GOTHENBURG MONOGRAPHS IN LINGUISTICS 9

Aspects of Swedish Speech Rhythm

by

ANDERS ERIKSSON

Department of Linguistics
University of Göteborg
1991

GOTHENBURG MONOGRAPHS IN LINGUISTICS 9

Aspects of Swedish Speech Rhythm

DOCTORAL DISSERTATION

publicly defended in Stora Hörsalen, Humanisten,
University of Göteborg, Renströmsgatan 6, Göteborg,
on September 28th, 1991 at 10.00 a.m.,
for the degree of Doctor of Philosophy.

by

ANDERS ERIKSSON

Department of Linguistics
University of Göteborg
1991

ABSTRACT

This study examines some aspects of speech rhythm, with particular reference to Swedish. A background to the problem area is given and some fundamental problems pointed out. Some theoretical issues are also studied. The question of how to describe and model interstress interval duration is addressed. It is shown, using published data from five languages, that interstress interval duration can be described as a linear function of the number of syllables. Languages seem to fall into two classes, however. It is suggested that this is due to differences in the duration of stressed syllables. It is also shown that a linear growth in interstress interval duration, as a function of the number of syllables in the interval, does not preclude the existence of interval-internal temporal compensations.

Speech rhythm in Swedish is studied experimentally in both production and perception. In the production study, the hypothesis that interstress interval duration may be described as a linear function is tested on a recorded material consisting of 5 sentences read by 30 speakers. An analysis of the results gives support for the hypothesis. The possible existence of compression of syllables, as a function of interval length, is also studied, but no significant effect is found.

The perception part of the study describes two sets of experiments. In one type of experiment the locations of stress beats in a phrase of read poetry are studied. Stress beats are found to be closely associated with the onsets of the stressed vowels. Duration perception of interstress intervals is also studied in a series of experiments, in which stimuli and experimental conditions are varied. Duration perception is shown to be quite accurate, indicating that subjects are capable of determining that interstress intervals are of unequal durations in speech.

Key words: speech rhythm, Swedish, stress-timing, syllable-timing, mora-timing, isochrony, stress beats, duration perception.

ISBN 91-628-0489-8
ISSN 00346-6248

Printed in Sweden
Kompendietryckeriet, Kållered, 1992

Table of contents

II Temporal regularity in speech production.

III Temporal regularity in speech perception.

IV Discussion and suggestions for further research.

8 Summary of the results, discussion, and suggestions for further research. 199

List of tables.

List of figures.

Acknowledgements

First of all I would like to thank my two supervisors; my main supervisor Jens Allwood and my assistant supervisor Bertil Lyberg.

Jens Allwood must be thanked because he was the one who encouraged me to take up linguistics in the first place. He has helped and encouraged my work in many ways. His many critical, but constructive, suggestions have helped greatly in improving all sections of this work. I am particularly grateful for his valuable suggestions for improvement concerning logical structure.

Bertil Lyberg has been my assistant supervisor. His many ideas, particularly concerning methodological questions in connection with the experimental parts, have been of great help. He has also provided me with several of the more important papers from which I have drawn information.

Roeland van Hout has been my guide into the wonderland of statistical analysis. Roeland has unselfishly devoted numerous hours of his time helping me with statistical problems. Naturally I take full personal responsibility for any mistakes that may still remain

Eva Strangert has provided valuable help in the form of discussions of models and ideas. She has kindly given me access to some of her own data on which I have tried my ideas, particularly during the early stages. She has also helped me by suggesting relevant literature on many occasions.

I would like to thank Olle Engstrand, head of the phonetics department at the University of Stockholm, for giving me full access to all the technical facilities of the phonetics laboratory, where most of the analysis and preparation of stimuli for the perception experiments was done.

I want to thank Una Cunningham-Andersson for her careful proof reading of the text and many valuable suggestions for clarifications.

Among the many friends and colleagues, who have helped and encouraged me during my work, there are a few that I would like to mention particularly: Rolf Lindgren for never tiring in helping me, and discussing with me all sorts of technical and phonetic problems; Sven Strömqvist for encouragement and help, particularly on methodological questions; Robert Bannert for his kind interest and encouragement; Elisabeth Ahlsén and Joakim Nivre read preliminary versions of parts of the manuscript and provided many valuable suggestions for improvement; Jaan Kaja wrote some of the computer programs I needed for the analyses in Chapter 4; Mats Dufberg helped me on numerous occasions when the desk-top program and I could not quite agree on who was to be the master; Lennart Andersson kept the spirit of scepticism high by hardly ever believing in anything; Gunilla Wetter and Tore Hellberg have made life easier by helping me with all sorts of practical matters.

Overview of the study.

The present study examines some aspects of speech rhythm, with particular reference to Swedish. The focus of the study is on regularities and irregularities in the temporal domain. The study addresses some theoretical and methodological questions in connection with the investigation of speech rhythm. It also contains two empirical studies of regularity in production and in perception.

The study is divided into four parts. The first part (Chapters 1—3) provides a theoretical, methodological, and historical background for the latter parts which are experimental. Part II (Chapter 4) is an experimental study of regularity and variation of interstress intervals and syllables in speech production and Part III (Chapters 5—7) a study of speech perception. Part III will primarily be concerned with determining what the possibilities are of correctly perceiving interstress interval durations in speech. Part IV (Chapter 8) contains a summary and discussion, and some suggestions for further research.

Chapter 1 presents some concepts and results from general rhythm research, which were found useful also in the analysis of speech rhythm done in this study.

In Chapter 2 previous research on speech rhythm is reviewed and discussed. An attempt is made to evaluate the original claims made about the rhythms of languages by Pike and others against the results obtained in experimental research done over the years.

Chapter 3 is a discussion of some theoretical and methodological issues. A linear model of interstress interval duration as a function of the number of syllables in the interval is suggested and its validity is tested on published data from five languages. The gradual compression of syllables as a function of increased interval length has often been put forward as a test for tendencies to isochrony in certain languages. Does such a tendency exclude the possibility that interstress intervals grow with a constant amount per added syllable? This question is examined in one of the sections. The perception of syllable 'beats', perceptual isochrony, and the possibility of measuring regularity are also among the questions discussed in this chapter.

Chapter 4 presents the results from a study of speech production. Five different sentences read by 30 speakers are examined with respect to interstress interval and syllable durations. The hypothesis put forward in Chapter 3, that interstress interval duration may be described as a linear function of the number of syllables in the interval, is tested on the material. Whether or not a similar relationship holds between interval duration and the number of phonemic segments is also studied. Variation is examined with respect to individual differences, and sex and age differences. Syllable durations are examined as a function of the number of segments as well as stress and syllable position. An attempt is made to determine how different variables interact to produce interstress interval durations.

Chapter 5 contains a brief description and discussion of some theoretical and methodological questions in connection with time and duration perception. It was found that some knowledge of these questions was essential for the planning and understanding of the perception experiments presented in the following chapters.

Chapter 6 describes a perception experiment in which the locations of stress beats in a phrase of poetry read aloud are studied. The method used in the experiment is perceptual matching of stressed syllables and clicks copied into the phrase. The main aim of the study is to determine how precisely it is possible for subjects to determine the locations of stressed syllables. Locations are also examined as a function of phonetic context.

In Chapter 7, duration perception of interstress intervals is studied in a series of experiments. Subjects were presented with the interstress intervals from the phrase used in the stress beat experiment, under two different conditions. The task was to judge the durations of the interstress intervals. Identical tests, where the stimulus material was noise, were also made to obtain reference material. The aim of the study was to establish some kind of 'just noticeable difference' for interstress interval durations in speech.

In Chapter 8, the experimental results are summarized. The questions of regularity in production and perception are discussed against the background of the empirical results in this and other studies. Based on the results obtained here and the results from other studies, some suggestions for future research are made.

Part I

Background and theoretical considerations

Chapter 1

Rhythm.

Rhythm plays an important role in human behaviour and in the way we interpret the world around us. This is reflected in language through the many expressions explicitly referring to rhythm, 'rhythm of the seasons', 'daily rhythms', etc. The role of rhythm is most clearly reflected, however, in the central role rhythmic behaviour plays in all cultures, and has done, as far as we know, in all times. Singing, dancing and playing music are very important activities in all cultures and in these activities, rhythm is a very important aspect, perhaps the single most important one.

Aristotle and Plato went as far as claiming rhythm to be one of the very qualities that make us human and that there is a direct correspondence between rhythm and character.

"And whereas animals have no sense of order and disorder in movement ('rhythm' and 'harmony', as we call it), we human beings have been made sensitive to both and can enjoy them" (Plato, *Laws*, book II, p. 87)

"Rhythm alone, without harmony, is the means in the dancer's imitations; for even he, by the rhythms of his attitudes, may represent men's characters, as well as what they do and suffer." (Aristotle, *Poetics*, p. 24)

And one only has to mention Pythagoras to remind oneself of the central role music played in classic education.

The interest in rhythm has not diminished over the years. The study of rhythm has in consequence caught the interest of many scientists. As a background to this study of speech rhythm, I would like to discuss rhythm in more general terms and mention some results from general rhythm research that are relevant in this context.

The common denominator between those things that we tend to call rhythmical is little more than a certain regularity of occurrence. The perception of rhythm often includes the subdivision of a series of events into groups, but this is not always necessary. In its simplest form, a rhythm need not consist of more than one element, recurring with a certain regularity.

The experience of rhythm can be grouped into two subgroups, those rhythms that we tend to construct from our knowledge of the world, like the rhythmic alternation between day and night, and those rhythms that can be immediately perceived, like musical rhythms. What this difference means is closely related to what the psychologists call 'the psychological present'. Fraisse (1978) describes 'the psychological present' as *"the temporal field in which a series of events is rendered present and integrated into a unique perception"* (p. 204). I must admit that I do not find the meaning of this altogether clear, but I interpret it as meaning the time span within which a group of events is immediately perceived as belonging together and forming a unity (a group). The psychological present is closely connected with what is called 'short term memory'. Fraisse seems to regard the two concepts as more or less synonymous. *"This duration limit corresponds to what has been called the 'psychological present'.... This phenomenon is also called 'short-term storage'"* (Fraisse 1982, p. 158). One can perhaps say then that we may perceive rhythms directly if the events that form a rhythmic group occur within a time span not exceeding the limits of the short term memory. The size of this time span seems to be in the order of 4—5 seconds. I quote Fraisse (1956) again, summing up a review of different findings: *"En effet il semble y avoir une durée totale limite, quelque soient la nature du groupement et le nombre des éléments. Cette durée totale semble de l'ordre de 4 à 5 secondes"* (p. 17). This is also the limit he found in one of his own studies of the production of rhythms. *"Nous avons trouvé que la dureé des groupes de frappes les plus longs, telle que la perception de l'unité ne disparaisse pas, était aussi de 4 à 5 sec."* (p. 17).

The difference between the two forms of rhythm perception, constructed rhythms vs. immediately perceived ones, may be one of degree rather than an absolute one. It is the latter form of rhythms, those which occur within the time domain of the psychological present, that will be the main concern in this study.

1.1 Definitions of rhythm.

There have been many attempts to define rhythm in more precise terms. All definitions, however, are built on the regular occurrence of some event or events and some kind of

structuring (grouping) of these events. They may differ, though, with respect to which of the two sides, temporal regularity or temporal structuring, that is emphasised. A quotation from Woodrow (1951) may serve as an example of a definition of the first kind (although he mentions both sides).

"By rhythm in the psychological sense, is meant the perception of a series of stimuli as a series of groups of stimuli. The successive groups are ordinarily of similar pattern and experienced as repetitive. Each group is perceived as a whole and therefore has a length lying within the psychological present." (p. 1232)

A definition giving priority to the structure is that by Allen (1972).

"Rhythm is the structure of intervals in a succession of events." (p. 72)

One might suspect that Allen's definition is inspired by the way we talk about musical rhythms, where the structure of events is precisely what one uses to characterize musical pieces rhythmically, as waltzes, foxtrots, and so on. However, whether one chooses to emphasise regularity or structure, the common element is always the occurrence of events in a succession.

1.2 Rhythm is an activity by the subject.

One could say with Fraisse that rhythm is a phenomenon that has to do with perception only. Rhythm is a construction by the listener; *"all perceived rhythm is the result of an activity by the subject since, physically there are only successions"* (Fraisse, 1982, p. 156). In a sense this is true, of course, but in this creative process there is also a strong correlation between the character of the stimulus and the resulting perception. First of all, not all types of stimuli give rise to a perception of rhythm, not even if they mean successions of events. The rate and the regularity of successions, among other things, play a role (see 1.4). It is also the case that although several different types of successions all may be described as rhythmical, the perception of these rhythms may differ widely in character. One has only to consider the different musical rhythms to make this clear. A waltz and a tango are both perceived as rhythmical in a general sense but there are also differences. In fact these differences are marked enough for us to be able to classify them as belonging to different categories. There are, thus, interesting differences between rhythms that have as much to do with differences in the structure of the stimuli as with the constructions by the subject.

1.3 Temporal and structural aspects of rhythms.

Although this investigation is about rhythm and regularity in the temporal domain, it should be pointed out that we may also talk about aspects of rhythm that are not primarily temporal in character. Repeated visual patterns, for example, are sometimes described as rhythmical.

This shows that certain repetitive structures can be perceived as rhythmical although there is no temporal element present. It seems possible to perceive even non temporal repetitions as rhythmical. But more importantly in the particular context of this study, most temporal rhythms also have a quality of structural regularity which is equally important, sometimes even more so. In language, both temporal and structural regularity are clearly reflected in poetry. Particularly in classical poetry there are very strict rules concerning the structure of metric feet and lines of a poem. A line may be required to consist of a certain number of feet and each foot must be of a certain type. In the case that all feet have the same structure, say iambic, this will result in lines with perfect structural regularity, and we may talk about the poem as having an iambic rhythm. In most cases this will also result in a certain temporal regularity if the poem is read aloud, but that is a different aspect and the connection need not be very strong. As the test phrase used in the experimental studies in Chapters 6 and 7 clearly demonstrates, feet in a line of poetry with structurally very similar feet may vary considerably with respect to durations. And it is quite possible for structurally different feet, that is feet containing different numbers of syllables, to be similar in duration. So the connection between structural regularity and temporal regularity is not a necessary one. A structurally regular sequence may display a high degree of temporal irregularity and a sequence of structurally unequal elements may display a high degree of temporal regularity. It is conceivable that both these aspects play important roles in rhythm perception. In the study of speech rhythm, one would like to know how these factors interact. Is a structurally regular, but temporally irregular, sequence of feet, perceived as more or less rhythmical than a temporally regular sequence of structurally different feet? My intuitions would tend to favour the structural side. Again, using music as an example, the difference between a waltz and a tango is not that tangos are more regular but that the structures of the recurring elements are different. An interesting question is just how irregular the tempo of a waltz may be before its character as a waltz is completely lost. My guess would be that constancy of structure is more important for the perception of a certain rhythmic character than temporal regularity. There may be a correlation between the complexity of the structure and the resistance to temporal distortion, but to my knowledge these factors have not been studied. These questions are, of course, highly relevant in the study of speech rhythm since different languages have different syllable structures and, perhaps even more important, different distributions of these structures. This may very well result in different rhythm perceptions even if mean durations of feet and syllables are fairly similar.

1.4 Subjective rhythmization and grouping.

The tendency to perceive groups among a series of events is very strong. Even when presented with a perfectly regular sequence of identical stimuli there is a strong tendency for subjects to perceive the stimuli in groups. This phenomenon was known more than a century ago, and has been called *'subjective rhythmization'*. A simple reflection of this is

the way the sound of a clock is described in many languages. Although the sound, usually produced by a pendulum, by the very nature of the clock, forms a perfectly regular sequence the beats are, nevertheless, thought of as coming in pairs. The sound is described as: Tic-tac (French), ticktack (Swedish), tick-tock (English).

Like the immediate perception of groups, subjective rhythmization is also limited to a certain time interval. Bolton (1894) found, in a classical study of this kind, that subjective rhythmization took place when stimulus beats were presented at rates varying between 115 ms between beats and 1580 ms between beats. These values are probably far too precise since there is a great deal of individual variation, but other investigations have confirmed Bolton's general results. Subjective rhythmization seems to be possible only if the durations between successive sounds are in the range 0.1 to 2 seconds. There also seems to be some connection between the rate at which stimuli occur and the number of elements in the perceived groups. There is not a very high degree of agreement between the figures I have seen reported (Woodrow, 1951; Fraisse, 1978). The general tendency, however, seems to be that the faster the rate between stimuli the more elements are perceived as being grouped together.

Subjective rhythmization occurs when subjects are presented with a sequence of stimuli that are regularly spaced and identical. One can, however, influence the perception of groups by introducing different types of accents on some of the stimuli. The introduction of an accent normally results in one of two opposite perceptions of structure. The accents can be perceived as beginning the groups or they can be perceived as ending them. This effect is a complex one that depends, among other things, on the type of accent, durations between sounds, and the number of elements in the group. There is also considerable inter-individual variation. If one considers the simplest type, a regular sequence where every second element is accented, the following seems to hold true for most subjects: if the accented element is louder or higher in pitch, it is perceived as beginning the group and if it is longer in duration, it is perceived as ending the group (Woodrow, 1951; Fraisse, 1956, 1978, 1982). The effect can probably be generalized to groups of more than two elements (Fraisse, 1982; Allen, 1975). Some linguists (Wenk and Wioland, 1982; Allen, 1975) have suggested that this phenomenon may explain why we perceive stressed syllables as beginning interstress intervals in some languages, where stress is mainly a function of pitch changes, (e.g. English) and ending them in others, where stressed syllables are longer, (e.g. French). The idea may have some truth in it but must be considered a rather weak one. It is, for instance, the case that stressed syllables in languages with marked stress, like English and Swedish, are normally also longer. The whole complex needs much further investigation but the phenomena as such no doubt have relevance for the study of rhythm perception in language.

1.5 Personal tempo and preferred tempo.

Psychologists have studied many different kinds of spontaneous movements and measured the frequencies with which they occur. The frequency in this type of behaviour varies between individuals. It is, therefore, often referred to as *'personal tempo'*. (The term 'spontaneous tempo' is also used.) Most of these studies are not relevant in this context, but at least one type of result is worth mentioning. Frischeisen-Köhler (1933) and Mishima (1951–1952) (cited in Fraisse, 1982) have measured spontaneous tempo in tapping. In these studies, the lengths of the intervals between taps varied between 380 ms and 880 ms. Fraisse says: *"One can assert that a duration of 600 ms is the most representative"* (p. 153). This is a piece of information one should keep in mind. In many studies of rhythm perception, finger tapping is used as a means to represent rhythm. In that context it is important to know what kind of tapping subjects would be likely to produce in the absence of any outside stimulus rhythm or if the influence of the stimulus is weak.

Preferred tempo is the rate of a succession of events that subjects, when asked to judge, find most natural. Fraisse (1982), reviewing several investigations, claims that an interval of 600 ms between events seems to be the most frequently reported one. Now, this is the same rhythm as that of the spontaneous tempo for tapping mentioned above. It would be natural to assume that the two should be correlated for a particular individual. But this does not seem to be the case. In a study of the correlation between personal tempo and preferred tempo Mishima (1965, cited in Fraisse, 1982) found a correlation of only .40.

1.6 Rhythmic synchronization and rhythmic anticipation.

It is very common in human behaviour to react with some kind of body movement as a response to rhythmic stimuli. People tap their fingers, stomp their feet or rock their bodies to the rhythms of music, and they do it in synchrony with the rhythm of the stimulus. This ability to synchronize movement with an outside stimulus has been subject to many studies. The ability to synchronize has an interesting property that makes it an exception to other forms of reactions to stimuli. Normally a reaction to a stimulus succeeds the stimulus by some time interval—the reaction time. In synchronization this is not the case. Movements in time with a rhythmic stimulus are almost simultaneous with the beats of the stimulus rhythm. In fact, studies have shown that the accompanying movements tend to precede the stimulus (Miyake 1902, King 1962). In finger tapping, taps tend to precede the stimulus by some 30 ms (Fraisse, 1966). This shows that the taps cannot be simple reactions to the stimuli. The explanation proposed is that subjects anticipate the beats, using the durations between successive beats as the predictor. There are quite a number of results that support such an interpretation. In an interesting experiment by Fraisse and Voillaume (1971), subjects were told to synchronize to a stimulus beat. But the set-up was such that the

subjects' own taps initiated the 'stimulus sounds'. Sounds and taps were thus perfectly simultaneous. The result was that the subjects accelerated the tempo as if they were trying to anticipate their own tapping. If asked to follow the beats instead of preceding them, subjects find the task very difficult, particularly if intervals are shorter than one second (Fraisse, 1966). They also find it difficult to insert 'extra taps' between the beats (Fraisse and Erlich, 1955). It is even possible to synchronize to a sequence of beats which are not equidistant. In an experiment, Erlich (1958) asked subjects to tap to accelerating or decelerating sequences. Synchronization was possible but the precision decreased with the rate of change in tempo (rates of change varied between 10 and 100 ms/beat and initial intervals between 700 ms and 2000 ms). Finally, subjects establish their synchronization very rapidly. Three taps is usually enough to find the right rhythm (Fraisse, 1966). All these findings are significant for the evaluation of rhythm experiments involving tapping that have been made to investigate speech rhythm.

Even from this very brief overview of some of the work in general rhythm research it should be clear that the questions dealt with are highly relevant for the study of speech rhythm as well. It should be observed, however, that most of the results have been obtained using simple stimuli like clicks or simple tones. One should, therefore, use caution in generalizing these results to speech. Speech stimuli are far more complex than tones or clicks. Complex and unforeseeable differences may result when speech is used instead of tones. There is, however, no reason to doubt that the general principles that govern rhythm production and perception are the same whether the stimulus is speech, music or tones.

An interesting observation is that a duration of about 500—600 ms between events pops up in many different contexts. Typical values in many spontaneous activities (like walking) fall in that range. It is representative for intervals in personal tempo and also typical for preferred tempo. In synchronization tasks, subjects find it easiest to synchronize if stimuli are presented at rates in that region. In tapping tasks without an outside stimulus, tapping is least variable in that range (Fraisse, 1956). It is also the time interval that is perceived with the greatest precision. Durations of this magnitude also seem to play a role in speech. In a study by Dauer (1983) of interstress interval durations in five different languages the mean durations were all in the range 380—510 ms and, as I will show in Chapter 4, comparable data from Swedish give interval durations in the range 500—700 ms. Allen (1975) reports similar values from his own (1972) study, 300 ms to 600 ms, and that of Abe (1967), 400 ms to 700 ms. Most intervals in the study by Shen and Peterson (1962) also fall into roughly the same range.

Now the reason for the importance of this time interval has yet to be found. It could be an indication of the existence of some timing mechanism operating with a frequency in that range, but other explanations are also possible. There can be complex interactions of several different timing mechanisms that often, but not always, result in similar frequencies. For example, Lenneberg's (1967) hypothesis that the basic time unit in the motor programming of speech production is in the order of 160 ± 20 ms, would be in agreement with this

idea. Typical syllable durations are of that order and typical interstress interval lengths are 3—4 syllables, thus resulting in interstress interval durations of 500—600 ms. Data from a study by Faure, Hirst, and Chafcouloff (1980) agree with this view. They estimate the typical durations of unstressed syllables to be 140 ms and stressed ones 220 ms. Mean durations for 2—4 syllable intervals (87% of the cases) fall in the range 358 to 685 ms. But further research will have to be done to approach a solution to these questions.

Even if the results obtained in experiments using simple stimuli cannot be generalized to speech, most of the techniques used in the experiments can. Obvious examples are the study of perceptual grouping of speech or speech-like stimuli, synchronization of speech to non-speech stimuli, comparing personal tempo to behaviour in tapping tasks and speech behaviour etc. The study of perception of interstress interval durations in Chapter 7 is an example of using a technique previously used only for non-speech stimuli on speech.

Chapter 2

Speech rhythm—an overview of previous studies.

It would be surprising if the general tendencies towards regularity and rhythm in human behaviour were not reflected in language. It is not obvious, however, what exactly rhythm should mean when we talk about speech. Nor is it immediately clear at what level of speech production or perception one should look for the rhythmic units. In this chapter, I will give a brief overview of some of the work that has been done to study speech rhythm, in both production and perception. Some of the questions concerning speech rhythm pose considerable theoretical and methodological problems. In this overview, these problems will be dealt with only briefly but I will return to them in more detail in Chapter 3 which contains a discussion of some theoretical and methodological issues.

2.1 The role of the syllable.

Human intuitions about language have a long tradition of connecting speech rhythm with syllables. The whole theory of meter in poetry, the foundations of which were laid in antiquity, is based on the idea of syllables as the rhythmic building blocks. It is felt that syllables, particularly stressed syllables, are somehow the carriers of the rhythmic beats in speech.

Now, in normal speech there is a continuous flow of sounds that are all part of some syllable. Still we may often perceive the particular rhythmic qualities as a succession of more or less discrete 'beats'. It is as if there were certain 'points' in time at which these

beats occur. "Stress is felt to occur at a certain definite *point* in the syllable; that is to say, it is not felt to have any appreciable duration" (Classe, 1939, p. 17).

This need not imply, however, that there is any particular acoustic correlate to this perception. It could be a perceptual illusion, caused by the organization of the perceptual system, or it could be a 'construction of the mind'. The illusion, or whatever it is, is, however, strong enough for many researchers to have tried experimentally to find possible physical correlates to which the perception can be connected.

The first such study that I know of was made almost a century ago by Miyake (1902). In his study, subjects were told to read syllables while tapping on a telegraph key 'in time' with the reading. It goes without saying that the state of the measurement techniques in those days introduced severe limitations on what could be studied experimentally. Among other things, registrations were made with a kymograph which restricted the use of sounds to those which could be reliably identified on the prints. The method was very time consuming and only short sequences of sound could be registered. The study by Miyake included only monosyllables beginning or ending with the consonants /m/, /p/, and /h/, and the vowel was always /a/. The taps were found to precede the vowel onsets by some 50 to 140 ms depending on context; 143 ms for /pa/ and 52 for /a/, the rest of the values falling in between those values.

Classe (1939), in an extensive study of speech rhythm, also studied the perceptual location of syllables. The experimental set-up he used was almost identical to that used by Miyake. The speech material in Classe's study was lines of poetry. Subjects read the lines while tapping on a telegraph key to stressed syllables. Classe's reason for using poetry is worth noting: *"Verse was selected in preference to prose as allowing a freer feeling of rhythm and thus being less likely to interfere with the hand movements which are more evenly distributed than in similar experiments with prose"* (p. 24). What Classe means, if I interpret him correctly, is that the variable rhythm of ordinary prose might conflict with the tendency to tap regularly. This may seem as a strange limitation to introduce in the experiments. But Classe believed that stress was primarily a psychological phenomenon connected with the speaker. He seems to have believed that the hand movements and the movements of the articulatory organs were reactions to the 'same' inner stimulus and should therefore in principle be synchronized. The irregular rhythm of ordinary prose might make this more difficult but only for purely mechanical reasons that are of little consequence for the phenomenon of stress itself. Whatever one's personal reactions to these ideas are, questions of this nature are certainly something one must consider very carefully whenever motor responses are used.

The general result from Classe's experiments was: *"the stress occurs somewhere in the course of the emission of all the consonants considered, with the exception of /b/ and /h/, in the case of which the stress occurs in the course of the following vowel"* (p. 32). In his study too we may see the influence of the limitations introduced by the experimental

apparatus. Classe explicitly states (p. 25) that the reason why the releases of consonants were used as reference points was that those points were generally easy to identify on the prints.

If the results are compared with those obtained by Miyake they agree in their general tendency. Stress beats occur in the vicinity of the vowel onsets. It is possible, using Classe's data, to state the locations a little more precisely. The complete set of initial consonants is {/t/, /b/, /s/, /th/, /r/, /m/, /j/, /h/}. When the consonant is /b/ or /h/ the average beat location is some distance into the vowel (38 ms from the release in /b/ and 14 ms from the vowel onset in /h/). For the rest of the cases, placement is fairly uniform (13 ms before the plosion in /t/ and 29 ms on average before the vowel onset for the rest). (Values are corrected for a systematic mechanical error of .01 s reported by Classe).

These early experiments were restricted in many ways by what was at all technically possible to do at the time. The reason why finger tapping was the only means used to mark rhythmic beats was the technical limitations, but the same technique has also been used in later experiments. More recent techniques that have been used are placing an acoustical 'marker' so that it perceptually coincides with a given syllable or judging whether a marker placed near a syllable is simultaneous with the syllable or not.

Allen (1972) has shown, rather convincingly, in a series of experiments, using all these methods, that the perceived syllable beats of spoken English are closely connected with the onsets of the vowels in the syllables. Other studies have produced comparable results. Lindblom (1970) and Rapp (1971) have made similar studies using nonsense syllables and Swedish subjects. In these studies, subjects were told to read words to a metronome. The subjects adjusted their readings so that the onsets of the vowels were close to the metronome beats.

There is not complete agreement between different studies about the exact syllable beat locations. There is some general agreement that the onset of the vowel plays a significant role but the exact locations proposed may deviate from these onsets by some amount for different syllables and between studies. The deviations found are, however, rather small. The values reported above in connection with Classe's work may be seen as typical.

There also exists an alternative view of what it is that accounts for the perception of rhythmic beats in speech; the theory of p-centres. The phenomenon was discovered by Marcus (1975) when preparing word lists to be used in a psychological experiment. Marcus wanted the lists to sound as if the words came at perfectly regular intervals in time. He tried different alignment criteria, like word onsets and vowel onsets, but the resulting lists did not seem to display the desired property. Through trial and error he finally managed to construct lists that sounded perfectly regular. But when analysing the lists acoustically, no feature in the signal could be found that recurred regularly. Neither word nor vowel onsets were evenly spaced. Marcus came to the conclusion that some other quality of the word accounted for its perceptual moment of occurrence and decided to call it the

perceptual centre (p-centre) of the word. Later studies replicating Marcus' experiment have confirmed his results in a general way (Morton, Marcus, and Frankish, 1976, Marcus, 1981, Fowler, 1979). Neither word onsets nor vowel onsets seem to work as the points of alignment for perceptually regular word lists.

In a typical experiment (Morton, Marcus, and Frankish, 1976), relative word onset irregularities were measured in perceptually isochronous lists of spoken digits. ('*Isochrony*' means 'equal time'. An isochronous list is a list where words or syllables come at equal intervals. This concept will be discussed more in the following sections.) Judging from their diagram, the range of deviations from vowel onset isochrony was around 70 ms. Marcus (1981) also found p-centre locations which varied within a range of approximately 80 ms. Fowler (1979) reports the greatest deviation to be about 60 ms. These are the maximum deviations for any combination of syllables. On the average deviations are about half of those values. It seems reasonable to suggest 30—40 ms as a representative average value.

An interesting experiment comparing word list isochrony in production and perception has been done by Fowler (1979). In one experiment she had a male speaker read sequences of monosyllables. The syllables were of the type 'Cad', with C $\in$ {#, /b/, /m/, /n/, /t/, /s/}. Two different types of lists were constructed—homogeneous lists consisting of repetitions of the same syllable and lists of syllables alternating between two types. The subject was told to speak *"at a slow rhythmic rate, stressing every syllable"* (p. 377). The results were in agreement with those obtained by Marcus and Morton. For the homogeneous lists, word onsets were nearly isochronous, but for the alternating lists, word onset anisochronies appeared. In a second experiment, Fowler used the 12 most anisochronous utterances from the first experiment. In addition, she used manipulated versions of these same utterances. The manipulations meant that silence was added or deleted at relevant places between the syllables in order to make word onsets isochronous. The original utterances together with the manipulated ones were presented to listeners who were to decide which of two utterances that sounded the most rhythmical. The natural, anisochronous, utterances were chosen significantly more often as the more rhythmically regular ones. The author concludes that *"when asked to produce isochronous sequences, talkers generate precisely the acoustic anisochronies that listeners require in order to hear a sequence as isochronous"* (p. 375).

Results obtained in p-centre experiments have been used as evidence against the view that vowel onsets are the closest correlates of the rhythmic beats in speech. But this does not follow in any obvious way. First of all, there is no proof that normal speech is perceived as perfectly isochronous. Thus, if vowel onsets do not come at perfectly regular intervals that does not mean that they cannot be the relevant carriers of rhythmic beats. Secondly, there is a long way to go to prove that the conditions that are necessary to produce perfectly isochronous lists of isolated words must be the same as those that determine normal

continuous speech production. Theoretical issues in connection with stress beats as well as p-centres will be discussed in more detail in section 3.4.

2.2 Isochrony.

In metric theory, the *foot* occupies a central place. A foot consists of a stressed syllable and a number of unstressed ones. It has become customary to talk about 'feet' in the analysis of temporal properties of all types of speech. In the analysis of speech rhythm in general, a 'foot' usually means a sequence of syllables consisting of a stressed syllable and all following unstressed syllables up to the next stressed syllable. These sequences are also often referred to as *'interstress intervals'* (ISI). In speech rhythm research, speech is often analysed in terms of closed syllables (VC) because of the central role played by the onsets of vowels in the perception of rhythmic beats in speech. In this respect the analysis is somewhat different from what is normally the case in metric theory.

In view of the importance of structure in the analysis of musical rhythm and poetry, one would expect that the study of structure would also play a major role in speech rhythm research. In the earliest writings about speech rhythm, this was also the case. When, in *The art of rhetoric*, Aristotle talks about different speech styles, he uses the same metric concepts to describe them as he uses when talking about poetry:

"Of the different rhythms the heroic is dignified, but lacking the harmony of ordinary conversation; the iambic is the language of the many, wherefore of all metres it is the most used in common speech; ... The trochaic is too much like the cordax; this is clear from the tetrameters, which form a tripping rhythm. There remains the paean, used by rhetoricians from the time of Thrasymachus ..." (pp. 383—385)

But looking back at the development in speech rhythm research during the last century or so, one discovers that the study of structural aspects of this kind has played a very subordinate role. Studies of speech rhythm have instead almost exclusively dealt with strictly temporal aspects of rhythm. In fact, the scope has been even more restricted than that. Most studies have been concerned with one of two central topics; the possible existence of isochrony in certain languages, particularly English, and, in later years, the rhythmic classification of languages into one of the two categories *stress-timed* and *syllable-timed.*

The term 'isochrony' has been used in speech rhythm research to refer to the intuition that stressed syllables come at equal intervals in time irrespective of the number of intervening unstressed syllables; interstress interval durations are equal. This is said to be the case for certain languages called 'stress-timed'. In this view there is also another type of language, called 'syllable-timed', the rhythm of which is characterized by equally spaced syllables; syllable durations are equal. The terms 'stress-timing' and 'syllable-timing' were introduced by Pike (1945).

The idea that stresses in English come at equal intervals in time dates back to the works of the influential English 18th century phonetician Joshua Steele. In his analysis of English prosody he partitions the speech stream into temporal units using a notation which is similar to that used in musical notation. Each 'bar' consists of one stressed syllable and a number of unstressed syllables (although he calls them 'heavy' and 'light'). The structure of an interval is given in musical notation like 2/4 or 3/4. The bars are supposed to occupy the same amount of time in spoken language. They come as regularly as 'the swings of a pendulum'. For some reason this idea seems to have caught on. It recurs in the writings of other linguists following Steele, and is accepted, more or less uncritically, by most of them. It has to be said, though, in all fairness, that Steele and his followers *did* consider the structure of the intervals as responsible for different types of rhythm. But isochrony was taken more or less for granted. One also has to say in their defense that up till somewhere around the turn of this century there were no possibilities to study these things instrumentally. Intuition and perception were the only tools available.

