well as the /iː/ of the word bilen in the sentence frame were measured. For segmentation and measurements, see 3.7.

5.2.3 Data and interpretations

The duration of the vowels and consonants in the test sequences and the following /iː/ is shown in Figure 5-4 for the three speakers.

![Figure 5-4](image-url)

Figure 5-4. Effects of boundary insertion: The durations of vowels and consonants in two test sequences. (Durations of the following /iː/ are included.)

Although measures were taken to eliminate the influence of the boundary insertion, durational differences nevertheless existed between the two test sequences. They were, however, small, ranging from about 50 ms (EEJ) to 0 ms (ES), if the entire interval is considered.
The lengthening of /e/, the first segment of *erbilen was evident for all of the speakers. Moreover, EEJ and CH both lengthened the /ø:/ of böt though EEJ more than CH. EEJ also lengthened the /t/ after /ø:/ and the /t/ after /ø:/ in the monosyllabic test interval for CH and ES in the study of the quantity distinction (5.1.3). That is, the interval between the explosion of the consonant and the following vowel onset was increased.

It seems that this lengthening of EEJ was the result of the boundary insertion. This is evident from the fact that EEJ did not exhibit the same kind of lengthening in the preceding study.

The temporal realization of the boundary may be the result of an effort to distinguish between the two test sequences. Similar adjustments to achieve contrast effects have been observed by Lehiste (1973). She studied what might be called "phonetic disambiguation", that is, the means by which speakers differentiate between syntactically ambiguous sentences (see 1.4). Lehiste's result indicated that the disambiguation was achieved either by the insertion of a pause, or by the lengthening of one or more segments before the boundary. Lehiste assumed these alternatives to be equally possible means of achieving the same end: an increase of the interstress interval. In a later study (Lehiste et al., 1976) it was concluded that such an increase would signal the presence of a boundary because it leads to a deviation from an expected pattern. That is, listeners expect a certain amount of temporal regularity between stressed syllables, and irregularities therefore signal structural complexities.

This interpretation is neither supported, nor should it be rejected on the basis of the present data. It seems, however, that lengthening at a syntactic boundary might just as well imply a boundary between two production units. That is, syntactic boundaries might sometimes take precedence over stressed syllables as criteria for grouping syllables together.
However, the absence of any extensive lengthening of bötı for CH and ES in the present study as compared to the very long, perhaps lengthened, monosyllabic interval in the preceding one (5.1.3) points to the possibility of a certain independence between rhythm and syntax. This seems a reasonable conclusion, since if the boundary insertion had really influenced rhythmic patterning, the same kind of durational differences as for byt/byter in the preceding study would be expected for bötı/böter in the present one. The optionality of the marking of syntactic boundaries is additionally supported by the fact that the test situation (lists with sentences in which both test sequences occurred) would most reasonably favor a contrastive rendering of the alternatives.

5.3 Conclusions

The first experiment showed the stressed vowel (phonologically short or long) and the following consonant to be affected in a similar way, when the test interval increased in length. As a consequence, the relation between the vowel and the consonant was preserved in spite of the variation in the number of following syllables. Minor exceptions to this pattern occurred in the monosyllabic interval.

Irrespective of the phonological length of the vowel, the stressed syllable (VC-sequence) appeared as a stable temporal unit. This stability of the stressed VC-unit makes it a reliable basis for rhythmic grouping.

The second experiment demonstrated an interaction between stress and boundaries as a basis for grouping syllables into rhythmic units.
6 EFFECTS OF VARIATIONS IN SPEECH TEMPO

6.1 Purpose

The study concerns variations in overall speech tempo. These variations may be seen as intentionally guided by the speaker's perception of the communicative demands in the speech situation (2.4). Thus, adjustment of speech tempo is in this respect relatively independent of the linguistic form of the utterance.

In the present study the tempo was manipulated by instructing the subjects to speak at a fast, normal, and slow rate. The purpose was to investigate how temporal transformations were made when speech changed from normal to fast and slow rates, respectively. The observations concerned the total duration of the test interval and the duration of stressed and unstressed syllables within the interval.

The following hypotheses were tested:

(1) The duration of the test interval changes proportionally with a change of speech rate (cf. Kozhevnikov & Chistovich, 1965, p.82-87). This proportionality should hold irrespective of the number of syllables in the test interval.

(2) The distinction (temporal relation) between stressed and unstressed syllables is assumed to be unaffected by variations of speech rate.

6.2 Procedure

The study was based on three types of reading of the basic material (3.4.2), normal, fast, and slow, by three of the speakers (CH, EEJ, ES).
The normal data are identical to the data on the basic material presented in Chapter 4. "Normal", therefore, refers to the rate preferred by each speaker when not specifically asked to adjust her speech tempo. "Fast" and "slow" refer respectively to the data produced when the speakers were instructed to read either as fast or as slowly as they could. The speakers were also instructed to pronounce the sentences fluently and without reductions, that is, to avoid segment omissions at the fast rate, and hesitations and pauses at the slow rate.

Measurements were made of the duration of the test interval as well as of the duration of the syllables (VC-sequences). For segmentation and measurements, see 3.7.

6.3 Results: The total duration of the test interval

The absolute durations for normal, fast, and slow speech of course differed from speaker to speaker. However, these differences are unimportant for the problem at issue, which concerns rather the intraspeaker variation in speech rate.

Table 6-1 shows the grand mean of the durations of the five test intervals for each rate and speaker. The relative variability, as measured by the coefficient of variation, is fairly similar for the different rates and speakers. The normal rate gives the least variability.

Table 6-1. Grand means (M) and coefficients of variation (v) for the durations of five test intervals for fast, normal, and slow speech rate.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Speech rate:</th>
<th>FAST</th>
<th>NORMAL</th>
<th>SLOW</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M  v</td>
<td>M</td>
<td>v</td>
<td>M</td>
</tr>
<tr>
<td>CH</td>
<td>411 6</td>
<td>466 4</td>
<td></td>
<td>702 7</td>
</tr>
<tr>
<td>EEJ</td>
<td>435 5</td>
<td>559 4</td>
<td></td>
<td>775 8</td>
</tr>
<tr>
<td>ES</td>
<td>427 6</td>
<td>560 3</td>
<td></td>
<td>931 7</td>
</tr>
</tbody>
</table>
It appears that the step between fast and normal speech was much smaller than that between normal and slow speech for all three speakers. This is not surprising, since there are far more obvious limits to fast speech than to slow.

The transformations from normal to fast and slow rates for the entire range of test intervals are illustrated in Figure 6-1 (a-c). The plot shows the linear relationship between the corresponding five test intervals at the different rates. In general the data confirm the hypothesized proportional change of the duration of interstress intervals as a function of speech rate. At a fast rate, the test interval was proportionally reduced by 6 and 17 per cent for CH and ES, as compared to normal rate. For EEJ, however, the reduction was nil. The proportional increase from normal to slow rate was 36, 61, and 58 per cent for CH, EEJ, and ES. The Y-intercepts indicate small deviations from a simple proportionality (Table 6-2) that will be commented on below (6.4.2). The observations in general thus confirm the first hypothesis.

Table 6-2. The slopes and Y-intercepts of the linear function relating the durations of the test intervals at normal speech rate to those of fast and slow rates, respectively.

<table>
<thead>
<tr>
<th>Subjects</th>
<th>Speech rate: NORMAL-FAST</th>
<th></th>
<th>Speech rate: NORMAL-SLOW</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>slope</td>
<td>intercept</td>
<td>slope</td>
<td>intercept</td>
</tr>
<tr>
<td>CH</td>
<td>.94</td>
<td>-28</td>
<td>1.36</td>
<td>67</td>
</tr>
<tr>
<td>EEJ</td>
<td>1.00</td>
<td>-123</td>
<td>1.61</td>
<td>-126</td>
</tr>
<tr>
<td>ES</td>
<td>.83</td>
<td>-41</td>
<td>1.58</td>
<td>46</td>
</tr>
</tbody>
</table>
Figure 6-1 (a-c). The linear functions relating the durations of the test intervals at normal speech rate to those of fast and slow rates, respectively.
6.4 Syllable durations

6.4.1 The duration of the unstressed syllables

Figure 6-2 (a-c) shows the increase of the duration of the unstressed subset of the test interval as a function of the number of syllables it contains. Two obvious tendencies appear from the data. First, the unstressed syllables were proportionally reduced, or lengthened, when speech rate was speeded up or slowed down. The step between normal and fast speed was less marked than that between normal and slow for all subjects. The first hypothesis is accordingly true also for the unstressed subset of syllables.

Secondly, the linear increase in duration of the unstressed set of syllables indicates a very simple growth of the test interval: Let $U$ be the duration of the set of unstressed syllables and $N$ the number of unstressed syllables. Then

$$U = c \times N,$$

and its equivalent

$$\bar{U} = \frac{U}{N} = c,$$

where $c$ is a constant, gives a good description of the data. Thus the mean duration of the unstressed syllables is constant when the number of syllables increases from 1 to 4. This holds good for CH and EEJ. For ES it is necessary to include an additive constant, $a$, which increases concomitantly with $c$ across rates: $U = c \times N + a$. The coefficients of the best fitting lines appear in Table 6-3.

The mean duration of the unstressed syllables is apparently relatively independent of the segmental composition of the syllables, which varies considerably in the test material. Constancy, therefore, was attained by temporal adjustments between the unstressed syllables within the test interval. The adjustments of the unstressed vowels demonstrated in Section 4.4.1 may be an expression of this tendency to a constant mean duration of unstressed syllables for a given speech rate.
Figure 6-2 (a-c). The duration of the set of unstressed syllables as a function of the number of syllables for fast, normal, and slow speech rates.
Table 6-3. The slopes ($c$) and intercepts ($a$) of the linear function between the number and the duration of the set of unstressed syllables ($U = c \times N + a; \ N = 1 \ldots 4$) for different speech rates and subjects.

<table>
<thead>
<tr>
<th>Speech rate:</th>
<th>FAST</th>
<th>NORMAL</th>
<th>SLOW</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects</td>
<td>$c$</td>
<td>$a$</td>
<td>$c$</td>
</tr>
<tr>
<td>CH</td>
<td>111</td>
<td>3</td>
<td>121</td>
</tr>
<tr>
<td>EEJ</td>
<td>111</td>
<td>-4</td>
<td>117</td>
</tr>
<tr>
<td>ES</td>
<td>92</td>
<td>25</td>
<td>112</td>
</tr>
</tbody>
</table>

Predictions for the temporal relation between stressed and unstressed syllables. The simple expression derived for the duration of the set of unstressed syllables obviously has an implication for the distinction between the stressed syllable and the unstressed ones in the stress group. In Section 2.5 this distinction was assumed to be the primary basis for the partitioning of the phrase into stress groups. It follows from the assumption that the ratio of the duration of any of the unstressed syllables, $U_i$, to the duration of the stressed syllable, $S$, must be less than unity,

$$\frac{U_i}{S} < 1.$$  

Assume further that the ratio between the mean duration of the unstressed syllables and the duration of the stressed syllable is approximately constant for a given rate of speech,

$$\frac{\bar{U}}{S} \approx k.$$  

Since $U$ can be substituted by the constant $c$, see above, then

$$\frac{c}{S} \approx k.$$  

Thus $S$ should be approximately constant for a given speech rate, when $N$ increases from 1 to 4 unstressed syllables.
6.4.2 The duration of the stressed syllable

Figure 6-3 (a-c) shows the duration of the stressed syllable as a function of the number of following unstressed syllables for each speech rate and subject. Disregarding for the moment the duration of the stressed syllable in monosyllabic intervals, we find relatively stable durations across different sizes of the stress group for each speech rate. The level is dependent on the rate of speech, as expected.

The exceptions to an approximate constant duration mainly concern the slow speech rate. There was a gradual decrease of duration when unstressed syllables were added, especially for ES, but also for CH. However, a small gradual decrease seems to be a general characteristic of ES, as the curves for all speech rates show this tendency. Apart from these deviations, however, the general tendency is towards a constant duration of the stressed syllable. The stressed syllable then contributed to the overall duration of the interval, mainly by adding a constant duration when the stress group contained one or more unstressed syllables.

Generalizing from the observations of normal speech rate in the preceding chapters (see 4.4; 5.1.3), it may be assumed that both the vowel and the consonant contributed to this approximate constancy. However, the deviations from constancy should be associated with the consonant rather than with the vowel. This is apparent from Figure 5-1 in which a greater and more gradual decrease occurs for the consonant as compared to the vowel. EEJ is an exception to this pattern.