Although the claims of isochrony made by Steele and many following him were based on perception, it is clear that they meant that the regularity they thought they heard was a characteristic of speech production. I have found no indication that they considered the possibility that the impression might have been a perceptual illusion. On the contrary, one gets the impression that they believed in a rather direct correspondence between what they heard and what was produced. I think it is correct to say that the claims of isochrony were primarily meant to say something about speech production. This is also reflected in the activities in the experimental field. When it became possible to study speech rhythm instrumentally, phoneticians started to try to verify the isochrony hypothesis by measurements of the speech signal.

With the arrival of the kymograph such measurements became possible. As I mentioned above, the first studies that have direct relevance for the study of speech rhythm were concerned with perceived syllable locations. The first study that I know of that attempted to measure interstress intervals was Classe (1939), mentioned above. In his study, Classe also addresses the question of isochrony. He points out that the 'ear' is a poor instrument for measuring time objectively and he wants to see if the impression that English speech is isochronous is met by any corresponding regularity in the speech signal. He therefore recorded and measured, using the kymograph, numerous phrases and sentences read by 13 different speakers. Although Classe seems to have believed in isochrony as some underlying principle, it became clear to him in the course of his work, that interstress interval duration is not in general independent of the number of syllables and that isochrony therefore occurs only under very special circumstances.

"... perfect isochronism can only be realized when very definite conditions are fulfilled. These are:

(a) Similarity of phonetic structure of the groups, including number of syllables.

(b) Similarity of grammatical structure of the groups, and similarity of connexion between the groups.

These conditions are comparatively seldom met with in ordinary speech ...." (p. 100)

No more experimental investigations appeared for a period of time after Classe's study. Why this was so I am not sure, but one may guess that with the technical means that were available, studies of this kind were bound to be very tedious and time consuming, something that may have discouraged many from taking up the question.

2.3 Stress-timing and syllable-timing.

In 1945, the American linguist Kenneth Pike published a book, *The Intonation of American English*, which came to have a considerable influence on speech rhythm research. His book is not, strictly speaking, a scientific treatise. His main concern is the teaching of American intonation to non-American (mainly Hispanic) students, and the book is actually a revised version of an earlier text book.

The ideas are put forward in the form of statements, much like those of traditional style normative grammars. They are not presented as hypotheses meant to be subject to empirical tests, or as the result of empirical studies, by himself or by others. Reading his book today, one finds it surprising that his ideas got the attention they did, but for whatever reason, they *did*, and they influenced speech rhythm research to a considerable extent in the following decades. So, in order to understand the present situation, it is necessary to be reasonably familiar with his ideas. I will begin by quoting some passages from his book:

"The units tend to follow one another in such a way that the lapse of time between the beginning of their prominent syllables is somewhat uniform. ... The tendency toward uniform spacing of stresses in material which has uneven numbers of syllables within its rhythm groups can be achieved only by destroying any possibility of even time spacing of syllables. ... the syllables of the longer ones are crushed together, and pronounced more rapidly ... its length (interstress interval, my remark) *is largely dependent upon the presence of one strong stress, rather than upon the specific number of syllables. ... Many non-English languages (Spanish, for instance) tend to use a rhythm which is more closely related to the syllable than the regular stress-timed type of English; in this case, it is the syllables, instead of the stresses, which tend to come at more-or-less evenly recurrent intervals—so that, as a result, phrases with extra syllables take proportionately more time"* (pp. 34—35)

His ideas on the different rhythmic types were not completely original. Lloyd James (1940) had used the terms *'machine-gun rhythm' and 'morse-code rhythm'* five years earlier to describe different speech styles much along the same lines of thought. But it was Pike's terminology that became accepted.

Below, I have tried to reformulate Pike's ideas as statements that may be subjected to empirical testing. To this end, I have strengthened them a little, making them a little more categorical to make it easier to see how they might be tested.

Languages fall into one or the other of two distinct rhythmical categories: languages with a stress-timed rhythm and those with a syllable-timed rhythm.

Stress-timed languages are characterised by the fact that:

1) Interstress interval durations are equal.

and as corollaries to 1)

2) Interstress interval duration is independent of the number of syllables

3) Syllable durations are gradually compressed as the number of syllables increases

Syllable-timed languages are characterised by the fact that:

4) Syllable durations are equal.

and as a corollary to 4)

5) Interstress interval duration is proportional to the number of syllables

Now, these hypotheses are not completely independent. If 1) is true then so are 2) and 3). But the reverse is not necessarily true. If 1) is false, 2) or 3) may still be true. The same type of relation holds between 4) and 5). These remarks should be enough to make it clear that it makes sense to test all the above hypotheses separately as well as in combination. If this seems a little 'technical' for the moment it is my hope that the situation will become clearer as I expand these ideas further in the discussions which will follow.

2.3.1 Studies of stress-timing.

With respect to the first hypothesis, it was seen above that Classe's work had already cast some doubt on its validity. In the following decades, several more studies were made and below I will review a few of those, trying to compare the results with the hypotheses based on Pike's ideas. In the studies reviewed here, and in similar ones, the questions of stress-timing and isochrony are not usually kept apart and in this particular context I will also regard them as more or less synonymous.

Unfortunately, there is no commonly agreed upon framework within which to carry out experimental studies in this area. The results from different studies are, therefore, not always comparable. With respect to the question of whether intervals are equal or not, different authors usually present their results so that they may be compared with the results obtained by others. Regarding the question of how interstress interval duration depends on the number of syllables, this is not always the case. Many authors do not address this question and those who do usually only present mean values of total durations. A

reasonable first approximation to model a systematic dependency would, for example, be a linear model, something which may be tested using linear regression. But most authors do not carry out such an analysis, nor do they always present their data in such a way that an analysis may be done by the reader. Where it has been possible, and for reasons that will be more fully explained in 3.2, I have used linear regression on the data to obtain a measure of how interval duration depends on the number of syllables. With respect to the question of whether there is any temporal adjustment of syllable durations there is even less consistency among reports, making comparisons difficult, and most authors do not address the question at all. The differences in the way authors address the problems will be apparent in the presentation below, but this is only a reflection of the current situation and choosing other studies or adding more would not change this impression in any significant way. The papers are chosen so as to span over a certain period of time.

Shen and Peterson (1962) (referred to in Lehiste, 1977) studied interstress intervals in English prose texts read by three readers. They differentiate between primary and secondary stresses and recognize only one primary stress per sentence. In their measurements, they examine all possible interstress intervals between the two types of stresses and find a considerable range of durations in all comparisons. On the basis of these results, the authors concluded that the isochrony hypothesis must be rejected.

Bolinger (1965) let six speakers read two English sentences. Stresses were identified and the distances between stresses measured. On the basis of these measurements Bolinger concludes that: *"The results give little support to the idea of isochronous rhythm."* (p. 167). He also concludes that the number of syllables in an interstress interval seems to be the determining factor for its duration, but makes no attempt to describe the form of this dependency. (Using linear regression on figures published in a table gives a rate of increase per added syllable of about 100 ms if average values of the pooled results are used.) Bolinger also addresses the question of temporal adjustment but only in a few special cases and not as a function of interval length.

Faure, Hirst, and Chafcouloff (1980) studied recordings of English sentences read by three subjects. Altogether 114 intervals were measured. Interval duration was found to correlate significantly with the number of syllables in the interval and the authors conclude that *" ... it is simply not true that stressed syllables are separated by even "roughly equal" intervals of time"* (p. 73). Interstress interval duration increased by a nearly constant amount as a function of the number of syllables. The authors show that, assuming durations of 220 ms for the stressed syllables and 140 ms for the unstressed syllables, these values predict interstress interval duration almost perfectly. From this they suggest that assuming fixed syllable durations for stressed and unstressed syllables may be a relevant model, but no attempt is made to test this model any further against empirical data.

Nakatani, O'Connor, and Aston (1981) used reiterant speech mimicking English words and interstress intervals with varying number of syllables. They found that duration

increased approximately linearly as a function of the number of syllables. With respect to the question of isochrony they conclude that: *"No evidence to support even a liberal interpretation of isochrony was found in this study"* (p. 103). They present interval durations as a function of the number of syllables in the form of diagrams. There is no numerical data in the paper on which to base a regression analysis of interval durations, but judging from the diagrams the increase in duration per added syllable varies between 95 and 165 for different subjects with a mean of approximately 140 ms. Syllable durations were studied with respect to syllable position. Phrase-final and word-final position had the effect of increasing syllable duration. Syllable duration as a function of position within intervals of different lengths are presented in the form of diagrams. As far as it is possible to tell from the diagrams, there is no temporal adjustment in syllable duration as a function of interval length.

Dauer (1983) collected duration data from recorded readings in five different languages in a comparative study. The languages examined were, English, Thai, Spanish, Greek, and Italian. Interstress interval durations were not equal in any of the languages. Moreover the results showed no significant differences in mean durations or standard deviations between the languages compared. Durations increased as a function of the number of syllables for all of the languages in the study. Interestingly, the increase in interval duration per added syllable was approximately the same for all languages (Dauer reports 110 ms per added syllable as a representative value) although Spanish, which is assumed to be a syllable-timed language, was included. Syllable duration as a function of interval length was not studied.

Strangert (1985) has made a study of Swedish speech rhythm. Interstress intervals with varying number of syllables embedded in a carrier phrase were studied. Interstress intervals were not constant but showed an approximately linear increase of about 100 ms per added syllable. The duration of the stressed syllable as a function of the number of syllables in the interval was studied. Duration showed a sharp drop between monosyllabic intervals and polysyllabic intervals. But for intervals with 2 to 4 syllables there was no consistent effect. Unstressed syllable duration also seems to be independent of interval length. In one of the experiments, only the vowel of the stressed syllable was studied with respect to temporal adjustment. In this case, there seems to be a more gradual dependency on interval length but the effect is small and the greatest difference again seems to be between monosyllabic intervals and polysyllabic ones.

In a study of lines of Icelandic poetry meant to consist of isochronous metric feet, Lehiste (1990) found feet not to be of equal durations. Duration was found to be a function of the number of syllables. The growth per added syllable in metric feet varied between 88 ms and 148 ms for five speakers, with an average of 106 ms (calculated by myself using linear regression on the basis of published data). Lehiste addresses the question of temporal compensation comparing the variance in phrase duration with the sum of the variances in

feet and pauses. Using this method she found no evidence for temporal compensation. No information is given on syllable durations.

The results from the different studies with respect to the different hypotheses regarding isochrony or stress-timing may now be summarized.

1) Interstress intervals are *not* found to be equal. In none of the above cited studies, nor any other study I have come across, have intervals been found to be equal.

2) Interstress interval duration is *not* independent of the number of syllables. Whenever this question is studied, duration is found to increase as a function of the number of syllables. There are seldom any attempts to describe the function, but calculations done by myself, using published data, suggest that a linear model fits data to a very close approximation. I will come back to this question in 3.2 where it will be discussed in more detail.

3) With respect to the question of compression of syllables, or any other compensatory mechanisms, no very definite conclusions may be drawn from these or other studies. It is true that in some studies such effects seem to be present. In other studies, on the other hand they are not. Some tendencies to temporal adjustment were found in Strangert's study. In the other studies above such effects were either not found or not studied. There are other studies where such effects have been found, however (e.g. Fowler, 1977, Fourakis and Monahan, 1988). Some of these will be discussed in detail in 3.3 in connection with theoretical and methodological questions. It may be said at this point, however, that the results have usually been obtained in experiments involving test intervals in a carrier phrase and carefully selected material. (This is the case with Strangert's study too.) The implications for normal speech are therefore not obvious. It should also be said, though, that when an effect is found it means that syllables become shorter as a function of increasing word or interstress interval length. Thus, there seems to be some weak evidence for an effect, at least under certain conditions, but more studies are needed before it is possible to draw any definite conclusions about such processes.

There are obviously a number of methodological questions one has to resolve in connection with studies like the ones above. And it is not always easy to evaluate the results. Some of the problems of this kind will be discussed in detail in Chapter 3. I would, however, like to mention briefly at this point, two questions that have a direct bearing on the results discussed above. The first question is about how interstress intervals are measured and the other concerns the evaluation of the results.

If one wants to see whether the intuitions that stresses seem to appear at regular intervals has any counterpart in the speech signal, it is necessary to define the measurement criteria rather precisely. Stressed syllables have considerable duration. To say that one should measure the distances between stressed syllables is not precise enough.

As I have mentioned above, experimental studies of syllable locations in perception have indicated that the onset of the vowel in the syllable is closely connected with its perceptual moment of occurrence. This is also reflected in most studies. Interstress interval durations are measured as the distances between vowel onsets. To study the question of isochrony in production, however, one is not obliged to use a criterion that has any particular relevance in perception. One could simply measure distances between some regularly recurring feature in the signal, say vowel onsets, consonant onsets, segment centres, or the like. But since rhythm seems to play such a significant role in perception, one would prefer a criterion that has some relevance also in perception. For this reason, the most commonly used criterion is vowel-onset to vowel-onset. This is also the case for some, but not all, of the studies cited above.

About Classe I am not absolutely sure, but he seems to have based his measurements on the results of his syllable beat experiments, which means that syllable boundaries come at, or very near, the vowel onsets. Bolinger uses a radically different segmentation criterion. He measures interstress intervals as the distance between the 'centres' of the stressed syllables. Still another type of segmentation criterion is used by Lehiste. She uses a morphologically based syllable definition meaning that syllables often start with consonants or consonant clusters. It is interesting to note, however, that with respect to the question of whether interstress intervals are equal or not, the results are the same regardless of the segmentation criterion; intervals are *not* equal. And in studies where interstress interval durations have been studied as a function of the number of syllables, duration has been found to be a monotonic increasing function of the number of syllables.

As the reader may recall, there is also the theory of p-centres to consider. But first of all, no one has been able to say how p-centres should be calculated for continuous speech. And secondly, the displacements of p-centre locations relative to word or vowel onsets found in word list experiments are far too small to result in isochrony even if one assumes that p-centres are consistently displaced in the direction of greater isochrony. It is therefore highly unlikely that an analysis using p-centres would change the conclusions arrived at in the above studies in any significant way.

Statements about isochrony are often very vague. A statement by Halliday (1967) may serve as a typical example: *"there is a tendency for salient syllables to occur at roughly regular intervals of time"* (p. 12). Now, statements like this are much too vague to be of any scientific use. The reason is, of course, that they are not possible either to verify or falsify. Whatever the measurements tell us, it is always possible to claim that one finds the durations *roughly equal*, or in extreme cases that there is at least a *tendency* for them to be so.

It is possible, however, to give statements like the above a reasonable interpretation that *may* be verified or falsified. One may say that given the status of the measurement techniques and criteria used at present, intervals are *roughly equal* in the sense that the

differences we find are of the same order as possible experimental errors. Now, given that interpretation, statements about intervals being *roughly equal* when measured in the speech signal are clearly false. As was shown above, experimental studies have found that interstress intervals vary considerably as a function of the number of syllables. A factor of 100—150 ms per added syllable is a common finding, meaning that the range of durations when 1 to 5-syllable intervals are measured is in the order of 500 ms or more. This is clearly far beyond any possible measurement errors. In fact measurement in itself is no problem any more. Modern equipment permits segmentation with errors down to 1 ms or less. The problem instead lies in the selection of segmentation criterion. But as was shown above, all the studies arrived at approximately the same results irrespective of syllable segmentation criterion (vowel to vowel onset, metric-morphological, or 'middle of the syllable'). Those who are familiar with this type of work know that there is also another type of decision problem. Where are the segment boundaries? This may be particularly problematic when neighbouring segments are voiced. But even allowing 'decision errors' of the order of, say 1—3 glottal pulses, the errors will still not amount to more than possibly 50 ms or so in extreme cases which again is well below the range of interval durations found. One may therefore conclude that there is no basis for claiming that the interstress interval durations that have been found by measurements in the speech signal are even *roughly equal.*

A reader with only a slight knowledge of speech rhythm research must have concluded by now that I must have missed something when I say that experimental studies have failed to lend any support to the isochrony hypothesis. Aren't there in fact quite a number of studies meant to show that there are at least tendencies, even strong tendencies, to isochrony in English? The answer is: *Yes, there are* and *No, they don't.* Statements claiming that English is a *more or less* isochronous language are quite common. But first of all, these claims are usually found in text books, presumably following in the pedagogical footsteps of Kenneth Pike, where the claims are merely stated but no evidence is presented. Secondly, where the claims are backed by some type of evidence the quality of the argument is such that the claims cannot really be taken seriously. Below I will make this point clear by citing two representative examples.

Uldall (1971) recorded a reading of the text 'The north wind and the sun'. Stresses were marked in a manuscript while listening to the recording and the durations of the interstress intervals were then measured. Based on the results of these measurements Uldall drew the following conclusions:

"There is a tendency to isochronism in this speaker, in his style of speaking. ... it will be seen that rather more than half, 57%, of the filled feet fall into a group from 38.5 cs. to 52 cs. It can be seen that the averages increase from 1- to 2- to 3-syllable feet, but not very strikingly. The 4-syllable feet, on the other hand, are much longer than the others. ... We

can conclude that there is a strong tendency to isochronism, in this speaker's reading style, but that 4-syllable feet do not follow this tendency." (p. 208)

One wonders of course by what standard of comparison Uldall is able to determine that the difference between 385 ms and 520 ms is 'small' (and why are 4-syllable feet, 720 ms on the average, excluded when the question of isochrony is decided). All one has to say to upset the whole 'proof' is that one finds the increase in foot duration as a function of the number of syllables 'very striking'. From that follows, using Uldallian logic, that there is a strong tendency in this speaker *not* to speak isochronously. But more importantly, perhaps, looking at the published figures it is immediately clear that interstress intervals increase in a very systematic way as a function of the number of syllables. The increase is, in fact, linear to a very close approximation. A regression analysis on 1- to 4-syllable intervals results in a regression coefficient of .94 (based on non pre-pausal intervals) and a 'slope' of 109 ms per syllable. From this may be concluded firstly that 4-syllable feet do *not* seem to be so out of place as Uldall claims and, secondly that intervals increase in duration by approximately 110 ms per added syllable, perfectly in line with results from other studies. So much for Uldall's 'strong tendency'.

The second example is from one of the better known figures in linguistics, David Abercrombie (1965).

"Let us consider the utterance: 'This is the house that Jack built'. It contained, as I spoke it then, four stress-pulses, and they occurred on the syllables 'this', 'house', 'Jack' and 'built'. Now if I say the sentence again, and while I do so tap with a pencil on the table every time there is a stress-pulse, the taps will be unmistakably isochronous, showing that the stressed syllables are too." (p. 18)

Now, for this method to show anything at all with respect to isochrony one would first of all have to show that the taps correspond to the stresses in a representative way, which is far from obvious. Secondly, one would have to measure the time intervals between taps to show that they are indeed *unmistakably isochronous*. Abercrombie does none of this, but merely makes the statement as if it were perfectly obvious and uncontroversial.

There is a whole 'school' of linguists and phoneticians who talk about speech rhythm in this casual way, presenting claims about isochrony. (For more examples see Catford, 1977, Halliday, 1967, 1970, O'Connor, 1973, Ladefoged, 1975) In my opinion these 'studies' do not deserve a place in a serious discussion of speech rhythm.

As the discussion above demonstrates, attempts to find any evidence for isochrony in the speech signal have failed. Now, this does not preclude the existence of isochronous events at some other level of speech production, say for instance the neuro-muscular level. Some attempts have been made to explore this possibility. The first attempt (to my knowledge) to look for isochronous events at the neuro-muscular level is Stetson's (1951, 1st ed. 1928) theory of chest pulses, often referred to, particularly in earlier discussions of isochrony.

He proposed that isochrony in speech has its origin in the activity of the intercostal muscles. He claimed to have found in experiments that regular contractions of the muscles accompanied the production of syllables. The contractions that could be connected with stressed syllables were particularly strong. He called these contractions chest pulses. These ideas became very popular among believers in isochrony and were widely used and cited in their writings as an explanation of isochrony. Now, even if Stetson himself believed in isochrony this is not an indispensable part of his theory. If there is no isochrony it may still be the case that stressed syllables are caused by chest pulses. Later investigations have shown, however, that this does not seem to be the case. Ladefoged (1967) in a detailed study of neuro-muscular activity and subglottal pressure during speech could find no, or very weak, support for Stetson's claims. *"It is quite clear that there is no simple correlation between intercostal activity and syllables. ... there is certainly insufficient basis for a chest pulse theory of the syllable in normal speech."* (pp. 20 and 47) And Adams (1979), taking a new, and even more thorough look at possible correlations between chest pulses and stressed syllables concludes: *"A physiological definition of stress based upon internal intercostal muscular activity is untenable."* (p. 192)

Few other attempts have been made to study speech rhythm from this angle but there exists an interesting study of neuro-muscular activity in connection with research on p-centres which may have some relevance in this context.

Studies of p-centres have so far failed to connect them with any acoustical correlates in the speech signal. That has lead some researchers to look for correlates to p-centres in another domain. In an attempt to find isochronous articulatory correlates in the neuro-muscular domain, Tuller and Fowler (1980) combined acoustic and electromyographic studies of subjects reading word lists *"as if speaking in time to a metronome"* (p. 279). The words in the lists were nonsense-syllables chosen so that they would involve activity of the orbicularis oris muscle. Simultaneous recordings were made of the EMG activity and the acoustic signal. Anisochronies in the acoustic signal were comparable to those obtained in other similar p-centre experiments.

Interestingly, onsets in EMG activity were significantly more regular than the corresponding acoustic onsets. In the words of the authors, the results *"suggest that, when asked to produce isochronous monosyllabic utterances, talkers comply by producing isochronous articulatory gestures"* (p. 281).

In the context of speech rhythm research, it must be noted, however, that these results may have very little to tell us about rhythm and isochrony in continuous speech. The task of reading a list of isolated words consciously aiming for isochrony is a very particular type of speech activity which may have little, or nothing, to do with normal speech. The authors are aware of this. They point out that the results are to be interpreted in the context of p-centre studies. In normal utterances involving both stressed and unstressed syllables departures from isochrony are much larger. *"The conclusions of the present study cannot*

be generalized to these departures without investigation of utterances that include unstressed syllables" (p. 282).

2.3.2 Studies of syllable-timing.

Pike's claim that there is also another rhythmic type, syllable-timed rhythm, characteristic of Spanish and 'many other' languages, has also been studied empirically. This type of rhythm was to be characterized by the equality of syllable durations instead of interstress interval durations. How this hypothesis stands up to empirical testing will be discussed in the following.

In connection with such studies it may be said, first of all, that speakers of the languages called 'syllable-timed' have often felt uncomfortable about the description, resulting in papers with titles like *'Is Spanish really syllable-timed?'* (Pointon, 1980) and *'Is French really syllable-timed?'* (Wenk and Wioland, 1982). It is a fact that most linguists who advocate the classification are native speakers of English, and both Pointon and Wenk and Wioland explicitly suggest that the native language of the investigators might have influenced the classifications. Be that as it may, empirical studies of languages claimed to be syllable-timed have been as unsuccessful in finding any solid evidence for the validity of the claim as was shown to be the case with claims of isochrony in the languages discussed above. In the following, I will review a few studies which I believe to be representative. The studies below are only concerned with Spanish and French. This seems appropriate, however, since Spanish and French are among the most often cited examples of syllable-timed languages.

There are several factors that may condition syllable duration; stress, position, structure, and type (open/closed) being the most important ones. To falsify the hypothesis that all syllables have the same duration it is enough to show that some syllables differ with respect to durations. But studies of languages meant to be syllable-timed are usually much more detailed, comparing syllable durations under all, or most of the above mentioned conditions. In the examination of the syllable-timing hypothesis made here, the different sources of variation will not be considered in any detail, only the basic question of whether the hypothesis is a tenable one or not. The most important contrast is usually that between stressed and unstressed syllables. The review below will, therefore, mainly be concerned with this contrast.

A paper by Delattre (1966) comparing syllable durations in four languages, among them Spanish and English, will be reviewed in the following section where some comparative studies are reported. To facilitate an evaluation of the results reported here, it may help to know that Delattre found stressed/unstressed ratios of 1.30:1 for Spanish and 1.60:1 for English.

Pointon (1980) has not made an experimental study of his own but has reviewed and compared the results from a number of older studies. The studies that Pointon has examined

are those of Navarro Tomás (1916, 1917, 1918, 1922), Gili y Gaya (1940), Delattre (1966), and Olsen (1972). The studies are primarily concerned with syllable durations, which are studied as a function of position, syllable structure, and stress. Considerable variation in syllable duration is found in all studies, the greatest range of durations (2.3:1) being reported by Olsen. The ratios of stressed vs. unstressed syllable durations are in the region of 1.3:1 in all studies. On the basis of his review, Pointon concludes: *"The figures clearly show ... that a classical syllable-timed rhythm is not being used"* (p. 300).

Manrique and Signorini (1983) have made a very thorough study of durations at different levels of Argentine Spanish. Their data base was recorded sentences; 120 sentences from one male speaker (S1) and an additional set of sentences read by three other male speakers whose results were pooled together (S3). Durations of segments, syllables, and interstress intervals were studied. The variation in duration was studied with respect to such factors as structure, type, and position. The most relevant of their results in this particular context are the results concerning syllable and interstress durations. Syllable durations were affected by several factors. The ratio of the average durations for stressed and unstressed syllables in non pre-pausal positions was 1.58:1 for S1 and 1.32:1 (average) for S3. Other factors that had an influence on syllable durations were position within the phrase and word type. With respect to the question of syllable-timing the authors conclude: *"These data do not give support to the claims that Spanish is 'isosyllabic'"* (p. 126). I have used their data to compute average durations of interstress intervals as a function of the number of syllables and also regression equations based on these averages (see also figure 1.1). These computations reveal that interstress interval durations increase monotonically as a function of the number of syllables. The 'growth rate' calculated as the slope of the regression equation (all data pooled) is approximately 94 ms per added syllable. Moreover, if the average durations of interstress intervals are used, the regression is almost perfectly linear ($r = .992$). The rate is thus nearly the same as that of the stress-timed languages discussed above (100 ms to 140 ms), perhaps even a bit slower. (cf. the discussion in 3.2) A regression analysis also shows that interstress interval duration is *not* proportional to the number of syllables, in contradiction with another claim about syllable-timed languages. (The constant term is around 100 ms.) The data thus lend no support to the claim that Spanish is syllable-timed. The authors offer an interesting suggestion, however, why Spanish may be heard as syllable-timed by many: *"The high occurrence of the syllable type CV and the different manner of reduction of unstressed vowels make the syllables perceptually more clear cut than in other languages"* (p. 127).

Wenk and Wioland, (1982) have studied speech rhythm in French, which is another of those languages claimed to be syllable-timed. They studied French speech rhythm from a number of aspects, physical as well as perceptual, after which they had to conclude: *"Clearly, French is not 'syllable-timed'. Far from being turned out with machine gun equality, French syllables are produced and perceived in rhythmic groups, just as those of English ..."* (p. 214). They also suggest that French may not fit into any of the two

categories with its different realization of accented syllables, making stressed syllables appear to end stress groups as opposed to English where stresses are normally perceived as beginning stress groups. They suggest that 'leader-timed' and 'trailer-timed' may be better descriptions of the two rhythmic types.

From the studies of syllable-timing reviewed above, it seems clear that, as far as temporal factors are reflected in the speech signal in the two languages most often claimed to be syllable-timed, there is as little empirical basis for the construct of syllable-timing as there is for that of stress-timing. The studies reviewed above are but a small sample of all the studies done. But as far as I know, no single study has been able to demonstrate that syllables are of equal duration in any language.

Japanese is often claimed to belong to a third rhythmic category characterized by *mora-timing*. But since *morae* are syllables, although defined in a way which is a bit different from the way syllables are defined in languages like English, mora-timing must be regarded as a special case of syllable-timing. Most of the discussion above therefore also applies to mora-timing, including the division between those who *claim* that Japanese is mora-timed since each mora *"takes about the same length of time"* and those who study the question and find that *"Japanese does not seem to follow the prediction of a strict version of the mora-timing hypothesis"* (Hoequist, 1983a, p.26). The phonetic reality of the mora has also been questioned (Beckman, 1982).

2.3.3 Comparative studies.

Pike mentioned English and Spanish as examples of a stress-timed and a syllable-timed language. References to other languages belonging to one of these groups or the other are often seen, particularly in text books. Among those claimed to be stress-timed are, Russian (Abercrombie, 1967; O'Connor, 1973), Arabic (Abercrombie, 1967), and Thai (Luangthongkum, 1977, cited in Dauer, 1983) whereas French (Abercrombie, 1967; Corder, 1973; Ladefoged, 1975), Yoruba (Abercrombie, 1967; Corder, 1973) and Telugu (Abercrombie, 1967) have been considered as syllable-timed.

Now, as was shown above, there seems to be little basis for maintaining a classification based on claims originally put forward by Pike. It may still be the case, though, that interesting inter-language differences with respect to speech rhythm exist in both production and perception. One may, for example, hypothesize that although interstress intervals are not equal in English, they are more similar in English than in Spanish. If such differences can be reliably established, they may still form the basis of a classification. In view of the fact that the original ideas actually meant to say something about how different languages compare with respect to rhythm, one would expect to find a large of number of such comparative studies. But this is not the case. There are very few such studies.

The two studies of interstress interval durations by Roach (1982) and Dauer (1983) discussed below are in fact the only ones of that kind that I have come across. Speech

rhythm perception in a comparative perspective has been studied by Miller (1984); again, the only one of its kind that I have found. Some comparative results may also be found in a study by Strangert (1985), although the aim of that study is not to test the validity of the classification. There are also a few studies of syllable durations. The most extensive seems to be that by Delattre (1966). All these studies will be discussed in the following. I make no claim that this list is exhaustive, more studies probably exist, but there is certainly a striking contrast between the number of single-language studies one may find and the number of comparative ones.

The relative scarcity of comparative studies is somewhat surprising since such studies often provide the best clues in the analysis of a problem. There are, however, quite a number of serious methodological difficulties one has to find a reasonable solution to if one wants to make a comparative study. Obvious problems are: finding comparative material, finding representative speakers, agreeing on a number of analysis criteria such as the marking of stresses, segmentation criteria etc. It is probably also necessary for several linguists to cooperate in such an undertaking, since native knowledge of a language is often needed in the analysis of data. However, if one seriously wants to study speech rhythm in a comparative perspective there is no way around these problems. In view of all these difficulties, however, one might guess that many linguists have hesitated to embark on such studies. That may be the explanation why so few have been done.

An attempt to test various claims about stress-timed and syllable-timed languages in a comparative perspective has been made by Roach (1982). For his study, Roach chose three languages that have been claimed to be stress-timed (English, Russian and Arabic) and three that have been claimed to be syllable-timed (French, Telugu and Yoruba). Recordings of about two minutes of casual speech by six speakers (one from each language) were analysed. Three claims about the difference between stress-timed and syllable-timed languages put forward by Abercrombie (1967) were tested: (i) There is considerable variation in syllable length in a language spoken with stress-timed rhythm whereas in a language spoken with a syllable-timed rhythm, the syllables tend to be equal in length (ii) In syllable-timed languages, stress-pulses are unevenly spaced (iii) In syllable-timed languages, interstress intervals will tend to be longer in proportion to the number of syllables they contain, whereas such a tendency should be absent (or weaker) in stress-timed languages.

The first claim was tested by comparing the standard deviations of syllable durations in the six languages. No significant differences were found between the two groups of languages. To test the second claim, relative deviations from a hypothesized perfectly regular beat were calculated for each tone group in each language. The actual durations were measured and compared with the hypothesized ones and the differences in per cent were calculated. As a measure of regularity for each language, the variance of the differences between actual and hypothesized durations was used. Again, the claim found no support. In fact, the variation was significantly greater for English than for any of the

other languages, in complete contradiction with the claim. Finally the deviations from regularity were correlated with the number of syllables for each language to test for a dependency of regularity on the number of syllables. Again, the prediction of the claim was not met. The author concludes: *"The results reported here give no support to the idea that one could assign a language to one of the two categories on the basis of measurement of time intervals in speech. Consequently one is obliged to conclude that the basis for the distinction is auditory and subjective"* (p. 78). He suggests that such factors as syllable structure and vowel reduction may be at least partly responsible for the perceptual differences.

The study by Dauer is similar to that by Roach. By comparing interstress interval durations in different languages, she wanted to see if there were any significant differences in this respect between languages considered as stress-timed and those considered as syllable-timed. The languages investigated were English and Thai, classified as stress-timed, Spanish, classified as syllable-timed, and Greek and Italian, both unclassified. Eleven native speakers, 2 English, 1 Thai, 3 Spanish, 3 Greek, and 2 Italian, were recorded reading a prose text. A native speaker and a phonetician listened independently to the recordings and marked the stressed syllables in a script. Agreement is reported to be good. Interstress interval durations were then measured from mingograph recordings. Based on these measurements average interstress interval durations were calculated. It turned out that mean durations as well as standard deviations were almost identical for all five languages. Dauer draws the following conclusion: *"Consequently, we can conclude that the difference between English, a stress-timed language, and Spanish, a syllable-timed language, has nothing to do with the durations of interstress intervals. Furthermore, stresses recur no more regularly in English than they do in any other language with clearly definable stress. Rather, what these data reflect appears to be universal properties of temporal organization in language."* (p. 54).

Like Roach, Dauer comes to the conclusion that the explanation for the impression of different rhythms may lie in differences in language structure. She also questions the use of the word 'timing' in this context if, as the results seem to indicate, the differences are not in the timing of interstress intervals but due to other factors. She sums up by saying: *"If rhythmic grouping takes place in all languages, then the differences summed up by the terms 'stress-timed' and 'syllable-timed' refer to what goes on within rhythmic groups"* (p. 60).

The results from the studies by Roach and Dauer may be summarized as follows: No significant differences were found between the two groups regarding 1) mean interstress interval durations 2) standard deviations of interstress interval durations, or 3) regularity as measured by the deviation from an idealized regular rhythm. One must conclude that neither of the studies lends any support for the relevance of a classification based on interstress interval durations.

Strangert's results are not immediately comparable with those by Dauer and Roach. The comparative part of her study is not concerned with the question of whether the stress-timing vs. syllable-timing dichotomy is tenable, but only with putting some of her findings in the Swedish study in a comparative perspective. The emphasis is on compression and the influence of syllable structure, but these questions will not be discussed here. I will instead mention the results concerning interstress interval durations. Since the results are presented in the form of diagrams, no exact figures may be given. The following conclusions are based on calculations using the diagrams and indirect evidence from other types of tables than interstress interval durations. The languages compared are Swedish, Finnish, and Spanish. Interstress intervals increase as a function of the number of syllables for all languages in what seems to be an approximately linear fashion. The increase per syllable for Finnish seems to be somewhat greater than for Swedish and Spanish (approximately 135 ms vs. 100—110 ms). The only marked deviation from a strictly linear increase in interstress interval duration as a function of the number of syllables is in the behaviour of stressed monosyllables in Swedish which are especially long (350—400 ms vs. 250 ms for Spanish). Bolinger (1965) made a similar observation in his study of English, and suggested that the particular status of the monosyllable may play an important role for the perception of English speech rhythm. Strangert's results suggest that the same may be the case for Swedish. Syllable durations were also compared. The stressed/unstressed contrast was found to be greatest for Swedish (1.82:1) and lowest for Spanish (1.20:1, only CV-syllables). It should be pointed out, however, that the results are based on readings by only one speaker for each language and that differences may be caused by language specific properties as well as inter-individual variation. Another experimental condition that may have had some influence on the results is the use of test words in a carrier phrase. For reasons which will be discussed in 3.3 this method introduces some problems connected with the generalizability of the results to other types of speech.

There is also an often cited comparative study of syllable durations by Delattre (1966). The languages compared in the study are English, German, Spanish, and French. Syllable durations are studied as a function of stress, position (within the phrase) and type (open/closed). All conditions have an effect on average durations and all the effects work in the same direction in all languages. Stressed syllables are longer than unstressed syllables, final syllables are longer than non-final syllables and closed syllables are longer than open ones. If the contrast stressed/unstressed is considered, the greatest ratio is for French (!) with 1.78:1 and the lowest for Spanish 1.30:1 with the two other languages falling in between (English, 1.60:1; German, 1.44:1). It may be seen in his data that the origin of the difference in the stressed/unstressed contrast is in both the stressed and unstressed syllables. In Spanish the stressed syllables are somewhat shorter than in English (257 ms vs. 298 ms) and the unstressed ones somewhat longer (198 ms vs. 186 ms). German syllable durations fall in between for both types. French has the second longest stressed syllables (293 ms), almost the same as in English, and by far the shortest unstressed ones

(165 ms). Thus, the contrast between stressed and unstressed syllable duration is most marked in French, which is particularly noteworthy since French has been claimed to be a syllable-timed language.

The effect of phrase-final position follows approximately the same pattern. It is most marked in French and English (1.78:1 and 1.53:1 respectively) and least marked in Spanish (1.17:1).

The effect of syllable type (open/closed) is in the order of 1.30:1 for all languages.

Inter-language differences are not tested for significance nor are any figures given for variances. It is, therefore, difficult to appreciate how much weight one should attach to individual differences.

The differences between languages may seem rather small (perhaps not even significant in some cases) if the factors that condition duration are considered one at a time. But the effects of different factors may combine. This means that when one considers the accumulated effect of stress, position, and type, the differences may be on quite another scale, 3.39:1 for English against 1.77:1 for Spanish. One may, thus, find contrasts in English which are almost twice those of Spanish. It is conceivable that, although average values for different types of syllable contrasts do not seem to reveal any remarkable differences between languages, the potential for much greater contrasts between individual syllables in languages like English as compared to Spanish may play an important role in the perception of their respective rhythmical characters.

The studies I have discussed so far have been studies of temporal properties as reflected in the speech signal. It was found that, with respect to studies of interstress interval durations, there was little support for the claims concerning stress-timing and syllable-timing. Delattre's and Strangert's results indicate some differences at the syllable level. When syllable durations are compared, there seem to be differences between languages with respect to absolute durations as well as the durational contrast between stressed and unstressed syllables. It is, however, difficult to see how languages could be placed in different categories using these results. Languages seem to place themselves on a continuum of contrast values rather than falling into any distinct categories.

This does not, however, exclude the possibility that there may be some truth in the claims if they are considered as claims about perception only. One might ask, for example, if the languages of the world can be reliably classified as stress-timed or syllable-timed on perceptual grounds, for example by the use of listening panels. Considering the important role of perception in any discussion of speech rhythm one would expect that there would be quite a number of such studies. But there are not. I know of only one (Miller, 1984), and even if there may be other studies that I have not come across there are certainly not many studies of this kind.

In Miller's study seven languages were studied—Arabic, meant to be stress-timed, Spanish, Japanese, Indonesian, and Yoruba, considered by many as examples of syllable-timed languages, and Polish and Finnish, which, to my knowledge, have not been put forward as belonging to either of these categories. Recorded samples from the languages in two speech styles, reading aloud and casual speech, were used. The subjects in the tests were asked to judge whether the samples sounded stress-timed or syllable-timed and also indicate the strength of the tendency to one category or the other. Four groups of subjects took part in the study—English and French phoneticians (EP, FP in Table 2.1), and English and French non-phoneticians (EN, FN). The results can be summarized as follows: Arabic was classified as strongly stress-timed by all groups. Spanish was classified as strongly stress-timed by the English phoneticians and French non-phoneticians and as showing a tendency towards stress-timing by French phoneticians. Indonesian was classified as strongly syllable-timed by the phoneticians. Polish was classified as strongly stress-timed by English phoneticians and strongly syllable-timed by English non-phoneticians. Japanese was felt to be syllable-timed by English non-phoneticians. Yoruba was identified as syllable-timed by the phoneticians. All other judgements meant no clear tendency in either direction. Two things should be noted in particular: in only half of the cases are there any significant tendencies in any particular direction, and even when there are, they may go in opposite directions as the case of Polish shows. There is a reasonably strong (more than half of the groups being significantly in favour of one category or the other), and unambiguous, tendency for only two of the languages—Arabic and Spanish. For Arabic this tendency goes in the predicted direction. But for Spanish, traditionally considered as syllable-timed, two groups out of four perceived it as strongly stress-timed and a third one as having a tendency to stress-timing. An overview of the general results is found in Table 2.1.

The reservations one might have against an investigation of this kind lie in the very nature of the task. How can one be sure that the subjects are really listening for what one assumes

Table 2.1. An overview of the results (by groups) on the perception of stress-timing and syllable-timing found by Miller (1984). 'x' indicates a strong tendency and '?' a weak tendency.

	Stress-timed				Syllable-timed				Undecided			
	EP	FP	EN	FN	EP	FP	EN	FN	EP	FP	EN	FN
Arabic	x	x	x	x								
Polish	x						x			x		x
Finnish									x	x	x	x
Spanish	x	?		x							x	
Japanese							x		x	x		x
Indonesian					x	x					x	x
Yoruba					x	?					x	x

that they are listening for? This is particularly problematic with naive subjects. How does one describe what is meant by the two rhythm categories without giving away information that may bias the results? With phoneticians the problem is somewhat the opposite. Is it really possible for them to disregard whatever knowledge they may have about which category a particular language *should* belong to? (The case for Spanish indicates, however, that this may be possible) But even taking these reservations into account the results strongly suggest that the perceptual classification too may rest on very shaky ground.

It should, however, be pointed out that the study of speech rhythm perception need not be limited to the traditional framework of syllable-timing and stress-timing. Interesting research on the perception of rhythm, that may prove relevant for the study of speech rhythm, has, for example, been carried out in music psychology. In one such experiment, Gabrielsson (1973a) asked subjects to rate different musical stimuli using adjectives like 'simple', 'varied', 'wild', 'pulsating', 'aggressive' etc. to characterize different rhythmical impressions. The adjectives were chosen from a data base of adjectives suggested by professional musicians and musicology students as suitable for the description of musical rhythm. Based on factor analysis of these ratings Gabrielsson suggested a set of dimensions to characterize musical rhythm. The dimensions were grouped into three main groups expressing 1) structural properties, 2) character of movement and 3) emotional aspects. In another series of experiments, Gabrielsson (1973b, 1973c) presented subjects with pairs of rhythmic stimuli and asked them to rate the similarity between the two. The dimensions found in the different experiments were 1) meter, 2) rapidity, 3) uniformity—variation, 4) forward movement, 5) accent on the first beat, and 6) duration pattern of the sound events. Most of these aspects seem highly relevant also for the characterization of speech rhythm. It is conceivable that these and other similar techniques could be applied to the study of speech rhythm and that they may reveal new and interesting aspects of the differences and similarities in the rhythms of languages.

Chapter 3

Speech rhythm—some theoretical and methodological issues.

In this chapter, I will discuss in further detail some theoretical and methodological questions that arise in connection with the study of speech rhythm. As was shown above, attempts to find evidence for isochrony in the speech signal in any language have failed. Although it cannot be ruled out that there may be some correlate that no one has thought of as yet that will turn out to be isochronous this seems highly unlikely in view of the numerous attempts to find such correlates that have been made. To some extent, the same may be said about attempts to find isochrony at the neuro-muscular level, although this possibility is far less studied. This being the case, various modifications of the theory have been suggested. What those modifications mean is usually some relativization of the concept allowing for degrees of isochrony, e.g. that no language is perfectly isochronous, but there are nevertheless differences between languages meaning that interstress intervals are more equal in, say English than in Spanish. Now, it was shown above that, if mean durations are considered, this does not seem to be the case. Studies by Dauer (1983) and Roach (1982) failed to show any such differences. But there may be other ways of modelling the data which may result in interesting differences. Two possibilities will be considered. The first one is a model that I have called 'relative durations' which has been used by Hill, Jassem, and Witten (1979) among others. Their views will be considered in 3.1. In 3.2 I will reanalyse the data presented in Dauer (1983) in an alternative way which seems to place languages in two separate categories.

A consequence of the original idea that interstress intervals should be isochronous, is that this would require gradual compression of syllables as the number of syllables in the interval increases. The question of syllable compression as a function the number of syllables has been examined in some studies, but usually in the context of test words in a carrier phrase. This type of situation may not be altogether representative for what goes on in normal speech production. I will discuss this question in 3.3. But I will begin 3.3 by discussing another theoretical question concerning syllable compression which to my knowledge has not been examined before and which has implications for the study of compression. It is obvious that the possibilities of compression must be closely linked to the total durations of interstress intervals. If these are constant then compression is of course necessary. But now that we know that they are not, we must ask what the precise correlation is between interval duration and possible internal compositions of these intervals. Is it possible, for example, that there may be compression in syllables if interval durations grow with a constant amount, say 120 ms, per added syllable? Nakatani, O'Connor, and Aston (1981) say in the evaluation of their results that *"Isochrony predicts that big words are spoken more rapidly than little words. ... Thus ... the interstress interval increases with word size when words are concatenated. So again, a negatively accelerated relationship ... is needed to support isochrony"* (p. 103). ('Big words' and 'little words' differ in the number of syllables using their terminology) What they mean, if I interpret them correctly, is that if there is any compression in the syllables this must necessarily show up as a negatively accelerated increase in interstress interval duration. That is, intervals will grow as a function of the number of syllables at a rate which is slower than a linear increase. This seems reasonable. But is it necessarily true?

Sections 3.4 and 3.5 will be concerned with methodological questions in connection with speech rhythm perception.

The theory of stress beats will be evaluated against that of p-centres in 3.4.

It has been suggested that *"isochrony ... is primarily a perceptual phenomenon"* (Lehiste, 1977, p. 253). The implication is that although interstress intervals are not equal they are nevertheless perceived as equal. Now, this question has not been sufficiently examined. One may see statements claiming that interstress intervals are perceived as more or less equal, but whether this is actually the case, and if so under what circumstances this illusion arises has not been studied empirically to any extent. And the results obtained by Miller (1984), discussed in 2.3.3, indicate that questions of this sort may not have any simple answers. If for the sake of argument, however, we assume that intervals are actually perceived as equal, there are several ways in which this can happen. One obvious such possibility is that the differences are too small to be detected in perception. One may not automatically assume that if intervals are unequal, say 450 ms and 600 ms respectively, they are also perceived as unequal. Our duration perception may not be accurate enough an instrument to determine whether they are or not, in which case we may assume that they are equal. This solution has been suggested by Lehiste (1977): *"if you cannot tell them*

apart, they must be alike" (p. 257). She bases her suggestion on a study of duration perception (Lehiste, 1973) in which she found subjects' duration discrimination of inter-stress intervals to be rather poor (see 5.1.6). This possibility will be also explored in detail in my own study of perception in Chapters 6 and 7, and in Chapter 5, several theoretical aspects of duration perception will be discussed. In 3.5, a different view of the perception of interstress interval durations will be examined, however. It has been claimed that the reason why people have suggested that some languages are isochronous is that we hear speech in a different way from non-speech auditory stimuli. There seems to take place some perceptual regularisation which makes us 'perceive speech as more regular than it really is'. This is sometimes referred to as 'perceptual isochrony'. There are studies in which empirical evidence for such a regularisation tendency is claimed to have been found. But there are also studies which question this view. Some of these studies will be reviewed and discussed.

In 3.6, I will discuss the use of nonsense syllables in comparative studies of speech rhythm and finally, in 3.7, the possibility of constructing an objective measure of regularity by which the regularity of intervals may be evaluated and compared.

3.1 Relative durations.

Some authors have attempted to describe degrees of isochrony by the use of relative durations rather than absolute ones. The technique means essentially that one uses the duration of the stressed monosyllable as the 'yard stick' by which polysyllabic interstress intervals are measured. This way of representing interval duration is often given in a graphical form displaying regressions lines of relative durations as a function of the number of syllables. Different relative durations will appear as lines with different slopes. These differences are meant to tell us something about the relative isochrony of a particular language, speech style, or rhythmic unit. Strangert (1985) has used this method, among others, to describe her Swedish data and Hill, Jassem, and Witten (1979) and Jassem, Hill, and Witten (1984) have used it for their English data.

There may not be anything technically wrong with this method, but it does not add any new information either. On the contrary, there is a loss of important information. An example will clarify this point. Suppose that we have a language in which durations for 1- to 4-syllable intervals are 300, 375, 450, and 525 ms respectively. Represented as relative durations the corresponding figures will be 1.00, 1.25, 1.50, and, 1.75. It should now be obvious that if we would like to recover the original figures from the relative durations alone this is no longer possible. There are virtually infinitely many sets of duration values that would have the same relative duration figures. Let us examine two of these; L1 = {200, 250, 300, 350} and L2 = {400, 500, 600, 700}. It is obvious that if the sets L1 and L2 are thought of as representing interstress interval durations in two languages as a function of the number of syllables they are drastically different in a number of ways. Stressed

monosyllables are twice as long in L2. Interstress interval durations grow twice as fast as a function of the number of syllables in L2 and the range of durations is also twice as large. These facts are totally obscured by using relative durations. The two languages will instead be classified as having the same degree of isochrony. Whether this is the relevant description is far from obvious.

A more serious objection, perhaps, is that the description of interstress intervals depends entirely on the duration of the stressed monosyllable. Whether the stressed monosyllables should be given this status or not is of course something that one may debate but in the case of relative durations, that debate is already over with. The assumption is integrated in the measure itself.

In the following section, I will discuss a different approach to the problem of describing differences in the rhythmic structure of languages, as reflected by the interstress interval durations.

3.2 Interstress interval duration as a function of the number of syllables.

The isochrony hypothesis, in its original formulation, meant that interstress interval duration is independent of the number of syllables. Now, it is perfectly clear that this is not a correct assumption. Interstress interval durations *do* depend on the number of syllables in both 'stress-timed' and 'syllable-timed' languages. This does not mean, however, that they necessarily depend on the number of syllables in the same way. There may be interesting differences in the way they depend on the number of syllables. In the following discussion, I will use data from the study by Dauer (1983) mentioned in 2.3.1 to explore this possibility.

In Dauer's analysis of the temporal properties of interstress intervals, mean durations and standard deviations were compared and no significant differences were found between the languages. But the way these durations depend on the number of syllables may also be a relevant issue. Dauer is clearly aware of this: *"the increase in average duration of an interstress interval due to the addition of another syllable was similar in all speech samples (about 11 cs): note the close correspondence between intervals in English ... containing one to four syllables and intervals in Spanish ... and Greek ... containing two to five syllables."* (p. 54)

What I am going to say below is not anything dramatically new compared to Dauer's observations. All the relevant information is contained in her statement, but to some extent only implicitly so. The analysis I will give below offers, in my view, a clearer and more explicit presentation of the same basic facts. And also, although the dependencies of interval durations on the number of syllables are indeed similar, there is nevertheless also an interesting difference which becomes more apparent the way I present the data.

Table 3.1. Mean durations of interstress intervals (in ms) as a function of the number of syllables for the five different languages in Dauer's study.

No. of syllables	1	2	3	4
	Duration (ms)			
English	295	410	520	598
Thai	300	420	550	580
Spanish	183	328	440	543
Greek	213	313	416	525
Italian	215	320	425	530

Table 3.2. Linear regression equations based on the durations in Table 3.1. 'I' is the interval duration and 'N' is the number of syllables.

English	I = 201 + N*102	r = .996
Thai	I = 220 + N* 97	r = .973
Spanish	I = 76 + N*119	r = .997
Greek	I = 107 + N*104	r = 1.000
Italian	I = 110 + N*105	r = 1.000

In Table 3.1, I have summarized Dauer's data for interval durations for 1- to 4-syllable intervals. Average durations are worked out from the data presented in the paper. The reason for limiting the number of syllables to 4 is that there are not reliable data for more than 4-syllable intervals for all the languages. (By 'not reliable' I mean that duration figures are based on less than 4 occurrences.) For Thai there are in fact reliable data only for 1- to 3-syllable intervals. This is a minor point, however, and nothing of what I am going to say would be different in any important way by making a different decision.

Using these figures, it is possible to compute linear regression equations for the different languages. These equations are presented in Table 3.2.

How are these equations to be interpreted? Well, they somehow model the durations of intervals as a function of the number of syllables. The 'slope' of a line tells us how much is added to the duration by adding a syllable and the constant term indicates that the increase is not directly proportional but has an 'initial value'. What could be the cause for this initial value? Well, it is most likely due to the longer durations of stressed syllables. These equations can, thus, be seen as models for interstress interval durations that tell us that the duration is a linear function of the number of syllables plus an added constant duration due

Table 3.3. Linear regression equations for the individual subjects taking part in Dauer's study. The equations are based on the mean durations for all interval types of which there are more than 4 occurrences.

English		Italian	
AK	I = 209 + N* 89, r = .992	LM	I = 102 + N*103, r = 1.000
SD	I = 188 + N*118, r = .999	GC	I = 118 + N*107, r = 1.000
Spanish		Greek	
JF(1)	I = 84 + N*101, r = .999	OS	I = 103 + N* 93, r = .996
JF(2)	I = 99 + N*110, r = 1.000	DM	I = 57 + N*122, r = .992
GP	I = 90 + N*134, r = .990	KM	I = 130 + N*106, r = .997
Thai			
LU	I = 173 + N*125, r = 1.000		

to the lengthening of stressed syllables. Given this model, some rather striking observations can be made. First of all, the regression coefficients tell us that the fit between the model and data is almost perfect. To assume that interstress interval durations increase linearly as a function of the number of syllables, predicts the durations almost perfectly. This is only true for averages, of course, but it is nevertheless striking. Another observation is the small differences in the slopes of the lines. What this means is that intervals seem to grow at approximately the same rate (105 ms per added syllable on the average) for all five languages. And, finally, the main difference between the languages seems to lie in the constant term, presumably connected with the extra duration of stressed syllables. Here the variation is greater but not random. It seems that the constant terms roughly fall into one of two categories; for English and Thai around 200 ms and around 100 ms for the other languages. Now, just to show that this was not due to any particular choice of data on which to compute these equations, I have also processed the data in a slightly different way. In Table, 3.3 I have computed regression lines for the individuals participating in Dauer's study. The same general picture emerges from the individual equations. The added duration per syllable varies around a mean of roughly 110 ms. It is to be noted that the inter-individual variation is greater than that between languages. The impression one might have from Table 3.2, that the addition per syllable was somewhat greater for Spanish, is seen to be due to the deviation in one particular speaker. Also, the added duration due to stress again seems fall into two classes, 199 and 173 for English and Thai, and 91 ms, 97 ms and 110 ms for Spanish, Greek and Italian (mean values). Now, based on these figures it is possible to propose a very simple model for all languages. The average duration (I) of an interval as a function of the number of syllables (N) can be described by the formula:

$$I = k + 100*N$$

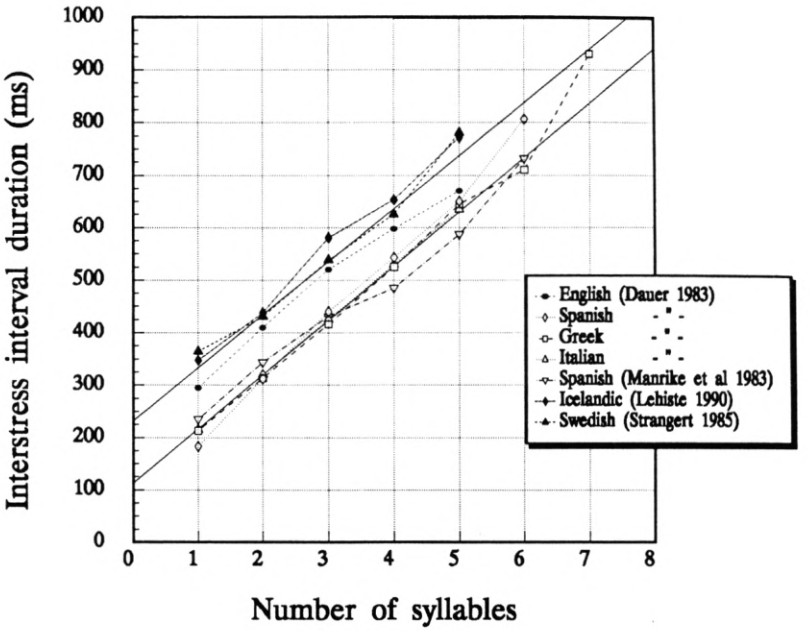

Figure 3.1. Interstress interval duration as a function of the number of syllables for the languages in Dauer (1983) and some of the languages discussed in 2.3.1. Data points in the diagram are based on average values computed from figures published in the papers referred to in the legend. The two straight lines are regression lines based on average durations for the two groups taken separately.

where 'k' is a constant, 200 ms for 'stress-timed' languages and 100 ms for 'syllable-timed' ones. The durations predicted by the formula can be compared with the average durations in Table 3.1. It turns out that the greatest relative deviation from the values predicted by this very simple formula is for trisyllabic intervals in Spanish and Thai, where the deviation is 10%, and less for all other cases. Now if this holds true when tested on a larger material, it would mean that interstress interval durations grow at the same rate as a function of the number of syllables in all languages but that the different durations of stressed syllables make the total durations different.

In Figure 3.1, I have illustrated interstress interval duration as a function of the number of syllables for the languages analysed above. In addition, I have added data from some of the studies discussed in 2.3.1. From the diagram, the pattern I have pointed out above is clearly visible, although the picture becomes a bit scattered for the longest intervals. The two straight lines in the diagram represent regression lines based on the average duration values in each group. It may be seen that the lines are almost perfectly parallel, indicating that the growth rate is the same in both groups. It may also be seen that the intercepts are in good agreement with the predictions, based on the simple formula above. For the regression lines in the figure, the values for the 'slope' agree almost perfectly with the

predicted 100 ms (101 ms for the 'stress-timed' group and 103 ms for the 'syllable-timed' one), but the intercept values are somewhat higher than the predicted 100 ms and 200 ms (112 ms and 232 ms respectively).

It should be pointed out, perhaps, that in order to say anything with certainty about universals, as well as language differences, one must have access to a larger data base, since inter-individual variation may be considerable, which will be clearly demonstrated in my own study in Chapter 4, and most of the data used above is based on the results of studies of only 1—3 speakers. But I *do* think that there is a very strong suggestion in these results that this is a line worth exploring further.

Another question to consider is the fact that descriptions of the total durations of interstress intervals do not permit any definite conclusions to be drawn about what processes that might operate within the intervals. For instance, it is quite possible for there to be compression tendencies in syllables as the number of syllables grows without it necessarily having any detectable effect on the growth of total durations. This will be shown in the following section (3.3). With respect to language comparisons it may very well be the case that languages which appear to behave in a similar way at interstress interval level will turn out to be subject to very different interstress interval-internal processes.

3.3 Compression of syllables.

As was demonstrated above, interstress interval durations increase linearly to a very good approximation. It would seem that if intervals increase by a constant amount per added syllable, this would indicate that there is no compression of syllables inside these intervals. As I pointed out above, this also seems to be an assumption that some authors make. But as I will demonstrate in the following, this is not necessarily true. The implication of this result is, of course, that it is necessary to study durations at both levels, interstress intervals as well as syllables.

I will carry out my demonstration by example rather than as a formal proof. It does constitute a kind of proof, however, in the sense that one counter example is enough to upset a general rule. Those so inclined may construct a more formal and general proof, all according to personal taste.

Let us assume that interstress interval durations in a particular language can be described as a linear function of the number of syllables. The duration of an interstress interval with 'n' syllables (I_n) can thus be described by the formula:

$$I_n = a + b*n \qquad (1)$$

where 'a' and 'b' are constants. If we want to make explicit the fact that the durations are composed of the duration of a stressed syllable plus the duration of a number of unstressed ones we could rewrite the formula as:

$$I_n = \alpha + \beta*(n-1) \qquad (2)$$

The interpretation of the constants 'α' and 'β' will be the durations of stressed and unstressed syllables respectively, in the case of constant durations for stressed and unstressed syllables. Applying the formula to speech data, I have found the fit to be very good. Regression coefficients of .95 or higher are common. One interpretation of this fact is that the interstress interval durations are composed of stressed and unstressed syllables of constant but unequal durations. But suppose that we do not make these assumptions about syllable durations. Let us then see what possibilities there are, still given that the linear function shall be a correct description of the overall durations. Now, given that there are no restrictions at all on the durations of the component syllables of an interval of a given length, then there are virtually infinitely many possible combinations of durations of the component durations that would all add up to a given total duration. There is no need, though, to consider the problem in such a general way since we want to apply it to speech and we know that there are severe restrictions on what durations that may come into question. So let us simplify the problem, using our intuitions about what durations and relations one is likely to find in real speech. I will also restrict the generality of the problem by making the following assumptions: 1) Only mean values for interstress interval durations and syllable durations will be considered. This means that I will not worry about the durations of specific phonemes but only consider mean values over a fairly large number of examples, thus obscuring differences in detail. 2) The duration of stressed and unstressed syllables will be considered constant for an interval of a given length. That is, stressed and unstressed syllables will be considered to have certain (constant) durations for, say, all 4-syllable intervals, which need not, however, be the same as those for intervals of other lengths. Given those restrictions, the duration of an n-syllable interval (i_n) can be given by the formula:

$$i_n = S_n + U_n*(n-1) \qquad (3)$$

where 'S_n' and 'U_n' are the durations of stressed and unstressed syllables for a given interval length (n). The indices on 'S' and 'U' are there to remind us that the durations are functions of the number of syllables in the interval. (Note that this is not the regression equation any more.)

In section 3.2, I showed that the formula assuming interval durations of a stress-timed language, say English, to increase by a constant amount (100 ms) per added syllable and with an extra duration of 200 ms to compensate for the stressed syllable fit the data almost

perfectly. Using these data we can express interval durations for that language as a function of the number of syllables as:

$$I_n = 300 + 100*(n-1) \qquad (4)$$

This equation gives the durations 300 ms, 400 ms, 500 ms, 600 ms, and 700 ms for 1- to 5-syllable intervals respectively. Now, let us explore equation (3) above under the assumption that interval duration should be the same as that expressed by equation (4). That is, what kind of variation in 'S_n' and 'U_n' is permitted if, for every n, we require that $i_n = I_n$.

With respect to the compression of syllables there are four possible processes that could operate inside the intervals: 1) there is no compression of syllables, 2) there is compression in the stressed syllables but not in the unstressed ones, 3) there is compression in the unstressed syllables but not in the stressed ones, 4) there is compression in stressed as well as in unstressed syllables. I will consider each of these possibilities using numerical examples, under the assumption that $i_n = I_n$, and using the total duration values computed from equation (4). The examples will be restricted to intervals with 1 to 5 syllables but can, of course, be generalized to intervals of any size.

1) *There is no compression.* This possibility has already been discussed. In the case that stressed syllables (S_n) are 300 ms and unstressed syllables (U_n) are 100 ms the conditions are satisfied and there is no compression in either type of syllable.

2) *Stressed syllables are compressed but not unstressed ones.* Suppose stressed syllables (S_n) have the durations 300 ms, 290 ms, 280 ms, 270 ms, and 260 ms respectively for 1- to 5-syllable intervals. Then if unstressed syllables (U_n) are always 110 ms, the conditions are again satisfied. It is, thus, possible for there to be compression in the stressed syllables as the interval grows without giving up the requirement that interval durations should grow linearly.

3) *Stressed syllables have constant durations but unstressed ones are compressed.* Now, if stressed syllables (S_n) are of the same duration, it follows that the duration of the monosyllable is that duration, that is 300 ms. But if that is the case there is no possibility for any compression in the unstressed syllables (U_n) without violating the 'constant duration increase' requirement. All unstressed syllables will in this case be 100 ms. It is thus not possible for there to be any compression in the unstressed syllables if all stressed syllables (including the monosyllable) are equal.

4) *Both stressed and unstressed syllables are compressed as the number of syllables in the interval grows.* If stressed syllables (S_n) have the durations 300 ms, 275 ms, 260 ms, 249 ms, and 240 ms for 1- to 5-syllable intervals and if the corresponding unstressed syllables (U_n) have the durations 125 ms, 120 ms, 117 ms, and 115 ms, then again the requirements are fulfilled. It is thus possible for there to be compression in both stressed and unstressed syllables while at the same time interval durations increase linearly.

We can, thus, see that even in this simple case, with all its restrictions, there are possibilities for constancy of syllable durations as well as two types of compressions without abandoning the requirement that interval durations should grow linearly as a function of the number of syllables. Nakatani *et al.* (1981) mention that the durations in their data are in some cases slightly positively accelerated and indicate that this is a tendency in quite the opposite direction of compression. Not even this is necessarily true, however, which the following series of duration values demonstrates. If stressed syllables are 300 ms, 265 ms, 250 ms, 245 ms, and 244 ms, and the corresponding unstressed syllable durations 130 ms, 123 ms, 120 ms, and 118 ms, this will result in an accelerating series of durations, 300 ms, 395 ms (= +95 ms), 496 ms (= +101 ms), 605 ms (= +109 ms), and 716 ms (= +111 ms), in spite of the fact that there is compression in both stressed and unstressed syllables.

What these examples show is that to be able say anything about compression tendencies inside interstress intervals, one has to examine these inner processes very carefully. Data about interstress interval durations alone are not enough to decide this questions.

A possible effect, which I have not mentioned, is that of final lengthening within intervals. It cannot be said to be established with any certainty that such an effect exists, but there are some results that indicate that this may be the case. In the study presented in Chapter 4, the results seem to indicate that there is a final lengthening within intervals. The significance of the effect is not established beyond doubt but it is there and must be considered as a possibility. As was mentioned in 2.3.1, Strangert found stressed syllables in monosyllabic intervals to be longer than those in polysyllabic intervals. This does not constitute a proof that there is a final lengthening effect but the result is exactly what one would predict assuming that there is one. If, for the sake of argument, we assume that there is such an effect, further possibilities arise. Among other things, it would always seem as if unstressed syllables were compressed (even when they are not) and average durations for unstressed syllables would be longer than the 'growth factor'. (An example of how this may work can be found in 4.7.) Comparing mean durations found for unstressed syllables with the size of the linear increase (around 100 ms, found here, as compared to 186 ms for English found by Delattre, 1966), this does indeed seem to be the case. But these remarks must not be seen as attempts to prove anything, only as suggestions that these possibilities should be further explored.

The question of temporal adjustment as a function of context has been studied with respect to segment duration as a function of word length in a number of studies (Lindblom and Rapp, 1973; Lindblom, Lyberg, and Holmgren, 1981). It has been suggested that segment duration may be a function of position as well as word length. But these results are not immediately applicable to the question of temporal adjustment of syllable duration as a function of interval length in interstress intervals. How and if syllables are actually shortened as a function of increased interstress interval length has not been the subject of many studies. There are a few such studies and their results indicate that there may be

shortening processes at interstress level as well. But the experimental conditions were such that it is not immediately clear how the results should be interpreted with respect to the question of interstress interval length as an independent duration conditioning factor. In the following I will review and discuss two studies which I believe to be representative.

Fowler (1977) has studied stressed vowel duration as a function of interstress interval length. Six sets of carrier phrases, each in six versions with different target intervals were read by one speaker. One of the sets was:

1. The *FACT* started the argument.
2. The *FACT has* started the argument.
3. The *FACTor* started the argument.
4. The *FACT has re*started the argument.
5. The *FACTor re*started the argument.
6. The *FACTory* started the argument.

The target interval is italicized and the stressed syllable in the target is in capitals. The other five sets had a similar construction. The test variable was the duration of the vowel in the stressed syllable in the target. In sentence type 1 the stressed syllable is immediately followed by another stressed syllable, in types 2 and 3 it is followed by one unstressed syllable and in types 4, 5 and 6 by two unstressed syllables. The idea was to see if the stressed vowel in the target was systematically shortened as a function of the number of following unstressed syllables.

For the 6 types, the mean durations of the target vowels are, 118 ms, 106 ms, 103 ms, 106 ms, 98 ms, and 93 ms. In the case when there is no following unstressed syllable the target duration is thus 118 ms. When the target is followed by 1 unstressed syllable the mean duration for the target is approximately 105 ms, and with 2 following unstressed syllables 99 ms. It thus seems as if the vowel duration is shortened as a function of the number of following unstressed syllables. Fowler reports the overall differences to be significant (ANOVA, $p < .005$). A few observations can be made that should invite some caution when interpreting these and similar results. First of all the differences are small. In fact none of them are significant if one uses the figures published. Neither the overall results (ANOVA), nor pairwise t-tests show any significance. Of course analysing the whole set of data (5 repetitions for each sentence) makes it possible to discriminate between smaller differences. Also, looking at the results in more detail, they are far from unambiguous. For the particular sentence cited above, the duration of the vowel in type 1 (0 following unstressed syllables) is actually shorter than that in type 4 (2 following syllables) (113 ms vs. 117 ms). For two other sentences the respective durations in type 1 and 4 are almost identical (132 ms vs. 130 ms, and 62 ms vs. 60 ms). The most drastic differences are present in contrasts like 'know' vs. 'noticing' and 'Dave' vs. 'Davidson'. One might

Table 3.4. The mean durations in milliseconds of the target vowels in some of the test words used in Fowler's study. Target vowels are italicized.

Target word	No. of syllables in the interval		
	1	2	3
f*a*ct	113	105	117
kn*ow*	122	120	92
D*a*ve	132	122	120
J*a*n	147	122	120
d*a*rk	132	113	130
P*e*te	62	55	60
Mean	118	106	106

question if 'know' and 'no' in 'noticing' is the *same* syllable, but more importantly whether shortening effects in words should not be treated separately from possible shortening in interstress intervals. It does not follow automatically that if there are shortening effects in words, that for example makes the vowel in 'Davidson' shorter than that in 'Dave', the same must be true for interstress intervals. One would certainly have reason to expect such an effect, but its existence will nevertheless have to be shown independently.

If one treats those sentences where the words are the same, a slightly different picture emerges (Table 3.4). The difference in means between the three conditions is rather small. We can also see that vowels in monosyllabic intervals are markedly longer than in trisyllabic intervals in only two of the cases (know and Jan). For the rest of the sentences the difference is negligible. Another noteworthy fact is that there is no difference in means between di- and trisyllabic intervals. If there is any shortening involved it seems to be primarily between monosyllabic intervals and disyllabic ones. A possible explanation for this could be the following: a general problem with carrier phrase/target word types of experiments is that it is difficult to avoid focusing the target in such readings. Particularly if the same subject reads the phrases over and over again, it is difficult not to think of the changing targets as marked in some way and place focus on them. This could introduce a lengthening effect which would affect the stressed syllable more the shorter the target is. If this is the case, then, what seems to be a gradual shortening could instead be a gradual decrease in the lengthening effect on the stressed syllable in the focused 'word' as a function of word length, and not an effect of the number of syllables in the interval in general. It should also be noted that this type of task is rather 'laboratorish' and may not have a lot to say about more normal reading tasks, let alone conversational speech.

Perhaps I must also say that I am not trying to play down the importance of Fowler's experiment. All I am saying is that the interpretation is not unambiguous, and the extension of the results to speech in general is not obvious.

In a study similar to Fowler's, Fourakis and Monahan (1988) studied the effect of metric foot structure on syllable durations. Their approach is somewhat different, mainly in that they analyse phrases in terms of metric feet rather than interstress intervals. This difference is not important here, however, and it is possible to reinterpret their data using the interstress interval concept instead if one wishes.

The test variables in their study were syllable durations. Like Fowler they used a carrier phrase within which a test interval was varied but all intervals were analysed. They found shortening effects as a function of context which in some cases were significant. The stressed syllable in the target was significantly shorter when it was followed by two unstressed syllables than when it was followed by only one (372 ms vs. 396 ms). But here again, we may not exclude the possibility of a focus intonation effect. There was, however, a similar effect also outside the target. When the first stressed syllable (not in the target) was followed by two unstressed syllables it was significantly shorter than when it was followed by only one (376 ms vs. 400 ms).