Asymptotic durations. There are different asymptotic levels of the duration of the stressed syllable for slow, normal, and fast speech. Note that these differences between rates may explain the deviation from a zero intercept in Figure 6-1. The asymptotic value (constant duration) is larger at a slow as compared to a normal speech rate, which leads to a greater increase of the Y-values than the X-values of the function. Thus a positive Y-intercept will result. This is confirmed by the data of two of the speakers (CH and ES). The oppo-
Figure 6-3 (a-c). The duration of the stressed syllable as a function of the number of following unstressed syllables for fast, normal, and slow speech rates.
site effect, and a negative Y-intercept, will be expected for the function that relates normal to fast speech rate. The data for all three speakers are compatible with this explanation.

The similar tendency for the duration of the stressed vowel to reach an asymptote upon the addition of unstressed syllables after the stressed one was interpreted by Klatt (1973) and Lindblom et al. (1976) as a consequence of constraints on compressibility set by the articulatory mechanism (see 1.3.1; 4.4.1). However, while articulatory constraints are without doubt reflected in the temporal structure of speech, the present data point to the distinction between the stressed and the unstressed syllable as a basic determinant of asymptotic durations. To maintain this distinction, different speech rates should be associated with different asymptotic values.

6.5 Speech rate and the distinction between stressed and unstressed syllables

To find out whether the hypothesis of a constant relationship could be borne out by the present data, ratios between the mean unstressed syllable duration and the stressed syllable duration were calculated for each speaker and speech rate. The ratios are plotted against the number of unstressed syllables in the test interval in Figure 6-4 (a-c), separately for each speaker.

The general tendency is clearly consonant with the hypothesis. Unstressed to stressed syllable duration ratios are about 50 percent (Table 6-4). However, for each of the speakers, there was some variation with speech rate. This variation was not systematic across subjects. ES showed the greatest deviations from constancy. These deviations were mainly restricted to the first unstressed syllable which was much longer than predicted for slow and normal speech. The other perturbations in the curves are related to the inherent characteristics of the different syllables in the test interval, and to the patterning of rhythmic alternation between unstressed syllables (4.4.2).
Figure 6-4 (a-c). Ratios of the mean unstressed syllable duration ($U$) to the duration of the stressed syllable ($S$) for test intervals with 1-4 unstressed syllables at fast, normal, and slow speech rates.

Table 6-4. Ratios ($k$) between the mean duration of the unstressed syllables and the duration of the stressed syllable for different speech rates and speakers (means of 4 sizes of test intervals).

<table>
<thead>
<tr>
<th>Speech rate:</th>
<th>FAST</th>
<th>NORMAL</th>
<th>SLOW</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjects</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CH</td>
<td>.57</td>
<td>.62</td>
<td>.63</td>
<td>.61</td>
</tr>
<tr>
<td>EEJ</td>
<td>.54</td>
<td>.44</td>
<td>.52</td>
<td>.50</td>
</tr>
<tr>
<td>ES</td>
<td>(.66)</td>
<td>(.59)</td>
<td>.50</td>
<td>(.58)</td>
</tr>
</tbody>
</table>
One might speculate that the stability of the VC-unit, resulting in almost identical durations of VC: - and V:C-segments (5.1.5), is an expression of the tendency to maintain the stressed/unstressed syllable distinction. Moreover, it seems reasonable that the presence of the quantity distinction should be the basis for the relatively small ratio between the mean unstressed and stressed syllable durations.

6.6 Conclusions

The two hypotheses presented in the beginning of this chapter were both confirmed. The duration of the test interval changed proportionally with variations in speech rate. Specifically, the unstressed syllables were found to contribute to this change by adding a constant increase of duration to the interval.

The duration of the stressed syllables, on the other hand, tended to level out at different values for different speech rates, when the interstress interval contained more than one syllable. Thus they contributed to the duration of the test interval mainly by a constant duration. This constancy in relation to the constancy of the mean duration of the unstressed syllable was seen to be the basis for the distinction between stressed and unstressed syllables, reflected in the ratios of about 50 per cent of mean unstressed to stressed syllable durations.

The different asymptotic values for the stressed syllable at normal as compared to fast and slow speech may be seen as an expression for a tendency to maintain the stressed/unstressed syllable distinction. The asymptotic values appear when there is a risk that the distinction would otherwise be lost. This points to the importance of the distinction between stressed and unstressed syllables which may be seen as the basis for rhythmic grouping. The result therefore supports the general assumption for the distinction between the stressed and unstressed syllables in section 2.5.
7 EFFECTS OF CONTEXT

7.1 Purpose

The study includes two experiments. In Experiment I one purpose was to study the effects on the test interval of the number of syllables in the surrounding stress groups. The presence of such effects would indicate a planning mechanism with a larger domain than the single interstress interval—either a phrase or a part of a phrase (see 2.4). Such a mechanism would most reasonably impose some kind of structure (regularity) on the elements within its domain. This could be attained by adjustments of processing rate at the planning stage of speech.

In accordance with the assumptions in section 2.5 two types of adjustments may be hypothesized:

(1) One is a uniform adjustment of the processing rate due to the length of the phrase. Thus, longer phrases, in terms of the number of syllables, would be processed at a faster speed than short phrases, if a limited capacity of processing is assumed. This implies an inverse relation between the number of syllables in the phrase (context) and the duration of the test interval. Further, positive correlations would be expected between the durations of the adjacent elements (interstress intervals, syllables etc.) of the phrase. The context effect should in this respect be equivalent to the effects of intentional variations in speech rate, as elaborated in Chapter 6, though the planning operations could be different. The study by Lindblom & Rapp (1973) gives some weak support for adjustments of the duration of stressed vowels to phrase length. Lehiste (1980b) found similar adjustments of word duration.

(2) The other is a variable adjustment of processing rate between adjacent stress groups in order to attain a regular recurrence of stressed syllables. For example, monosyllabic interstress intervals
should be processed at a slower speed, i.e. have longer durations, when adjacent to a stress group with many syllables than one with just a few. This type of adjustment depends on the sizes of adjacent stress groups and leads in general to a compensatory lengthening of short stress groups and a shortening of long ones. Thus, negative correlations would be expected between the durations of adjacent interstress intervals. Some support for this type of adjustment comes from Cutler (1980) and Wenk & Wioland (1982).

The two types of adjustments are presented schematically below. The diagram shows the relative change in the duration of the test interval of a given size (e.g. 3 syllables) when the surrounding stress groups contain 1, 3, and 5 syllables:

- **(1) PHRASE-LENGTH ADJUSTMENTS**
  - 1
  - 3
  - 5

- **(2) STRESS-GROUP ADJUSTMENTS**
  - 1
  - 3
  - 5

It should be noted that these effects often work in opposite directions but are sometimes positively related. Generally, it is only possible to study their net effect. The effects are contrasted in Table 7-1.
Table 7-1. Predicted relative effects on the test interval (arbitrary units) according to the hypotheses of a phrase-length adjustment \((C+n)\) and a stress-group adjustment \((C-n)\), respectively. (The cells denote \((C+n)/(C-n)\), where negative (positive) values indicate shortening (lengthening) effects.)

<table>
<thead>
<tr>
<th>Test interval ((n = \text{number of syllables}))</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Context ((C = \text{number of syllables}))</td>
<td>1</td>
<td>-2/0</td>
<td>-3/-1</td>
<td>-4/-2</td>
<td>-5/-3</td>
</tr>
<tr>
<td>(3)</td>
<td>-4/2</td>
<td>-5/1</td>
<td>-6/0</td>
<td>-7/-1</td>
<td>-8/-2</td>
</tr>
<tr>
<td>(5)</td>
<td>-6/4</td>
<td>-7/3</td>
<td>-8/2</td>
<td>-9/1</td>
<td>-10/0</td>
</tr>
</tbody>
</table>

The purpose was also to examine the consequences of contextual variation on the distinction between stressed and unstressed syllables. The following predictions were made:

(3) The duration of the stressed and unstressed syllables may vary, but a constant relationship between the stressed and the mean unstressed syllable durations can be predicted for the uniform adjustment effect. This prediction is in accordance with the results of the variations of speech rate in Chapter 6.

For the variable adjustment effect due to different sizes of adjacent stress groups, the outcome is more difficult to predict. The effect presupposes compensatory changes in the durations of stressed and unstressed syllables for each separate stress group in the phrase. A constant ratio across the different stress groups is possible but seems to demand a more complicated form of planning than for the uniform case of adjustment.

In Experiment II the effects of preceding as compared to following stress groups were studied. The sizes of the preceding and the following stress groups were therefore varied orthogonally. Earlier studies of Swedish have shown influences on vowel duration of both preceding and following syllables (or words), though the effects of following elements were shown to be strongest (Lindblom & Rapp, 1973; Lindblom et al., 1976). The following predictions were made:
(4) According to the hypothesis of variable adjustment of the stress groups, combined effects of preceding and following context elements could theoretically be less than, equal to, or greater than the sum of each separate effect, depending on the size relations between the test group and the preceding and following context groups.

The hypothesis of a uniform temporal adjustment, as formulated above, predicts no asymmetrical effects of preceding and following contexts. These cases, as well as any combined cases, produce their effects through the length of the phrase, irrespective of its composition. The only limiting factor would be extremely long phrases, where some form of partitioning of the phrase must be presupposed in the planning stage. This partitioning could introduce discontinuities in the otherwise uniform processing of the phrase.

7.2 Experiment I

7.2.1 Procedure

All five speakers participated in the test.

The material included the basic material (3.4.2) and two sets of supplementary sentences in which the sentence frame of the basic material was exchanged for stress groups with one and five syllables. Thus, the material consisted of 15 sentences.

The different sentence frames were:

(1) B̄ill ------ B̄ōs bōll.
'Bill ------ Bo's ball.'

(3) Bōnderna ----- bīlen i Bērgamo.
'The peasants ----- the car in Bergamo.'

(5) Bārabbasarna ----- bērberisen med bārytonerna.
'The Barabbases ----- the berberry with the barytones.'
The test interval was, as in the basic material: böt, byter, byter väl, byter välan, byter välan då.

Measurements were made of the test interval and the initial inter-stress interval. For segmentation and measurements, see 3.7.

The data for the basic material are identical to the data presented in Chapter 4.

7.2.2 Results: Effects of the size of the surrounding stress groups

Figure 7-1 (a-e) shows the effects on the test interval of the size of the surrounding stress groups for each of the five speakers. The test interval contains from 1 to 5 syllables and the surrounding intervals either 1, 3, or 5 syllables.

A phrase-length effect should show up as a significant main effect of the context factor, where an increasing size of the surrounding stress group should lead to decreasing durations of the test interval (cf. Table 7-1). Four subjects showed this effect. Especially EE3, and to a certain extent BH, demonstrated a simple main effect of phrase length, while for ES and EEN the picture was complicated by an interaction between context and test interval. The deviations from a simple net effect of phrase length were mainly located at small sizes of context (C=1,3) and test intervals (n=1,2). Possibly the phrase-length effect is indiscriminate to these small sizes, or else the variable adjustment effect is relatively stronger there. The latter effect was clearly present in the speech of CH, where there were significant trends toward a positive relation between context size and the duration of the test interval.

Thus it seems that the strongest stress-group adjustments occurred when the size of the adjacent stress groups differed from that of the test interval by either one or two syllables. In addition, the effects were stronger in shorter phrases. These indications are contrary to the quantitative predictions in Table 7-1. However, since
the observed effects only demonstrate the relative dominance of one type of adjustment over another, it is in general not possible to infer the absolute strength of the adjustment effects. For example, either a small phrase-length effect or a strong stress-group effect may lead to the same observed net effect.

The different patterns produced by the speakers point to the existence of consistent interindividual strategies of adjustment. CH represented one extreme, making adjustments primarily in order to preserve the regularity of stress groups. EEJ, on the other hand, made adjustments to the regularity of phrases. This points to a certain inherent flexibility of the articulatory planning, resulting in different speech styles.

Figure 7-1 (a-e).
The duration of the test interval as a function of the number of syllables in the context and test interval.
Phrase length and the duration of the first stress group. A positive relation between phrase length and speech rate should show up already in the first stress group. The processing of the initial part of the phrase would possibly be less affected by stress-groups adjustments than later parts. Measurements demonstrated a significant positive relation between the duration of the first stress group and the number of following syllables for three speakers (EEJ, ES, EEN) across all levels of context. For BH the result was inconclusive, while for CH the relation was reversed. The latter observation was in agreement with the stress-group adjustments of CH in the test interval.

The results presented so far concerned differences in mean durations between the experimental conditions. What about the trial-to-trial variability in speech rate for each condition? A positive correlation between the durations of the first (context) and second (test) stress groups would indicate a uniform speech rate, while a negative correlation would indicate either a stress-group adjustment or measurement errors (2.2). Unfortunately, the adjustment factors work in opposite directions so it is only possible to infer their relative power. The product moment correlations between the durations of the first and second stress groups were computed for all subjects and conditions separately. The results showed that 67 out of 75 correlations were positive. Five of the negative correlations were obtained for CH. The mean correlations were: .10 (CH), .32 (EEJ), .38 (ES), .34 (EN), and .64 (BH). Thus, the variability in the data is partly explained by the rate factor. Indications of variable adjustments effects or measurement errors were not significant.