In addition to the analysis of durations of syllables, Fourakis and Monahan also treat the questions of whether the process of adding syllables is a purely concatenative process in a different way. In this analysis they commit an error, however, that I will report because it has some methodological importance. Their line of reasoning is the following: if adding syllables is a purely concatenative process, then the addition of one or more syllables will not alter the syllable rate (measured as the number of syllables produced per unit time). Foot rate, on the other hand, will decrease because the average foot duration will be longer. Thus, there will be no correlation between syllable rate and foot rate if phrases with differing number of syllables are compared. This is not correct, however. In the particular context of this and similar experiments it is (for obvious reasons) the number of unstressed syllables which is varied. Since they are shorter in duration, this means that adding more such syllables will decrease mean duration. This will show up as an increase in syllable rate. The prediction one would make is thus not that syllable and foot rate will be uncorrelated but that there will be a (slight) negative correlation between the two. This is also exactly what Fourakis and Monahan get. From that they draw the conclusion that the process is not strictly concatenative. But this is not a valid conclusion. A closer inspection of their data shows that for the simplest sentence the average syllable duration is 243 ms. The average of the 'added' syllables, on the other hand is only 166 ms. From these values alone the negative correlation follows. Their results may be used in another way, however, that *does* lend some support to their claim. If one predicts that the durations of the phrases with more syllables should be the same as the durations of shorter ones plus the duration of the added syllables then it turns out that this is not true for most of the cases. The sentences with more syllables are some 40 ms shorter, on the average, than they 'should' be assuming a purely additive process. The effect is very small, of course, and it is not possible here to say anything about its significance, but it does go in the direction predicted by the authors.

In spite of the methodological objections one might have about the two studies by Fowler, and Fourakis and Monahan their results must nevertheless be taken seriously. They indicate that there may indeed be some shortening processes operating within feet as a function of the number of unstressed syllables. But the effect seems rather small and one would like to see more studies were the possible confusion involved with using the target—carrier paradigm is avoided, to get a clearer picture.

A radically different approach to the study of possible shortening effects working in the direction of greater regularity has been taken by Cutler (1980). She uses data from syllable omission errors in an attempt to demonstrate that there are tendencies to isochrony in production. The assumption is, of course, that if syllable omission errors systematically have the effect of making speech more regular this would indicate an underlying force striving for isochrony. In a corpus of 28 sentences containing syllable omission errors she found that, in 24 of them, the errors tended to make the sentences more regular than the target utterances would otherwise have been. The measure of regularity was the variation in the number of syllables per interstress interval. No physical measurements were made. This is a weakness but, since interstress interval durations increase as a function of the number of syllables, a more even number of syllables will, in most cases, lead to a greater physical regularity. I have more serious doubt concerning the inference that these errors show that there is an underlying tendency towards regularity. Other equally plausible explanations exist. In the examples used by Cutler, omission errors occur in long and complicated structures—'bicential' for 'bicentennial', 'interlocker' for 'interlocutor', 'metrolitan' for 'metropolitan' an so on. It is reasonable to assume that the longer and more complicated a sequence of syllables is, the greater the probability of making an omission error. And if an omission error occurs in an interval with a greater than average number of syllables, this will automatically result in a greater regularity, without there having to be any underlying regularizing force. The greater regularity would then be a purely secondary effect. This makes the use of syllable omission errors as evidence for regularization problematic and one must evaluate data very carefully. To summarize, one may say that the data in Cutler's study are compatible with, but not necessarily evidence for, regularization. There are other equally plausible explanations.

3.4 Stress-beats vs. perceptual centres.

The study by Allen (1972), mentioned in 2.1, is an example of an experimental study of stress beat perception using modern techniques. In a series of experiments, using different response types, Allen tried to determine where subjects perceived those events to be that gave the impression of being the 'beats' in the perception of speech rhythm. The speech material in all three experiments was utterances selected from a recording of conversational speech between three speakers. To establish the perceptual degree of stress for the individual syllables in the utterances, Allen asked five of his colleagues to transcribe the

utterances and mark the stresses. (The markings were not entirely consistent.) Each syllable was assigned a stress score based on the total number of stress scores for all markers.

Allen carried out three experiments in order to establish the perceptual locations of the syllable beats. In all experiments, the test sentences were presented repeatedly to the subjects. The task was to determine the location of a given syllable. Three different techniques were tried in the experiments. In the first experiment, subjects were instructed to tap with a finger each time they heard a given syllable. Tap locations were registered electronically. Subjects tapped 50 or 100 times to each syllable. In the second experiment the subjects could place a sharp click anywhere within an interval of 1.5 seconds around a given syllable with the aid of a control knob. When the subjects felt satisfied that the click matched the beat of the test syllable the procedure was repeated or a new syllable was chosen. In the third experiment, an audible click was placed somewhere within an interval of 600 ms centred around the average click placements derived in experiment 2 for each of the syllables to be tested. Subjects were asked to determine whether the click 'hit the beat' or not.

The results from the three experiments were in general agreement. In the author's own words: *"Both clicks and taps are placed by subjects in the general region of the onset of the nuclear vowels of the stressed syllables but before the vowel onset by an amount correlated with the length of the consonant(s) preceding; both clicks and taps are more reliably placed on stressed syllables than on unstressed syllables. ... But tapping locations are different for different subjects, and these tapping biases make this behaviour less suitable as a beat locating device than click matching, which does not exhibit inter-subject differences."* (p. 189)

Experiments of the type conducted by Allen mean producing or matching a click or some other acoustical marker to a speech stimulus. A technique working in the opposite direction could also be used. Subjects could be asked to produce speech, syllables or words, to some acoustic marker stimulus. This technique was used in the experiments by Lindblom (1970, cited in Rapp, 1971) and Rapp (1971) mentioned above. In those experiments subjects were asked to read words so that they matched the clicks of a metronome pulse.

Lindblom asked subjects to read nonsense syllables in synchrony with a constant stimulus pulse. The syllables were aCa: with C ∈ {/t/, /s/, /d/, /n/, /l/}. Subjects timed their readings so that the onsets of the second (stressed) vowel came close to the stimulus pulses. There were some systematic deviations, however. For the voiceless consonants, the beat was placed, on average, 20 ms after the onset of the vowel vs. 50 ms for the voiced ones.

Rapp's study is very similar to Lindblom's and may be seen as an extension of his study. In Rapp's study the stimulus pulse was a series of equidistant pulses presented via earphones at a rate of 2 per second. Three subjects were instructed to synchronize their reading with every second pulse. The 'words' were again nonsense syllables of the type aCa:, with the set of consonants and consonant clusters extended to {/s/, /t/, /d/, /l/, /n/, /st/,

/str/}. The syllables were read by three male Swedish speakers. They synchronized their readings so that the pulses came at or before the onsets of the second vowel. Exact figures are not given, only diagrams, but judging from the diagrams, click placements were in the interval – 80 ms to 0 ms relative to the vowel onset. There was considerable inter-individual variation. One subject placed the pulses on average some 55 ms earlier than the other two. Standard deviations ranged between 20 ms and 45 ms. There was some general agreement, though, between subjects with respect to the influence of phonetic context. Subjects tended to place the beats earlier the longer the duration of the consonant(s). There seemed to be an approximately linear correlation between placement relative to the vowel onset and the duration of the consonant(s). In Rapp's study too, the voiced/voiceless contrast seems to have had an effect, pulses being placed later when the consonant was voiced.

The technique of letting subjects read to the clicks of a metronome was also used by Fox and Lehiste (1987). In one of their experiments three subjects were asked to read stressed CVC-syllables to a metronome at a rate of one per second. They were instructed to read as isochronously as possible. All syllables started with a fricative (/s/) and ended with a voiceless stop (/t/), the vowel being the variable. The results were in agreement with those obtained by Allen and by Rapp. Subjects read the lists so that the metronome clicks came to be placed very close to the onset of the vowel. The mean values for the click locations were all in the interval –10 ms to +25 ms relative to the vowel onset. Vowel duration seemed to influence click location. Longer duration tended to place the click later. Vowel quality on the other hand did not seem to have any effect. It should be noted, however, that the direction of the correlation between click placement and vowel duration obtained when average placements are used holds for only two of the subjects when individual results are considered. For the third subject there is a correlation in the opposite direction.

Both production and perception experiments involving some kind of matching between an acoustical marker and speech seem to suggest that the onset of the vowel plays a central role. For the match to be acceptable the marker should be placed at, or very near the onset, although factors like phonetic context may influence placement to some extent. There seems, thus, to be rather strong evidence for connecting the perceptual stress beat with the vowel onsets. Results from research on temporal perception in other areas than speech generally agree with this view.

Hirsh (1959) carried out a series of experiments in which subjects were to decide the order between two stimuli, by telling which of the two came first. Stimuli were, clicks, noise bursts, and tones. Onset differences in the range 10 ms to 20 ms were enough for 75% correct responses. The results from one of the experiments form an interesting contrast with the results obtained by Fox and Lehiste reported above. A brief click sound was presented together with tones of different durations (20 ms, 50 ms, and 100 ms), and rise times (7 ms and 15 ms). For both rise time conditions the click had to precede the longer tone by about 10 ms to be perceived as simultaneous. In other words, the longer tone seemed

to start a little earlier. This is the opposite result to that obtained by Fox and Lehiste, but the experimental conditions are of course quite different.

Evidence to support the idea that sound onsets play a crucial role can also be found in another field. In ensemble music, and in singing to accompaniment, the synchronization of instruments and voices is of course crucial. In a series of studies, Rasch (1978, 1979, 1981) studied the synchronization of instrument voices in ensemble music and the perception of simultaneous tones. The standard deviations for tone onset asynchronies in ensemble music were in the range 30 ms to 50 ms. Deviations of this order go unnoticed in normal music listening. In a laboratory situation, however, when only two tones are presented, differences of this order may be perceived. Larger onset differences, in the order of 100 ms to 200 ms, that occasionally occur in ensemble music can be perceived if listened for carefully. Similar evidence for the role of tone onsets and vocal onsets in accompanied singing comes from experiments in music synthesis. In experiments with the synchronization of synthetic voices and music, it has been found by Sundberg (1989, and personal communication) that the best sounding versions are those where vowel onsets are synchronized with tone onsets.

All these different experiments seem to suggest that the onsets of tones and vowels play a crucial role in perceiving their 'moment of occurrence'. Other factors that may influence the perception to some extent may be particular for speech. In studies involving linguistic material, the results by Fox and Lehiste, reported above, suggest an influence of vowel duration. However the effect is small (typically in the order of 10 to 30 ms), and the results were not unambiguous. No similar effect has been found in music, as far as I know. According to Sundberg (personal communication) the synchronization of onsets is crucial, but the offsets do not seem to play any similar role. This means, for instance, that a long and a short tone will be perceived as simultaneous if their onsets are. With respect to a possible effect of the consonants preceding the vowel, no comparative material exists, as far as I know, in any other field.

In many experiments trying to establish the locations of stress beats in speech, stress beats have typically been placed some 30 ms or so before the onset of the vowel. This is particularly true when the experiments have involved some kind of motor response, usually finger tapping (Miyake, 1902; Classe, 1939; Allen, 1972), but has also been found in other studies (Rapp, 1971). The explanation proposed is generally that the phonetic context influences the perception of where the beat is. A positive correlation between the duration of the preceding consonants and beat location has been demonstrated in some of the experiments. This may very well be the case. There exist, however, experimental results from other areas of rhythm research that indicate a need for caution in the interpretation of these correlations. As was mentioned in section 1.6, it has long been known that subjects, when asked to synchronize tapping to a stimulus rhythm consisting of a sequence of tones or clicks, tend to tap a little earlier than the stimulus beat; a phenomenon known as rhythmic anticipation. A typical value for anticipation is 30 ms before the beat. It should be noted

that this is of precisely the same order of magnitude as the 'anticipations' in the stress beat experiments. In the experiments where the anticipation effect has been studied, there can, of course, be no influence of any phonetic context since the stimulus beats were identical short tones or clicks and the subjects tapped before the onsets of these sounds. One must therefore ask if an anticipation effect cannot also be involved in the experiments on stress beat placements in speech. As was also mentioned in section 1.6, subjects establish their perception of inter-beat distances very quickly, often within three beats. It would, thus, be perfectly possible for subjects to establish some kind of anticipatory behaviour in most of the experiments on stress beats in language. If there are strong correlations between some specific phonetic variable, say the duration of the prevocalic consonant(s), this can perhaps not be explained by anticipation alone. But the possible influence of an anticipation effect must definitely be considered.

In the study by Allen (1972), described above, he observed in the tapping experiment that there was great inter-subject variation in mean tap placement. In an attempt to 'calibrate' the subjects, he carried out a control experiment with neutral stimulus material. In this experiment, the subjects tapped to the last three of a sequence of four clicks, 800 ms apart. The results from this experiment showed that subjects tapped earlier than the clicks by roughly the same amount as they did with the speech stimuli. For the three subjects, the average amount of time by which their taps preceded the clicks were 15 ms, 19 ms, and 65 ms. The corresponding values for tapping to syllables were approximately 6 ms, 23 ms and 37 ms. But this means that the effect found, that the beat locations tend to precede the onsets of the vowels is cancelled out completely if one assumes an anticipation effect in tapping to syllables. The interpretation of the correlations (between the amount by which subjects place stress beats before the onset and the duration of the consonants), which, by the way, are very moderate, and only significant for two of the three subjects, also become questionable in this light. In fact, if the mean value for anticipation (33 ms) is subtracted from the means over subjects for the different consonant categories, only two mean tap placements (voiceless stops and clusters) precede the onset of the vowel. For the other 7 categories, the mean tap locations are in the vowel.

If, and how, anticipation also plays a role in matching tasks is not easy to say. But it is quite conceivable that subjects have a tendency in these tasks to accept early click placements more easily if they coincide with their expectations. If one assumes that there is such an effect and of the same order, then for the subjects in Allen's second experiment, the click placements precede the vowel onsets for only one of the subjects if anticipation is subtracted.

In experiments where stress beats have been found to precede the vowel onset by some amount, influence of the phonetic context is usually offered as the explanation. But, as I have indicated, using Allen's data as an example, at least in some types of experiments the effect may be an artefact of the experimental situation, through the anticipation effect. My

calculations do not prove that, but they strongly suggest that the possibility be taken seriously.

The results of the experiments discussed above have significant implications for the study of speech rhythm. They mean that the onsets of the vowels should play a decisive role in the analysis of speech rhythm, particularly in perception. In analogy with music, the distances between onsets ought to be relevant measures of interstress interval durations from a perceptual point of view. Other factors, like the duration of the vowel and the duration of the preceding consonants, may be relevant. But the deviations from the vowel onsets that have been found are typically small compared to typical interstress interval durations—in the order of 30 ms compared with 300 ms to 800 ms. It is, therefore, reasonable to assume that the distances between successive vowel onsets give us a fairly accurate picture of what intervals our perception of interstress interval and syllable durations are based on.

Proponents of the theory of p-centres as the solutions to what it is in the speech signal that accounts for our perception of speech rhythm would not agree with the views on stress beats I have expressed above. On the contrary, they often dismiss the idea that vowel onsets could be the underlying acoustic cue to rhythmic stress beats in rather strong terms—"*These findings disconfirm a hypothesis about listeners' judgements of rhythmicity in speech ... namely that listeners base rhythmicity judgements on the intervals between the onsets of acoustic energy of successive syllables*" (Fowler, 1979, p. 375), "*Experiments on the perception of speech have also shown that listeners' judgements of rhythmicity are not based upon intervals between the onset of acoustic energy of successive syllables*" (Fox and Lehiste, 1987, p. 1). But there are several reasons to be cautious. The results in p-centre experiments have been obtained under conditions very far removed from natural continuous speech. First of all, these results have been obtained under conditions where the task was explicitly connected to isochrony, in the production experiments often even with a metronome present. In the perception experiments, subjects were explicitly told to manipulate stimuli to obtain the greatest possible isochrony or pick out the stimulus that sounded the most isochronous. There is no evidence that isochrony is this important in normal continuous speech production or speech perception. Secondly, all experiments on p-centres have been carried out on lists of isolated words, often nonsense syllables, and it has not been possible to say how p-centres could be defined for continuous speech, the main reason for this being that it has not been possible to find any correlate in the speech signal that can be reliably connected with the p-centres. In fact, we do not even know whether the p-centre concept has any relevance at all in connection with continuous speech. This also means that it is not possible to compare p-centre locations with stress beat locations or any other correlate to rhythm one might want to suggest for continuous speech.

The case for stress beats, connecting the rhythmic beats in speech with vowel onsets, is much stronger. Here, many of the results have been obtained using continuous speech. The exact stress beat location may be influenced by phonetic context but, compared to normal

interstress interval durations, the deviations found are very small. Also, the experiments involving accompanied singing (e.g. Sundberg, 1989), indicating that tone and vowel onsets should coincide for one to perceive that singing is in time with the music, argue strongly in favour of the importance of the onset. In ensemble music, the onset is again the relevant point of synchronization. And finally, in experiments on the perception of simultaneous tones and noise bursts the most important cue to simultaneity is simultaneity of onsets.

Since the possibility of accepting p-centres as a serious alternative describing the rhythmic beats of speech hinges critically on the possibility of defining p-centres for continuous speech, I will close this section by describing a suggestion by Marcus (1981) how this may be done. He tried, in a series of experiments, to find correlates that would make it possible to establish locations for p-centres. He was not able to suggest any absolute locations for the p-centres but found some interesting correlations. The material used in the experiments was monosyllables of the C_1VC_2 type (where the Cs are single consonants, consonant clusters, or null). For this type of stimulus, the relative p-centre locations were strongly correlated with the durations of the initial consonants, but also with the vowel durations. Increasing consonant duration tended to place the p-centre earlier in the word while increasing vowel duration moved the p-centre to a later point in time. Based on these findings, Marcus proposed a model which defines relative p-centre location as a function of context for monosyllables of the type used.

$$P = \alpha X + \beta Y + k$$

The equation expresses the p-centre location (P) relative to the word onset. 'X' is the duration of the initial consonant(s) and 'Y' the duration of the vowel and the final consonant(s). 'α' and 'β' are parameters, expressing the relative influence of vowel and consonant durations, and 'k' an arbitrary constant, expressing the fact that p-centre locations are relative. Parameter values for (α,β) of (.65,.25) accounted for most of the variation in p-centre location in Marcus' study. Eling, Marshall, and van Galen (1980) obtained a correlation of .88 using the same model for Dutch digits (with (α,β) = (.58,.25)).

It is, however, far from obvious if and how this model can be applied to continuous speech. Marcus suggests a way of generalizing the model to include continuous speech in the form of a differential equation for p-centre location.

$$dP = \alpha dC + \beta dV$$

or if the p-centre is measured relative to the vowel onset

$$dP_V = -(1 - \alpha)dC + \beta dV$$

But this equation can be solved only if one has some means of estimating the durational changes. One must also establish some 'anchoring point' in time if one wants to determine

absolute locations. It is far from obvious how this can be accomplished. The formula could be applied to continuous speech, of course, if one regards the stressed syllables as 'words' and the unstressed syllables as 'silence'. The interstress interval durations could then be calculated by first using vowel to vowel onsets and then adjusting these durations using the formula. But even if this works technically, I am not convinced that this is the relevant analysis.

3.5 Perceptual isochrony—"Speech is heard as more regular than it really is".

As I have mentioned above, the idea that some languages (particularly English) are isochronous has a long tradition. For obvious reasons, the ideas were originally based only on intuitions. Steele obviously thought he heard stresses appear at regular intervals. Now that we know that there is little basis for claiming that the speech signal has any such regularity, one might ask if we nevertheless perceive it as regular. Some researchers claim that we do. Now, a simple explanation why one perceives things to be equal is that one is unable to tell the difference. It may, for example, be the case that we perceive interstress intervals as being equal because our perceptual system is unable to detect the physical differences there are. This is a hypothesis I will try to approach in two of my experimental studies (Chapters 6 and 7). The research I will review and discuss in this section has taken a different approach to the perception of regularity in speech. It has not been primarily concerned with the role of duration perception as such but has addressed the perception of regularity more directly. A more structural approach, one might perhaps say. In some of these studies, results have been obtained that have been interpreted as evidence that speech stimuli are perceived as more regular than non-speech stimuli with identical temporal structures. Speech is 'regularized', as it were, in perception. Although, if this principle has some truth in it, it cannot be ruled out that it is applicable to speech in general; the original studies were done using English speech stimuli only, and aimed at explaining why English may be perceived as isochronous. The phenomenon has been referred to as 'perceptual isochrony'.

In this section, I will describe and discuss the results of some experiments that have been carried out in order to study this question. Some are meant to demonstrate the existence of regularization in the perception of speech, particularly English, and some cast serious doubt on the whole idea.

All experiments that I know of are in principle of the same kind. Subjects are asked to reproduce the rhythm of a speech stimulus by some means other than speech. The experimental techniques used are finger-tapping or arranging a sequence of tones or noise bursts so that the intervals between them replicate as closely as possible what the subjects believe to be the corresponding intervals in the speech stimulus. In most of these experiments the subjects' responses have shown greater regularity than the corresponding speech

stimuli. This has been interpreted as evidence that we perceive speech as more regular than its physical properties would suggest. But the results from the experiments are not unambiguous and they are not easy to interpret. I will review a few representative studies in some detail and try to evaluate their results.

Often cited studies are those by Darwin and Donovan (Darwin and Donovan, 1980; Donovan and Darwin, 1979). They performed a whole series of experiments varying the conditions in different ways. The general format of the individual experiment was that subjects listened to some stimulus, speech or non-speech, and then tried to adjust the distances between a set of adjustable noise bursts so that the pattern matched the perceived pattern of the speech stimulus. The non-speech stimuli they used were of two kinds; a sequence of brief tones or a sequence of isolated synthetic syllables. The speech stimuli were varied with respect to intonation, syntactic structure and the number of tone groups. The subjects could listen to the stimulus as many times as they liked, during their attempts to match them, but were not able to hear the stimulus and their own responses simultaneously. It had been observed in one of the first experiments that subjects often repeated the stimulus phrase during the experiment. To facilitate the task and to be able to see whether the subjects' own repetitions of the stimuli were more regular than the stimuli, subjects were explicitly encouraged to repeat the stimulus sentences during their attempts to match the rhythm in some of the subsequent experiments. These repetitions were recorded and analysed separately.

The general result of the experiments was that the sequences of matching noises were more regular than the stimulus when the stimulus was speech, but not in the case of non-speech. Subjects' own repetitions, which were analysed separately, were not significantly more regular than the stimulus sentences. Intonation did not seem to have any influence in general, but when it signalled the number of tone groups, one vs. two, it did. Stimulus versions with two tone groups were not significantly regularized. The syntactic structure did not have any effect. Donovan and Darwin draw the conclusion that we perceive speech to be more regular than it really is, but not so for non-speech. But the results are not unambiguous. In one of their experiments, 3 responses out of 8 are not more regular and one is even less regular. It is also far from clear that the results show what the authors claim they show, namely that we perceive speech as more regular than it is. Another, perhaps more likely, interpretation is that the subjects tap or match in a generally regular fashion because they are unable to distinguish the exact rhythm, or the exact durations, because of the complexity of the stimulus. Some evidence for this can be found already in Donovan and Darwin's own study. As I mentioned, no significant regularization took place when the intonation broke up the sentence into two tone groups. This meant that the subjects had to match two groups of two syllables instead of a sequence of four. It cannot be ruled out that this facilitates the task by imposing a simple hierarchical structure—two groups of two instead of one group of four. It should also be noted that when the non-speech stimulus was a sequence of isolated synthetic syllables, the regula-

rization did not occur. But this type of stimulus is essentially a speech stimulus although with most of the speech information removed, thus making it simpler. But more convincing evidence, pointing in the direction of a relation between degree of complexity and the tendency to regularization, can be found in two investigations that I will now review.

The first one is a study by Bell and Fowler (1984). They tried to replicate the results obtained by Donovan and Darwin in a similar experiment. Two groups of subjects were told to tap to speech and non-speech stimuli. The speech stimuli were two sentences and the non-speech stimuli were sequences of tone bursts with the same inter-tone distances as the distances between the stressed syllables in the sentences. The results seemed to confirm Donovan and Darwin's results in the sense that tapping to the sentences was more regular than the corresponding sequences of syllables, while the tapping to the tone sequences quite closely matched the actual intervals. When the results were analysed in more detail, however, an interesting observation could be made. The subjects' responses seemed to fall into two groups. In one group (9 of 16 subjects) the subjects tapped out the rhythm of the sentences preserving the durations of the stimuli. In the other group, they tapped more regularly than the rhythm of the stimulus. Two alternative explanations are proposed by the authors. One explanation is that the group that regularized listened to speech stimuli in a different way, 'speech mode', while the other group tried to process the non-speech properties of the stimulus. The other explanation is simply that the groups differed in their abilities to perceive rhythm correctly. The 'good group' who were able to perceive correctly the sentence rhythm also tapped correctly to it while the others, not being able to perceive the rhythm very precisely, tapped in a generally regular fashion. The reason they all did 'better' in the non-speech task could be that the rhythm of that type of stimulus is easier to follow. In a control experiment, Bell and Fowler tried to test the two alternative hypotheses but no conclusive results were obtained. Neither hypothesis found any clear support. Another question Bell and Fowler raise is to what extent, and in what sense, the non-speech analogues in these experiments are really analogous to the speech stimuli they are meant to correspond to. This is certainly also a question that needs careful consideration.

Another interesting experiment that seems to support the view that regularization is a function of the complexity of the task was done by Scott, Isard, and Boysson-Bardies (1985). In an experiment similar to that by Donovan and Darwin, they wanted to compare subjects' responses to stimulus sentences in two different languages, English and French. The idea of perceptual isochrony is that, although some languages (e.g. English) are not physically isochronous, we perceive them as such. Some kind of perceptual regularization takes place. Now, for the perceptual distinction between stress-timed and syllable-timed languages to have any validity, French, which is traditionally considered to be a syllable-timed language, should not possess this perceptual quality. The stimuli in the experiments were sentences in English and French, with varying interstress interval durations. As a control, there were also matching non-speech stimuli (noise bursts). Scott *et al.* also wanted to see whether the native language of the listener might have any effect on the perception

of speech rhythm. In other words; whether English listeners perceive English as isochronous only because of some general rhythmic property of English or because, having English as their mother tongue, they are used to perceiving English in a special way. If this latter alternative should be the case, at least to some degree, one would expect French and English listeners to differ with respect to their perception of English speech rhythm. The tests were therefore carried out with two different groups, having English and French as their native languages. In order to minimize influence of the subjects' knowledge of the 'foreign' language, subjects were chosen so that their knowledge of the other language was as insignificant as was practically possible. (English subjects were for instance unable to translate any of the French sentences.) Both groups were presented with sentences from both languages as well as the non-speech controls. The results were in complete contradiction with both of the assumptions. English subjects regularized the French stimuli as much as they did the English. French subjects regularized the French stimuli even more than the English ones. Both groups regularized noise bursts significantly less. The prediction that English subjects might regularize more because of some acquired tendency to perceive language as regular was thus not met. The two groups did differ, however, in a rather unexpected way. The French subjects regularized significantly more. Also, the French subjects tapped at a slower rate (although both groups overestimated the gaps between the target syllables). How these differences are to be interpreted is not clear. The authors suggest that French subjects simply found the task more unnatural. If this is the case, it still remains to be explained why this should be so. The authors compare their results with those obtained by Darwin and Donovan. They conclude that *"It is clear from the results presented here that Donovan and Darwin's results do not represent evidence for any more perceptual isochrony in English than in French, and as such do not constitute support to the claim that English has an underlying isochronous rhythm (unless French is claimed to have one as well)"* (p. 160). What is not clear, however, is why listeners regularize their taps to speech stimuli significantly more than they do to non-speech stimuli. The authors propose two alternative explanations. Either the phenomenon is language bound, but not to any specific language, or it is the differing complexities of the tasks that makes the difference—*"The speech stimuli are acoustically more complex than the corresponding sequences of noise-bursts, and the memory load in remembering a sentence is greater than that in remembering four noise bursts."* (p. 160). In a second experiment, the authors try to decide between these two hypotheses. In this experiment, the five English sentences and matched noise-burst sequences from the first experiment were used. In addition they made non-speech, but speech-like, stimuli by distorting the English utterances beyond intelligibility. The utterances retained some of their prosodic qualities, however, but *"outside the target syllables, segmental information was severely degraded"* (p. 161). Nine English subjects took part in this experiment. In their responses, they regularized the normal utterances, as predicted, but also the distorted speech-like ones. They did not regularize the noise-burst sequences. Their responses to normal speech did not differ significantly from their responses to the degraded speech. The results of the experiments

are perhaps best summarized in the authors' own words: *"The results of this experiment show that the phenomenon of regularization is not even specific to speech, but extends to other unintelligible noises with some speech-like properties. They raise the possibility that the subjects are not actually doing anything very interesting at all—that they are simply exhibiting a response bias toward evenly spaced taps when the task becomes difficult."* (p. 161)

The idea that we hear speech as more regular than it physically is is an interesting thought. If it were true it could help explain why so many linguists have believed in the isochrony hypothesis, and still do. The effect should be particularly noticeable when we listen to the so called stress-timed languages. But as the experiments performed by Scott, Isard, and Boysson-Bardies show, this does not seem to be the case. The effect does not even seem to be limited to speech but appears in tapping to other complex stimuli as well. The explanation offered is that the complexity of the stimulus makes the rhythm difficult to perceive and that subjects, uncertain about the precise rhythm they are supposed to tap to, respond in a generally regular way. The fact that some subjects (in Bell and Fowler's study) are able to reproduce the rhythms correctly while others regularize is also an indication in that direction. This hypothesis is in good agreement with the results from the experiments on general rhythm perception and production. As the investigations I discussed in Chapter 1 show, we have a spontaneous tendency to rhythmically regular behaviour. When the stimulus becomes so difficult to follow that we become uncertain about what it is that we are supposed to follow it seems natural that we resort to a general regular tapping.

In the experiments carried out by Scott, Isard, and Boysson-Bardies, it was observed that subjects tended to respond with longer inter-tap intervals than the interstress intervals of the corresponding stimuli. They give no figures for the respective durations, so we do not know what this means in absolute terms. Donovan and Darwin do not report any figures either but they present their results in diagrams. These diagrams are very crude if one wants to say something about the exact figures, but some observations may be made. As far as I am able to see, the subjects in most, but not all, of these experiments show the same tendency to tap or match in a slower tempo. Scott *et al.* proposed that the reason might be that subjects found the task unnatural and therefore behaved in this, perhaps, hesitant way. Looking at the diagrams in the paper by Donovan and Darwin, another, in my view more interesting, possibility comes to mind. It seems, at least visually, as if there might be some centralizing tendency, particularly in the case where subjects regularize the most. The responses, moreover, seem to be taps or matching with intervals in the order of 500 ms. Now, as I reported in section 1.5 this is precisely the kind of interval duration found in experiments on personal tempo. This offers a different kind of explanation as to why subjects responses show different (but not necessary always longer) interval durations. If they are not able to perceive the rhythm accurately enough to tap veridically, they tend to tap, more or less regularly, in their own personal tempo. Now these ideas are just speculations based upon very impressionistic evidence, but the nice thing about them is

that they can be tested. If I am right, then subjects' responses to complex stimuli should show a tendency towards intervals determined by their personal tempo when the task becomes sufficiently difficult.

So, is the regularization effect found in experiments of this type then an artefact of the experimental situation as Scott *et al.* indicate? Well, it seems clear to me that there is no solid evidence to support the view that we perceive speech as being more regular than it really is, other than, perhaps, in the very general sense that we interpret as regular that which we cannot decide whether it is or not. In particular, these experiments provide no basis for claiming that the so called stress-timed languages possess any particular rhythmic properties that make them seem more regular.

3.6 What does it mean to study Japanese speech rhythm?

For technical reasons, nonsense syllables are used in many phonetic investigations. By doing so one is able to reduce variation and control perceptual stimuli or production more closely. By choosing an appropriate syllable structure, analysis of the results, particularly the acoustic analysis, is often facilitated. This is fine in many ways, but when trying to compare perception of the rhythms of different languages this technique presents a problem. I will use an investigation into speech rhythm and p-centres by Hoequist (1983c) to illustrate my point.

Hoequist wanted to be able to compare the speech rhythms of three different languages, English, Spanish, and Japanese. To do so, he used the same technique that had been used in the experiments on p-centres by Marcus, Fowler and others described above. Subjects were told to produce isochronous strings of syllables under two conditions—speaking to the clicks of a metronome and speaking without any external time keeper. The syllables used were *a, ma, ba, pa,* and *sa.* The sequences to be spoken were alternations between the following pairs: *a–ba, ma–ba,* and *pa–sa.* Subjects in the experiment were native speakers of English (4), Spanish (6), and Japanese (4). The subjects were students at Yale University. The Spanish and Japanese subjects came from English-language classes at the university. The results of the experiments agreed basically with those of other experiments (e.g. Morton, Marcus, and Frankish 1976; Fowler, 1979). I have some methodological caveats concerning this study but that is of no importance here. The important question is instead what the results obtained in this experiment tell us about language differences. I will summarize the experiment in Hoequist's own words:

The experiment described here uses rhythmically produced strings of nonsense syllables to investigate whether the P-centre effect behaves the same in English (a stress-timed language) as in two non-stress-timed languages, Spanish (syllable-timed) and Japanese (mora-timed). (p. 370)

The cross-language similarities are obvious. The P-centre effect is present in all three languages investigated, and apparently present to a similar degree. ... The P-centre effect, which had previously been demonstrated only for English (a stress-timed language), is shown to be present in syllable- and mora-timed languages as well. (p. 375)

Is it now? I must admit that it is not entirely clear to me how the sequence of nonsense syllables *ma–ba* becomes a Japanese utterance. Nor is it obvious that the same utterance produced by an American is English. What these results tell us may be relevant if one wants to study some aspect of articulation in general. It is an interesting finding in its own right that, when asked to produce nonsense syllables, in this particular context, native speakers of different languages seem to behave the same way. But does it really tell us anything about *Japanese speech rhythm*? Hoequist is not the only one using nonsense syllables in the experiments and trying to generalize from the results to a specific language. I have only used his particular paper as an example because I wanted to dramatize the problem a bit by comparing *ma–ba* uttered to a metronome with conversational Japanese. But the problem is there in all similar investigations. What does an experiment using nonsense syllables tell us about the characteristics of a particular language? If it is language specific properties we are after, it would seem reasonable to use samples from the specific languages as the material. Even if we want to say something about language universals, it still seems necessary to be able to describe the different languages before one can tell whether they are similar or not.

3.7 Can temporal regularity be measured?

A great deal of speech rhythm research has been concerned with regularity. Interstress intervals in speech production have been claimed to be regular, or at least *roughly* so. Some attempts have been made to talk about regularity in more well defined terms. I think the concept of relative durations may be seen as such an attempt, although, in my opinion, perhaps not a wholly successful one. Another attempt to compare interstress interval regularity between languages was the use of the arithmetic means and standard deviations in the studies by Roach (1982), and Dauer (1983). But, as we saw, the attempt failed to show any significant differences. Now, there is nothing technically wrong with these measures. They are well enough defined. But it is also the case that, when using a measure like the relative durations or arithmetic mean, a lot of information that may be relevant in the context of rhythm research, even from the point of view of regularity, is lost. An few examples may clarify what I mean.

One way of describing a set of successive intervals is by stating their sizes and their order. What measure we use to describe their sizes is irrelevant in this context as long as it is well defined. Suppose, for the sake of argument then, that we describe one sequence as 11:6:1 and another 7:6:5 in some arbitrarily chosen, but well defined, measure. Now, using the arithmetic mean on these figures we may see that it does not differentiate between the two

sequences. The mean will be 6 in both cases. From this point of view both sequences are equally regular. One may very well accept this of course but I think most people would feel slightly uneasy about such a decision. At least from an intuitive point of view one would like to say that the first sequence is more irregular. How about using the standard deviation as a measure? Well, obviously that will correspond to intuition in a better way assigning the value 5 to the first sequence and 1 to the second. So, perhaps the standard deviation is the measure to use then. But what about the sequence 70:60:50? Is that sequence really so much more irregular than 7:6:5? The standard deviation is certainly greater. Ten times greater in fact.

What the little exercise above is meant to demonstrate is only that it is very easy to construct a measure by which one may measure the regularity of a given sequence. The problem is to do so in such a way that the results we obtain when we use the measure correspond in any reasonable way to our intuitions about what such a measure should tell us. Ideally we would like to be able to measure regularity in such a way that the results correspond to our intuitions about the physical world in the same way, and to the same degree, that our use of centimetres to measure length does, or the use of milliseconds to measure time as in most of the studies discussed above.

One might ask if one should not also think of factors of perceptual relevance so that the results of applying the measure correspond to perceived regularity as closely as possible. Perhaps, but I think this is the wrong way to go, just as we do not allow time perception to have any influence on how we measure clock time. This does not preclude the use of clock time in experiments on time perception. On the contrary, the existence of a physical measure makes it possible to talk about perceived time in much more precise terms, comparing the time intervals we may measure with the corresponding perceived durations.

In the following, I will briefly discuss some properties I think may be considered as desirable properties of a regularity measure. I will also describe two attempts that have been made to construct such measures, and finally report three of the comparatively few studies there are on the perception of regularity.

I would like to propose that a measure of regularity should fulfil the following requirements:

1) The measure should define an 'absolute zero'.

2) The measure should be insensitive to order.

3) The measure should be insensitive to the absolute sizes of intervals.

4) The measure should be independent of the length of the sequence.

The meaning of requirement 1) is very simple. It means that the measure should define an absolutely regular sequence. What characterizes such a sequence may not be totally uncontroversial, but I can find no really significant objection to saying that a sequence of

perfectly equal intervals is also perfectly regular. The measures I am going to discuss below assign the value 0 to a perfectly regular sequence and higher values for irregular sequences. This is not a necessary choice, but it does have some appeal.

Insensitivity to order is perhaps more controversial. Should the sequence 1:2:3 be regarded as equally regular as 3:2:1? In my opinion it should. My reasons are again negative rather than positive. I can find no really strong arguments against the idea. If one regards the measure as a measure of how the sizes of the intervals compare with each other their respective sizes should be the relevant comparison but hardly the order.

The third requirement has to do with structure. I would like to think of a regularity measure as saying something about the structure of intervals. In this respect the intervals mentioned above, 7:6:5 and 70:60:50, should be regarded as equally regular. The *relative* sizes of intervals should be regarded rather than the absolute ones.

The last point is perhaps the easiest to accept but technically it is probably the most difficult to fulfil. At any rate, none of the measures discussed below does and I have no suggestion how it might be done in a reasonable way.

The most important quality of any measure seems to me to be that the regularities of different sequences may be compared. What this means with respect to the measure is that it should be possible to use the measure to order sequences with respect to regularity. There seems to be little point, on the other hand, in trying to obtain some absolute measure of the size of a certain regularity.

The ideas I have put forward here do not seem to be altogether original. The two attempts I have found that address this question are both more or less constructed along these lines. Below, I will review and discuss these attempts.

In the two studies of speech rhythm perception by Darwin and Donovan discussed above (3.5) subjects were asked to match, subjectively, a sequence of noise pulses to the stressed syllables in a test sentence in one experiment and tap their fingers to the test sentences in another. The durations of the intervals between the pulses or taps were then compared to the durations of interstress intervals in the sentence. What Darwin and Donovan wanted to know was whether the tapping or placing of noise pulses would be more or less regular than the sequence of stresses in the stimulus sentence. In order to determine that they constructed a test variable 't'.

$$t = \sum_{i=1}^{n-1} \left\{ \left| 1 - \frac{a_i}{a_{i+1}} \right| - \left| 1 - \frac{p_i}{p_{i+1}} \right| \right\}$$

In the formula, the a:s are the durations of interstress intervals in the test sentence and the p:s the corresponding durations derived from the sequences of noise pulses or finger-tapping. (This measure is, strictly speaking, a measure of regularity difference rather than

regularity. The measure is composed of the sum of differences between a regularity measure for actual and measured durations. By setting all perceived durations to 1 the measure can be transformed into a direct regularity measure. This distinction is, however, of no importance here.) The test variable provides a means of testing regularity but it violates the requirement of insensitivity to order that I suggested above as a desirable property. If compared to a perfectly regular sequence, the sequence 1:2:3 gets the score 0.833 while the sequence 3:2:1 gets the score 1.500. As I mentioned above, I find it difficult to motivate why they should be regarded as unequal and there is no indication that Darwin and Donovan are aware of this. There is no discussion of these questions in their paper.

A test variable that is not sensitive to order has been suggested by Scott, Isard, and Boysson-Bardies (1985). The test variable that they propose is:

$$t = \sum_{1 \leq i < j \leq n} \left| \ln \frac{d_i}{d_j} \right|$$

The variable meets the first three of the requirements suggested above. A perfectly regular sequence, say 1:1:1, receives the score 0. The measure is insensitive to absolute durations. The sequences 1:2:3 and 6:12:18 receive the same scores (2.197). It is also insensitive to order. The above mentioned sequences, 1:2:3 and 3:2:1 both receive the same score, 2.197.

This attempt at operationalizing regularity has been criticized by Benguerel (1986). His criticism can be summarized as follows. 1) If sequences of three durations are tested, the score does not depend on the middle value but only the two extremes. 2) By using a logarithmic measure, small differences tend to be overrated and large ones underrated. 3) The measure yields to high irregularity values. Beginning with the last point, it can be easily dismissed on the ground that it does not seem possible at present to say anything about what would count as a large or a small score in absolute terms. Moreover the scores are meant to make possible the comparison of different sequences but not to assign absolute values to regularity.

The two other points carry more weight. In a reply to Benguerel, Scott, Isard, and Boysson-Bardies (1986) maintain that their proposed measure is a sound one. It meets what they regard as three important criteria. It is symmetric (that is insensitive to order), it is generalizable to sequences of arbitrary length and it is insensitive to absolute durations. They admit that in the case of three intervals, the intermediate value plays no role, but do not regard this as a disadvantage. A consequence of the property is, for example, that the sequences 1:1.5:3 and 1:2.5:3 receive the same scores. They say that they can find no reason why one would regard one of these types as more regular than the other. One tends to agree with this view. In defense of their use of logarithms, they claim that it is necessary in order to meet the requirement that the measure should be insensitive to absolute durations. This is not correct, however. Any measure using relative durations would have this property. The use of logarithms is important in order to meet another one of their criteria, however,

the independence of order. Without the logarithms the variable would be sensitive to order in the same way that the variable used by Darwin and Donovan is. It is true, as Benguerel has observed, that using logarithms means overrating small differences and underrating large ones, if by that one means that small deviations around some mean contribute relatively more to the score than large ones would. It is difficult, however, to decide whether this is really a undesirable property.

Another critical objection one might raise, though, is that the measure does not make any sense if one wants to compare two sequences containing different numbers of intervals. It is true, as the authors claim, that the measure is generalizable to sequences of arbitrary length, but the scores obtained from two sequences of different lengths are not comparable in any obvious way. The sequences 1:2:3 and 1:2:3:4 would yield different scores, the four-interval sequence yielding the higher one. But how is this difference to be interpreted? One must find some way of normalizing for length so that two sequences of different lengths can be said to be equal in regularity, but it is not obvious by what criterion this normalization should be made. Using the measure in its present form, it only makes sense to compare sequences of equal lengths.

In spite of the weak points discussed above, I would still argue that the measure proposed by Scott *et al.* is a useful one, and the best one presently available. I have decided to use this variable to test for differences in regularity in my own study (Chapters 4 and 6).

Not a lot is known, it seems, about the perception of regularity. Would what may reasonably be described as a regular sequence in a physical sense also be perceived as such? And is regularity in perception really independent of order? In the following, I will briefly review some results which may give an indication of which direction the answers to such questions may take.

With respect to the question of perceptually regular sequences, one may hypothesize that such phenomena as final lengthening may come in and influence perception to some degree. Whereas the sequence 1:1:1 may be regarded as a physically regular sequence, a slightly modified sequence like 1:1:1.1 might perhaps *sound* more regular. There exists in fact at least one study that indicates that expectations of final lengthening might have precisely this kind of influence on what is perceived as a regular sequence. Benguerel and D'Arcy (1986) used accelerating, isochronous and decelerating sequences of clicks or syllables as stimuli in a series of experiments aimed at studying under what circumstances subjects perceived the sequences as regular. The results showed that subjects judged sequences as regular over a range of 'time-warping' conditions but that there was a slight bias in favour of decelerating sequences in the sense that median values for the sequences judged as regular corresponded to decelerating sequences. Three groups of subjects with three different native languages, English, Japanese, and French participated in the experiments. An interesting outcome of the experiments was that there seemed to be no influence of native language on the perception of what is to be considered a regular sequence.

An interesting variable in connection with the perception of regularity is the size of the difference limen for perception of irregularity. In the experiments by Benguerel and D'Arcy reported above, it seemed as if the deviations that created a perceptible irregularity were of the same order as those found as just noticeable differences in duration discrimination experiments. Other studies confirm this view. Hirsh, Monahan, Grant, and Singh (1990), and Monahan and Hirsh (1990) in studies of the perception of rhythmic auditory patterns found perceptual thresholds of the same order as those in other duration discrimination experiments (Weber fractions of .06—.08 for a base duration of 200 ms). For short inter-tone intervals, however, there seemed to be position effects as well. It thus seems as if the perceptual thresholds found in duration discrimination experiments may serve as a good first approximation for what may be perceived as a deviation from regularity in rhythm perception as well. (A discussion of duration perception may be found in Chapter 5.)

Another question one might ask is whether perception of regularity is independent of order. It seems reasonable, from a technical point of view, that there should be no difference in regularity between the two sequences 1:2:3 and 3:2:1. But again, this is not necessarily true for perception. As was mentioned above, Benguerel and D'Arcy found a slight bias in favour of decelerating sequences. Effects of that type may result in the sequence 1:2:3 being perceived as less irregular than 3:2:1.

And finally, from a technical point of view, it seems reasonable that only the relative durations of intervals should be considered. Thus the sequences 1:2:3 and 10:20:30 should be considered to show the same degree of regularity. But this is not necessarily true for perception. It may be the case, for example, that when durations approach the limits of the 'psychological present' sensitivity to durational differences changes. If perception of regularity depends on duration perception in the same way as in duration discrimination tasks, then the prediction would be that sensitivity to irregularity would be less for longer durations (absolute values).

What I hope to have shown with these examples is that the question of how to quantify regularity is highly relevant for the study of speech rhythm in both production and perception.

Part II

Temporal regularity in speech production.

Chapter 4

A study of prose reading.

This part of the study is concerned with speech rhythm in speech production. It was pointed out in the discussion in Part I that speech production may be studied at different levels (e.g. neuro-muscular, speech signal) and that one cannot rule out the possibility of arriving at somewhat different results, for example with respect to regularity, depending on which aspect of speech production one studies. There is no reason in principle to favour one aspect before another, all must be given careful consideration. But for practical reasons, if for nothing else, it is not possible to cover all possible aspects in every study. One has to limit the scope one way or the other. In the study presented below, I have limited the scope to the study of temporal phenomena as they are reflected in the speech signal. Now, although there is certainly a close connection between what goes on in articulation and the acoustic result of these processes there is not necessarily any perfect isomorphism. Without, however, in any way playing down the importance of a thorough discussion of such matters I will not make them an issue here. I will simply study properties of the speech signal and relate the results of my own analysis to those obtained by others adopting a similar approach. Another decision one must make is about the kind of speech material to be used in the study. The data base on which this study is built is a corpus of read sentences. The particular speech style that the analysis is concerned with is thus reading aloud. This is of course a limitation with respect to the generality of the results. To what extent the results may be generalized to other types of speech, for example spontaneous speech, is unclear. But the choice of speech style also has its advantages. It brings under control a number of factors which are difficult or impossible to control in spontaneous speech. In this study,

one such factor is to have identical material, from several speakers so that individual differences and group differences may be examined. Another obvious advantage is that by using this type of material the results become comparable to the results from other studies, the vast majority of which have been done on material of this type. Having said this, however, I also want to say that now that there are results from quite a number of studies of this type of speech it seems urgent to turn our attention also to other types, particularly spontaneous speech.

The study presented here also forms a background to the perception experiments presented in Part III. To be able to draw any conclusions about the possibility of detecting temporal irregularities in speech one must have a clear idea of the size of these irregularities.

4.1 Aim of the study.

The experimental study presented below is a detailed study of durations of interstress intervals and syllables in a small corpus of recorded sentences, with particular emphasis on the questions of variation in interval duration, dependency of interval duration on interval length, and compression of syllables as a function of interval length. The aim is to suggest models by which to describe durations at interstress interval level and to test the linear model proposed in 3.2. The variation of intervals is also studied as a background to the study of stress beat and duration perception presented in Chapters 6 and 7.

The first question to be examined is variation. As was shown in 2.3.1 and 3.2, it is quite clear that interstress intervals are not even 'roughly equal'. But to show this is only a first step in a description of interstress interval variation. One would like to be able to describe the variation in more detail. What are typical interstress interval durations, and what is the range of the variation one is to expect? These questions are of course relevant to speech rhythm perception which will be examined in Part III of the study. But they are also relevant in connection with speech production, particularly in the light of findings by others (Dauer, 1983; Roach, 1982) which suggest that languages may not differ significantly with respect to mean interval durations or variation in interval duration. More data will help to clarify whether these findings have a more universal status.

It was mentioned in 2.3.1, and further elaborated in 3.2, that a model predicting a linear increase in interstress interval duration as a function of the number of syllables seems to be a very good first approximation. The hypothesis that intervals increase linearly will be tested on the material analysed here. But I will also study interval duration as a function of the number of phonemic segments to see if a similar model holds for segments as well. Data from two other studies of Swedish (Fant, Kruckenberg, and Nord, 1989; Fant and Kruckenberg, 1989) indicate that that may be the case.

The third main question addressed in this study is that of compression of syllables. This question is closely connected with the discussion of whether there are any tendencies to

regularity, particularly in the supposedly stress-timed languages. Compression has been regarded as important evidence for such a tendency. But, as was shown in 2.3.1 and 3.3, the question is far from simple and the results obtained so far are not unambiguous. In this study, I will attempt to see if any such tendencies may be shown to exist in the material used here. But the scope may also be widened a bit. How syllables combine to make up interstress intervals is interesting in its own right regardless of whether there is compression or not. In connection with studying the question of possible compression tendencies, the temporal organization of interstress intervals will be studied also from other points of view. Phrase-final lengthening will be considered but also the possibility of a similar effect in interval final position.

In addition, a few other points will be touched upon briefly. Since the data base includes material from speakers of both sexes as well different age groups, the influence of these parameters will be looked at in some contexts. Studies involving accurate segmentation of the speech signal are very time consuming. One would, therefore, like to be able to limit the amount of data that needs to be analysed in some way. By making comparisons between the groups, it was hoped to be able to say something about how critical such variables as sex and age are if one wants to generalize the results. But also the size of the groups has some methodological interest. Given the considerable individual variation, clearly seen in the re-analysis of Dauer's data (3.2), one has to ask how many subjects must be included in a study of this kind if one wants the results to converge towards some kind of means that may reasonably be thought of as representative for a particular language.

Another factor which will be given some consideration is the use of one of the regularity measures described in 3.7. The measure will be applied to interstress interval data for the sentences used, in an attempt to see if any interesting information may be obtained by using such a measure.

4.2 Speech material used in the study.

The speech material used in this study comes from a corpus of sentences developed by Korsan-Bengtsen (1973) for use in audiometry tests. The corpus consists of 550 sentences. From this corpus a subset of 30 sentences was chosen. These sentences were used to prepare 5 lists of sentences with 10 sentences in each list. The first five sentences in each list were identical. Altogether 300 sentences were recorded with 30 speakers. Each speaker read one of the lists. The original purpose of the recordings was to use them as test material in a comparative evaluation of different speech coding techniques. From the recorded material, the subset consisting of the five sentences included in all the lists was chosen for the analysis done below.

All sentences are such that they prompt a reading with four main stresses. That is, each sentence should normally contain four interstress intervals. This is also the case in the recordings used here. The number of syllables in the intervals varies between 2 and 4. Each

sentence was read by 30 subjects, 10 children, 10 male speakers, and 10 female speakers. One sentence, read by one of the children, was omitted however, for technical reasons. The data on which this study is based is thus drawn from 149 recorded sentences, consisting of altogether 596 interstress intervals; 447 intervals in non phrase-final position and 149 phrase final intervals. The total number of syllables is 1220 syllables in non phrase-final intervals and 328 syllables in final intervals.

The five sentences used in the study are given below together with phonetic transcriptions and a translation:

1. *Torpet hade blommor och gräs på taket*
   t'ɔrpəthadəbl'ɷm,ɷrɔgr'ɛ:spɔt'ɑ:kət
   The cabin had flowers and grass on the roof
2. *Många trivs med att vandra i fjällen*
   m'ɔŋ,atr'i:vsmedatv'andraɪfj'ɛl,ɛn
   Many enjoy hiking in the mountains
3. *Isen kan omöjligt bära en vuxen*
   'i:senkan'ɷ:m,øjlɪtb'æ:ranv'ɵks,ən
   The ice cannot possibly support an adult
4. *Sikten är ganska skymd i kurvan*
   s'ɪktəneg'ansk,afj'ymdɪk'ərv,an
   Visibility is rather bad in the curve
5. *Bussens förare fick körkortet indraget*
   b'ɵsənsf'œ:rarefɪkç'œk,oʈət'ɪndr,ɑ:gət
   The bus driver was deprived of his driver's license

4.3 Subjects.

Subjects in this study were 20 adults, 10 male and 10 female, and 10 children. The adult speakers were employees at the Swedish Telecom head office in Farsta. Their ages ranged between 20 and 62 years of age. The median age was 38 years in both groups. The children, 4 boys and 6 girls, were fifth graders from a local primary school. They were 10 or 11 years old (median age 11). None of the speakers had any known speech disorders.

4.4 Recording.

The recordings were made in an anechoic room at the Swedish Telecom acoustics laboratory in Farsta using high quality analogue recording equipment (Brüel & Kjær 4145 microphones and a Telefunken M 12, open reel tape recorder). Subjects were given a script

with the written sentences. They were instructed to read them as naturally as possible ("as if they were talking to someone").

4.5 Analysis.

Analysis was done at the Swedish Telecom laboratories in Älvsjö using Symbolics computers and the Spire signal analysis system. The recorded material was sampled into the computer and each sentence was placed in a separate file. The Spire system, originally developed at M.I.T. for signal processing and analysis, was then used to analyse the recordings. The Spire system contains a wealth of features which permit very precise labelling of segments in the speech files. Figure 4.1 shows a typical layout used in the transcription process.

It is not always easy to determine exactly where the segment boundaries should be. In this investigation, however, the important time points were considered to be the onsets of the vowels. The reason for this decision is the role these onsets have been shown to play in the perception of speech rhythm (see 3.4). Other authors have made the same decision (e.g. Dauer. 1983; Strangert, 1985), but as was shown in 2.3.1 it is by no means the only possible choice. To determine the onsets did not prove very problematic.

The speech files with transcriptions could be used for automatic computation of the durations of the segments defined by the transcriptions. A special program was also written for automatic computation of syllable durations, defined as the distance from vowel onset to vowel onset. (The program was written by Jaan Kaja, Swedish Telecom Laboratories, Älvsjö.)

A more difficult question arose in connection with determining how many segments were contained in a particular syllable or interstress interval. Should one count the number of segments as the number of segments that 'ought to' be there according to the text or by some other criterion? Since this study means to say something about timing in production (including articulation, to the extent that the speech signal reflects articulation) it was decided to count as segments only those segments that could be reliably identified in the speech signal. The rationale for this decision was the belief that timing should be most closely connected with what is actually produced in the articulation process rather than what may be underlying the process at some other level of speech production. But this is of course a controversial question, and open to debate.

Determining if a syllable is to be counted as stressed or not is not always easy. Sometimes, particularly in the case of spontaneous speech, it is often not possible to reach any decision at all. But in this case, due to the nature of the sentences and the task, marking stresses did not present a problem. The sentences are constructed in such a way that they prompt a reading style where readers stress the sentences uniformly and with easily identifiable stresses. Stressed syllables were marked in the transcriptions while listening to the

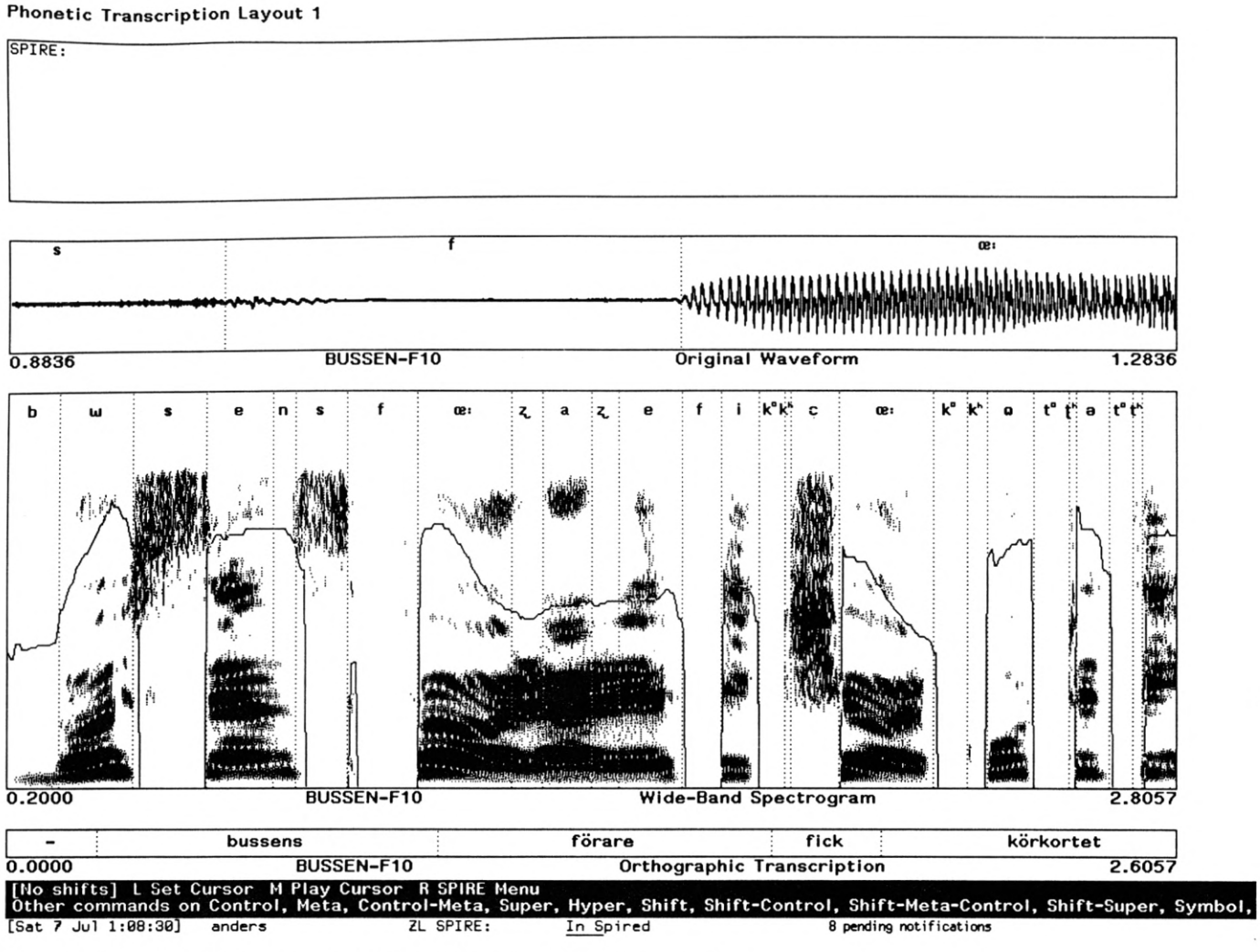

Figure 4.1. A typical SPIRE-layout used for the transcription of the sentences studied.

recordings. Marking the stressed syllables was done by the author and two co-workers engaged in a speech recognition project. There was no disagreement about which syllables that were stressed.

4.6 Results.

The first half of the analysis of the results will be concerned with interstress interval durations and the relation between these durations and the number of syllables or phonemic segments in the interval. The second half will deal with interval-internal processes. An attempt will be made to describe the temporal structure of the interstress intervals as a function of its syllable components. In this context, the question of syllable compression will also be examined as well as processes like phrase-final lengthening, and possible lengthening effects on syllables in final positions of intervals which are not phrase-final.

It was decided to analyse phrase-final lengthening separately. The analysis therefore begins with data from the first three intervals in each sentence. These results are then compared with data from the final intervals to be able to determine the magnitude of any final-lengthening effect.

4.6.1 Regularity.

The first factor to be analysed is interstress interval duration, in terms of mean interval duration, standard deviation, and range. The results are summarized in Table 4.1. Mean interstress interval duration for all subjects is 580 ms, and mean durations are very similar for the three groups of speakers (596 ms, 571 ms, and 573 ms) . An analysis of variance of the durations by group reveals no significant difference between groups. Neither sex nor age seems to be a differentiating factor. If an analysis of variance using the mean values is made over individuals, however, there are significant differences between subjects within all three groups ($P < .001$). Figure 4.2 shows the distribution of durations in the form of a histogram. As can be seen in the diagram, the distribution is unimodal but slightly skewed (skewness = .870).

Standard deviation and range are measures of the variation in interval durations. The range is perhaps the most relevant measure of variation since it gives an idea of what kind of durational contrasts one may find between intervals in a phrase. As can be expected, the two measures are highly correlated ($r = .929$). The distribution of ranges is shown in Figure 4.3. Their distribution also seems to be unimodal with a very slight skewness (.593).

It is clear from the values in Table 4.1, and Figures 4.2 and 4.3, that the variation in interstress interval duration may be considerable. Subject F5 shows the lowest variation with a range 203 ms. As will be shown in Chapter 7, deviations of this order (compared to a mean duration of 475 ms) are clearly detectable in perception. All other ranges are greater, with subject M2 forming the upper limit with a range of 545 ms.

Table 4.1. A summary of interstress interval duration data in ms for sentences 1 to 5. Only the first three intervals are considered. Phrase final intervals are excluded. 'Range' is the difference between the longest and the shortest interval (not necessarily adjacent). 'N' is the number of intervals on which the means and standard deviations are based.

Subject	Mean	Std Dev	Range	Min.	Max.	N
C 1	590.5	75.5	250	447	697	15
C 2	514.6	89.6	257	395	652	15
C 3	611.5	99.0	324	440	764	15
C 4	647.7	114.0	380	467	847	15
C 5	550.8	134.2	433	376	809	15
C 6	526.3	93.9	317	419	736	15
C 7	708.2	102.7	370	516	886	15
C 8	641.3	91.2	326	493	819	15
C 9	632.0	131.4	351	482	833	12
C10	539.8	94.4	325	438	763	15
All children	595.6	116.9	510	376	886	147
M 1	486.2	100.6	336	302	638	15
M 2	861.5	167.5	545	629	1174	15
M 3	455.0	69.2	261	359	620	15
M 4	512.3	91.4	320	405	725	15
M 5	527.9	87.2	325	370	695	15
M 6	584.6	125.0	374	396	770	15
M 7	547.0	134.6	375	386	761	15
M 8	695.2	118.8	414	513	927	15
M 9	519.5	100.6	328	381	709	15
M10	516.9	92.4	357	374	731	15
All males	570.6	158.4	872	302	1174	150
F 1	532.7	81.4	265	433	698	15
F 2	570.7	114.7	340	399	739	15
F 3	642.9	115.7	368	447	815	15
F 4	533.3	89.1	311	386	697	15
F 5	474.8	61.9	203	383	586	15
F 6	539.5	85.8	332	381	713	15
F 7	516.8	109.1	360	343	703	15
F 8	562.7	127.4	397	397	794	15
F 9	673.5	151.9	476	458	934	15
F10	685.7	148.9	470	472	942	15
All females	573.3	127.8	599	343	942	150
All subjects	579.7	135.8	872	302	1174	447

The results obtained here may be compared with those obtained in other studies. In the study of Swedish prose reading by Fant and Kruckenberg (1989), mentioned above, mean duration of intervals was 548 ms with a total range of 750 ms. These results compare well with those obtained here. The slightly lower mean value may be due to the fact that the material in Fant and Kruckenberg's study was a longer prose text while the material used here was isolated sentences, but it may also be a function of individual variation.

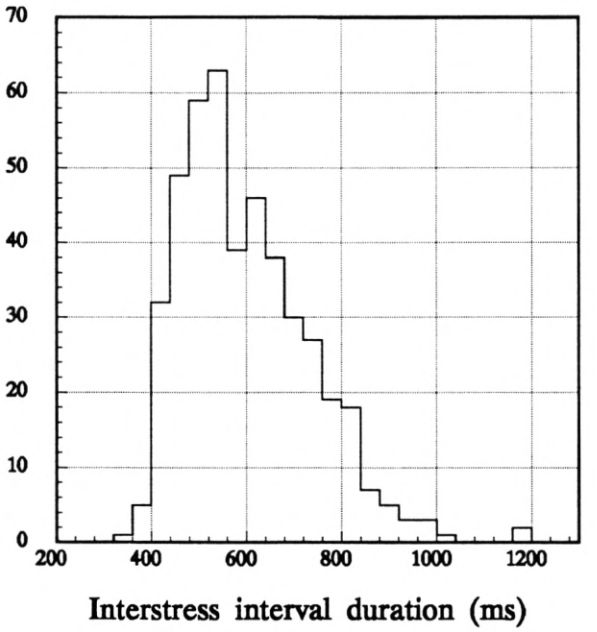

Figure 4.2. Histogram showing the distribution of interstress interval durations for all subjects and sentences. The distribution is built on a total of 447 interval durations.

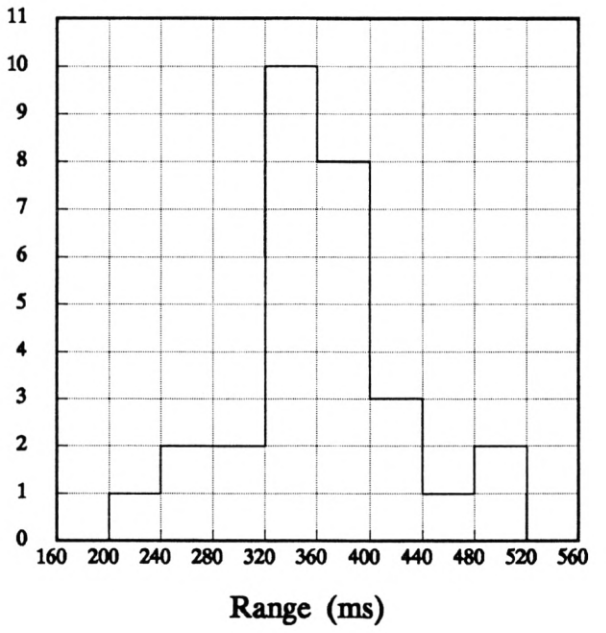

Figure 4.3. The distribution of ranges of interstress interval durations for all 30 subjects.

In studies of English, comparable but slightly lower mean values have been obtained. In the study by Bolinger, discussed in 2.3.1, the mean value is approximately 500 ms (calculated by myself using Bolinger's tables). And Faure, Hirst, and Chafcouloff (1980) obtained a mean value of 476 ms and standard deviations of around 200 ms for two speakers.

In the study by Dauer (1983) (see 2.3.3 and 3.2), where data from 5 languages were compared, mean ISI durations are very similar for the different languages, with 380 ms being the shortest (Thai) and 530 ms the longest (English). Most means are around 450 ms. The standard deviations vary between 131 ms and 176 ms. 150 ms may be regarded as a typical value.

If the results from these studies are compared with the results obtained above and by Fant and Kruckenberg (1989), one finds that the mean durations for Swedish seem to be somewhat higher. Whether these differences reflect differences between languages cannot be determined with any certainty, however, since the speech material used, as well as subject characteristics, may well be responsible for differences of that order.

A question one might ask is if and how the variation depends on speech rate. If there are any tendencies to isochrony one might hypothesize that the slower the speech rate the more room there will be for manipulations of durations to accomplish isochrony. How one is to show such a tendency is, however, far from clear. Range and standard deviations were used above to describe the variation in interstress interval duration. Now, considering range as a relevant measure of ISI variation it seems almost trivial to propose that range and speech rate should be negatively correlated. That is, the slower the speech the greater the variation, in absolute terms. This is indeed also the case. If speech rate is defined as syllable rate (computed as the inverse of mean syllable duration) then range and speech rate are significantly correlated ($r = -.702$). But this does not seem very interesting. What one would like to know is if the variation, in relative terms, varies with speech rate. This is particularly relevant if one also takes perceptual considerations into account since there is rather strong evidence that relative, rather than absolute, contrasts determine whether a difference is perceptible or not. (This will be discussed in detail in Chapter 5.) A measure of relative variation that one might then propose is range as a proportion of mean ISI duration. I have tried this idea on the material used in the study and the result may be seen in Figure 4.4.

As may be seen from the diagram there is no correlation at all between relative range and mean ISI duration ($r = .006$, $P = .97$). If syllable rate is used instead as a measure of speech rate the result is exactly the same. So if one accepts the concept of relative range as saying something about relative regularity there is no correlation between regularity and speech rate. In particular, speech does not seem to be relatively more regular in slower speech.

I have also tried to look at this question from a slightly different angle. The possibility of measuring regularity was discussed in section 3.7. Of the two measures discussed, I will use the one proposed by Scott. Isard, and Boysson-Bardies (1986) as the measure of

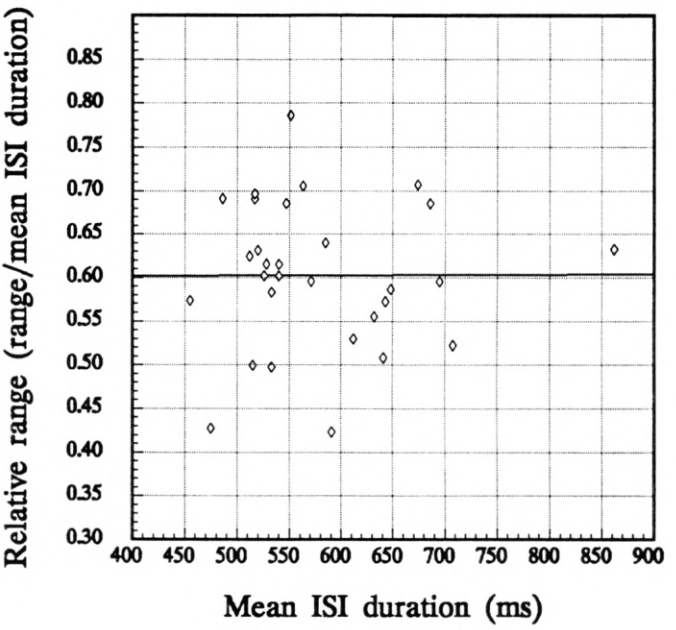

Figure 4.4. Relative range as a function of mean interstress interval duration. The line represents a linear regression analysis of the data.

regularity in the following analysis. Since the value of the measure is 0 for absolute equality of durations and greater than 0 for unequal durations, it might be thought of as a measure of irregularity (*i*-score).

$$i = \sum_{1 \le i < j \le n} \left| \ln \frac{d_i}{d_j} \right|$$

If this measure is applied to the data in this study, the first observation is that there are no significant differences between groups (ANOVA, 2 df, F = 1.553, P = .215). And the correlation between the irregularity score and the syllable rate (as defined above) is not significant (r = –.178, P = .35). Put in other words, this means that only about 3% of the variation in regularity, as expressed by the i-score, is explained by syllable rate. Regularity in this sense does not seem to depend on speech rate either.