Reexamination of earlier data. The main results have consequences for the interpretation of the data on segment durations presented in Chapters 4 and 5. It will be recalled that both vowel and consonant durations were longest in the monosyllabic interval, next longest in the bisyllabic interval and either somewhat shorter than or about equal to the bisyllabic duration in the polysyllabic intervals. Also, the first unstressed vowel and following consonant were longest in the bisyllabic interval (see Figure 4-2 and 5-1). This pattern was characteristic for all of the speakers to some degree, though for EEJ
the vowels and consonants in the monosyllabic interval were not very different from those in longer intervals. The observations were interpreted as due to a tendency to make stressed vowels recur regularly.

The durational changes are in effect what would be expected because of contextual influences. In the trisyllabic context vowels and consonants within the monosyllabic intervals should both be long, most reasonably lengthened, to compensate for the length differences between the test interval and the surrounding stress groups. Further, segments in bisyllabic intervals should be longer than in polysyllabic ones. Contextual adjustments may thus very well be the basis for the "shortening" patterns observed in stress groups of increasing length.

Some individual differences may now be explained, especially between EEJ and the other speakers. For EEJ the durations of the vowel and the consonant in the monosyllabic interval were almost the same as those in the polysyllabic intervals (4.4; 5.1.3). However, this is to be expected, as a long monosyllabic interval should be a reflection of adjustments to the regularity of stress groups. Such adjustments were found not to be characteristic of EEJ.

The existence of contextual effects such as those demonstrated in the present study might also explain some of the effects observed in other investigations. For example in the study by Lyberg (1977), the shorter stressed vowel in the word Dágobert, 'a Christian name', as compared to Dågen, 'the day', seems reasonable against this background. Two sentences used by Lyberg (see 1.2.2) were:

(a) Finúrlige Dågen berömmer idåg.
(b) Finúrlige Dágobert berömmer Dåg.

Disregarding word boundaries and counting the number of syllables from one stressed to the next in the two sentences the following patterns appear: (a) 333, (b) 342. Such a difference would be reflected in a shorter duration of the vowel in (b) as compared to (a). As the
two sentences contain the same number of syllables, this difference is not attributable to a phrase-length adjustment.

If contextual temporal effects are strong, which they appear to have been, the choice of the frame within which test elements are inserted should have consequences for the temporal effects to be observed. Even a constant test frame clearly introduces variation. The evaluation of the temporal effects, therefore, should include the contextual conditions of the test item.

### 7.2.3 Contextual variations and the distinction between stressed and unstressed syllables

How do the contextual variations influence syllable durations and the relation between stressed and unstressed syllables?

First, a phrase-length adjustment implies a constant rate of change for all the elements in the phrase. The uniform increase or decrease of the duration of stress groups should be associated with shorter and longer phrases, respectively. Within a larger contextual frame stressed and unstressed syllables should be shorter than in a more restricted frame. For stressed syllables this should be reflected in lower asymptotic values in long as compared to short phrases. However, the ratio between stressed and unstressed syllables should be unchanged.

Secondly, stress group adjustments would imply local rate variations over the different stress groups in a phrase, resulting in increased or decreased durations of interstress intervals. Such an increase (or decrease) should require a corresponding increase (or decrease) of the duration of both stressed and unstressed syllables in order to maintain the distinction between stressed and unstressed syllables. In this case, however, asymptotic values of the stressed syllables in long phrases should be higher than in shorter ones.
The duration of stressed syllables. The duration of stressed syllables as a function of the number of following unstressed ones is shown in Figure 7-2 (a-e) for the five speakers.

As expected, EEJ showed a clear phrase-length effect. In her speech the length of the contextual frame is reflected in the different levels of the curves. The shortest duration of the stressed syllable is associated with the longest contextual frame and vice versa. Thus, the duration of the stressed syllable changed in the same way as the duration of the test interval for EEJ.

Figure 7-2 (a-e). The duration of the stressed syllable as a function of the number of following unstressed ones (1, 3, or 5 syllables in the surrounding stress groups).
(b) EEJ

![Graph showing stressed syllable duration vs. number of following unstressed syllables for EEJ.]

(c) ES

![Graph showing stressed syllable duration vs. number of following unstressed syllables for ES.]

NUMBER OF FOLLOWING UNSTRESSED SYLLABLES
The other speakers exhibited phrase length effects to a varying degree. For EEN and BH the longest contextual frame is associated with the shortest duration of the stressed syllable, while the curves for the two shorter contextual frames are more or less identical. The data for CH and ES demonstrate only marginal effects of phrase length. For CH the predicted stress group adjustments seem to be reflected throughout the range of contextual frames. This is apparent from the fact that the longest contextual frame produced the longest stressed syllables, except for monosyllabic test intervals.

A very close correspondence between the durational patterns of stressed syllables and interstress intervals apparently characterized the speech of EEJ (cf. Figure 7-1). This was true for BH, too. Such a correspondence cannot be observed for the other speakers. It therefore seems that adjustments to stress groups should be the basis for this discrepancy. This would show up as instabilities in the relations between the durations of stressed and unstressed syllables.

**Ratios of unstressed to stressed syllables.** Figure 7-3 (a-e) shows the ratio of the mean duration of the unstressed syllables to the duration of the stressed one for different sizes of the test interval and contextual frame. As expected, the ratios are most stable for EEJ and BH. However, for the other speakers too the ratios are relatively stable. Table 7-2 shows that the mean ratios vary between .44 (for EEJ) and .63 (for CH).

**Table 7-2.** Ratios of the mean duration of the unstressed syllables to the duration of the stressed syllable for different sizes of context (means of 4 sizes of test interval).

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Context 1</th>
<th>Context 3</th>
<th>Context 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>CH</td>
<td>.63</td>
<td>.63</td>
<td>.61</td>
</tr>
<tr>
<td>EEJ</td>
<td>.46</td>
<td>.45</td>
<td>.44</td>
</tr>
<tr>
<td>ES</td>
<td>.57</td>
<td>.59</td>
<td>.53</td>
</tr>
<tr>
<td>EEN</td>
<td>.61</td>
<td>.55</td>
<td>.61</td>
</tr>
<tr>
<td>BH</td>
<td>.54</td>
<td>.52</td>
<td>.51</td>
</tr>
</tbody>
</table>
Figure 7-3 (a-e). The ratio of the mean unstressed syllable duration to the stressed syllable duration for different sizes of test interval (1-4 unstressed syllables) and context (1, 3, or 5 syllables).
In conclusion, though stress-group adjustments appear to introduce special complexities, the durational relation between stressed and unstressed syllables is preserved across different contexts.

7.3 Experiment II

7.3.1 Procedure

CH, EEJ, and ES participated in the test.

The number of syllables in the preceding and following stress groups as well as in the test interval was varied in a factorial design. The material consisted of 18 sentences, all with three stress groups and the test interval in the middle position. The test interval contained either the one-syllable bö́t, or the five-syllable býter vä́lan dǻ from the basic material (3.4.2) The preceding and following stress groups were either Bill, 'the Christian name', bö́nderna, 'the peasants', or Barabbasarna, 'the Barabbases'.

Thus, the sentences were constructed from the following inventory:

<table>
<thead>
<tr>
<th>First stress group</th>
<th>Test interval</th>
<th>Third stress group</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) B̥ill</td>
<td>(1) bö́t</td>
<td>(1) B̥ill</td>
</tr>
<tr>
<td>(3) B̥önderna</td>
<td>(5) býter vä́lan dǻ</td>
<td>(3) B̥önderna</td>
</tr>
<tr>
<td>(5) B̥arabbasarna</td>
<td></td>
<td>(5) B̥arabbasarna</td>
</tr>
</tbody>
</table>
7.3.2 Results: Effects of preceding versus following stress groups

The duration of the test interval, when the preceding and following stress groups contained either one, three, or five syllables in all possible combinations is shown in Figure 7-4 (a-c) for each speaker. The upper part of the figures shows the duration of the one-syllable test interval (böt), and the lower part the duration of the five-syllable interval (byter välan dá).

There were significant main effects of both context factors, the number of preceding and following syllables, for all speakers. In general the duration of the test interval decreased with an increasing number of syllables in the context, a result that replicates the phrase-length effect obtained in Experiment I. Note that the hypothesis of stress-group adjustment predicts the duration to be positively related to the number of preceding and following syllables.

For the one-syllable test interval (böt) the effect of the following context was stronger than that of the preceding one. For the five-syllable test interval there was a weak tendency to a stronger effect from the preceding context, except in the data of ES.

However, there is some clear evidence of an interaction between the context factors. For example, in the case of the one-syllable test interval the duration was not affected by an increase from one to three preceding syllables when the following stress group included just one syllable. This result was obtained in the very similar data for CH and ES, though not replicated for EEJ. The same insensitivity to increments in the preceding context from three to five syllables was true of all subjects when there were five syllables in the following stress group.

For byter välan dá, the five-syllable test interval, the patterns were more varied. The different speakers exhibited more differences in detail. This is not unexpected, as longer stress groups should permit more varied strategies of planning.
Figure 7-4 (a-c). The duration of the test interval as a function of the number of syllables in the preceding and following stress groups for test intervals with one and five syllables (böt/ byter välan då).
Phrase-length versus stress-group adjustments. Obviously, the data point to strong effects from phrase length, but the different effects of preceding and following contexts as well as the interaction effects indicate a need for qualification of the hypothesis.

Figure 7-5 (a-c) allows more detailed comparison of critical conditions. Consider again the issue of phrase-length versus stress-group adjustments. In the one-syllable case the duration of the test interval decreased with increasing context, instead of adjusting (increasing) to the surrounding context. For the five-syllable test interval the hypothesis of stress-group adjustment predicts the greatest durational decrease in the condition 151, and smaller adjustments (decreases in duration) with increasing context. The opposite trend was true. Note also that combining the effects of preceding and following contexts (313, 515 and 353, 555) led to stronger decreases than for single effects (e.g. 113 or 311 vs. 313).

Thus, there was no clear evidence of stress-group adjustment in this experiment. A certain discrepancy between these data and those in Experiment I therefore exists. It concerns the durational change of the one-syllable interval when surrounded by stress groups with one and three syllables, respectively. In Experiment I there was an increase of duration of the one-syllable interval at the step between one and three syllables in the surrounding stress groups (i.e. between the conditions 1111 and 3133). In the present data the duration of this interval was either unchanged (because of the preceding stress group) or decreased (because of the following stress group). The phrase-length effect was somewhat stronger.

This discrepancy between the results of Experiments I and II may be related to the size of the different sentences used. Experiment I used sentences with four stress groups. Suppose there are limits to the domain of planning. Then possibly sentences with four stress groups could be partitioned and processed in two parts. Context effects would be strong within but not between the parts. If now the 3133-sentence is processed in two steps (31 and 33), and the phrase-length effect is small for the 31-chunk, which even the data of Ex-
Figure 7-5 (a-c). The duration of the test intervals for various combinations of preceding and following stress groups (test intervals with one and five syllables).
DURATION OF TEST INTERVAL

NUMBER OF SYLLABLES IN CONTEXT
(PRECEDING, FOLLOWING, PRECEDING/FOLLOWING)

102
NUMBER OF SYLLABLES IN CONTEXT (PRECEDING, FOLLOWING, PRECEDING/FOLLOWING)

DURATION OF TEST INTERVAL
periment II suggest, then the adjustment of the test interval to the preceding stress-group could surpass the phrase-length effect. This ad hoc reasoning reinforces the need to qualify the phrase-length hypothesis with respect to its planning domain.

But why is there a difference at all between preceding and following contexts when the phrase is short, and there should be no need to use more than one cycle of planning or processing? Comparing conditions with the same phrase length (113 - 311; 315 - 513; 115 - 313 - 511) indicates almost invariably the same pattern: the processing rate depends on the distribution of information in the phrase. More syllables in the following stress group leads to faster processing than corresponding increments in the preceding one.

For the five-syllable test interval, the picture is more varied. The comparisons (153 - 351; 355 - 553; 155 - 353 - 551) yield a slightly stronger effect from the preceding context for CH. The same tendency is true for the data of EEJ when the phrase is long (n=11). In the other cases the comparisons indicate little or no effect from the following stress group.

The conception of the phrase-length effect as an approximately uniform rate of processing depending on the information load may still be retained. Then it is necessary to explain how the distribution of information in the phrase influences the "computation" of processing speed. Similar results have been reported and even rationalized by Lindblom et al. (1976). The different effects of preceding and following syllables were explained as being due to an asymmetry of the space in the short-term motor memory (1.3.1).