The two measures of relative regularity tried above are neither conventionally agreed upon nor very well tested. One must, therefore, interpret the results with caution. But the results were included because they may have something relevant to say and at least they may provide some 'food for thought' with respect to a problem that should receive some further attention.

4.6.2 Interstress interval duration as a function of the number of syllables.

The analysis will now turn to questions of how interstress interval durations are conditioned by such factors as the number of syllables in the interval, the number of segments, position, and inter-individual differences.

It was shown in 3.2 that a linear model to describe how interstress interval durations depend on the number of syllables in the interval seems to fit available data very well. This is, at least, the case if mean values for durations are used. In this section, ISI durations will be analysed assuming a linear model for interval increase as a function of the number of syllables. Linear regression will be used as a measure of how well data agrees with the assumption.

Mean durations for the three groups as well as for all subjects pooled together are given in Table 4.2. As can be seen, the durations increase monotonically as a function of the number of syllables. The correlation between duration and the number of syllables is significant (Pearson, $r = .527$, $P < .001$).

Table 4.2. Mean ISI durations in ms as a function of the number of syllables.

Subjects	No.syll.	Mean	SD	N
Children				
	2	528.3	85.3	60
	3	620.0	107.3	72
	4	747.5	79.1	15
Male				
	2	497.9	121.6	50
	3	576.8	154.0	80
	4	727.7	141.0	20
Female				
	2	491.7	83.5	59
	3	595.2	109.5	73
	4	751.9	98.2	18
All				
	2	506.5	97.7	169
	3	596.6	127.2	225
	4	741.5	110.6	53

There are small differences in mean durations between the three groups but none of the differences are significant (ANOVA, P > .09). If the mean durations are used to calculate regression equations, the following equations result:

Children	I = 303.1 + 109.6*N	r = .996
Male	I = 256.1 + 114.9*N	r = .984
Female	I = 225.6 + 130.1*N	r = .993
All	I = 262.4 + 117.5*N	r = .991

The regression coefficients indicate that when average duration values are considered, the linearity is almost perfect. As always when one uses averages, there is some loss of information. In this case the durations of intervals with 2, 3, and 4 syllables are given the same weight regardless of the number of occurrences. It can be seen in Table 4.2 that 4-syllable intervals are less frequent. The influence of their durations may therefore be overestimated if mean durations are used. This may be particularly misleading if results from individual speakers are regarded, since there may in some cases be only one occurrence of a 4-syllable interval. In the following, regression equations have therefore been computed using the whole set of duration data, not averages.

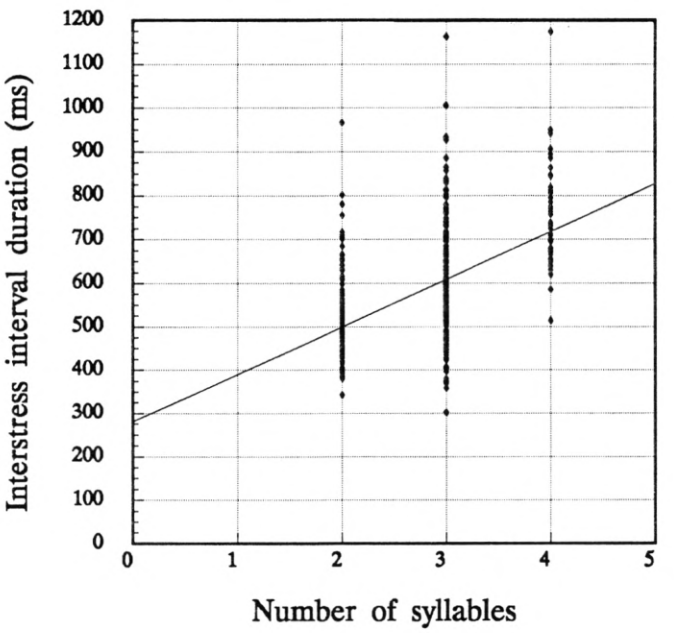

Figure 4.5. Interstress interval duration as a function of the number of syllables. The regression line is based on all the duration values.

Table 4.3. Linear regression equations based on the interstress interval durations of the first three intervals. ($P < .01$ for $r > .64$, $P < .05$ for $r > .53$, for individual subjects)

C 1	I = 398.2 + 70.4*N	r = .553
C 2	I = 298.4 + 79.1*N	r = .621
C 3	I = 301.6 + 113.4*N	r = .806
C 4	I = 248.4 + 149.8*N	r = .811
C 5	I = 269.3 + 105.5*N	r = .593
C 6	I = 223.8 + 110.7*N	r = .830
C 7	I = 448.6 + 97.4*N	r = .686
C 8	I = 360.3 + 102.8*N	r = .669
C 9	I = 294.7 + 130.6*N	r = .664
C10	I = 258.8 + 105.4*N	r = .689
All children	I = 317.6 + 103.2*N	r = .572
M 1	I = 240.5 + 87.8*N	r = .590
M 2	I = 436.0 + 152.0*N	r = .613
M 3	I = 254.2 + 71.7*N	r = .701
M 4	I = 221.4 + 103.9*N	r = .769
M 5	I = 270.8 + 91.8*N	r = .712
M 6	I = 173.9 + 146.7*N	r = .794
M 7	I = 217.6 + 117.7*N	r = .591
M 8	I = 467.9 + 81.2*N	r = .462
M 9	I = 222.6 + 106.0*N	r = .713
M10	I = 236.1 + 100.3*N	r = .734
All males	I = 274.1 + 105.9*N	r = .438
F 1	I = 320.0 + 76.0*N	r = .631
F 2	I = 221.1 + 127.9*N	r = .785
F 3	I = 273.3 + 132.0*N	r = .772
F 4	I = 283.5 + 93.7*N	r = .649
F 5	I = 287.3 + 70.3*N	r = .701
F 6	I = 269.5 + 101.3*N	r = .854
F 7	I = 204.1 + 114.4*N	r = .738
F 8	I = 162.9 + 150.0*N	r = .852
F 9	I = 278.2 + 141.2*N	r = .629
F10	I = 258.0 + 156.5*N	r = .740
All females	I = 240.6 + 122.0*N	r = .635
All subjects	I = 280.6 + 109.2*N	r = .527

Table 4.3 shows regression equations for individual subjects as well as for each of the three groups and all subjects pooled together. It may be seen that the equations for the groups, and for all subjects pooled, differ slightly from the ones based on mean values. The differences are not dramatic but should be noted. The reason why regression coefficients are lower is, of course, the variation in durations of intervals with a given number of syllables.

There is a considerable range of intercept values as well as slope values. The impression from the analysis of Dauer's (1983) data in 3.2, indicating that inter-individual variation may be considerable is confirmed. Intercept values vary between 163 and 468 ms and slope

values between 70 and 157 ms/syllable. The result for all subjects taken together is in very good agreement with the results obtained from data from the 'stress-timed' languages presented in Figure 3.1. Figure 4.5 shows the duration data and a regression line based on the whole set of data. The variation is obvious but also the dependency between durations and the number of syllables.

4.6.3 Interstress interval duration as a function of the number of phonemic segments.

In the previous section, it was shown that interstress interval durations depend significantly on the number of syllables in the interval (Pearson, $r = .527$, $P < .001$). The dependency on the number of phonemic segments is even stronger (Pearson, $r = .620$, $P < .001$). Obviously the number of syllables and the number of segments in an interstress interval are also significantly correlated (Pearson, $r = .844$, $P < .001$). In this section, I will analyse in more detail how durations depend on the number of segments.

As can be seen in Table 4.4, the number of segments in an interval is roughly proportional to the number of syllables. Since ISI duration is an approximately linear function of the number of syllables it is conceivable that it is also a linear function of the number of segments. (This may seem a trivial conclusion, but note that whereas the dependency follows from the mutual correlations it does *not* follow that the function must necessarily be linear.)

Mean ISI duration as a function of the number of segments is shown in Table 4.5. The following regression equations are based on the average values presented in Table 4.5.

Children	I = 320.4 + 42.1*N	r = .879
Male	I = 195.5 + 54.6*N	r = .942
Female	I = 265.1 + 45.1*N	r = .942
All	I = 211.0 + 53.2*N	r = .951

An inspection of the duration values in Table 4.5 shows that interstress interval durations increase monotonically as a function of the number of segments with the exception of the values for 10- and 11-segment intervals. These duration values all come from the first interval in sentence 1. It is, of course, not clear how representative they are of 10—11 segment intervals. To obtain better balanced regression equations, however, I have used regression on all intervals to obtain the equations in Table 4.6. This is particularly important when regression equations for individual subjects are computed where the 'rarer' types may only occur once. But, as can be seen in the table, the differences compared to the equations presented above are by no means dramatic.

Table 4.4. Mean number of segments in the interstress interval as a function of the number of syllables.

No.syll.	Mean	SD	Segm./syll.	N
2	5.41	.583	2.71	169
3	7.09	.887	2.36	225
4	9.83	1.105	2.46	53

Table 4.5. Mean durations of interstress intervals in ms as a function of the number of segments.

Subjects	No.segm.	Mean	SD	N
Children				
	5	484.9	63.4	41
	6	593.5	86.5	25
	7	600.5	86.1	43
	8	680.0	95.8	22
	9	783.7	75.3	7
	10	742.3	77.8	4
	11	718.0	85.2	5
Male				
	4	381.0	0.0	1
	5	462.0	94.0	40
	6	533.1	136.3	27
	7	575.6	145.0	41
	8	648.0	130.0	21
	9	784.8	166.5	10
	10	726.3	125.0	3
	11	729.1	107.9	7
Female				
	5	464.5	72.1	37
	6	549.9	83.5	34
	7	583.6	124.9	37
	8	622.4	95.0	22
	9	737.7	110.7	10
	10	673.0	0.0	1
	11	752.3	94.7	9
All				
	4	381.0	0.0	1
	5	470.7	77.7	118
	6	557.3	105.3	86
	7	586.9	120.0	121
	8	650.2	108.7	65
	9	767.1	125.2	27
	10	727.6	87.2	8
	11	736.4	93.5	21

Table 4.6. Linear regression equations based on interstress interval durations as a function of the number of segments. ($P < .01$ for $r > .64$, $P < .05$ for $r > .53$, for individual subjects)

C 1	I = 321.9 + 39.9*N	r = .786
C 2	I = 232.2 + 41.1*N	r = .773
C 3	I = 262.5 + 50.3*N	r = .869
C 4	I = 232.4 + 61.7*N	r = .879
C 5	I = 179.5 + 55.7*N	r = .713
C 6	I = 242.5 + 43.0*N	r = .790
C 7	I = 369.4 + 50.3*N	r = .704
C 8	I = 282.3 + 53.3*N	r = .869
C 9	I = 135.0 + 74.6*N	r = .850
C10	I = 176.4 + 54.5*N	r = .807
All children	I = 247.2 + 51.7*N	r = .680
M 1	I = 168.2 + 46.8*N	r = .770
M 2	I = 479.4 + 53.1*N	r = .564
M 3	I = 198.7 + 39.6*N	r = .806
M 4	I = 222.8 + 43.4*N	r = .776
M 5	I = 221.0 + 47.5*N	r = .793
M 6	I = 152.8 + 61.7*N	r = .875
M 7	I = 173.0 + 54.5*N	r = .664
M 8	I = 396.4 + 42.7*N	r = .592
M 9	I = 201.1 + 46.8*N	r = .848
M10	I = 171.6 + 52.3*N	r = .928
All males	I = 186.6 + 56.6*N	r = .577
F 1	I = 254.9 + 40.5*N	r = .816
F 2	I = 242.6 + 47.3*N	r = .722
F 3	I = 265.4 + 55.0*N	r = .780
F 4	I = 241.9 + 43.3*N	r = .811
F 5	I = 277.3 + 29.0*N	r = .777
F 6	I = 252.3 + 42.2*N	r = .836
F 7	I = 130.4 + 58.5*N	r = .806
F 8	I = 161.3 + 58.5*N	r = .811
F 9	I = 262.0 + 59.9*N	r = .647
F10	I = 238.8 + 65.1*N	r = .755
All females	I = 226.1 + 50.9*N	r = .646
All subjects	I = 220.6 + 53.0*N	r = .620

Some interesting observations can be made. First of all, it may be seen that the regression coefficients are greater than those for the regression on syllables presented in Table 4.3. The difference is significant (Pairwise t-test, $P < .05$). The number of segments in an interval is, thus, a better predictor of interval duration than the number of syllables. It may also be seen that the assumption that ISI duration is a linear function of the number of segments seems to be justified. This may not be particularly surprising, it is rather what one would expect, but it is interesting to see the assumption confirmed. The values for the slopes indicate how much an interstress interval increases, on average, per added segment.

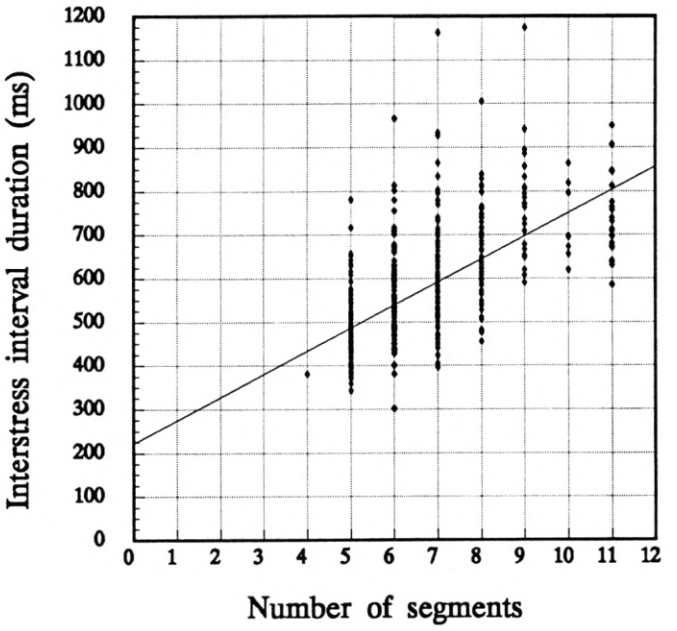

Figure 4.6. Interstress interval duration as a function of the number of segments. The regression line is based on all the duration values.

The average increment is 53 ms per segment for the whole group. Interestingly enough, this is exactly the same result as that obtained by Fant, Kruckenberg, and Nord (1989) in a study of a prose text read aloud. In Fant and Kruckenberg, (1989) the slope of the regression line based on data for five speakers is 55 ms/segment. An increase of around 55 ms per added segment thus seems to be a representative value for Swedish prose reading.

Another observation in the analysis of the dependency of durations on segments as well as syllables is that, although inter-individual differences are considerable, there are very small differences between the groups. It seems as if the values converge towards some mean provided the results from a sufficiently large group of subjects are pooled. This is encouraging, of course, with respect to the search for language specific properties.

In Figure 4.6, a plot of the pooled results and a regression line based on these results are shown. The monotonic increase as well as the considerable variation in individual interval durations is evident.

4.6.4. Syllable durations.

In the analysis of interstress interval durations above I have excluded the fourth, final, interval in each sentence from the analysis. The reason was that I did not want the possible occurrence of a phrase-final lengthening effect to interfere with the results, preferring to treat that question separately.

I will continue to postpone the question of phrase final lengthening for a while and only consider syllable data from the first three intervals in each sentence. This will make it possible to compare interval duration as the sum of syllable durations and with the linear regression analysis made in section 4.6.2.

Mean durations for syllables are presented in Table 4.7. As can be seen, the durations of stressed and unstressed syllables differ slightly between the groups but the differences are not significant (ANOVA, $P > .05$). With respect to the durations of stressed and unstressed syllables I will, therefore, in the following use data from all three groups pooled together. This becomes particularly important in the analysis of the finer details when the number of occurrences may be small.

ISI durations as a function of the number of syllables were expressed in section 4.6.2 as regression equations (see Table 4.3). The regression coefficients, using the average values, were close to 1, indicating that the function is linear to a close approximation. And even using the whole set of data, the deviation from linearity is quite small. We may thus assume

Table 4.7. Mean syllable durations and standard deviations in ms for all syllables in non phrase final interstress intervals.

	Mean	SD	N
Children	221.1	78.8	396
unstressed	184.9	67.3	249
stressed	282.5	55.3	147
Male	206.8	88.2	415
unstressed	171.3	76.8	265
stressed	269.5	70.4	150
Female	210.3	76.7	409
unstressed	175.9	63.2	259
stressed	269.7	59.8	150
All	212.6	81.6	1220
unstressed	177.2	69.6	773
stressed	273.8	62.4	447

that the model that predicts a linear increase in ISI duration as a function of the number of syllables is a fairly accurate one. But comparing the predictions by this linear model with the syllable duration values in Table 4.7, one realizes that something needs to be added to the model. If the regression model is interpreted as the addition of syllables of constant durations, long stressed and shorter unstressed ones, the prediction one would make based on the regression equation for the data from all subjects is that a stressed syllable is 389.8 ms on the average and an unstressed one 109.2 ms. But by looking at the duration values in table 4.7 one realizes that this is obviously not the case. Average durations are 273.8 ms for stressed syllables and 177.2 ms for unstressed ones (= 1.55:1, cf. 2.3.3).

Assuming a model based on average syllable durations which assumes stressed syllables with constant durations (273.8 ms) to which are added unstressed syllables, also of constant duration (177.2), the predicted ISI durations grow at a much faster rate than was found to be the case in the previous section. The interstress interval durations predicted by such a model would be 451 ms, 628 ms, and 805 ms, for 2- to 4-syllable intervals compared to measured averages of 506 ms, 600 ms, and 744 ms, or those predicted by the regression equation, 499 ms, 608 ms, and 717 ms. In the following analysis, this apparent discrepancy will be examined in some detail.

As is the case with interstress intervals, the number of segments must certainly be a strong factor in determining the duration of syllables. I will disregard this factor for a moment, however, and look at three other candidates that may influence syllable duration; stress, the number of syllables in the interstress interval (interval length), and syllable position within the interval. Table 4.8 shows an edited printout from a statistical analysis of syllable durations where these variables are separated. (In early versions of the analysis 'foot' was used as a synonym of 'interstress interval'. This is reflected in Table 4.8 and some of the following tables.)

In Table 4.9 the most important data from Table 4.8 is summarized. I have tried to reconstruct the internal composition of interstress intervals with varying number of syllables assuming the average values to be representative (data from phrase-final intervals is not included).

The values in Table 4.9 seem to indicate that syllable durations vary under all three conditions. Stressed syllables are longer than unstressed ones and they also seem to get successively shorter as a function of the number of syllables in the interval. Interval-final unstressed syllables are longer than medial unstressed syllables, and there may also be a tendency for them to be shorter in longer intervals, although this tendency is not unambiguous.

Now, regardless of whether these differences are significant or not and, if they are, what may be the explanation for it, they may help to explain the seeming contradiction between the increase in durations suggested by the regression equations and that suggested by mean syllable durations. As was mentioned above, the regression equations show that there is

Table 4.8. An edited output from the SPSS/PC+™ statistical package showing syllable durations as a function of stress, syllable position, and the number of syllables in the interval.

Summaries of	SYLLDUR	Syllable duration (ms)
By levels of	FOOTPOS	Position of foot in the phrase
	STRESS	Stressed/unstressed
	SYLLPOS	Position of syllable in the foot
	NOSYLL	Number of syllables in the foot

Variable	Value Label	Mean	Std Dev	Cases
For Entire Population	225.2	80.9	1548	
FOOTPOS	0 non phrase-final	212.6	81.6	1220
STRESS	1 stressed	273.8	62.4	447
SYLLPOS	0 non foot-final	273.8	62.4	447
NOSYLL	2	281.1	69.2	173
NOSYLL	3	272.2	59.5	222
NOSYLL	4	256.5	44.8	52
STRESS	0 unstressed	177.2	69.6	773
SYLLPOS	0 non foot-final	142.6	60.2	326
NOSYLL	3	141.2	65.7	222
NOSYLL	4	145.7	46.7	104
SYLLPOS	1 foot-final	202.4	65.0	447
NOSYLL	2	224.6	72.7	173
NOSYLL	3	186.6	59.4	222
NOSYLL	4	196.2	31.2	52
FOOTPOS	1 phrase final	272.2	57.7	328
STRESS	1 stressed	281.8	52.8	149
SYLLPOS	0 non foot-final	281.8	52.8	149
NOSYLL	2	276.4	54.8	119
NOSYLL	3	303.2	37.6	30
STRESS	0 unstressed	264.2	60.4	179
SYLLPOS	0 non foot-final	229.1	43.0	30
NOSYLL	3	229.1	43.0	30
SYLLPOS	1 foot-final	271.2	61.0	149
NOSYLL	2	260.0	58.3	119
NOSYLL	3	315.7	51.4	30

an increase in interval duration of approximately 110 ms per added syllable. This also agrees very well with the increase suggested by the mean ISI durations in Table 4.9 (120 ms the average). None of the syllable types in Table 4.9 is as short as that. Now, the explanation for this apparent paradox is fairly obvious from the values in Table 4.9, but the graphical presentation of the same basic facts presented in Figure 4.7 is perhaps a better illustration. It can be seen in the figure that it is primarily the reduced durations of the

Table 4.9. Mean syllable durations based on the values in Table 4.8. Phrase-final intervals are not included. 'Mean', the average ISI durations, is added for comparison.

		non ISI-final			ISI-final		
		Stressed		Unstressed		Total	Mean
NOSYLL	2	281.1	–	–	224.6	505.7	506.5
NOSYLL	3	272.2	–	141.2	186.6	600.0	596.6
NOSYLL	4	256.5	145.7	145.7	196.2	744.1	741.5

stressed syllables in combination with shortening of the final unstressed syllable between 2- and 3-syllable intervals, which creates the 'extra room' needed while keeping down the increase in total duration. An additional factor is the slightly accelerating total duration. If the increase were perfectly linear then 4-syllable intervals would be some 40 ms shorter. But even so, syllable durations would be considerably longer than the rate of increase would suggest. Now, it must be stressed that while these duration values explain perfectly satisfactorily how interstress interval durations are composed in this particular material, they tell us nothing at all about the causes behind these processes. Indeed they do not even tell us that there is necessarily any other 'process' involved than mere coincidence.

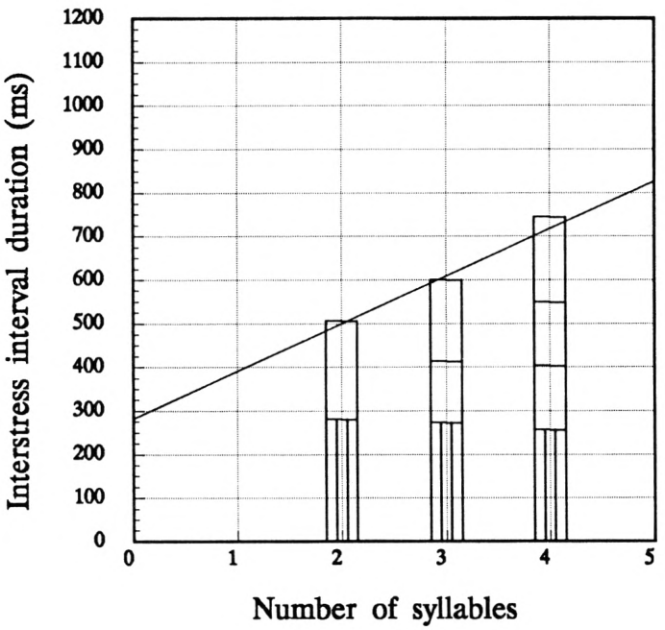

Figure 4.7. Interstress intervals decomposed into syllables. The regression line is the same as that in Figure 4.5 based on all ISI durations pooled together. Note how closely the regression line approximates the total durations based on average syllable durations.

The division of syllables into the three categories, stressed, unstressed medial, and unstressed final syllables was motivated by the need to compare ISI durations with the internal composition of the intervals. The following analysis will show, however, that this classification may also be based on the different inherent characteristics of the syllables.

An analysis of variance made using the data underlying the average durations in Table 4.9 may be summarized as follows:

1) Duration is highly correlated with stress. Durations for stressed syllables are significantly longer than unstressed syllable durations. They are also significantly longer than the interval-final unstressed syllables (ANOVA, $P < .001$).

2) Position within the interstress interval also plays a significant role. A Tukey-HSD test reveals that syllable durations are significantly different for all three positions ($P < .05$). There is an interaction here, of course, between stress and position since initial syllables are always stressed.

3) With respect to the influence of the number of syllables in the interval, the situation is less clear. The apparent dependency on interval length for the stressed syllables only reaches a significance level of .038, and a Tukey-HSD analysis reveals that only the difference between 2- and 4-syllable intervals is significant at that level. There is no significant length effect in the medial syllables. The final syllables, however, also differ significantly, but again it turns out that it is the deviant duration in one of the cases (2-syllable intervals) that is responsible for the significance. The possible existence of a genuine length effect will be discussed further below.

A graphic illustration, in the form of histograms, of the distributions of syllable durations for the three syllable types may be found in Figure 4.8.

Now, the analysis has so far been concerned only with establishing whether the differences with respect to stress, position, and interstress interval length are significant, but no attempt has yet been made to explain what may be the cause of the differences. I will now proceed with a closer study of the internal composition of syllables to try, if possible, to find causal explanations for the variation noted above. As was shown above, interstress interval durations depend critically on the number of segments in the interval. It is conceivable that syllable durations too must depend on the number of segments in a similar way. It is therefore necessary to examine more closely what form this dependency takes at syllable level.

Table 4.10 is a counterpart to Table 4.8, only now the dependent variable is the number of segments.

It can be seen in Table 4.10 that there is considerable variation in the number of segments per syllable for syllables in different positions in the table. I have summarized data from non phrase-final intervals in Table 4.11. The table is analogous to Table 4.9.

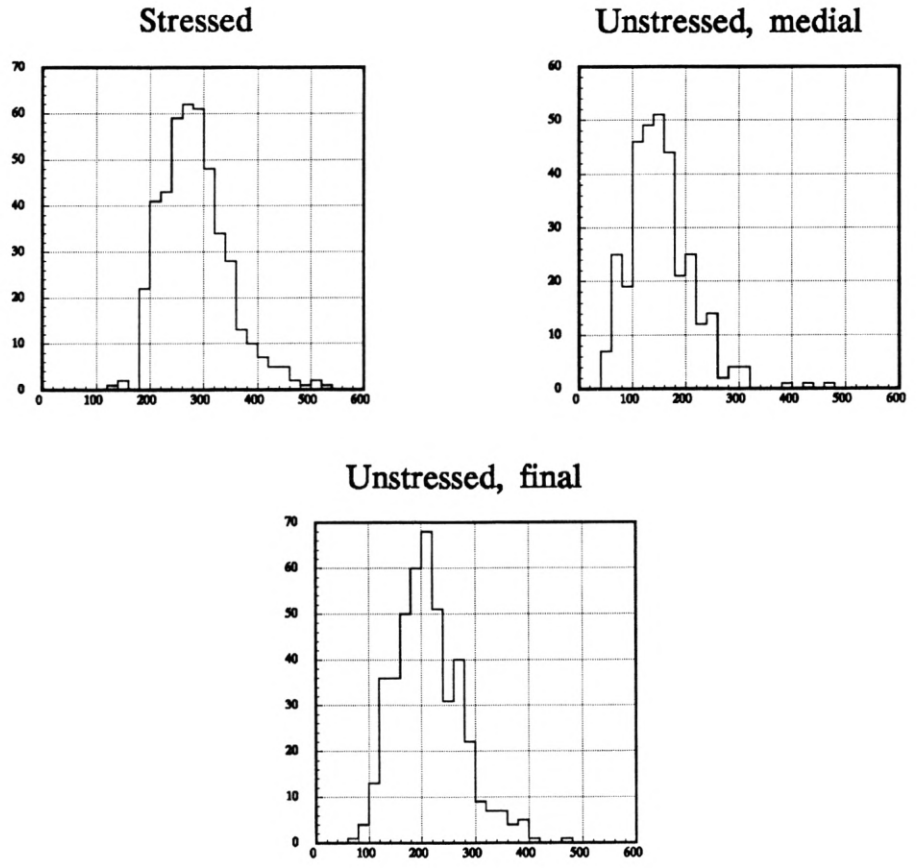

Figure 4.8. Histograms showing the distributions of syllable durations (in ms) for stressed, unstressed medial, and unstressed final syllables in all non phrase-final interstress intervals. Mean syllable durations for the three groups (273.8 ms, 142.6 ms, and 202.4 ms respectively) differ significantly.

It is quite clear that the number of segments per syllable varies with both syllable position and the number of syllables in the interstress interval.

If there should be any temporal adjustment in interstress intervals and syllables which depend on such factors as stress, syllable position, and interval length, there are at least two possible ways in which this could be achieved. There could be an adjustment of the number of syllables or segments (a process suggested by Cutler 1980, see 3.3) as a function of these factors or there could be temporal adjustments of syllable or segment durations or both processes. I will first explore the possibility of any systematic adjustments of the number of segments.

There is a difficult problem involved if one wants to address the question of reductions of segments. To be able to say that a reduction has occurred one must, of course, be able

Table 4.10. An edited output from the SPSS/PC+™ statistical package showing the number of segments per syllable as a function of stress, syllable position, and the number of syllables in the interstress interval.

Summaries of	SEGMENTS	No. of segments in the syllable
By levels of	FOOTPOS	Position of foot in the phrase
	STRESS	Stressed/unstressed
	SYLLPOS	Position of syllable in the foot
	NOSYLL	Number of syllables in the foot

Variable	Value Label	Mean	Std Dev	Cases
For Entire Population		2.460	.697	1548
FOOTPOS	0 non phrase-final	2.485	.708	1220
STRESS	1 stressed	2.635	.730	447
SYLLPOS	0 non foot-final	2.635	.730	447
NOSYLL	2	2.786	.767	173
NOSYLL	3	2.536	.728	222
NOSYLL	4	2.558	.502	52
STRESS	0 unstressed	2.398	.680	773
SYLLPOS	0 non foot-final	2.086	.646	326
NOSYLL	3	2.036	.724	222
NOSYLL	4	2.192	.420	104
SYLLPOS	1 foot-final	2.626	.611	447
NOSYLL	2	2.642	.738	173
NOSYLL	3	2.541	.535	222
NOSYLL	4	2.942	.235	52
FOOTPOS	1 phrase final	2.366	.645	328
STRESS	1 stressed	2.805	.751	149
SYLLPOS	0 non foot-final	2.805	.751	149
NOSYLL	2	2.504	.502	119
NOSYLL	3	4.000	.000	30
STRESS	0 unstressed	2.000	.000	179
SYLLPOS	0 non foot-final	2.000	.000	30
NOSYLL	3	2.000	.000	30
SYLLPOS	1 foot-final	2.000	.000	149
NOSYLL	2	2.000	.000	119
NOSYLL	3	2.000	.000	30

to determine how many segments there *should* have been, had there been no reductions. How one should attack this problem is far from obvious. What should be the standard of comparison, the target? Usually one uses some idealized rendering of a text as the target, but whether this is always relevant is highly doubtful. In this particular case I have nevertheless chosen that method. My reason is primarily that the material used here is an example of rather careful reading where reductions are few. There are in fact a number of

Table 4.11. Mean number of segments of the syllables in Table 4.10.

		non ISI-final			ISI-final		
		Stressed		Unstressed		Total	Mean
NOSYLL	2	2.79	–	–	2.64	5.43	5.41
NOSYLL	3	2.54	–	2.04	2.54	7.12	7.09
NOSYLL	4	2.56	2.19	2.19	2.94	9.88	9.83
MEAN		2.64		2.09	2.63		

readings which contain no reductions at all compared to an idealized target. It thus does not seem too unreasonable to use 'perfect' readings as the comparison. But I would like to underline once more that it is far from obvious that this is a relevant analysis. There is no proof that in the case where subjects have not pronounced all theoretically possible segments, those segments were actually ever present at any level of their speech planning. Having made these reservations, however, I will proceed with the analysis as if the question were not quite this controversial.

The distribution of the mean number of segments in syllables in different positions based on an idealized realization of the texts would be: 2.73, 2.08, and 2.87 segments per syllable for stressed, unstressed medial, and unstressed final syllables respectively. Two things are immediately clear; 1) the variation in the number of syllables found in the data (see Table 4.11) is a reflection of the same kind of variation in the 'target' material and, 2) the mean number of segments per syllable in stressed and final positions is greater in the target than in the production data. Medial syllables on the other hand do not seem to be reduced. But a closer inspection of the data shows that these figures are not immediately comparable. For unstressed medial syllables, 31 of 357 syllables are missing altogether. For obvious reasons, data from these syllables therefore do not appear in the statistics in tables 4.10—4.13. The distributions of syllables with respect to syllable position and interstress interval length are, thus, not identical, although the difference is small. I have, therefore, chosen a different way of treating data. Assuming an ideal rendering, interval by interval, to be the target it is possible to compute for every syllable in each interval how many segments there *should* be. The number of 'realized ' segments for each syllable may also be computed and compared with the maximum possible number of segments, the target. If a syllable is missing altogether it will, in this particular context, be regarded as a syllable with 0 segments. The result of such an analysis is presented in Table 4.12.

The first observation is that there is a relatively low degree of reduction. Overall, only about 6.5% of all theoretically possible segments are 'missing'. The distribution follows that of the target material very closely. Compared to the figures presented in Table 4.11, the main difference is that the number of segments for the medial syllables is lower (1.90 vs. 2.09). This is due to the inclusion of the 0-segment syllables as described above. The

Table 4.12. The number of segments per syllable for different positions and interstress interval lengths compared to an idealized reading with no reductions. 'Reduction' is the proportion of segments in the actual readings compared to the ideal. 'Rank' is the rank according to the degree of reduction.

		Segments/syllable			
		Target	Real	Reduction	Rank
Stressed		2.7338	2.6353	.9640	
NOSYLL	2	2.7987	2.7718	.9904	2
NOSYLL	3	2.7531	2.5858	.9392	1
NOSYLL	4	2.4915	2.4915	1.0000	3
Medial		2.0812	1.9048	.9152	
NOSYLL	2				
NOSYLL	3	2.0000	1.8075	.9038	1
NOSYLL	4	2.2458	2.1017	.9358	2
Final		2.8658	2.6264	.9165	
NOSYLL	2	2.6040	2.5839	.9923	3
NOSYLL	3	2.9958	2.5732	.8589	1
NOSYLL	4	3.0000	2.9492	.9831	2
All		2.5947	2.4237	.9341	
NOSYLL	2	2.7013	2.6779	.9913	3
NOSYLL	3	2.5830	2.3222	.8990	1
NOSYLL	4	2.5388	2.4526	.9660	2

degree of reduction is not uniform, however. First of all, it should be noted that stressed syllables are less reduced. Unstressed syllables seem to be reduced to the same degree in both medial and final position. If there is a systematic influence of interstress interval length, one would expect it to be a monotonic increasing or decreasing function of interval length. This is clearly not the case. The ranks show that it is the trisyllabic interstress intervals which are most reduced. For 2- and 4-syllable intervals there are contradictory trends for stressed and final syllables. The conclusion must be that there is no determinable systematic effect of interstress interval length. Unless, of course, one proposes that trisyllabic interstress intervals are always more reduced. But there is no theoretical basis to suggest such a solution. The most reasonable explanation must be that the reductions are caused by the particular segmental structures of the syllables and interstress intervals. An inspection of the material shows that the reductions occur precisely at those places one would predict on the basis of what one finds when analysing casual speech. This is in itself no argument against the idea that precisely these types of reductions occur in order to achieve some overall temporal adjustment. But as the data presented here show, this does not seem to be the case. The number of segments per syllable (and ISI, cf. last part of Table

4.12) as a function of interstress interval length is not influenced in any systematic way by these reductions. And more particularly, there is no tendency for the reductions to result in systematically shorter syllables or interstress intervals as a function of the number of syllables.