An alternative conception is that the relative effects of the phrase-length and stress-group adjustments vary with the position in the phrase.Possibly the stress-group adjustment is more sensitive to the preceding stress group and is stronger in short phrases. Then the uniform phrase-length effect may be counteracted more by variations in the preceding context than by the following one.
Such an explanation would also apply to very long phrases which should be partitioned into almost independent, short subphrases. But in general these special interaction effects were quite small in the present experiment as compared to the considerable main effect of the phrase-length factor.

7.4 Conclusions

The contextual variations produced strong effects of phrase length on the duration of the test interval. This result may be interpreted in terms of the assumed planning mechanism with a larger domain than the single interstress interval. However, the concept of planning domain needs to be qualified, since the effects of the preceding and following sentence context were different. These asymmetrical effects may be related to an interaction between the phrase-length and the stress-group adjustments. There was actually certain evidence in the data for such competing adjustments.
8 INTRODUCTION TO THE CROSS-LANGUAGE STUDY

8.1 Purpose and outline

The purpose of the cross-language comparisons is to explore into the generality of the assumptions summarized in Chapter 2. Thus at the outset it will be assumed that differences in language structure constitute the main factor responsible for temporal patterning of languages characterized as stress-timed and syllable-timed (cf. Dauer, 1983). Therefore, contrary to what seems to have been the basis for this dichotomization (the assumption of different coordination patterns of the breathing muscles, see 1.2.1) similar rhythmic constraints are assumed for different languages. Rhythmic patterning in different languages as traced in the temporal structure of the speech signal should thus display common characteristics. The patterning, however, should be modified with regard to structural properties of each specific language, such as characteristics of syllable structure and quantity.

These assumptions will necessarily need to be tested against data from a great number of languages. The present explorative study aims to contribute to this issue with some data from three languages with marked differences of rhythm: Swedish, Spanish, and Finnish. The study is in two respects more limited than the Swedish main study; the experimental variation is restricted, and for each language the observations are based on just one speaker.

The comparisons, just as as in the Swedish main study, will be focussed on the duration of interstress intervals and their internal temporal structure. In Chapter 9 the purpose is to examine the temporal effects of an increase of the number of syllables in the stress group (cf. Chapter 4). Similar effects in different languages should imply similar rhythmic mechanisms at work. Interactions between the number of syllables and language-specific properties, especially stress and quantity, will also be considered.
In Chapter 10 word stress and its realization in the different languages is investigated. The reason, naturally, is that the rhythmic patterning of speech appears to be tied to the stressed syllable. In Swedish such syllables are prominent and clearly differentiated from unstressed ones (cf. 6.5), while in other languages it appears that stressed and unstressed syllables are differently related. One would reasonably expect such differences to have consequences for the impression of rhythm in a language.

The following sections of the present chapter include an account of the principles for the selection of languages, sketches of the main results of earlier research on rhythm in Finnish and Spanish, comparisons of structural properties of Finnish, Spanish and Swedish, and finally, a description of the general experimental frame for the comparative study.

8.2 The selection of languages for comparison

From the analysis in the preceding chapters (4-7) Swedish appears similar to languages referred to as stress-timed. (For comparisons of the rhythm of Swedish and some other languages, see Strangert, 1981.) To offer a contrast to Swedish rhythm, Spanish was chosen as a representative for languages referred to as syllable-timed. Though objections could without doubt be raised to this characterization of Spanish (see the review of research on Spanish rhythm in 8.3.1), it was nevertheless considered sufficiently distinct from Swedish to justify its inclusion in the study. Finnish, the third language, was included to represent languages with a more indeterminate rhythm (O'Connor, 1973, p.240; Miller, 1984).

These languages were also chosen because of their structural differences (8.4). Such differences, however, would be expected between rhythmically different languages according to the assumptions in section 8.1 and 2.5.
8.3 Main results from earlier research related to rhythm

Some studies of Spanish and Finnish will be reviewed. For studies based on Swedish, see section 1.2 and 1.3.

8.3.1 Spanish

The rhythmic patterning of Spanish. Most studies on rhythm in Spanish (e.g. Dauer, 1979, 1983; Pointon, 1980; Manrique & Signorini, 1983; Hoequist, 1983) have been inspired by the fact that it has long been considered a syllable-timed language (Pike, 1946). These studies have in fact shown that syllable-timing in any strict sense does not occur. Syllable duration, just as in other languages, appears to be determined by stress, position and syllable structure (e.g. Delattre, 1966).

Manrique & Signorini (1983) measured interstress intervals in an Argentine Spanish sentence material. Their data included intervals with two to six syllables ranging between 220 and 800 ms. The close correspondence to the range of 200 - 800 ms assumed by Allen (1975) to characterize stress-timed languages, was especially noted. Furthermore, durational differences between differently-sized stress groups were in many cases found to be small. Stress groups differing by one or two syllables differed in duration by at most 100 ms in about half the sentences. The greater durational differences in the other half were attributed to interacting syntactic and semantic factors. The general conclusions drawn were that the data revealed tendencies to stress-timed rhythm. Previous assumptions of syllable-timing in Spanish were explained as being due to the preponderance of CV-syllables and the specific manner of reduction (to /a/ of vowels in unstressed positions, see 1.4).

Durational correlates of stress. Delattre (1966) compared syllable durations in English, German, Spanish, and French. For Spanish the average durational difference between stressed and unstressed sylla-
bles was expressed as a ratio of 1.3:1. This ratio is relatively low compared to, for example, English with a corresponding ratio of 1.6:1. The maximal difference was found between a stressed, closed, final syllable and an unstressed, open, non-final one. In this case the Spanish ratio was 1.8:1, although compared to English (3.4:1) it was still quite low. On the other hand, almost identical durations were found for stressed and unstressed non-final open syllables. Here the ratios were 1.1:1 and 1.6:1 for Spanish and English respectively. The generally lower ratios for Spanish point to less differentiation between stressed and unstressed syllables as compared to English (cf. the classification of English as a stress-timed language). Specifically, the low ratio of non-final open syllables (stressed and unstressed) contributes to making stressed and unstressed syllables more similar.

Olsen (1972) reports for her Mexican Spanish speaker somewhat higher ratios than Delattre. (It is unclear whether Delattre's data are based on European or Latin American Spanish. It is therefore not possible to decide whether the differences should be described as individual or geographical.) However, it seems that the greater differentiation between stressed and unstressed syllables as observed by Olsen corroborate with those found in other studies of Latin American Spanish (Manrique & Signorini, 1983; Hoequist, 1983). However, the Latin American Spanish data nevertheless reveal more equal syllable durations than those in English.

8.3.2 Finnish

Dependencies between segment durations and larger units of speech.
Word-length dependencies of segment duration in Finnish were examined by Lehtonen (1970) in a study mainly devoted to the durational correlates of Finnish quantity. The study was based on measured durations of phonologically long and short vowels and consonants appearing in different combinations in words within sentence frames. Lehtonen found only restricted evidence for word-length effects. He therefore concluded that "word length in Finnish has no significant effects on
the duration of the individual sound, when the words appear in the same sentence stress position" (Lehtonen, 1970, p.144).

According to Lehtonen (1970, p.138-141), durational changes observed in earlier studies of rhythm in Finnish should be attributed to utterance length rather than word length. This interpretation was supported in a later study (Lehtonen, 1974). Even the length dependence of segment durations observed in isolated words (see Iivonen, 1974) is according to Lehtonen (1974) conditioned by the length of the phrase rather than the word, since isolated words are in effect one-word phrases. Further, Lehtonen (1974) considers the durational changes to be lengthening rather than shortening effects. Segments are assumed to "reach" their normal duration when phrases (or words) increase in length. Segment durations in short phrases are therefore lengthened, the conditioning factor of which is assumed to be phrase-end signalling.

**Durational correlates of stress.** Because of the fixed initial position of Finnish word stress, it is difficult to determine its phonetic manifestations. This is probably the reason why Lehtonen in his investigation of the temporal structure of Finnish restricted himself to the study of the effects of sentence stress on different word patterns. According to Lehtonen (1970, p.145-147) sentence stress in Finnish is manifested as an increase of duration. This increase affects the word-initial syllable and the following unstressed one. Most of the extra duration falls on the segments following the first vowel. Thus, the increase of duration is greatest in the second vowel (32 per cent), next greatest in the intervocalic consonant (20 per cent), and smallest in the first vowel (8 per cent) in words of the (C)V.CV and (C)VCVV types.

When more varied structures were considered, the greatest increase was observed in the phonetically long vowels (33 per cent), the next greatest in the intervocalic consonant, irrespective of its length (about 25 per cent), and the smallest in the phonologically short vowels (15 per cent). In addition, the increase of the phonologically long first-syllable vowel was smaller than the increase of long sec-
ond-syllable vowels.

The reason for the relatively small increase of the word-initial vowel was assumed to be a consequence of its strong phonologic function. For example, a lengthening of a phonologically short vowel would shift it into a phonologically long one. The restricted lengthening, therefore, was seen as a means of preserving the temporal relations between the successive segments. Stress was thus manifested in a complex way owing to the interaction with quantity.

8.4 Cross-language comparisons of structural properties

8.4.1 Contrasting word stress, quantity, and syllable characteristics

Word stress. Word stress in Swedish (3.2) and Spanish has a distinctive function, that is, a shift of stress position within the word may be the sole basis for a difference of meaning. Word stress, therefore, has an important role in both languages. Also, in sentences there is a difference between words with stress and without in both Swedish and Spanish. Thus, for example in Spanish, nouns, pronouns, verbs, and adverbs are stressed, while articles, prepositions, and conjunctions are unstressed (Navarro Tomás, 1968, p.55).

In Finnish, on the other hand, stress is demarcative rather than distinctive, signalling word boundary, as stress is always on the first syllable of a word. This demarcation of word boundaries is important, as the synthetic character of Finnish morphology makes the word a more important unit than in languages of a more analytic character such as, for example, Swedish and Spanish (Karlsson, 1976, p.28). However, there is also in Finnish some words, e.g. ja 'and', which may be unstressed in sentence contexts.

In words with many syllables, stronger and weaker unstressed syllables tend to alternate. This is a phenomenon characterizing Spanish (Malmberg, 1966, p.99-103) and Swedish (3.2). In Finnish there is a
division of speech measures (stress groups) into measures consisting of a primary or secondary stressed syllable and a following unstressed one (Lehtonen, 1970, p.149-150). This subdivision may be associated with rhythmic alternation.

**Quantity.** Spanish, in contrast to Swedish and Finnish, lacks distinctive differences in quantity (Navarro Tomás, 1968, p.50).

In Swedish stressed vowels may be either phonologically long or short. The consonant following the vowel may also be either long or short. However, the complementarity between the stressed vowel and the following consonant restricts the sequences of vowel-consonant to VC: and V:C. Therefore, only vowel length is considered to be phonological. In unstressed positions the quantity distinction is neutralized (see 3.2).

Of the languages under study, Finnish has the most elaborated quantity system. Both vowels and consonants may be either phonologically long or short, and the length of the vowel is independent of the length of the consonant. Thus, vowel-consonant sequences like VC, VC:, V:C, and V:C: all occur in Finnish. Moreover, the quantity distinction is present in stressed and unstressed syllables alike (see Lehtonen, 1970, p.24-36).

**Syllable structure.** Most Spanish syllables consist of a single consonant and a vowel. The frequency of such CV-syllables amounts to 58 per cent within ordinary texts of Spanish, according to Navarro Tomás (1968, p.40-41). The next most frequent syllable type is CVC (27 per cent). However, the different syllable types seem to occur with different frequencies in stressed and unstressed positions. Olsen (1972), basing her study on Mexican Spanish, found stressed syllables to be 44 per cent closed, while unstressed ones were only 26 per cent closed. In final positions the difference was more marked (89 per cent closed stressed syllables, and 30 per cent closed unstressed, respectively).
Also in Finnish a simple syllable structure predominates. Because of phonotactic restrictions initial and final consonant clusters are not permitted. Exceptions occur, but are not very frequent. Besides, only certain combinations are permitted word-medially. Thus, though both open and closed syllables occur, the open CV syllable is the most frequent, and CVC the next most frequent type of syllable in Finnish (Häkkinen, 1982).

Swedish, accordingly, has the most complex syllable structure. Furthermore, the most complex structures are characteristic for stressed syllables, while unstressed syllables most often have a more simple structure (3.2).

**Segmental composition of syllables.** Not only the number of syllables, but also their segmental properties differ in stressed and unstressed syllables in Swedish. For example, only certain vowels (often /e/) occur in unstressed positions, and very often unstressed vowels reduce to schwa, or may even not be produced at all in causal speech (3.2). Unstressed syllables may therefore be considered "light" while stressed ones on the other hand are "heavy". This terminology is used by Vanderslice & Ladefoged (1972, p.820) to capture the difference between syllables with "reduced timing" and "full articulation", respectively.