One reason why medial syllables are shorter than the other two types is thus that they contain fewer segments. Whether this particular distribution of the number of segments is representative of Swedish syllable structure in general or particular to the material used here is not known at present. It is quite possible that there may be some systematic relation between syllable type and the mean number of segments (for a large enough number of cases) but this is something that will have to be decided in future research.

As has been pointed out above, durations seem to depend very strongly on the number of segments. Now that it is clear that the number of segments varies in different positions, as a function of the target material, the distribution of durations must be re-examined to see to what extent the different durations may be explained simply by the different number of segments in the syllable.

Simple regression analysis shows that syllable durations are significantly correlated with the number of segments in all three positions. Regression coefficients are .433, .623, and .665 ($P < .001$ in all cases) for stressed, unstressed medial, and unstressed final syllables respectively. The correlations are far from perfect, however, and one must again ask if interstress interval length may play a role. If the number of syllables in the interval is used as the independent variable, however, the correlations found are much lower. Correlation coefficients for stressed, medial, and final syllables are only .118, .035, and .219, respectively. Of these correlations only the last one is significant at the .01 level. It thus seems as if the number of segments in the syllable is the determining factor, at least for stressed and unstressed medial syllables. For ISI-final syllables there seems to be some effect of interval length. If multiple regression is used, including both factors, the picture is essentially the same. Multiple correlations coefficients are .437, .624, and .719, for stressed, medial, and final syllables respectively. It can, thus, be seen that for stressed syllables and for unstressed medial syllables using the number of syllables in the interstress interval as a second variable, the improvement in correlation is negligible. The contribution of the number of syllables does not come anywhere near significance in these cases ($P = .178$, and $P = .412$ respectively). Put in other words, this means that adding the number of syllables in the interval as a second variable, the resulting regression equation explains duration values no better than simple regression using only the number of segments does. For these syllable types, the results provide no evidence for any effect of interval length.

For unstressed ISI-final syllables the situation is less clear. There is some improvement by using both variables. The contribution of interval length is rather small (some 20 ms per syllable), but it is significant ($P < .01$). But an inspection of the material reveals that

some very long durations of syllables in 2-syllable intervals are responsible for the whole effect.

From the results, one is, therefore, obliged to conclude that there seems to be little basis for assuming an independent contribution of the number of syllables in an interstress interval to syllable duration. The number of segments in the syllable is the decisive factor for all syllable types with interstress interval length making only a non significant contribution.

This does not necessarily mean that the case for compression must definitely be closed. The contribution of ISI length is definitely not significant in a statistical sense. But looking at the figures, one may see that the contributions go in the predicted direction. The contribution of ISI length is –6 ms, –5, and –27 ms, per syllable in the interval for the three types of syllables. From these figures it may be understood why interval length makes little or no contribution if added as a variable to the regression equation, compared to the number of segments. However, one should not perhaps exclude the possibility of a trend in the direction of shorter syllables. But a trend of this magnitude will obviously be very difficult to document as a significant contribution given the overwhelming influence of the number of segments. These last remarks must not be interpreted as a suggestion that there *is* such a trend, but it may be of some interest to include information from which no definite conclusions may be drawn since the addition of many such inconclusive pieces of evidence from different studies may eventually be combined into a picture which begins to makes some sense.

Even if the discussion above shows that the number of segments in a syllable is the only factor which significantly contributes to syllable duration in the different types, this does not mean that syllable duration depends on the number of segments in the same way for all types. I will now examine the internal structures of the syllables in the three different positions to see if and how they may differ. In Table 4.13, syllable durations as a function of the number of segments in the syllable are shown.

It may be seen that there are quite noticeable differences in the durations of stressed and unstressed syllables, even given the same number of segments. There also seems to be a difference between medial and final syllables although this difference is small and may not be significant (20 ms for 1-segment syllables increasing to 29 ms for 3-segment syllables). Based on these durations it is possible to compute regression equations for the three types of syllables ($r = .43, .62$, and $.67$ respectively, $P < .001$ for all).

Stressed	Dur = 176.3 + 37.0*Segments
Medial	Dur = 21.5 + 58.5*Segments
Final	Dur = 16.7 + 70.7*Segments

Table 4.13. Syllable durations as a function of syllable position and the number of segments in the syllable.

Summaries of	SYLLDUR	Syllable duration (ms)
By levels of	SYLLPOS	Position of syllable in the foot
	SEGMENTS	No. of segments in the syllable

Variable	Value Label	Mean	Std Dev	Cases
For Entire Population		212.6	81.6	1220
SYLLPOS	1 foot-initial	273.8	62.4	447
SEGMENTS	2	246.5	51.9	230
SEGMENTS	3	299.1	55.9	150
SEGMENTS	4	311.2	66.2	67
SYLLPOS	2 foot-medial	142.6	60.2	326
SEGMENTS	1	78.3	26.2	55
SEGMENTS	2	138.4	47.0	188
SEGMENTS	3	194.8	57.6	83
SYLLPOS	3 foot-final	202.4	65.0	447
SEGMENTS	1	98.0	00.0	1
SEGMENTS	2	160.8	45.1	195
SEGMENTS	3	224.0	46.0	221
SEGMENTS	4	317.6	77.0	30

The difference in behaviour between stressed and unstressed syllables is further underlined by these equations. The most noticeable difference is the size of the constant term, no doubt indicating the role of the stressed vowel. The durations of unstressed syllables, on the other hand, seem to be more or less proportional to the number of segments; the constant term being only around 20 ms. Most of the difference in duration between unstressed medial and unstressed final syllables may be explained by the different number of syllables in the target material for these two positions. But given that the durations of unstressed syllables are approximately proportional to the number of segments, one might ask if the small difference in 'growth rate' is significant. It is not obvious how the difference in growth rate should be tested for significance, however. I have tried three different methods. As a first approximation, one may postulate that durations are simply proportional to the number of segments. If this is the case then growth rate becomes the mean duration per segment. A significance test on these means results in a significant difference. A slightly more sophisticated approach is to subtract the constant before dividing with the number of segments, assuming that there is some minor difference in mean durations between the initial vowels in each syllable and the following consonants. Using this method the difference again turns out to be significant ($P < .01$, in both tests). In the latter analysis the difference in growth rate comes out as 12.8 ms per segment. Given a mean number of

segments per syllable of about 2.5 for unstressed syllables this means a ISI-final lengthening effect in the order of 30 ms if an equal number of segments is assumed.

A somewhat less unorthodox method is testing for interaction between syllable position and syllable duration using multiple regression. Doing this, the hypothesized difference fares slightly less well. The interaction effect is not significant at quite the same level ($P = .023$), but there is a tendency in the hypothesized direction nevertheless. A cautious conclusion might, therefore, be that there seems to be a genuine ISI-final lengthening effect but that it is small and its significance cannot be said to have been established beyond doubt. There is a tendency, however, and that is enough for one to suggest that the possibility of a genuine effect should be further investigated.

4.6.5. Final lengthening.

The question of the effect of phrase final lengthening remains. Final interstress intervals are slightly longer than non final ones (597 ms vs. 580 ms, mean values). Due to variation, this difference is not, however, significant. It is hardly meaningful to compute regression equations on the number of syllables for final interstress intervals alone since all but one of them are disyllabic. A regression equation based on segments is slightly more meaningful and gives a hint as to where one should look for a possible difference between final interstress intervals and non-final ones with respect to duration ($r = .74$, $P < .001$).

Final ISIs $\quad I = 179.0 + 80.2*\text{Segments}$

If this equation is compared with the corresponding equations for non-final interstress intervals (see Table 4.6), it may be seen that whereas the constant term is of the same order of magnitude as many of those in the table (although somewhat smaller than the mean, 179 ms vs. 221 ms) the slope is considerably greater (80 ms vs. 53 ms). It is, in fact, greater than any of the individual slopes and nearly 30 ms/segment greater than the average. This is an indication that the difference may lie mainly in the behaviour of unstressed syllables. Because of the limited variation and relatively lesser number of final interstress intervals, not much more information may be derived from interstress interval durations alone. The regression equation must also be interpreted with caution for the same reason, and only be regarded as a 'hint' as to where to look for interesting differences. I will, therefore, turn to an analysis of syllable durations in the final interstress intervals to get a more detailed picture.

It was seen in the analysis of non phrase-final interstress intervals above that there were small differences with respect to reductions between stressed and unstressed syllables. In phrase-final interstress intervals there is no such difference. The explanation is that there are simply no reductions at all in final syllables. Not a single theoretically possible segment is 'missing' in any of the syllables.

Table 4.14. Syllable durations as a function of the number of segments in the syllable for phrase-final interstress intervals.

Variable	Value Label	Mean	Std Dev	Cases
For Entire Population		272.2	57.7	328
SYLLPOS	1 foot-initial	281.8	52.8	149
SEGMENTS	2	262.6	55.4	59
SEGMENTS	3	289.9	51.0	60
SEGMENTS	4	303.2	37.6	30
SYLLPOS	2 foot-medial	229.1	43.0	30
SEGMENTS	2	229.1	43.0	30
SYLLPOS	3 foot-final	271.2	61.0	149
SEGMENTS	2	271.2	61.0	149

Mean syllable duration as a function of syllable position and the number of segments in the syllable is shown in Table 4.14. First of all, it may be noted that the mean syllable duration for all phrase final syllables is longer than the corresponding duration for non phrase-final ones (272 ms vs. 213 ms, cf. Table 4.8).

In comparing phrase-final and non phrase-final interstress intervals, I will first consider stressed syllables. The mean duration for stressed syllables in non phrase final position is 273.8 ms. In phrase final interstress intervals, mean stressed syllable duration is 281.8 ms. The difference is not significant. In fact, durations of stressed syllables vary very little between different interstress intervals. Table 4.15 shows the durations of stressed syllables in different positions in the phrase. As can be seen, there is little variation in syllable durations. The differences are not significant.

Multiple regression using both the number of segments and interstress interval length as variables again shows that the determining factor is the number of segments, with interval length leaving only an insignificant contribution. The linear regression equation for syllable duration as a function of the number of segments for the phrase final stressed syllables is:

$$\text{Dur} = 222.2 + 21.6*\text{Segments}$$

($r = .30$) which is slightly different from the corresponding equation for all stressed syllables in non phrase final interstress intervals (Dur = 176.3 + 37.0*Segments). But there is some variation in regression equations for the three first interval positions with slope and constant term values that are higher as well as lower than in the equation above. This fact in combination with the finding above that there are no significant differences in duration as

Table 4.15. Mean durations and mean number of segments for stressed syllables as a function of ISI position.

ISI	Mean	SD	No.segm.	N
1	267.8	54.0	2.40	149
2	269.8	67.0	2.68	149
3	283.9	64.4	2.83	149
4	281.8	52.8	2.81	149

a function of position, may be used to support a (cautious) conclusion that stressed syllables in phrase final interstress intervals do not differ from stressed syllables in non phrase-final interstress intervals with respect to duration and internal composition. It must be pointed out, however, that none of the stressed syllables are ISI-final. Stressed syllables in absolute final position may very well display significantly different properties.

If unstressed syllables are considered, the situation is radically different. Mean durations are 87 ms longer for medial syllables and 69 ms longer for final ones, if compared with the corresponding syllable types in non phrase-final interstress intervals. This is much too crude a comparison, however, because of the markedly different segmental structure of syllables in final interstress intervals. All medial and final syllables in phrase-final interstress intervals contain only 2 segments. Comparing unstressed phrase-final syllables with only 2-segment syllables in non phrase-final interstress intervals may, therefore, provide a more representative comparison. If this is done, medial syllables are found to be 90.7 ms longer and final ones 110.4 ms longer. The lengthening of ISI-final syllables is probably an underestimate since the very last syllables are not followed by any new stressed syllable with the possibility of extra consonants preceding the stressed vowel as is the case with syllables in non phrase-final intervals.

An inspection of the source material shows that there is only one trisyllabic interval in final position. This means that there is only one occurrence of a phrase-final medial unstressed syllable for each speaker. This is, of course, very little to base any generalizations on. Moreover, the only medial syllable carries secondary stress, which is also likely to influence duration. In the sentence where this interval occurs, however, the third interval has exactly the same structure as the last one. Both interstress intervals are trisyllabic and the second syllable carries secondary stress and both medial syllables have the same phonetic structure, a vowel and a stop consonant. If these both interstress intervals are compared with respect to the medial unstressed syllable their durations turn out to be 160.2 ms and 229.1 ms respectively. That is the phrase-final syllable is about 69 ms longer. This may be a more representative value than the ones mentioned above. For obvious reasons it is not possible to carry this analysis much further. I will now only summarize the results.

It seems as if occurring in the phrase-final interval does not influence the duration of stressed syllables. Unstressed syllables are lengthened, however, interval-final syllables with around 100 ms (probably somewhat more) and interval-medial syllables with 70 ms. What may be said with some kind of certainty, of course, is only that those results seem to be true for the particular material analysed here.

4.7. Summary of results and conclusions.

It is now possible to summarize the results of this part of the study. But before I do so, I would like to make a brief methodological remark. The material used was not particularly designed to study the finer details of speech production under different conditions. The main aim of the study was to examine variation in interstress interval duration and how this variation depends on the number of syllables and phonemic segments in the interval. In addition, it was hoped to be able to say something about possible compression tendencies in interstress intervals. With respect to these questions, the material and the results may be regarded as reasonably representative of read speech. But the deeper down into details the study is carried, the greater is the risk that the material is no longer representative. A larger or different corpus may yield different results. This situation became particularly clear in the very last section when final lengthening was analysed in medial unstressed syllables based on only one type of syllable. One may perhaps wonder if it is worthwhile to carry the analysis to such a depth given the obvious risk that the results get less and less representative. My definite answer is yes. This study is not meant to provide the final answer to every conceivable question. Many more studies are needed to get a fuller picture. But by adding the results from many detailed and varied studies together, one may eventually approach a description which is representative for a large subset of spoken language and, at the same time, detailed. With those reservations made, I will now summarize the results without further reservations about their generalizability.

With respect to questions concerning regularity of interstress intervals, the most noteworthy result is perhaps the great variation. If the range of durations is used as a measure, a typical range for an individual subject is more than half of the mean ISI duration. This is true for a material with only five different sentences; it is to be expected that ranges may be even greater in a larger corpus. Standard deviations too are considerable, typically around 100 ms for mean interval durations of 500 ms, and 135 ms for all data pooled. Little support for any claims that ISI durations 'tend to be equal' is thus found in this material.

If the results are compared with those in other studies it can be seen that the mean values here are somewhat longer—580 ms compared to 548 ms in Fant and Kruckenberg's (1989) study of Swedish prose reading and around 500 in many other studies. The variation in terms of standard deviation found here seems to be comparable to those reported in other studies (perhaps even a bit lower). But whether these differences signify inter-language differences or just typical inter-study variation is not possible to say.

An attempt was made to see if regularity of interstress interval duration is correlated with speech rate. Two measures of regularity were used; relative range and the regularity measure proposed by Scott, Isard, and Boysson-Bardies (1985). When relative range (range/mean duration) was used, it turned out to be uncorrelated with mean interval duration (which is an indirect measure of speech rate). The irregularity measure was also uncorrelated with speech rate (measured in syllables/minute). Both attempts to find any correlation between speech rate and regularity thus gave negative results. In this sense, regularity does not seem to be a function of speech rate.

Interstress interval durations vary systematically as a function of the number of syllables and also as a function of the number of phonemic segments. The number of segments is the better predictor of interval duration. Both functions were found to be linear to a close approximation, confirming the hypothesis based on analyses of data found in other studies (see 2.3.1 and 3.2). The 'growth rate' per syllable based on the pooled data from all subjects is just under 110 ms per syllable. This agrees well with the results found in studies of so called stress-timed languages. If the number of segments is used, the rate is 53 ms per segment. Almost identical results have been obtained two other studies of Swedish made on comparative material (Fant, Kruckenberg, and Nord, 1989; Fant and Kruckenberg, 1989). This indicates that the number of segments is very stable predictor of interstress interval duration in read Swedish.

The question of syllable compression as a function of interval length was also addressed. One reason for studying syllable durations in the context of speech rhythm is to find if there are any compression tendencies in syllables that may work to make interstress intervals more equal than they would otherwise be. No real evidence for such tendencies was found, however. There were, in some cases, significant correlations between interval length, expressed as the number of syllables in the interval, and syllable duration but it was seen that these differences could almost entirely be explained by the different number of segments in the syllables. This factor in turn was shown to be primarily a function of the segmental structure of the target material. The conclusion must therefore be that there is, at least in the material studied here, little evidence to be found for any compression tendencies as a function of interval length. Other factors, primarily the syllable and segmental structure of the target material may fully explain syllable and interval durations. In this particular sense, adding syllables seems to be a basically concatenative process. It must be said, though, as a word of caution, that in one of the analyses a significant effect of interval length was found. The effect was rather weak and was caused only by the fact that interval final syllables in 2-syllable intervals were very long. This makes it doubtful whether the result is a genuine length effect or just what one may expect as a result of the normal variation one always finds in this type of study. The effect was, however, there and should be mentioned for completeness.

Data from three groups of subjects were studied; 10 and 11-year old school children, adult males and adult females. One of the reasons for varying subject parameters was to

be able to say how critical these parameters are for the reliability of the results. In several comparisons, there was considerable inter-subject variation but whenever the results from the three groups were compared, no significant differences were found. Sex and age do not seem to be critical factors. The critical factor instead seems to be the size of the group due to inter-individual variation.

A conclusion one may draw, although perhaps with a certain amount of caution, is that many of the variables studied here, mean ISI durations, increase in interval duration as a function of the number of syllables or segments, and syllable durations, all seem to converge towards certain mean values if a sufficiently large group is considered.

Further evidence pointing in the same direction is the finding that the increase in interstress interval duration as a function of the number of segments, found here is almost identical to the ones found in the two other studies of Swedish mentioned. The implication of these findings is that it does not seem crucial in a study of this type to choose subjects from any particular category but that the size of the group should not be too small.

Based on the results obtained in the analysis above it is possible to carry the analysis beyond the questions asked at the onset. In particular it is possible to give an account which is a little more detailed of how syllables combine into interstress intervals. I will use the remainder of this section for a brief discussion of this question.

Although interstress interval durations increase linearly with the number of syllables, by about 110 ms per added syllable, it was found that this was not accomplished by simply adding unstressed syllables of that duration. Unstressed syllables were shown to be considerably longer on the average than the 110 ms suggested by the 'growth rate' of interstress intervals. The explanation for this phenomenon was shown to be different combinations of syllable durations in intervals of different lengths. This underlines the necessity of studying the internal composition of interstress intervals very closely before any definite conclusions are drawn about underlying syllable durations. The durations of syllables cannot be deduced from interval durations in any simple way.

Syllables in different positions within an interstress interval and within the phrase were found to differ significantly with respect to duration and internal structure. Stressed syllables were found, not surprisingly, to be the longest. But they were also shown to have a radically different structure with the stressed vowels probably responsible for most of the duration, reflected in a large constant term in the regression equation. Unstressed syllable durations, on the other hand, were found to be roughly proportional to the number of segments. A possible interval-final lengthening effect was found, however. Part of the added duration in interval-final syllables was shown to be due to the fact that these syllables contained more segments. One explanation for this may lie in way the speech stream is divided up into syllables using vowel to vowel onsets as the dividing points. As a consequence, interval-final syllables will include initial consonants from the following stressed syllables. This may result in different segmental and durational structures in

comparison with unstressed medial syllables, but not a lot can be said with certainty about this. Further study is needed. But taking the different number of segments into account, there still seems to be a difference in the two types of syllables. Durations of syllables seem to increase at a slightly faster rate per added segment for interval-final syllables than for medial ones. The same type of effect has been found in another study of Swedish by Strangert (1988). Using syllable duration data drawn from a large data base of read Swedish, she found that unstressed syllables preceding stressed syllables were longer than unstressed syllables followed by additional unstressed ones. There is no mention of the number of segments in different positions, so a detailed comparison is not possible, but she proposes as an explanation for the effect that the consonant clusters preceding stressed vowels are more complex, thus adding to the durations of interval-final syllables of which they will be a part using the vowel-onset to vowel-onset syllable definition (cf. above). Another possible explanation proposed by Strangert is that the consonants preceding stressed vowels also may receive some lengthening. A rather striking observation one may make is that if the mean syllable durations published are used to work out the durational composition of a trisyllabic interval, expressing syllable durations as percentages of interstress interval durations, then the internal composition turns out to be exactly the same for Strangert's data (44%, 23%, and 32% respectively) and the data used in this study although absolute durations are somewhat shorter in Strangert's material. This may be a hint that the effect is real and perhaps rather general and stable. At least this is something that should receive some attention.

A considerable phrase-final lengthening in the order of 70—100 ms was found in unstressed syllables. Both absolute final syllables and medial syllables in phrase-final interstress intervals were affected.

Now, with all the results at hand it is possible to look anew at the relation between ISI growth rates and mean syllable durations for unstressed syllables whose durations are considerably greater than the growth rate. This phenomenon, which at first may seem contradictory, was seen to be the result of varying syllable durations. Assuming the results obtained above to be reasonably representative, it is possible to propose a more detailed model for interstress interval durations as a function of the number of syllables than the linear regression model.

In the following, I will use typical duration values found in the study above in an attempt to build a model that may help to explain how the durations of elements at different levels combine to produce the results found.

It was found above that stressed syllables were longer than unstressed ones. A small decrease as a function of interstress interval length was found, but in a multiple regression analysis the contribution of interstress interval length proved to be non significant ($P = .178$). Most of the variation may be explained by the number of segments in the target material. The durations of unstressed syllables turned out to be roughly proportional to the

number of segments but with a small contribution of the constant term in the order of 20 ms. ISI-final syllables were considerably longer than medial ones. Although most of the difference could be explained by the difference in the number of segments in the target syllables, an ISI-final lengthening effect also seemed to be present. This lengthening effect was in the order of 60 ms in absolute terms. The final lengthening per segment was shown to be about 15 ms.

Now, with respect to the relation between interstress interval duration, increase and mean durations of unstressed syllables, an ISI-final lengthening alone may to a considerable extent explain this relationship. The following example will demonstrate how this may work. Let us assume that there is an ISI-final lengthening effect in the order of 50 ms. Let us further assume that the lengthening only affects the last syllable of the interstress interval. If typical values found in this study are used, a stressed syllable may be assumed to be around 275 ms. The rate of increase in interstress interval duration may further be assumed to be about 110 ms per syllable. Based on these assumptions the following interstress interval durations would result for 1 to 4-syllable intervals: 325 ms, 435 ms, 545 ms, and 655 ms. This is some 50 ms less than the durations found above but that need not concern us. If mean durations for unstressed syllables are computed from these data they turn out to be 160 ms, 135 ms, and 127 ms respectively for 2- to 4-syllable intervals. As can be seen this is in all three cases considerably more than what one would assume by looking only at the increase per syllable in interstress interval durations. The same pattern of decreasing mean durations for unstressed syllables pooled together is also found in the data. Durations are 224 ms, 164 ms, and 163 ms, respectively for 2- to 4-syllable interstress intervals. The small difference between the durations for 3- and 4-syllable interstress intervals may be seen to be due to the fact that interstress interval durations are actually slightly positively accelerated (cf. Table 4.2). Note also how the decreasing sequence of durations in unstressed syllables seems to suggest a gradual compression as a function of interval length, although from the very construction of intervals we know that this is an illusion.

If we divide the unstressed syllables into medial and final syllables the durational pattern shown in Table 4.16 results.

The durations shown in Table 4.16 are not identical to the ones obtained in the study above but the general pattern the same. The simple addition of a postulated ISI-final lengthening effect is thus seen to bring the model in rather striking agreement with actual data (cf. Table 4.9).

There are also deviations, however, and they may need to be commented on. It may be seen that, in this model, the duration of medial syllables is numerically equal to the rate of increase in interstress interval durations. This is not the case for the medial syllables in the data. One factor that causes this difference is the fact that the durations of the stressed syllables are decreasing in longer interstress intervals. Another factor is the fact that the

Table 4.16. Theoretical distribution of durations based on the assumption of a constant stressed syllable, a constant ISI duration increase, and a constant interval-final lengthening.

		non ISI-final			ISI-final	
		Stressed		Unstressed		Total
NOSYLL	2	275	–	–	160	435
NOSYLL	3	275	–	110	160	545
NOSYLL	4	275	110	110	160	655

increase in interstress interval duration is slightly positively accelerated. These two factors create the 'extra room' needed to allow a greater duration in medial syllables. One may speculate that the ISI-final lengthening effect causes some lengthening also in the medial syllables. There is, at the present stage no basis for including such factors in the model, however. And there is no theoretical basis for assuming a curved relationship, say for example a second degree polynomial, between interstress interval durations and the number of syllables either. Even doing so the resulting correlation is not significantly higher. And as has been shown above, the decrease in stressed syllable duration as a function of the number of syllables is not significant.

An intriguing observation, pointed out in section 3.2, was that interstress interval durations seem to increase approximately linearly for a number of languages. The rate of increase turned out to be around 100 ms per syllable for all languages. But as is the case for the fragment of Swedish studied here, there are results that indicate that for the other languages as well, mean durations for unstressed syllables are greater than the rate of increase in interstress interval durations. In the material used by Manrike and Signorini (1983), discussed in 2.3.2, the rate of increase was a little less than 100 ms per syllable (as computed by myself from their published data) but the reported mean durations for unstressed syllables are around 150 ms. And Delattre (1966) reports durations of 155 ms for non phrase-final unstressed syllables in English, which may be compared with the rate of interstress interval duration increase (around 100 ms) that was found in Dauer's (1983) data. There is, thus, the possibility that the model for the composition of interstress interval durations suggested above has a wider generality than only as a model for the data presented in this study.

PART III

Temporal regularity in speech perception.

Chapter 5

Duration perception—a background.

Even though the perception of speech rhythm is a multidimensional phenomenon it seems reasonable to assume that the perception of durations must play a crucial role. To be able to approach an understanding of how we perceive speech rhythm it is, therefore, necessary to know more about how we perceive durations in the speech signal.

A great number of experimental studies of the perception of duration in *non-speech* material have been made, the first ones more than a century ago. Duration perception has been studied in different modalities and using many different techniques. The stimuli that have been used have, however, usually been of a very simple kind. Two basic types of stimuli have been used—filled intervals and empty intervals. In the case of studies of auditory perception, the empty intervals have generally been silences bounded by clicks or brief tones. Filled intervals have usually been bursts of noise or simple tones. Many different variables that may influence the perception of duration have been studied: interval durations, intensity, presentation order, inter-stimulus durations, physical and mental condition of the subject etc. Some studies also exist where the complexity of stimuli has been the variable studied. In the following sections, I will give a brief overview of some of these results and the experimental techniques used to obtain them.

In light of the wealth of information on the perception of duration in general it is surprising to find that comparatively few studies have considered the perception of duration in speech. This is even more surprising if one considers that most linguists would agree that the perception of duration is a very important aspect of speech perception. Of the studies that

exist, most have been concerned with duration at the phoneme level. But there are also a few examples of studies that have dealt with the perception of durations of interstress intervals as well. In the last section of this chapter, I will review some of these studies.

It goes without saying that the perception of durations in speech must be closely linked to the perception of durations in non-speech and there is no reason why the methods used in duration perception experiments on non-speech should not be applicable to speech stimuli as well. The theoretical issues concerning duration perception in general are also valid in the discussion of duration perception in speech. But it is also obvious that the results obtained in experiments using clicks or noises cannot automatically be generalized to speech. One cannot exclude the possibility that the perception of durations in such a complex type of stimulus as speech differs in important ways from the perception of durations in less complex stimuli. We know in fact, from existing studies, that the complexity of a stimulus may influence its perceived duration. Another complicating factor is the fact that in speech, the stimulus is not constant but varies continuously. This may have important effects on duration perception.

5.1 Some issues related to the study of time and duration perception.

The purpose of the following sections is not to give a comprehensive account of all the various aspects of duration perception or psychophysics in general. Such accounts can be found elsewhere (Guilford, 1954; Baird and Noma, 1978; Gescheider, 1985). What I will do instead is to give a brief account of some of the theoretical and methodological issues I had to deal with in connection with planning the experiments described in Chapters 6 and 7, and in the analysis of the results of these experiments. The first few sections will deal with the perception of time and duration in general. But even so the scope will be limited to those aspects that are most relevant for the study of speech. That means, for example, that, as far as experimental studies are concerned, the primary focus will be on those that deal with the perception of duration in auditory stimuli, although results from experiments using other types of stimuli (e.g. visual) will be used when it is felt to be relevant, for example in connection with methodological questions. In section 5.1.6, I will also review a few studies of duration perception in speech.

Closely connected with the perception of duration is the perception of time in general. One may say that duration perception is only a special case of time perception. In the following, I will not, therefore, keep these two questions apart but treat theories of time perception and duration perception together. A fundamental question in connection with time and duration perception is the problem of describing the form of what is called the psychophysical function (or law) for time perception. This question will be discussed in section 5.1.2.

5.1.1 Experimental methods.

In studying experimentally the perception of time and duration, four different classes of methods have been used—*verbal estimation, production, reproduction*, and *comparison*. In a verbal estimation experiment the task is to state, verbally, the duration of a given stimulus. Subjects may, for example, be asked to state the duration in seconds of a tone that is presented to them. In production experiments the task is reversed. Subjects are now told to produce by some means a given duration, say 1.5 seconds. Reproduction experiments mean that subjects reproduce an interval which has just been presented to them. 'Comparison', finally, means that subjects compare a test duration with some given reference duration. Either the reference duration is given first a number of times after which the test durations are presented or the test and reference durations are presented pairwise. The correlations between different methods have been found to be rather weak (Carlson and Feinberg, 1970). This means that if one wants to compare the results obtained in a particular experiment with those of another one must carefully consider the methods used.

Responses can be of two types, *scaling* and *discrimination*. 'Scaling' means that subjects assign a scale value or a category to the duration of a test stimulus. In verbal estimation tasks the scale used is often clock time. But other types of scales are also used. Two variants of scaling are *magnitude estimation* and *category-rating*. In the first type of task, the subjects assign a value to the durations in some other unit than clock time and in category-rating tasks, they assign the test durations to one of a set of given categories. In a discrimination task, stimuli are presented for direct comparison and the task is to decide which one is the longest (or shortest).

A very difficult theoretical question that I will not attempt to answer, but which I think one must mention, is the question of what exactly it is that one studies in experiments on subjective duration. When a subject states that one duration is half as long as another, what exactly does this mean? What one would like to say is obviously that there are two percepts in this individual's mind, one of which is twice as long as the other. But it is enough to state the problem this way to realize how speculative and uncertain such a theory must be. What we may study is, at best, the subject's interpretation of his perception, and it is obvious that there are many different forms that this may take. There are, however, at present, no better methods at our disposal than those mentioned above. But it is important to realize that the results from experiments on subjective time and subjective duration must be interpreted with an open mind. At present it seems to be very difficult to separate perception and interpretation. It is even possible that this is never fully possible. Thus, when in the following sections, I talk about perceived time and perceived duration the 'fuzziness' of these concepts should be kept in mind. In the context of this brief overview I will, however, use these concepts as if they were fully understood.

5.1.2 The psychophysical law for time perception.

What psychophysics is all about, is establishing relationships between stimuli that we perceive and the perceptions of these stimuli. But the aim is a little more specific. We are particularly interested in those stimuli which form continua and for which we may establish scales by which the size of a stimulus may be measured. A psychophysical law is a function which describes the relation between the scale by which the stimulus is measured and some 'internal scale' onto which stimuli of various sizes are supposed to be mapped. There are obviously a vast number of stimulus types which can be measured along scales that measure length, weight, light intensity, loudness, etc.

In psychophysics, 'size' is generally talked about as 'intensity'. The intensity of a stimulus is thus its scale value on the particular scale one uses to measure the stimuli; centimetres, dB, Hz, grams etc. The task of psychophysics may, thus, be expressed as finding the function that describes how the intensity of the stimulus is related to the intensity of the perception.

The description of psychophysics I have just given is obviously not one that behaviourists would agree with. They would instead say that psychophysical laws describe the relation between stimuli along some physical scale and responses along some other, equally physical, response scale. Whatever stand one takes in this controversy on metaphysics, however, the problem one faces is technically the same. What is the form of the function that maps one scale onto the other?

In theory, there are, naturally, infinitely many possible forms for such functions. Only a small subset of these possibilities have been considered. With a slight oversimplification one may describe the situation as follows: One has to make some assumption about the relative sizes of each step of a particular scale. The simplest two types are those where successive scale values form an arithmetic progression (length, weight, etc.) and those where successive steps form a geometric progression (loudness, luminance, etc.). Now, if we have two scales to compare there are obviously four possible ways in which one may form pairs of scales of the two basic types. These types are 1) arithmetic—arithmetic, 2) arithmetic—geometric, 3) geometric—arithmetic, and 4) geometric—geometric. The four mapping functions which correspond to these four combinations are 1) linear 2) logarithmic 3) exponential, and 4) power functions. The functions most often proposed to hold for data from psychophysical experiments are the logarithmic function and the power function. The logarithmic relationship is often referred to as Fechner's law. Fechner, who may be considered as the father of psychophysics proposed this relationship as early as 1860, in a by now classical book; *Elemente der Psychophysik*. I do not think one does Fechner any injustice by saying that his motives were wholly intuitive. His line of reasoning was very simple. He was aware of Weber's discoveries that just noticeable differences (JND) often seemed to be proportional to the absolute sizes of the stimuli. Now, Fechner simply

assumed that a JND on the physical side corresponded to a scale value of unity on the perceptual side. From that he was able to derive the logarithmic relationship by simply integrating both sides of a differential equation. Both these last steps involve a number of presuppositions which it is far from obvious that one is allowed to make, but this is the way classical psychophysics got its first 'law'. Later experiments have produced data that seem to favour a power function as the best description. This type of function has been claimed by Stevens (1957) to be a *"general psychophysical law relating subjective magnitude to stimulus magnitude"* (p. 153) and is, therefore, often referred to as Stevens' law. I will conclude this very brief description of the possible laws of perception by stating the two most favoured types.

Fechner's law with respect to time perception can be expressed as:

$$P = C_1 \ln S + C_2$$

where 'P' represents perceived time and 'S' stimulus time. C_1 and C_2 are constants which are related to the Weber constant and the perceptual threshold respectively.

A formulation of Stevens' law that I have borrowed from Eisler (1975, 1976) is the following:

$$\Psi = \alpha(\Phi - \Phi_0)^{\beta}$$

where 'Ψ' and 'Φ' denote subjective and physical time respectively and 'α', 'β', and 'Φ_0' are constants. In this equation, 'α' is related to the Weber fraction and 'Φ_0' to the perceptual threshold.

Now, little can be said, of course, with respect to whether one function or the other is the 'correct' psychophysical function. What can be determined to some degree, though, is how well a particular function describes stimulus-response correlations in a particular experiment. I will say something about this because it has some methodological importance.