Light syllables are not characteristic for either Finnish or Spanish. For example in Spanish, the same vowels occur in both stressed and unstressed syllables. The inherently long /a/ is as frequent in unstressed as in stressed syllables, according to Navarro Tomás (1968, p.47). In addition, the tendency to neutralization of unstressed /e/ is much weaker in Spanish than in Swedish (Malmberg, 1966, p.35). See also Dauer (1983).
8.4.2 Predictions for rhythmic patterning based on the structural analysis

The previous comparative analysis is summarized in simplified form in Table 8-1.

Table 8-1. The distribution of structural features across stressed and unstressed syllables and the function of stress in Swedish, Spanish, and Finnish.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantity</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Complex syllable structure</td>
<td>+</td>
<td>-</td>
<td>(+)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Light syllables</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Function of stress</td>
<td>distinctive</td>
<td>distinctive</td>
<td>demarcative</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Two major differences between Swedish, Spanish, and Finnish appear from the table: (1) The different functions of stress in Finnish as compared to Swedish and Spanish. (2) The extent to which stressed and unstressed syllables are structurally different in the three languages.

It has been suggested that in a language with a distinctive function of stress, this feature will be more prominent than in a language with demarcative stress (Dauer, 1983). Such prominence should contribute to the differentiation between stressed and unstressed syllables which seems to be the basis for the impression of stress-timed rhythm.

However, different structural properties of stressed and unstressed syllables may be more important for the impression of prominence.
Given that the stressed syllables are more complex and therefore longer in duration than the unstressed ones, this difference should also contribute to the temporal regularity between stressed syllables. This is so because unstressed syllables then add relatively little extra duration to that of the stressed syllable. This is not the case when there is a close correspondence between stressed and unstressed syllables.

Thus, we would expect stressed syllables to be more prominent and recur more regularly in Swedish than in Spanish and Finnish. Stressed syllables in Spanish should be more prominent and regular than in Finnish. Owing to the specific function of stress and the complexity of quantity, Finnish should exhibit more irregular temporal patterns than the other languages.

8.5 General experimental procedure

8.5.1 Material

For each language the material consisted of a set of semantically acceptable sentence frames in which different test words were inserted. Though most often the test words were real, some nonsense words had to be included to meet the requirements of the experimental conditions. The material in each language was chosen to facilitate cross-language comparisons. Thus, the segmental composition of the test words was as similar as possible for the different languages. Also, the same type of sentence frame was used. Owing to the different complexities of quantity in the three languages, the size of the material was different in Finnish, Spanish and Swedish. Apart from Spanish the material was restricted to initially stressed test words. The reason for including some finally stressed Spanish test words was that Spanish is assumed to have a "rising rhythm" with stressed syllables terminating stress groups (Navarro Tomás, 1968, p.59).
8.5.2 Speakers

Three speakers participated in the study, one for each of the languages represented. One of the speakers in the Swedish main study, ES, served as the Swedish speaker. The Spanish speaker was a businessman speaking the Castilian variant of European Spanish. The Finnish speaker, finally, was a female university teacher, representative of the variant of Standard Finnish spoken in Helsinki.

8.5.3 Recordings, registrations and measurements

All test sentences occurred together in random order on lists. The speakers read the sentences at their own preferred rate in a neutral non-emphatic way. They were specifically instructed to try to avoid giving specific prominence to the test word. They were also instructed to pause between each sentence to plan the next one as a whole before reading it out.

The sentences were recorded and analyzed in the same way as the material in the Swedish main study (3.6; 3.7).

The measurements included interstress intervals, defined as in the Swedish main study (3.7), for initially stressed test words. For finally stressed test words, which were included in the Spanish material, the duration of the test words was measured instead of interstress intervals. Also, the segment durations within interstress intervals were measured as well as the initial consonant in the test word excluded by definition from the interstress interval.

Segmentation was made according to the same criteria as in the Swedish main study, except in the case of /r/ which in this study was measured as a separate segment (3.7).

Also the same type of statistic calculation was made as in the Swedish main study (3.7). All calculations were based on six observations.
THE TEMPORAL STRUCTURE OF INTERSTRESS INTERVALS

9.1 Experiment I: Segment duration data

The purpose is to examine the effects on segment duration of an increase of the number of syllables in the interstress interval. The analysis is based on sets of test words of different lengths coinciding closely with interstress intervals. The material was restricted to initially stressed words, except for Spanish where finally stressed words were included as well (8.5.1).

9.1.1 Swedish

Material. The material consisted of 6 test words including phonologically long as well as short vowels produced within either one or the other of two sentence frames:

Péter säger ----- tåppert.
'Peter says ----- courageously.'
and
Péter säger ----- åmpert.
'Peter says ----- sharply.'

The choice between these frames was made dependent on the specific test words. This measure was taken to avoid different numbers of consonants at the boundary between the test word and the sentence frame. Thus words ending in a vowel were inserted into the first frame, and words ending in a consonant into the second one.

The experiment was based on measurement of the test words and the following initial /t/ in the sentence frame, when present. The test words were:
The durations of vowels and consonants in the six interstress intervals/test words are given in Table 9-1.

**Table 9-1.** Mean durations (M) and standard deviations (s) of vowels and consonants in interstress intervals/test words with one to three syllables (Swedish).

<table>
<thead>
<tr>
<th></th>
<th>t</th>
<th>a</th>
<th>k</th>
<th>e</th>
<th>n</th>
<th>a</th>
<th>(t)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>s</td>
<td>M</td>
<td>s</td>
<td>M</td>
<td>s</td>
<td>M</td>
<td>s</td>
</tr>
<tr>
<td>tack</td>
<td>132</td>
<td>3</td>
<td>118</td>
<td>7</td>
<td>249</td>
<td>39</td>
<td></td>
</tr>
<tr>
<td>tacken</td>
<td>118</td>
<td>10</td>
<td>98</td>
<td>3</td>
<td>192</td>
<td>7</td>
<td>81</td>
</tr>
<tr>
<td>tackena</td>
<td>118</td>
<td>3</td>
<td>98</td>
<td>9</td>
<td>181</td>
<td>7</td>
<td>70</td>
</tr>
<tr>
<td>tak</td>
<td>139</td>
<td>10</td>
<td>207</td>
<td>10</td>
<td>181</td>
<td>44</td>
<td></td>
</tr>
<tr>
<td>taken</td>
<td>123</td>
<td>9</td>
<td>178</td>
<td>9</td>
<td>138</td>
<td>5</td>
<td>71</td>
</tr>
<tr>
<td>takena</td>
<td>129</td>
<td>12</td>
<td>168</td>
<td>6</td>
<td>127</td>
<td>14</td>
<td>58</td>
</tr>
</tbody>
</table>

In Figure 9-1 the durations of vowels and consonants as a function of the number of syllables in the interstress interval/test words are shown separately for words with long and short stressed vowels.

The same type of durational changes were observed as in the Swedish main study. All segments were to some extent shortened as a function of the length of the interstress interval. However, the phonologically long vowel decreased in duration to a somewhat greater extent than the short one, and similarly the long consonant following the stressed vowel decreased more than the short one. The decrease of the long consonant, moreover, exceeded that of the phonologically long vowel.
Figure 9-1. The durations of vowels and consonants as a function of the number of syllables in the interstress interval/test word: phonologically short (upper part) and long stressed vowels (lower part) (Swedish).
9.1.2 Spanish

**Material.** Seven different test words were produced within the sentence frame:

Carlos dice ----- tanto.
'Carlos says ----- so much.'

The material included both initially and finally stressed words:

<table>
<thead>
<tr>
<th>Word</th>
<th>Pronunciation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>tá</td>
<td>/tâ/</td>
<td>'be careful!'</td>
</tr>
<tr>
<td>tapa</td>
<td>/täpa/</td>
<td>'cover'</td>
</tr>
<tr>
<td>táparo</td>
<td>/täparo/</td>
<td>'fool' (in Colombian Spanish)</td>
</tr>
<tr>
<td>a</td>
<td>/â/</td>
<td>'A' (the first letter in the alphabet)</td>
</tr>
<tr>
<td>pará</td>
<td>/pará/</td>
<td>'a nonsense word'</td>
</tr>
<tr>
<td>tapará</td>
<td>/tapará/</td>
<td>'will cover' (the future tense, 3rd pers. sing. of tapar 'to cover')</td>
</tr>
<tr>
<td>destapará</td>
<td>/destapará/</td>
<td>'will uncover' (the future tense, 3rd pers. sing. of destapar 'to uncover')</td>
</tr>
</tbody>
</table>

The experiment was based on measurements of the test word and the following initial /t/ in the sentence frame.

**Data.** The durations of vowels and consonants in the seven inter-stress intervals/testwords are given in Table 9-2. In Figure 9-2 the durations of vowels and consonants in the initially and finally stressed words, respectively, are plotted as a function of the number of syllables in the word (or stress group).

Examining first the upper part of Figure 9-2 (the initially stressed words) there was found to be a decrease of duration of the stressed vowel and the preceding consonant when the stress group increased in length. The most marked shortening, that of the stressed vowel, amounted to more than 40 ms on the addition of a second syllable, and additional shortening occurred when a third syllable was added. The gradual shortening of the consonant preceding the stressed vowel indicates dependencies across stress groups. Such dependencies are reconcilable with the assumptions in section 2.2 and 2.3.
Table 9-2. Mean durations (M) and standard deviations (s) of vowels and consonants in interstress intervals/test words with one to three syllables (Spanish).

<table>
<thead>
<tr>
<th></th>
<th>C (t)</th>
<th>V (t)</th>
<th>C (t)</th>
<th>V (t)</th>
<th>C (t)</th>
<th>V (t)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>s</td>
<td>M</td>
<td>s</td>
<td>M</td>
<td>s</td>
</tr>
<tr>
<td>tā</td>
<td>178</td>
<td>8</td>
<td>130</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tāpa</td>
<td>163</td>
<td>18</td>
<td>81</td>
<td>4</td>
<td>114</td>
<td>14</td>
</tr>
<tr>
<td>tāparo</td>
<td>154</td>
<td>19</td>
<td>72</td>
<td>8</td>
<td>113</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

What happened to the stressed syllable, then, when unstressed syllables were inserted before it? This is shown in the lower part of Figure 9-2. The data show two conflicting tendencies. First, all segments except /r/ and the stressed vowel decreased in duration on the addition of unstressed syllables before the stressed one. Secondly, the duration of the stressed vowel was gradually increased, when the word increased in length. This increase might indicate that the stressed vowel was rhythmically unconnected to the other segments in the word. The increase of duration, then, should reflect a stress group boundary before the stressed syllable.

If this is correct, the word dice in the sentence frame should also be affected, as it would then be the first part of the stress group with which the unstressed syllables of the test word are most clearly associated. This was in fact the case. The sequence ice was shortened successively from 283 to 248 ms, when unstressed syllables were added to the test word. The shortening effects in these data therefore point to rhythmic patterning as regulated by stress groups beginning with a stressed syllable and ending with unstressed ones, irrespec-
Figure 9-2. The duration of vowels and consonants in the initially (upper part) and finally (lower part) stressed test words as a function of the number of syllables in the inter-stress interval/test word (Spanish).
tive of word boundaries. However, it is not possible to decide whether the shortening was conditioned by the length of the stress group or the phrase, since both factors should yield the same result. The gradual increase of the stressed syllable vowel (the vocalic part of a monosyllabic interstress interval) on the other hand, should be attributed to adjustments to the length of the adjacent stress groups (cf. 2.5).

9.1.3 Finnish

Material. The Finnish test words were produced within the sentence frame:

Pekka sánoo ----- tänään.

'Pekka says ----- today.'

Owing to the complexity of the quantity distinction in Finnish, this material was the largest, consisting of 25 words divided into groups with two, three, and four syllables. Within each group all combinations of long and short first and second vowel and intervocalic consonant occurred. A single one-syllable word, taa, 'behind', was included, too.

The test words with two, three, and four syllables were:

Two syllables

<table>
<thead>
<tr>
<th>Syllable 1</th>
<th>Syllable 2</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>taka 1)</td>
<td>/tāka/</td>
<td>'rear back'</td>
</tr>
<tr>
<td>*taaka</td>
<td>/tā:ka/</td>
<td>(a nonsense word)</td>
</tr>
<tr>
<td>takka 2)</td>
<td>/tā:k:a/</td>
<td>'open fireplace' (nominative case)</td>
</tr>
<tr>
<td>takaa</td>
<td>/tā:ka:/</td>
<td>'from behind'</td>
</tr>
<tr>
<td>taakka 2)</td>
<td>/tā:kk:a/</td>
<td>'burden' (nominative case)</td>
</tr>
<tr>
<td>*taakka</td>
<td>/tā:kk:a:/</td>
<td>(a nonsense word)</td>
</tr>
<tr>
<td>takkaa 3)</td>
<td>/tā:k:a:/</td>
<td>'open fireplace' (partitive case)</td>
</tr>
<tr>
<td>taakkaa 3)</td>
<td>/tā:kk:a:/</td>
<td>'burden' (partitive case)</td>
</tr>
</tbody>
</table>

1) nominative case (in Standard Finnish used only as prefix)  
2) nominative case  
3) partitive case
Three syllables

two-syllable test words + /nsa/ 'his/her' (a possessive suffix)

Four syllables

two-syllable test words + /nsa/ + /han/ (an enclitic (pragmatic) particle with many functions)

*words marked by asterisks are nonsense words both isolated and when suffixes are added. *takansa and *takansahan are nonsense words, too.

The duration of the test words and the following initial /t/ in the sentence frame was measured.

Data. The durations of vowels and consonants in the first two syllables of the interstress intervals/test words as well as the durations of /nsa/ and /han/ and the following /t/ of the sentence frame are given in Table 9-3.

Comparing first the one-syllable word taa to the two-syllable words with a phonologically long first vowel, we find a systematic shortening of the first vowel in the two-syllable words:

<table>
<thead>
<tr>
<th></th>
<th>taaka 28 ms</th>
<th>* taakaa 9 ms</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>taakka 43 ms</td>
<td>taakkaa 31 ms</td>
</tr>
</tbody>
</table>

The shortening was clearly significant, except for the word *taakaa. The small decrease of the first vowel in this word might be associated with the quantity relations within the words, as a marked shortening would have led to disruptions of the quantity balance between the first and the second vowel. (A similar adjustment to the temporal relations between adjacent segments was found by Lehtonen (1970, p.145-147) for the manifestation of sentence stress, see 8.3.2.) There is no clear reason why the word taakkaa was shortened, but it seems that the explanation might lie in the different lengths of the consonants in the two words.
Table 9-3. Mean durations (M) and standard deviations (s) of vowels and consonants in interstress intervals/test words with one to four syllables (Finnish).

<table>
<thead>
<tr>
<th></th>
<th>t</th>
<th>a/aa</th>
<th>k/kk</th>
<th>a/aa</th>
<th>nsa</th>
<th>han</th>
<th>(t)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>s</td>
<td>M</td>
<td>s</td>
<td>M</td>
<td>s</td>
<td>M</td>
</tr>
<tr>
<td>taa</td>
<td>83</td>
<td>5</td>
<td>165</td>
<td>12</td>
<td>80</td>
<td>7</td>
<td>80</td>
</tr>
<tr>
<td>taka</td>
<td>74</td>
<td>5</td>
<td>56</td>
<td>2</td>
<td>93</td>
<td>13</td>
<td>80</td>
</tr>
<tr>
<td>taaka</td>
<td>75</td>
<td>5</td>
<td>137</td>
<td>12</td>
<td>66</td>
<td>7</td>
<td>44</td>
</tr>
<tr>
<td>takka</td>
<td>63</td>
<td>8</td>
<td>62</td>
<td>3</td>
<td>147</td>
<td>14</td>
<td>40</td>
</tr>
<tr>
<td>takaa</td>
<td>63</td>
<td>7</td>
<td>57</td>
<td>6</td>
<td>100</td>
<td>8</td>
<td>123</td>
</tr>
<tr>
<td>taakka</td>
<td>73</td>
<td>7</td>
<td>122</td>
<td>12</td>
<td>108</td>
<td>11</td>
<td>42</td>
</tr>
<tr>
<td>taakaa</td>
<td>72</td>
<td>9</td>
<td>156</td>
<td>10</td>
<td>89</td>
<td>4</td>
<td>116</td>
</tr>
<tr>
<td>takka</td>
<td>57</td>
<td>5</td>
<td>62</td>
<td>6</td>
<td>207</td>
<td>15</td>
<td>111</td>
</tr>
<tr>
<td>taakka</td>
<td>73</td>
<td>9</td>
<td>134</td>
<td>6</td>
<td>158</td>
<td>14</td>
<td>103</td>
</tr>
<tr>
<td>takansa</td>
<td>64</td>
<td>8</td>
<td>48</td>
<td>5</td>
<td>78</td>
<td>9</td>
<td>133</td>
</tr>
<tr>
<td>taakansa</td>
<td>71</td>
<td>9</td>
<td>136</td>
<td>9</td>
<td>53</td>
<td>3</td>
<td>48</td>
</tr>
<tr>
<td>takkansa</td>
<td>62</td>
<td>7</td>
<td>64</td>
<td>7</td>
<td>127</td>
<td>13</td>
<td>42</td>
</tr>
<tr>
<td>takaansa</td>
<td>66</td>
<td>7</td>
<td>48</td>
<td>5</td>
<td>86</td>
<td>6</td>
<td>103</td>
</tr>
<tr>
<td>takkansa</td>
<td>72</td>
<td>7</td>
<td>121</td>
<td>8</td>
<td>104</td>
<td>6</td>
<td>41</td>
</tr>
<tr>
<td>taakkaansa</td>
<td>76</td>
<td>9</td>
<td>138</td>
<td>6</td>
<td>78</td>
<td>5</td>
<td>85</td>
</tr>
<tr>
<td>takkaansa</td>
<td>58</td>
<td>6</td>
<td>63</td>
<td>4</td>
<td>177</td>
<td>12</td>
<td>75</td>
</tr>
<tr>
<td>taakkaansa</td>
<td>72</td>
<td>6</td>
<td>121</td>
<td>7</td>
<td>111</td>
<td>10</td>
<td>60</td>
</tr>
<tr>
<td>takansahan</td>
<td>65</td>
<td>4</td>
<td>51</td>
<td>6</td>
<td>78</td>
<td>8</td>
<td>74</td>
</tr>
<tr>
<td>taakansahan</td>
<td>77</td>
<td>10</td>
<td>113</td>
<td>11</td>
<td>57</td>
<td>5</td>
<td>48</td>
</tr>
<tr>
<td>takkansahan</td>
<td>63</td>
<td>5</td>
<td>58</td>
<td>5</td>
<td>127</td>
<td>12</td>
<td>46</td>
</tr>
<tr>
<td>takaansahan</td>
<td>69</td>
<td>7</td>
<td>48</td>
<td>5</td>
<td>88</td>
<td>3</td>
<td>120</td>
</tr>
<tr>
<td>taakkaansahan</td>
<td>73</td>
<td>10</td>
<td>102</td>
<td>9</td>
<td>110</td>
<td>13</td>
<td>44</td>
</tr>
<tr>
<td>takkaansahan</td>
<td>79</td>
<td>7</td>
<td>139</td>
<td>16</td>
<td>76</td>
<td>15</td>
<td>91</td>
</tr>
<tr>
<td>takkaansahan</td>
<td>57</td>
<td>11</td>
<td>58</td>
<td>5</td>
<td>160</td>
<td>7</td>
<td>82</td>
</tr>
<tr>
<td>taakkaansahan</td>
<td>73</td>
<td>5</td>
<td>115</td>
<td>4</td>
<td>119</td>
<td>4</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>80</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>81</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>69</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>68</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>71</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>73</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>77</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>64</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>72</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>163</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>145</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>138</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>160</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>147</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>141</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>134</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>134</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>190</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>178</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>193</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>199</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>204</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>206</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>56</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>50</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>53</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>51</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>53</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>51</td>
</tr>
</tbody>
</table>
Shortening quite obviously occurred also when syllables were added to words with two and three syllables. This is evident from Figure 9-3 in which the segment durations of the first two syllables are plotted as a function of the number of syllables in the word (stress group) for the different word types. However, the shortening varied considerably between the different segments. Most shortening occurred at the step between words with two and three syllables. The general tendency seems to be that the phonologically long vowels and consonants were shortened to a much greater extent than the phonologically short ones.

Exceptions occurred, however. In both *taaka and taakka the duration of the first vowel was almost identical in words with two or three syllables. This might be another indication (see above) that not only the length of vowels and consonants but also the quantity relations between them determine if and to what extent shortening will occur. Moreover, comparing the phonologically long vowels in the first and second syllable position of the word, for example *taaka - takaa and taakka - takkaa, we find the most marked shortening in the second vowel position. This is in accord with the assumption of Lehtonen (1970, p.146) that the duration of the first vowel is more clearly associated with phonemic factors than the second one.

The shortening observed in the present study is not compatible with the results of Lehtonen (1970, p.138-141). He found only unsystematic changes of the duration of segments upon the addition of syllables in a word in phrase-medial positions. However, on closer examination, it appears that the absence of shortening in Lehtonen's study is explainable in the light of the present data. The structures examined (Lehtonen, 1970, and also 1974) correspond closely to those in the present study where shortening was either restricted or absent. The study from 1974 was based mainly on comparisons of the first (phonologically short) vowels in words of different length. As the present study has shown shortening to be more or less absent in such vowels, Lehtonen's conclusions should be seen as a consequence of his choice of material.
Figure 9-3. The durations of vowels and consonants in two-syllable sequences as a function of the number of syllables in the interstress interval/test word. Data for 8 quantity types (Finnish).
The present data do not permit any explanations of the shortening effects. They may depend on adjustments to the stress group or the phrase (cf. 2.5).

Will the particle -nsa be shortened when -han is added, just as the segments in the two-syllable words were shortened when syllables were added to the stress group? Apparently, this was not the case; -nsa rather increased in length upon the addition of a syllable. The average lengthening for all eight words amounted to 19 ms (see Table 9-3). Though this could not be considered a great increase, it is systematic and present in all the words. This lengthening might be an effect of rhythmic alternation between unstressed syllables.

It seems that the observed lengthening is related to the concept of the Finnish "measure". According to Lehtonen (1970, p 148-152) longer stress groups (speech measures) in Finnish consist of smaller rhythmic units (measures) with a (primary or secondary) stressed and a following unstressed syllable. Thus, a four-syllable stress group consists of two of these smaller units (8.4.1) The lengthening in the present study at the step between stress groups, or words, with three and four syllables, may therefore be interpreted in terms of secondary stress assigned to the second unstressed syllable.

Thus there are two tendencies in the data: (1) A systematic decrease of duration upon the addition of syllables in the stress group, which holds for stress groups with less than four syllables and is attributable to adjustments to the stress group, or the phrase; and (2) a systematic increase of duration at the step from three to four syllables in a stress group, attributable to rhythmic alternation. In addition, the durational effects seem to depend on the length of the segments, since with few exceptions, only long vowels and consonants were shortened. Moreover, shortening did not occur at the expense of the quantity relations within the word.
9.1.4 Comparisons

We may now conclude that the durational effects on segments upon the variation of the length of the stress group are not very different for the three speakers.

Shortening in all three languages seemed to be guided by a principle according to which (phonologically or physically) long segments were shortened more than short ones. This was true for vowels as well as consonants. In Swedish and Finnish, therefore, the phonologically long vowels and consonants exhibited the most drastic shortening. In Spanish the greatest shortening occurred in the stressed vowel which is longer than the other segments in the stress group.

Also, lengthening effects associated with rhythmic alternation were observed in four-syllable Finnish words. Similar effects were not found for either Swedish or Spanish. In the Swedish material the longest words contained three syllables. Lengthening should not be expected since at least four syllables seems to be required for it to occur.

The absence of lengthening in Spanish cannot be similarly explained, as four-syllable words were included in the material. Rhythmic alternation, however, seems to occur also in Spanish (Malmberg, 1966, p.99-103). The very restricted observations in the present study, therefore, should not be seen as conclusive.

9.2 Experiment II: The duration of interstress intervals

Material. The study was based on the same material as in the study of the segmental effects of the length of stress groups (Experiment I). The Spanish sentences with finally stressed test words were not included in the study.

Data. The durations of interstress intervals in Swedish, Spanish, and Finnish are given in Figure 9-4. For Swedish and Finnish the data
have been separated for phonologically long and short first vowels, and for Finnish the different quantity types of the words are given separately.

Comparisons. Comparing first Swedish and Spanish, we find the most marked difference in the monosyllabic intervals - about 350-400 ms in Swedish and about 250 ms in Spanish. Very long monosyllabic intervals are a characteristic of Swedish. Similar long monosyllabic intervals have been assumed by Bolinger (1965) to be the most important reason for the characterization of English as a stress-timed language (see also Dauer, 1983). As a consequence the relatively short monosyllabic intervals of Spanish should contribute to its less stress-timed character.

Finnish presents a most varied picture with interstress intervals ranging from short to long, owing to the complexity of quantity. The single monosyllabic interval is quite short in spite of its phonologically long vowel. This short duration, however, reflects to a certain extent a relatively fast speech rate of the Finnish speaker. If the ratios of monosyllabic to polysyllabic intervals are calculated, this will give an indication of how the number of syllables influences the duration of the intervals. Low ratios, indicating weak influences, should imply a greater regularity of stressed syllables than high ratios pointing to stronger influences of the number of syllables.

The ratios of monosyllabic intervals to intervals with two and three syllables in Swedish amounts to 1:1.17 and 1:1.57, respectively, and the corresponding ratios for Spanish are 1:1.30 and 1:1.80.

The ratios of the Swedish speaker correspond to earlier data for Swedish (Strangert, 1981). However, the Spanish ratios are lower than would be expected in the light of data reported by Dauer (1979, 1983). In this study corresponding ratios of 1:1.68 and 1:2.26 were found. The reason for the disparity between the two studies is not
Figure 9-4. Durations of interstress intervals in Swedish, Spanish, and Finnish. The data are separated for phonologically long and short vowels (Swedish) and for different quantity types of the test words (Finnish).
NUMBER OF SYLLABLES IN TEST INTERVAL

DURATION OF ISI

FINNISH

CV:CV

CV:C:V

CV:C:V

CV:C:V:

 NUMBER OF SYLLABLES IN TEST INTERVAL

DURATION OF ISI

CVCV

CVC:V

CVCV:

CVC:V:

NUMBER OF SYLLABLES IN TEST INTERVAL
NUMBER OF SYLLABLES IN TEST INTERVAL

DURATION OF ISI

ms

CV: CVCV

CV:CV

CVC:V

CV:CV

CVCV:
clear, as different factors might have influenced the results in dif­ferent ways in the two studies. Dauer's study was based on readings of a prose text rather than separate sentences. Besides, speech rate seems to have been higher than in the present study.

The relatively small increase of the duration of interstress intervals on the addition of new syllables in both Swedish and Spanish should to a certain degree be attributed to the shortening of segment durations as observed in the preceding sections (9.1.1 and 9.1.2). Small durational differences between differently-sized intervals will be expected.

The shortening of segments in Finnish was observed to vary for the different quantity types of words. Most shortening occurred in phonologically long vowels and consonants. The most extensive shortening occurred in words with several phonologically long segments. This pattern is, as expected clearly reflected in the duration of interstress intervals. The ratio of the duration of the two-syllable CVVCCVV-word to the duration of words with three and four syllables, respectively, with the same quantity structure is 1:1.17 and 1:1.43. The corresponding ratios for the CVCV-words are 1:1.61 and 1:1.94.

Thus the extent to which shortening occurs on the addition of new syllables is reflected in the relative durations between differently-sized intervals. The somewhat less marked shortening of Spanish makes these intervals more different, while the marked shortening in Swedish leads to more similar durations of the interstress intervals. Finnish reveals a more variable pattern with different temporal relations between interstress intervals in words with different quantity structures. We may conclude that the ratio expressing the relations between interstress intervals of different sizes in Swedish, Spanish, and Finnish is one of the differentiating factors between the three languages. The absolute durational differences, most apparent in the monosyllabic intervals, are another important factor which clearly separates Swedish from Spanish and, at least partly, from Finnish.
9.3 Conclusions

In spite of certain differences between the speakers of the three languages, they all exhibited similar temporal adjustments. Segments within a stress group were shortened when unstressed syllables were added after the stressed one. However, it was not possible to decide to what extent the adjustments were conditioned by the length of the stress group or the phrase.

Owing to the limitations of the experiments (the restricted number of languages and speakers included), no safe conclusions can be drawn. Nevertheless, the data are compatible with the assumption that the interaction of constraints on language structure and speech processing to a certain extent determines rhythmic structure. Traces of similar processing constraints were observed for all three speakers. It was possible to attribute the differences to language-specific characteristics, especially of quantity and stress. These differences seemed to be brought about by adjustments of speech processing to the constraints of language structure, as temporal adjustments never occurred at the expense of important structural relations. This was especially evident for the Finnish speaker in whose material the complexity of the quantity distinction appeared to be the basis for the somewhat variable temporal adjustments.
The purpose is to examine the temporal effects of word stress. For Swedish and Spanish the experimental variation concerned the position of stress within the word. For Finnish, with a fixed position of stress, the durations of the same sequences of segments were compared in different positions of the word. Though not a straightforward test on stress, this procedure will permit certain comparisons with the manifestations of stress in Swedish and Spanish.

10.1 Swedish

Material. The study was based on three test words varying only with respect to the position of stress, which was on either the first, second, or third syllable. All stressed syllables contained a phonologically long vowel. The test words, including two nonsense words, were:

- takena /tá:kena/ 'the roofs' (a colloquial form)
- *takena /také:na/ (an nonsense word)
- *takena /takená:/ (a nonsense word)

The test frame was:

Péter ságer ----- tåppert.
'Peter says ----- courageously.'

All segments in the test words were measured.

Data. The mean durations of each vowel and consonant in the three test words are presented in Table 10-1 for the Swedish speaker.

Stress strongly determined the durational structure of the test words. The increase of the duration was most evident for the vowels.
Table 10-1. Mean durations (M) and standard deviations (s) for vowels and consonants in the test words. Stressed vowels are underlined (Swedish).

<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>V</th>
<th>C</th>
<th>V</th>
<th>C</th>
<th>V</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>s</td>
<td>M</td>
<td>s</td>
<td>M</td>
<td>s</td>
</tr>
<tr>
<td>tākena</td>
<td>129</td>
<td>12</td>
<td>168</td>
<td>6</td>
<td>127</td>
<td>14</td>
</tr>
<tr>
<td>takēna</td>
<td>102</td>
<td>8</td>
<td>72</td>
<td>5</td>
<td>148</td>
<td>5</td>
</tr>
<tr>
<td>takenā</td>
<td>102</td>
<td>10</td>
<td>83</td>
<td>15</td>
<td>108</td>
<td>4</td>
</tr>
</tbody>
</table>

The mean ratio of unstressed to stressed vowel duration is .41. The preceding and following consonants also changed with stress, though to a lesser degree than the vowel. The mean duration ratio of consonants, neither preceded nor immediately followed by a stressed vowel, to consonants preceding the the stressed vowel is .74. The corresponding ratio for consonants immediately following the stressed vowel is .81 (based on the durations of the first and second syllable).

When the mean duration of unstressed to stressed syllables (VC-units) is calculated for the two first syllables of the three words, a ratio of .55 is obtained. This is fairly close to the mean ratio of .59 for subject ES under the normal condition in the experiment on speech rate (6.5).

Thus it appears that stress was temporally manifested over several segments with the most extensive increase in the phonologically long vowel of the stressed syllable.

10.2 Spanish

Material. The test material contained three test words, structurally similar to the Swedish words and segmentally identical except for the final vowels:
táparo /táparo/ 'fool' (in Colombian Spanish)
tapara /tapára/ 'might cover' (the past tense subj., 1st or 3rd pers. sing. of tapar 'to cover'
tapará /tapará/ 'will cover' (the future tense, 3rd pers. sing. of tapar 'to cover')

The words occurred within the frame:

Carlos dice ---- tânto.
'Carlos says ---- so much.'

All segments of the test words were measured.

Data. Table 10-2 shows the mean ratios of each vowel and consonant in the three test words for the Spanish speaker.

Table 10-2. Mean durations (M) and standard deviations (s) for vowels and consonants in the test words. Stressed vowels are underlined (Spanish).

<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>V</th>
<th>C</th>
<th>V</th>
<th>C</th>
<th>V</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>s</td>
<td>M</td>
<td>s</td>
<td>M</td>
<td>s</td>
</tr>
<tr>
<td>tâparo</td>
<td>154</td>
<td>9</td>
<td>72</td>
<td>8</td>
<td>113</td>
<td>8</td>
</tr>
<tr>
<td>tapara</td>
<td>109</td>
<td>9</td>
<td>63</td>
<td>10</td>
<td>130</td>
<td>18</td>
</tr>
<tr>
<td>tapará</td>
<td>133</td>
<td>18</td>
<td>70</td>
<td>7</td>
<td>116</td>
<td>13</td>
</tr>
</tbody>
</table>

As may be seen, stress was temporally manifested in the vowel and in the preceding consonant. The consonant following the vowel appeared to be unaffected by stress. The mean ratio of unstressed to stressed vowel duration in the three words amounts to .81, and the mean ratio of consonants, not immediately preceding or following the stressed vowel, to consonants preceding the stressed vowel is .86

These ratios are not directly comparable to ratios obtained in other studies, which have been based on syllable durations. Therefore, the
ratios of unstressed to stressed CV- and VC-sequences respectively were calculated. The ratios were .83 (CV) and .88 (VC).

The mean ratio of .83 for unstressed to stressed CV-syllables for the Spanish speaker is somewhat higher than the ratio of .77 obtained by Delattre (1966) and Olsen (1972). (This ratio is converted from the identical ratio of 1.3:1 for stressed to unstressed syllable duration as given by Delattre and Olsen.) The ratio in the two earlier studies is based on both closed and open syllables. As closed syllables seem to occur more frequently in stressed than in unstressed positions, a lower ratio should be expected in a material restricted to open syllables. Therefore, the temporal manifestations of stress in the present study correspond quite well to the ratio obtained in the earlier investigations.

10.3 Finnish

The study of word stress in Finnish poses special problems owing to its fixed position in the initial syllable of the word. Consequently a different technique from that used in the studies of Swedish and Spanish had to be used. In spite of the fact that it cannot separate effects of position and stress, it permits some comparisons between Finnish and the other languages in the study.

Stress in Finnish should, if temporally manifested as in for example Swedish and Spanish, show up in the speech signal as an increase in the duration of a syllable. Conversely, shifting a specific syllable from a stressed to an unstressed position should lead to a decrease in duration of the syllable. Though it seems unlikely in the light of the complexity of quantity in Finnish that stress should be manifested in this simple manner, a testing of the prediction may still be considered worthwhile.

A set of six words were selected in which the same sequence of segments occurred in the first (stressed) and second (unstressed) position. Two of the words were not real Finnish words, though they con-
formed to Finnish rules for word formation.

The test words, with the sequences of test segments underlined, were:

*jatana /játana/ (a nonsense word)
kataja /kátaja/ 'juniper'
nakata /nákata/ 'throw'
kanava /kánava/ 'canal'
takana /tákana/ 'behind'
*tanaka /tánanaka/ (a nonsense word)

The words were produced within the sentence frame:

Pekka sánoo ----- tänäään.
'Pekka says ----- today.'

All segments in the test words were measured. However, the segments of /ja/ were treated as a unit, since the different parts could not be separated.

Table 10-3. Mean durations (M) and standard deviations (s) for vowels and consonants in the test words; test segments underlined. (/ja/ was measured as a unit.)(Finnish).

<table>
<thead>
<tr>
<th></th>
<th>1st pos.</th>
<th>2nd pos.</th>
<th>3rd pos.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>C</td>
<td>V</td>
<td>C</td>
</tr>
<tr>
<td>jatana</td>
<td>114*</td>
<td>4*</td>
<td>76</td>
</tr>
<tr>
<td>kataja</td>
<td>75</td>
<td>4</td>
<td>53</td>
</tr>
<tr>
<td>nakata</td>
<td>50</td>
<td>3</td>
<td>64</td>
</tr>
<tr>
<td>kanava</td>
<td>76</td>
<td>6</td>
<td>59</td>
</tr>
<tr>
<td>takana</td>
<td>63</td>
<td>12</td>
<td>53</td>
</tr>
<tr>
<td>tanaka</td>
<td>79</td>
<td>13</td>
<td>52</td>
</tr>
</tbody>
</table>

M 69 56 68 78 55 49
Exchanging first the mean durations for each position as indicated at the bottom line of the table, we get a rough estimate of the positional variation. It appears that this variation is small for the vowels as well as for the consonants, though the range of the vowels seems to be somewhat greater than that of the consonants. The duration of the vowels is shortest in the final position and longest in the medial (second) position. Thus it can be concluded that the predictions are not borne out. Shifting the test sequence from the first (stressed) position led to an increase of duration rather than the predicted decrease.

While the varying context of the vowels may have influenced the calculated mean durations for the different positions, this is not true for the positionally varied segment sequences. Their durations, when in the first and second syllable position respectively are shown in Figure 10-1.

Figure 10-1. The duration of 4 segment sequences in the first (V1 - upper part of the figures) and second (V2 - lower part of the figures) word position (Finnish).
The data contained in the figure confirm the observations that the vowels were longer in the second syllable position than in the first. The ratio for the first-syllable to the second-syllable vowel durations is .71.

Is this ratio simply an expression of positional differences, or has it something to do with stress? On the basis of the very similar observations of the temporal manifestations of sentence stress by Lehtonen (1970, p.145-147), it seems reasonable to assume that the durational differences are in effect stress-dependent (see 8.3.2). The reason why stress has very weak temporal correlates in the first vowel should be sought in the quantity distinction of Finnish, which it seems to be especially important to maintain in the first syllable. Quite obviously, an increase in the duration of the vowel due to stress should not be separable from an increase due to a shift from a phonologically short to long vowel. It follows that we cannot generalize from the present observations based on phonologically short first vowels to all types of Finnish words, since the length of the vowel determines how word stress is realized. Phonologically long vowels might well increase in duration without disrupting the quantity distinction. This effect due to sentence stress was found in Lehtonen's study.

These data thus support the assumptions of Lehtonen that (sentence) stress is temporally manifested in Finnish not merely in the stressed syllable, but also in the following one through a complex interaction with quantity.

10.4 Comparisons

It may now be concluded that stress is quite differently manifested in the three languages under observation. Comparing first Swedish and Spanish, we recall their different ratios for the durations of unstressed to stressed syllables (VC- or CV-sequences), being lower for the Swedish as compared to the Spanish data. Figure 10-2 illustrates these observations.
Irrespective of the absolute durational differences, which may partly be attributed to the individual speech rates of the two subjects, it appears that stress is more strongly manifested in Swedish than in Spanish. This is especially evident from the durations of stressed long vowels, but also from the durations of the immediately preceding and following consonants. As a consequence, the stressed and unstressed syllables (VC- and CV-sequences) are much more different in length in Swedish than in Spanish. The quantity distinction contributes to this differentiation between stressed and unstressed syllables in Swedish. The absence of quantity in Spanish should be the reason for the similar stressed and unstressed syllables.

The patterns observed in Figure 10-2 are most probably more similar than would have been the case if a more varied material had been used. This should be expected as the differentiation between stressed and unstressed syllables in Swedish would be even greater if closed syllables, so typical of Swedish, had been included.
However, in spite of these differences between Swedish and Spanish, the Swedish and the Spanish speaker seem to exhibit similar effects of stress manifested in the stressed vowel and the consonants preceding and following the stressed vowel. This is evident upon comparison with Finnish, in which stress presents a much more varied patterning of its temporal manifestations. The durational effects appear to be spread over the first two syllables of the word, though in such a way that the quantity relations between the first and second syllable are not disrupted.

Thus, just as for Swedish, quantity seems to interact with the manifestation of stress. The interaction is much more complex in Finnish than in Swedish, owing to the presence of the quantity distinction in all positions of the word.

Summarizing these comparisons we may characterize the three languages as representing three different types of manifestation of stress: (1) strongly manifested stress, centered to the stressed syllable (Swedish), (2) weaker manifested stress, centered to the stressed syllable (Spanish), and (3) weaker manifested stress, distributed over adjacent syllables (Finnish).

10.5 Conclusions

The different temporal relations between stressed and unstressed syllables in Swedish, Spanish, and Finnish, as reflected in the present data, should have consequences for perceived rhythmic differences. The impression of equal syllable durations in some languages and equal intervals between stressed syllables in others was the basis for the distinction between syllable-timed and stress-timed languages (Pike, 1946, p.35). However, the assumption of Pike, and later, Abercrombie (1967, p.97) of different timing mechanisms associated with specific breathing patterns has not been supported (1.2.1). The present data, and earlier studies (cf. Dauer, 1983) point rather to differences in language structure as the major determinant of rhythmic diversity.
11 CONCLUSIONS

11.1 Summary and conclusions

The study focussed on temporal aspects of rhythm in Swedish examined in a series of experiments. These experiments were reported in the Swedish main study and interpreted within a conceptual frame outlined on the basis of previous research on speech rhythm.

The basic variation throughout the Swedish main study concerned the number of syllables in phrase-medial stress groups. The first experiment aimed therefore to give a descriptive base-line for the following ones. The results indicated the presence of different types of temporal adjustments. There was a decrease of duration of stressed and unstressed vowels, when the number of syllables in the stress group was increased. This pattern was typical for stress groups with less than four syllables. In longer stress groups some unstressed syllables increased in duration, when other unstressed syllables were added to the preceding ones. This indicated a rhythmic alternation between unstressed syllables, the basis of which was taken to be a partitioning of longer stress groups into smaller rhythmic units.

On the assumption that stressed syllables are basic line-up points for grouping syllables together into rhythmic units, the properties of the stressed syllable were examined in the second experimental study. The data clearly indicated that, irrespective of the length of the stressed vowel, the stressed syllable (VC-sequence) appeared as a stable temporal unit across the variation of the number of syllables in the stress group. In addition, the plausibility of syntactic, or word, boundaries as alternative grouping criteria was considered. The results showed that boundaries as well as stressed syllables seemed to be possible bases for grouping.

The third experimental study concerned speech rate variations. It was hypothesized that these variations should affect the phrase as a whole through uniform changes of processing rate. It was also assumed that the temporal relation between stressed and unstressed syllables
would remain unchanged across the variations of speech rate. Apart from a minor deviation from this pattern, both predictions were borne out. However, the stability of the temporal relation between stressed and unstressed syllables was brought about by a non-uniform rate adjustment restricted to the stressed syllable. This deviation from the general tendency in the data demonstrated the interaction between language structure constraints and constraints on speech processing.

Finally, the temporal effects on phrase-medial stress groups of the length of surrounding stress groups were examined. The results indicated two types of adjustment to the contextual variation. There was a strong overall tendency to adjust the duration of stress groups to the length of the phrase, so that the longer the phrase the shorter the stress groups contained in it. These adjustments were seen as being due to uniform changes of processing rate over the phrase - that is, the same type of effect as observed in the experiment on variations of speech rate. The phrase-length adjustments were attributed to capacity restrictions on articulatory planning. In addition, there was a tendency for adjustments to be made between adjacent stress groups, occurring in such a way that differences in duration were counteracted. These adjustments were associated with local rate variations over the different stress groups in the phrase. In most cases they appeared in short phrases, and the strongest effects occurred when adjacent stress groups differed by either one or two syllables.

In a supplementary test, the respective effects of preceding and following stress groups were examined. Both effects were significant, though it appeared that the duration of a specific stress group was effected more by the length of following than the preceding stress group. At least this was true when the affected stress group was short. Phrase-length effects were found to be strong. However, there was no direct evidence for stress-group adjustments in this test. The discrepancy between the results of the two experiments was related to the size of the different sentences used. The differences might be explained, if a limited planning domain is assumed. Then longer phrases would be processed in parts and the stress-group adjustments
would vary with the length of the phrase.

The cross-language comparisons of Swedish, Spanish, and Finnish were undertaken to inquire into the generality of the assumed model. The study was explorative and, owing to the limited number of speakers and languages included, the conclusions should be seen as tentative.

The first experiment concerned the number of syllables in phrase-medial stress groups. The observed temporal effects closely resembled those in the Swedish main study. Vowels and consonants in stressed syllables, especially when phonologically long, tended to decrease in duration to a greater degree than short ones. However, temporal adjustments never occurred at the expense of important structural relations. This was clearly evident from the Finnish data. The results were interpreted in terms of interactions of language-specific structural constraints and language-independent constraints on processing.

In the second experiment the temporal manifestations of stress were compared for the three languages. Predictions about the differentiation between stressed and unstressed syllables, based on structural differences in the three languages, were confirmed in the temporal data. Differences of language structure were therefore suggested as the basis for the characterization of languages as rhythmically different.

11.2 Final remarks

The present study was restricted to the temporal correlates of the complex phenomenon of rhythmic structure. This naturally is a simplification, as rhythm without doubt may also be associated with other aspects of the speech signal. Intensity and fundamental frequency may be such features contributing to the prominence of the stressed syllable. Moreover, the rise-fall pattern of the fundamental frequency contour between one stressed syllable and the next clearly marks out stress groups. (Thorsen, 1979; Kohler, 1983). Kohler also
suggests that this pattern is an important cue for the perception of stressed syllables.

Secondly, experimental conditions model normal speech situations to some extent only. It might well be expected that an experimental situation contributes to more unnatural renderings of test sentences than would result if the same sentences occurred in normal speech situations. Readings of similar sentences from lists as in the present study may lead to some exaggeration of rhythm. However, there is no reason to expect that rhythm is qualitatively different in experimental and normal situations.

Finally, experiments often do not capture the normal interspeaker variation. In this respect, however, the present study was less constrained. Through the inclusion of five speakers in the Swedish main study a considerable flexibility of speech processing was revealed. The speakers in the study to a certain extent exhibited individual "rhythmic profiles".
BIBLIOGRAPHY


SAOB (1898 - ) = *Svenska Akademins Ordbok*. Lund.


<table>
<thead>
<tr>
<th>No.</th>
<th>Author</th>
<th>Title</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>26</td>
<td>Monica Loeb</td>
<td>Vonnegut’s Duty-Dance with Death-Theme and Structure in Slaughterhouse-Five</td>
<td>1979</td>
</tr>
<tr>
<td>27</td>
<td>Algot Hellbom</td>
<td>Ordbok över Njurundamålet.</td>
<td>1980</td>
</tr>
<tr>
<td>28</td>
<td>Kerstin Elert</td>
<td>Portraits of Women in Selected Novels by Virginia Woolf and E. M. Forster</td>
<td>1979</td>
</tr>
<tr>
<td>29</td>
<td>Birger Liljestrand</td>
<td>Strindbergs Mäster Olof-dramer. En studie i 1800-talets dramaspåk. II.</td>
<td>1980</td>
</tr>
<tr>
<td>30</td>
<td>Maria Bolander</td>
<td>Predikativens funktion i svenskan.</td>
<td>1980</td>
</tr>
<tr>
<td>31</td>
<td>Kerstin Elert</td>
<td>Portraits of Women in Selected Novels by Virginia Woolf and E. M. Forster</td>
<td>1979</td>
</tr>
<tr>
<td>32</td>
<td>Veronica Bonebrake</td>
<td>Historical Labial-Velar Changes in Germanic.</td>
<td>1979</td>
</tr>
<tr>
<td>33</td>
<td>Internordisk språkförståelse</td>
<td>Utg. av C.-C. Elert.</td>
<td>1981</td>
</tr>
<tr>
<td>34</td>
<td>Magnus von Platen</td>
<td>Privatinformation i skolan. En undervisningshistorisk studie.</td>
<td>1981</td>
</tr>
<tr>
<td>35</td>
<td>Marja Sjöberg</td>
<td>Brott och straff i Västernorrland 1861—90.</td>
<td>1981</td>
</tr>
<tr>
<td>37</td>
<td>Svenskans beskrivning 12</td>
<td>Utg. av C.-C. Elert och S. Fries.</td>
<td>1981</td>
</tr>
<tr>
<td>38</td>
<td>Ingeborg Nordin Hennel</td>
<td>Dömd och Glömd. En studie i Alfhild Agrells liv och dikt.</td>
<td>1981</td>
</tr>
<tr>
<td>40</td>
<td>Claes-Christian Elert</td>
<td>Ljud och ord i svenska.</td>
<td>1981</td>
</tr>
<tr>
<td>41</td>
<td>Rolf Hedquist</td>
<td>Språklig kommunikation med förbehåll. En studie i det politiska språkets pragmatik.</td>
<td>1981</td>
</tr>
<tr>
<td>42</td>
<td>Bo Sundin</td>
<td>Ingenjörsvetenskapens tidevarv. Ingenjörsvetenskapsakademins, Pappersmassekontoret, Metallografiska institutet och den teknologiska forskningen i början av 1900-talet.</td>
<td>1981</td>
</tr>
<tr>
<td>43</td>
<td>Readings on Unbounded Dependencies in Scandinavian Languages. Edited by Elisabet Engdahl and Eva Ejerhed.</td>
<td>1982</td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>Roger Qvarsell</td>
<td>Ordnad och behandling. Psykiatri och sinnessjukvård i Sverige under 1800-talets första hälft.</td>
<td>1982</td>
</tr>
<tr>
<td>48</td>
<td>Åke Jonsson</td>
<td>Den omsorgsfulle ordmålaren. Studier i Sven Jerrings radio-språk mot bakgrund av radions allmänna syn på språket under de första decennierna.</td>
<td>1982</td>
</tr>
<tr>
<td>49</td>
<td>Nordsvenska. Språkdrag i övre Norrlands tätorter. Utg. av C.-C. Elert och S. Fries.</td>
<td>1982</td>
<td></td>
</tr>
<tr>
<td>51</td>
<td>Maria Haar</td>
<td>The Phenomenon of the Grotesque in Modern Southern Fiction. Some Aspects of Its Form and Function.</td>
<td>1983</td>
</tr>
</tbody>
</table>


54. **Birger Liljestrand.** Tal i prosa. Om svenska författares anföringsteknik. 1983.


64. **Birgitta Hene.** »Den dyrkade Lasse och stackars lilla Lotta.« En syntaktisk-semantisk studie av personbeskrivande adjektiv och adverb i populära ungdomsböcker. Diss. 1984.


Försäljning och distribution:
Almqvist & Wiksell International, Box 45150, 104 30 Stockholm, Sweden.