I think it is correct to say that the function most often suggested is the power function. It is often said that there is substantial evidence to back up this view. After having read a fair number of these studies and the overviews by Eisler (1976) and Allan (1979) I feel somewhat sceptical about the strength of the evidence put forward in support of the power function hypothesis. It seems to me that a linear function could be used to describe the data equally well. Eisler reports exponents from no fewer than 111 different studies. The most striking observation is that almost all exponents reported are very close to 1, which means that the relation can equally well be described as a linear one. Allan, who studied a number of these investigations in detail, also points out the important fact that in none of the studies was there any attempt to compare the goodness of fit between the power function and a linear function. In a study by Kaner and Allan (reported in Allan 1979) where they tried both a linear and a power function to model their data, a linear function fitted their data

better than a power function for 23 out of 32 subjects. Their conclusion was that *"a simple linear function often fit data better than a power function, even in cases where the size of the exponent appears to rule out the linear function."* (Allan, 1979, p. 342) The constants of the power function are usually established by fitting a straight line to a log-log transformation of the data for subjective and physical time by the method of linear regression. Now, linear regression is not a very sensitive instrument by which to decide this question. In most cases, using a limited number of observations, it is possible to fit either a linear or a power function with very high regression coefficients. I have tested both linear and power function models using duration data from the interstress interval measurements presented in Chapter 4 and found that, in many cases, both models fit equally well: in some cases with regressions coefficients of .999 for both!

The size of the exponent of the power function has been used as another argument to support a power function model. An exponent of 1 means that the relation is linear and a significant deviation from 1 is taken to indicate that the relation is not linear. Again, this is not necessarily true when applied to small data sets. I have managed to find duration data to which it was possible to fit a linear as well as a power function with an exponent of .5, both with regression coefficients of .999! The data set does not even have to be particularly small or the exponent particularly close to 1 for a situation to arise where one cannot decide between a power function and a linear one on the basis of regression coefficients alone. For one of my data sets of interstress interval durations (447 cases), the exponent for a power function that fits the data is .488. Nevertheless both a linear and a power function fit the data equally well (r_{linear} = .527, r_{power} = .521). The inevitable conclusion from these observations seems to me to be that at present there is not enough evidence to decide the form of the psychometric function for time perception, but that a linear function seems to be a reasonably good first approximation. And a general lesson to be learnt from the examples I have just given is that one must exercise a great deal of caution in extrapolating the results from linear regression and similar mathematical methods to conclusions about functional relationships, the simple reason being that there are most of the time quite a number of model functions that will fit equally well given that the regression coefficient or some other type of similar technical criterion is the measure. If this is the case in a particular study, one should prefer the simplest function, which in this case is the linear one, following the sound theoretical maxim established by William of Ockham.

5.1.3 Just noticeable differences—Weber's law.

Discriminability between stimuli has often been found to depend on the durations of the stimuli compared. A frequent observation is that the least noticeable difference increases with duration. (In the following, I will use 'ΔT' to denote the 'just noticeable difference', JND.) As a measure of discriminability the standard deviation (SD) of the psychometric function for discrimination (or some fraction thereof) is normally used. In duration

discrimination tasks, the difference limen (DL), which is the interval within which 50% of the correct responses fall (DL = .67*SD), is often used as the JND. The choice depends, of course, on what one means by 'just noticeable'.

In many studies (see Allan, 1979 for a review), ΔT has been found to increase monotonically (but not necessarily linearly) with duration. What this means is that the minimal difference in duration between two stimuli needed for the difference to be perceptible increases as a function of the durations of the stimuli. It would be helpful for the understanding of duration discrimination if the form of this function could be described. To date there is, however, no general agreement on this point.

The simplest formulation of the discriminability function is Weber's law. This law states that discriminability is a constant fraction of the shortest of the durations compared. Expressed as a formula this is

$$\Delta T = kT_0$$

where 'ΔT' is the just noticeable difference, 'k' is a constant and 'T_0' is the shortest of the durations. Another way of expressing the same fact is to say that the *differential fraction*, $\Delta T/T_0$, is constant. From the results of many experiments, it is clear that a strong interpretation of Weber's law is untenable. If one computes the differential fraction as a function of duration it is not constant (ΔT does not increase linearly). A common result is that $\Delta T/T$ is minimal in the range from about 100 ms to 2000 ms and increases for durations that are shorter or longer. There is, thus, some support that Weber's law holds approximately for durations in that middle range, but the picture is not unambiguous, which I will demonstrate below by citing a few results. Weber's law has been claimed to hold for all modalities of perception, but because this study is about the perception of auditory stimuli, the selection of studies discussed below will have a bias towards studies where auditory stimuli have been used.

There exist a few studies that support Weber's law rather strongly. Getty (1975) used empty intervals bounded by clicks and two highly experienced subjects in a discrimination experiment. The interval durations used ranged between 50 and 3200 ms. In the range 400 ms to 2000 ms the differential seems to increase almost perfectly linearly as a function of duration, that is Weber's law seems to hold for this duration interval. Getty also proposes a slightly modified form of the function

$$Var(T) = k^2T^2 + V_R$$

where 'V_R' represents the stimulus-independent variance. (A very similar type of relationship has been proposed by Miller, 1947. There the constant term represented the contribution of the absolute threshold) The generalized form of the differential fraction will then be:

$$\frac{\Delta T}{T} = \sqrt{k^2 + \frac{V_R}{T^2}}$$

This differential fraction increases sharply for small values of 'T', but for larger values, the influence of 'V_R' becomes negligible and the generalized Weber fraction approaches a straight line. The generalized Weber function fits Getty's data almost perfectly for durations less than 2000 ms. For greater durations, however, the differential fraction increases faster than predicted by the formula.

Another study that lends support to Weber's law is one by Halpern and Darwin (1982). Stimulus material in their experiment was pulse trains of four clicks. The first three clicks were equally spaced in time with inter-click intervals (ICI) of 400—1450 ms. The fourth click came at the end of an interval that differed from the base ICI. Subjects were to decide whether the last click came 'early' or 'late'. The SD as a function of base ICI was an almost perfectly linear increasing function. The average differential fraction is .037

Other results give support for Weber's law in specific intervals, usually somewhere in the range 200—1000 ms. Blakely (1933, cited in Woodrow, 1951) obtained an approximately constant differential fraction of 0.08 for the interval 600—800 ms, using empty intervals as stimuli. With longer or shorter intervals the fraction increased. Stott (1933, cited in Woodrow, 1951) found values between 0.10 and 0.12 for tones in the range 0.4 to 2 s. Michon (1964) used interval trains marked by clicks as stimuli. He found very low and approximately constant differential fractions in two regions; 100—200 ms and 200—1000 ms. The values for these two ranges were 0.01 and 0.02 respectively. For durations lower than the shortest range and greater than the longest, however, the values increased steeply. Small and Campbell (1962) using filled intervals (noise and tones) with durations from 0.4 ms to 400 ms obtained a monotonically decreasing differential fraction, but in the region 40—400 ms it was approximately constant (0.19—0.18).

As I mentioned in the previous section, different authors prefer a linear or a power function to describe psychophysical functions, but the two alternatives are seldom compared using the same set of data. I will use a report on duration discrimination by Abel (1972b) where she uses a power function to describe her data to show that a linear function would fit her data equally well. She reports the results of a series of discrimination experiments using empty intervals bounded by markers (clicks) of varying intensity and duration. The inter-click intervals (ICI) used ranged from 0.63 to 640 ms. The data are presented in tables but also in the form of power functions derived from the data by using linear regression on the transformed data (log-log transformation). Abel, rather arbitrarily, only considers data from durations between 10 and 160 ms and obtains power functions with rather impressive 'goodness of fit' numbers. Assuming a linear relationship, however, the fit is almost as good; .973 vs. .995, .979 vs. .985, and .935 vs. .983, with the linear regression coefficients mentioned first. (I have recalculated her figures using the averages in the published tables, hence the slightly higher coefficients compared to the published figures.)

Now, the differentials for the first intervals, 0.63 and 1.25 ms, are clearly deviant but there is no apparent reason, looking at the data and the diagrams, why one should not test for a possible linear relationship for the rest of the range, 2.5 to 640 ms. If now one compares the goodness of fit between a linear function and a power function, it turns out that the fit is better for the linear functions for all three marker conditions.

The conclusion, after reinterpreting Abel's data, is that a linear function fits her data reasonably well (particularly for durations between 160 and 640 ms) and that they provide some support for the validity of Weber's law in that range. The differential fraction for interval durations between 160 and 640 ms is 0.22 on the average. In another study Abel (1972a) used intervals filled with noise. The average differential fraction for durations between 50 and 640 ms is approximately 0.10 (calculated from the diagram) and the relation seems to be roughly linear, again providing some support for Weber's law in that region.

But there are also results that are in genuine contradiction with the predictions made by Weber's law. Allan, Kristofferson, and Wiens (1971), using visual stimuli of two base durations, 50 and 100 ms, obtained no influence of duration on discrimination. Allan and Kristofferson (1974b) report on discrimination of light stimuli in the range 70 to 1020 ms. Again, discrimination was not a function of stimulus duration. It remained constant over a large range of stimulus durations. Kristofferson (1973, cited in Allan and Kristofferson, 1974a) has presented similar data for empty auditory intervals ranging from 100 to 2000 msec. In a study by Rousseau and Kristofferson (1973) using empty intervals marked by a light at the onset and a tone at the offset, and base durations from 100 to 2000 msec, discriminability was constant over the complete range durations.

The explanation offered by Allan and Kristofferson (1974b) is that the critical variable is the amount of practice given to subjects. *"We have found the amount of practice an O has with a particular set of duration values to be a critical variable. Inexperienced Os always yield functions which show discriminability to be a monotonic decreasing function of stimulus duration. Highly practiced Os often yield functions which show discriminability to remain constant over certain ranges of duration values"* (p. 439)

Another slightly more general explanation that one would like to test is that the difficulty of the task with respect to the ability of an individual subject influences the result. There exist results that, in my opinion, may point in that direction.

In a discrimination experiment, Lehiste (1979a) used sequences of four noise filled intervals separated by clicks. The sequences consisted of four equal intervals or three equal intervals and one test interval that was systematically varied in 10 ms steps. The base durations used were 300, 400, and 500 ms. The subjects were asked to mark the 'longest' or the 'shortest' interval. It turned out that the size and the variation of the JNDs were highly correlated with interval position. Both magnitude and variation was greatest for the first interval (60 to 100 ms), noticeably smaller for the second (40 to 80 ms) and smallest

for the third one (30 to 40 ms). For the last interval, JNDs rose again to roughly the same level as those for interval 2.

These results are very informative and to some extent also expected. As I mentioned in section 1.6, experiments on synchronization in tapping tasks have shown that subjects establish their synchronizations very rapidly. Three taps are usually enough. Fraisse (1966) showed in a synchronization experiment involving inter-tap durations of precisely the same magnitude as the interval durations in Lehiste's experiment that subjects were able to achieve synchrony with the stimulus rhythm within roughly 50 ms (in many cases much less) from the third tap. The interpretation of this result is that subjects base their expectations of where the next beat is to come on the durations between the previously heard ones. If this is correct it seems reasonable that they should also be able to detect deviations more easily from the third interval on.

Now, a very interesting observation is that the just noticeable difference in Lehiste's experiment does not seem to depend on the duration of the interval for the third interval. In the 'shortest' judgment task, JND is 30 ms for all three durations and in the 'longest' task it is 40 ms for 300 and 400 ms base durations and 30 ms for the 500 ms one. For the fourth interval there is no clear tendency either but for the first two intervals JND seems to increase with increasing base durations.

This is particularly interesting if compared with Allan and Kristofferson's (1974b) observation that the discriminability function may be a function of practice. If one replaces 'amount of practice' with 'difficulty of the task' the results agree very well. When the base rhythm is established the task is easier, thus the influence of base duration is reduced or cancelled. But these ideas must, of course, be tested more thoroughly before any definite conclusions may be drawn.

I have summarized the results from a few studies in Figure 5.1. The values on which the different curves are based are taken from the papers cited in the 'legend'. In some cases they are complemented by my own calculations. Where the authors do not mention explicitly how they have defined the Weber fraction I have assumed it to be .67*SD/T.

Some important pieces of information can be found by looking at the diagram. First of all, the curves clearly demonstrate the great variation in results from different experiments. Duration discrimination obviously depends on a number of factors in addition to duration. Another observation that can be made, and which is supported by findings in other investigations, is that differential fractions seem to increase at the end points of the range. For a given study they are often highest for the shortest and longest durations and lower in the mid range. The sharp increase for durations less than 50 ms is particularly noticeable. Also worth noting is the fact that differential fractions are reasonably constant in the duration range 100 ms to 1000 ms. This is particularly interesting in connection with duration perception in speech since the durations of most of the important elements of speech (phonemes, syllables, interstress intervals) fall in that range.

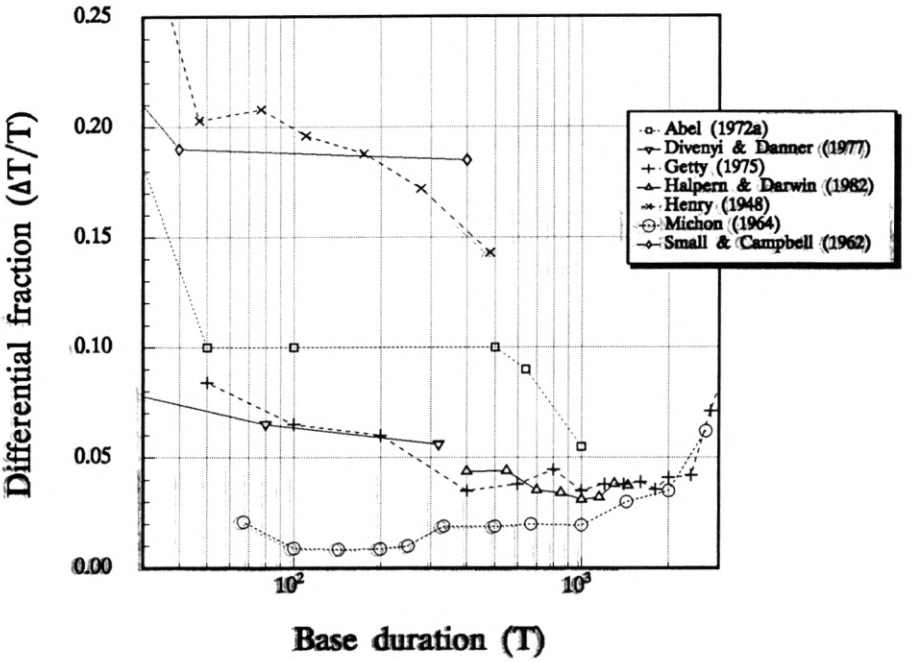

Figure 5.1. A graphical representation of the results in some of the studies of duration discrimination mentioned in the text. Curves are based on the values presented in the papers.

Numerical values of differential fractions from these studies and an additional few made on speech material can be found at the end of this chapter in connection with the discussion of duration perception in speech (5.1.6 and Table 5.1).

5.1.4 The influence of non-temporal factors on the perception of duration.

Non-temporal factors may also influence the perception of time and duration. The most studied of these factors is probably intensity. It has generally been found that an increase in intensity results in an increase in perceived duration.

Berglund, Berglund, Ekman, and Frankenhaeuser (1969) using 1 kHz tones with 50, 250, and 500 ms durations as stimuli in a magnitude estimation experiment found perceived duration to be a function of intensity. Subjective duration was an approximately linear function of intensity (expressed in dB) in the range 57 dB to 104 dB. The effect was most marked for the longest (500 ms) stimulus duration with an increase over the whole range of about 50%.

Zelkind (1973) using four different experimental techniques, direct estimation, comparison estimation, reproduction, and production, obtained similar results. Subjective duration increased as a function of intensity in all four conditions.

But there are also findings that contradict these results. Treisman (1963, cited in Allan, 1979) reported a decrease in both produced and reproduced intervals with increasing tone intensity.

If intensity may influence subjective duration, one would suspect that it might also influence discrimination of durations. This does not, however, seem to be the case. Henry (1948) used filled intervals with durations of 47, 77, and 277 ms, and intensities of 20, 40, 60, and 80 dB. He found the intensity variable to have very little effect on discrimination except for a slightly poorer result for the shortest stimulus and the lowest intensity. Abel (1972a) used noise filled intervals with a bandwidth of 3500 Hz and durations of 5, 40, and 320 ms to compare discrimination for two different sound levels; 65 and 85 dB. She found no difference in discrimination between the two sound levels. Creelman (1962) varied the intensities of 1 kHz tones against a noise background but found no influence of intensity on discrimination as long as the tones were clearly detectable. He concludes: *"Duration discrimination depends on sufficient intensity to mark the time unambiguously; it depends on detectability but not on loudness."* (p. 592) The results thus seem to indicate that the intensity must be above a certain level for the intervals to be clearly detectable but that above that level a further increase in intensity produces no appreciable effect on discrimination.

The stimulus type may also influence duration perception. It is well known that a filled interval is judged as longer than an empty interval of the same duration (the 'filled-duration illusion') (Craig, 1973, Thomas and Brown, 1974). The structure and complexity of the stimulus may also play a role. More complex stimuli are perceived as longer than less complex stimuli. (Schiffman and Bobko, 1974) and stimuli composed of many elements are perceived as longer than stimuli with few elements (Schiffman and Bobko, 1977).

Spectral characteristics of the stimuli may affect duration perception under certain circumstances. Burghardt (1973a) tested subjective duration of tones and noise pulses with frequencies from 200 Hz to 8 kHz by letting subjects adjust the duration of a variable test sound until it had the same apparent duration as a reference sound of 1 kHz. The frequency of the test sound had an effect on subjective duration. The subjective duration turned out to be shortest for a 3.2 kHz sound. For lower and higher frequencies the durations were perceived as longer. The effect was greater the shorter the sounds. For durations in the range 300 ms to 800 ms, however, the frequency seemed to have very little effect.

'Cognitive variables' may also play a role. In the studies referred to below visual stimuli were used, but it is conceivable that a similar influence of cognitive stimulus content may be present with auditory stimuli as well. In brief tachistoscopic presentations of real words, nonsense words, and blank fields Thomas and Weaver (1975) found that blank fields were judged shorter than both real words and nonsense words. Stimulus familiarity has also been shown to have an effect on duration perception, but results from different studies seem contradictory with respect to the direction of the influence. Warm, Greenberg and Dube

(1964), Warm and McCray (1969), and Devane (1974) all found significant interactions between stimulus familiarity and perceived duration. Presentation times of infrequent words were perceived as shorter than those of frequent words. The results thus seem to indicate that familiar stimuli are perceived as longer than unfamiliar ones. But results pointing in the opposite direction have also been obtained. Avant and Lyman (1975) and Avant, Lyman, and Antes (1975) found that familiar stimuli were perceived as shorter than unfamiliar ones.

It should be noted that there are some contradictory results with respect to all the variables discussed above. The studies on the influence of intensity and cognitive variables where opposite tendencies were found in different studies may serve as examples. One must, therefore, be careful not to draw too definite conclusions about the influence of a particular type of factor. This is particularly important with respect to the study of speech where all the above mentioned factors are normally present and, in addition, not constant over time but varying.

5.1.5 The time-order error.

It has long been known that in tasks where stimuli are presented sequentially, usually in pairs, the results are often asymmetric with respect to the order of the stimuli in each pair. For example it may look as if either the first or the second duration, in a pair of stimulus durations, is systematically overestimated. The phenomenon has been observed in different modalities and is usually referred to as the time-order error (TOE).

To say that one of the stimuli is overestimated is of course not very precise. One would like to express the time-order effect in a more well defined form. Allan and Kristofferson (1974a) suggest a formalism to express TOE for forced choice (FC) tasks involving the comparison of pairs of sequentially presented stimuli. Their definition is built on the probability of correctly discriminating between two stimuli in a pair as a function of the order of presentation. In a duration discrimination task where two stimuli of different durations are presented pairwise there are two possible presentation orders; the longer stimulus may come first or last. In Allan and Kristofferson's formalism, this is expressed as S_1S_0 and S_0S_1. Responses can also be of two kinds; either the first stimulus is judged as longer (R_{10}), or the second one is (R_{01}). If there is no time-order effect present the probabilities of correctly selecting the longer stimulus when it comes first, $P(R_{10} \mid S_1S_0)$, or last, $P(R_{01} \mid S_0S_1)$ should be the same. If, however, there is a time-order effect they are not. Now, the difference between those two probabilities can be used to express TOE.

$$TOE = P(R_{10} \mid S_1S_0) - P(R_{01} \mid S_0S_1)$$

The sign of the time-order error will be positive if the first duration is overestimated and negative if the second one is. If there is no TOE present, the value of the expression is of course 0. If all combinations of durations are presented in both orders then TOE simply

expresses the difference between 'longest first' and 'longest last' responses in proportion to the total number of responses.

In duration discrimination tasks it has been observed that the error is not constant but varies with many factors involved in an experiment. One of the first observations was that TOE varies with the duration of the stimuli used. Durations below a certain limit seemed to be overestimated while longer durations were underestimated. This led some researchers (e.g. Woodrow, 1935) to suggest that the perceptual duration of the first presented stimulus in a pair gradually approached that of some internal standard duration during the inter-stimulus interval (ISI). This means that the second stimulus is compared, not with the first one, but with a modified version of the first duration having approached the internal standard. This would explain why the first stimulus was overestimated for durations shorter than the standard and underestimated for longer durations. Other results showed, however, that the postulated standard duration was not constant even within the same subject. Stott (1935), for example, showed that the time-order error changed considerably as a function of subjects' practice. It has since also been shown that TOE may depend on ISI and on the mean durations of the set of stimuli used in a particular experiment. This has led some theorists to suggest that the cause of what seems to be a pure effect of the order of the stimuli in a pair, may be the result of an assimilation of the stimuli presented and the establishing of a durational standard based upon these durations. What this means is that subjects in the course of an experiment establish a standard of duration that is the weighed mean of the durations presented. Helson's (1964) adaptation level theory is a theory of that kind that has been successful in explaining some of the experimental results, but not all. Jamieson (1977), using his own results and those of Allan (1977), suggests that one should separate presentation order effects from assimilation effects. If I have understood him correctly, he means that TOE should be limited to those effects that can be attributed only to the order of stimuli in a pair while effects which range over the whole set of stimuli should not be included in TOE.

All the results discussed so far are effects that can be shown to depend, at least primarily, on the properties of the stimuli. The effects can be viewed as distortions of the temporal perception induced by the way stimuli are presented and their properties. But there is also evidence that there may be a subject related component involved. Allan, Kristofferson and Rice (1974), Carbotte (1973), and Creelman (1962) have presented results that have shown TOE also to be strongly subject dependent. This offers another possible explanation for TOE's. They may be the result of a response bias in the subject. Subjects may, for reasons that have little or nothing to do with time, simply have a tendency to prefer to choose the first or the second stimulus. This possibility has been referred to as the *response bias* theory of TOE. The theory has not been without critics, however, and there are also results that seem to indicate that there is no such effect (e.g. Hellström, 1977).

There are also studies in which the results show no apparent effect of TOE. Small and Campbell (1962) using filled intervals between .4 and 400 ms sum up their results with

respect to TOE: *"in the present study TOE is distinguished principally by its relative absence."* (p. 410)

What one must conclude after reviewing a number of investigations is that the phenomenon of TOE is not very well understood at present. First of all, the effect is not always present. Under certain circumstances there seems to be no TOE at all. Secondly, it is unclear to what extent it is subject specific. Some studies indicate that there is a strong subject specific component but others question that factor. The temporal factors that seem to be able to influence TOE are the duration of the inter-stimulus interval (ISI), the temporal order of stimuli in a pair, and the mean duration of the stimuli in a set. But again, these factors do not always produce an effect and when they do the effects are not consistently the same. The most likely explanation is that the effect is a combination of all these factors but that the weight of the different factors may depend on the experimental situation, the type and durations of stimuli used as well as subject specific variables. But the precise contributions of these factors are not known.

5.1.6 Duration perception in speech.

Of the research on duration perception in speech, surprisingly little has been done that has direct bearing on the perception of interstress interval durations. This is all the more surprising considering the extensive debate about speech rhythm, and particularly isochrony in perception, where this aspect is crucial. Most of the research on duration perception in speech has been concerned with the perception of durations at segment or syllable level. I will mention some results from this field in so far as I find them relevant for the study of speech rhythm and also present in more detail some of the few studies that have been concerned with durations at interstress interval level.

I think it is correct to say that a large proportion of the studies on segment durations in speech have been done in connection with research on speech synthesis. One of the goals in speech synthesis is to produce as natural sounding speech as possible. Therefore, much of the research on durations of segments has been concerned with what kind of variation in duration that is acceptable or what durational cues that are used to categorize phonemes. This means that the just noticeable differences that are reported are often based on acceptability rather than discriminability. Since one cannot assume that the largest acceptable difference is necessarily the same as the least perceptible one, the results from these experiments cannot immediately be used to decide perceptual thresholds for durations. The least perceptible difference may be considerably smaller than the range of acceptable variation. The results from these experiments are nevertheless quite relevant in the context of finding perceptual thresholds since at least they give us some idea of what kind of durational change that is perceptible at the segmental level. In the following, I will refer to all studies as studies of JND but indicate for each of them whether the criterion used was acceptability or a genuine measure of the least perceptible difference.

Huggins (1972a) studied just noticeable differences for the durations of four different consonants and one vowel in different contexts. The test variable was acceptability. Subjects were to say whether they found the segment to be 'normal', or too long or too short. The general result as expressed by the author was that *"subjects are much more sensitive to changes in vowel duration than to changes in consonant duration"* (p. 1270). The results are presented in the form of diagrams, so it is difficult to quote any exact figures, but judging from the diagrams and interpreting as the just noticeable differences the duration ranges within which segment durations seem to have passed as acceptable, a JND of about 15 ms seems to be representative for an average vowel duration of 110 ms. For the consonants, the JNDs are markedly longer: approximately 35 ms for the longest one (88 ms) and about 38 ms for the shortest ones (60 ms). This would correspond to relative JNDs of around .14 for the vowel and .40 to .63 for consonants. Whether there is any relation between discriminability and duration is not possible to say with any certainty.

Huggins reflection on the significance of his results for speech rhythm perception is worth noting. *"The jnd's for the vowel in the present experiments, which subjects often judged by changes in rhythm, correspond to about 2%—3% of these interstress intervals."* (p. 1277)

Nooteboom (1973) has carried out experiments to examine the internal representations of segment durations. To this end, he studied the stability of reproduction of durations of synthetic vowels. The results of his experiments may also have some relevance with respect to duration discrimination of vowel durations. In the first experiment, the subjects (non-naive) manipulated the duration of the medial vowel in a three syllable nonsense 'word' being produced by a speech synthesizer until they felt satisfied that the vowel had the 'right' duration. The target vowels were [a] and [a:]. Again the criterion was acceptability rather than perceptibility. Subjects repeated the manipulations 20 times to give information about the acceptability range. There is no correlation between standard deviations and durations if the results are pooled for all three subjects, but for each subject taken individually, standard deviations are lower for the shorter duration and for two of the three subjects differential fractions are reasonably constant (as calculated by myself on the basis of diagrams). Inter-subject variation is considerable, with differential fractions ranging from approximately .02 to .09.

In a second experiment, the influence of position in the word and the number of syllables in the word was investigated. What is relevant here is that, this time, the standard deviations varied between 12 and 25 ms for durations in the order of 130 ms to 200 ms. This would indicate differential fractions around .07.

Klatt and Cooper (1975) used the method of magnitude estimation with category rating responses to examine just noticeable differences of one vowel, the stressed vowel [i] of the word 'deal', and one consonant, the post-vocalic fricative [ʃ] of the word 'fish'. Stimuli were modified by deleting or duplicating portions of the waveform in the test segments.

The JNDs for the vowels were 41 ms on the average (range 22—59, average duration 236, mean $\Delta T/T = .18$). There is no apparent dependency on duration. For the fricative, there may be a slight dependency, since the JNDs for the two longest sounds are also the longest, 67 and 98 ms (average JND = 48, range = 25—98, mean duration = 128, mean $\Delta T/T = .35$). But the correlation is weak.

Bochner, Snell, and MacKenzie (1988) studied the ability to detect changes in the durations of vowels and tonal complexes, and in the duration of the closure in stop consonants and gaps in tonal complexes. Three normally hearing and seven severely hearing-impaired listeners served as subjects (I will only consider the results from the normally hearing in this context). The speech stimuli consisted of the vowels [i], [I], [u], [U], [a], [ʌ] and the consonants [p], [t], [k], and the tonal complexes consisted of digitally generated sinusoids at 0.5, 1, and 2 kHz. Vowel stimuli were presented in two contexts, CVC and isolated. The test method was a discrimination task with triplets of durations presented, two standards and one test duration.

Bochner *et al.* report difference limens without specifying exactly how they were calculated, but one may assume that the DLs are either equal to normal differential fractions or some constant fraction thereof. DLs for vowels ranged between 0.11 and 0.19 compared to 0.10—0.21 for tonal complexes.

There was no significant difference between speech and tonal complexes, either for filled or unfilled intervals. But there was a significant difference between filled and unfilled intervals for both stimulus types, unfilled intervals yielding the higher relative DLs. For durations over 100 ms, DLs seem to be approximately constant for filled intervals. For unfilled intervals there is a marked dependency on duration. DL decreases from about 1 for 25 ms durations to .25 for 150 ms durations.

As a summary of the results discussed so far, one may say that differential fractions generally seem to be of the same order as the ones found in experiments using non-speech stimuli of corresponding durations.

As for the prosodic role of duration, some aspects have been studied. Several studies have been published that show that durational means can be used to disambiguate otherwise ambiguous sentences. Lehiste, Olive, and Streeter (1976) have, for example, shown how duration can be used to disambiguate syntactically ambiguous English sentences. And Wenk and Wioland (1982) have shown how duration can be a cue for the disambiguation of semantically ambiguous sentences in French. The units on which these durational changes operate are not normally interstress intervals, but the time domains are of the same order, so when the disambiguations are successful this may give us some idea of what durational contrasts that are perceptible in this duration range. However, this type of analysis is not normally made. In fact, I do not know of any experiment where the perceptual threshold for duration changes in this context has been studied.

Duration perception of interstress intervals in speech has been examined in a few studies. In a series of experiments, Lehiste (1973) studied the regularity of interstress interval durations in both production and perception. The 17 sentences used in the experiments consisted of four mono- or disyllabic interstress intervals read by two speakers. An analysis of the results from the production experiment showed a considerable variation in interstress interval duration, also for intervals with the same number of syllables. For disyllabic interstress intervals in the first three positions, the range was 280—430 ms for one speaker and 283—441 for the other. The last interstress interval was considerably longer than the three first ones, indicating a possible final lengthening effect (355 vs. 585 for one speaker and 352 vs. 538 for the other, average values).

The recorded sentences from the production test were presented to subjects in a perception test. Each sentence was presented twice. On the first presentation, the task was to mark on a script the interstress interval they perceived to be the longest and the second time, the shortest one. The longest interval was identified as longest significantly above chance level for both speakers. For one of the speakers, the shortest interval was also reliably identified.

In a second experiment, speech and non-speech material was compared. The speech material was the same as in the first test but with some sentences, in which intervals were almost equal, removed. The non-speech material was produced using noise bursts with durations copied from the sentences. Stresses were marked as stronger bursts. The result showed that discrimination was significantly better for the non-speech material. But both the shortest and the longest durations were correctly identified above chance level for both types of material.

One of the speakers had produced several sentences where two or three interstress intervals had the same duration. In one of these sentences the first interstress interval was 260 ms and the following three all 550 ms. This sentence was used in a third experiment together with a non-speech equivalent. As could be expected the shortest interval was identified with almost perfect reliability. Interestingly enough, though, the last interstress interval received fewer 'longest' judgments and more 'shortest' judgments than the two preceding intervals although they all had the same duration. This tendency was even stronger in the non-speech material. Lehiste's interpretation of the result is that listeners expect the last interstress interval to have some extra length and when it does not, it is perceived as shorter than it really is. This seems very likely. It is a bit surprising, though, that the tendency was even stronger for non-speech stimuli. It could be an indication that final lengthening is a very general phenomenon. It should be noted that it is also found in music and is, perhaps, characteristic of a wide range of human behaviour.

Lehiste sums up the results from the three experiments by saying that she does not find the results very impressive and that the difference between the actual scores and perfect identification is greater than that between the scores and random answers. This may be so, but one would have liked to see some measure of discriminability (for example JND) to

Table 5.1. A summary of differential fractions reported in the studies mentioned in the text. Where the figures are not explicitely stated in the papers they have been calculated using diagrams etc.

Study	Method	Stimulus	Durations	ΔT	ΔT/T
Abel (1972a)	discrimination	noise, tones	50— 960	–	0.09
Abel (1972b)	discrimination	empty	20— 160	–	1—0.25
			160— 640	–	0.22(aver.)
Blakely (1933)	discrimination	empty	600— 800	–	0.07
			200—1500	–	<0.09
Bochner *et al.* (1988)	discrimination	tone-complex	25— 500	5—56	0.21—0.10
		vowels	75— 170	11—21	0.19—0.11
		empty	25— 150	22—35	0.9—0.2
		stop-closure	25— 150	26—30	1.2—0.2
Divenyi & Danner (1977)	discrimination	empty	25—320	–	0.06
Halpern & Darwin (1982)	discrimination	pulse-train	400—1450	18—54	0.04
Huggins (1972a)	naturalness	vowel	110(aver.)		0.14(aver.)
		stop-closure	60—90		0.40—0.63
Getty (1975)	discrimination	empty	400—2000	–	0.04
Klatt & Cooper (1975)	naturalness	vowel		236(aver.)	0.18(aver.)
		fricative	128(aver)		0.35(aver.)
Lehiste (1979)	discrimination	pulse-train	300— 500	30—100	0.06—0.30
Lunney (1974)	discrimination	pulse-train	30—1000	–	0.04—0.06
			1000—3000	–	0.06—0.12
Nooteboom (1973)	discrimination	vowels	50— 125	–	0.04(aver.)
Michon (1964)	discrimination	pulse-train	100— 200	–	0.01
			300—1000	–	0.02
Small & Campbell (1962)	discrimination	noise, tones	4— 40	–	0.35—0.19
			40— 400	7—70	0.19—0.18
Stott (1933)	discrimination	tones	400—2000	–	0.10—0.12

be able to evaluate more precisely what the results tell us about the possibility of detecting durational differences in speech.

Lehiste (1976) has also made an experiment that provides important information about the influence of non-temporal, auditory, properties on the perception of duration in speech or speech-like stimuli. She performed an experiment where she presented subjects with pairs of speech-like stimuli in a duration discrimination task. The stimuli consisted of synthesized versions of the vowel [a]. The parameter that was varied in the experiment was the F_0 contour within the vowel. Three types of F_0 contours were used, monotone, rising-falling, and falling-rising, and three different durations, 270, 300, and 330 ms. The F_0 inflection in the non-monotone stimuli was varied in 12 semi-tone steps for each type.

Stimuli were presented in pairs. One member in each pair was monotone and the other non-monotone, or both were monotone, and both had the same duration. Three interesting observations can be made studying the results: 1) There was a marked time-order effect. When both stimuli in a pair were monotone the first one was judged longer in 68.7% of the cases. 2) When one member had a changing F_0 contour, the number of 'longest' judgments for that member rose significantly. 3) The amount of change in F_0 seemed to contribute insignificantly. Only in 3 cases out of 12 was there a significant contribution of the amount of change on 'longer' judgements and the effect was not particularly dramatic.

As was mentioned above, results from another experiment by the same author (Lehiste, 1973) indicated that subjects tend to underestimate the last interstress interval in a sequence. One may therefore suggest an expectation of final lengthening as one possible explanation for the marked time-order effect. It is also interesting to note that a qualitative change in the stimulus makes a difference, but the degree of change seems to contribute insignificantly as long as the difference is perceptible.

In Table 5.1, I have made a summary of the results from some of the studies with particular emphasis on differential fractions and stimulus type.

Chapter 6

Stress beat perception in a phrase of read poetry.

In this chapter, I will report the result of an experimental study of stress beat perception. The emphasis will be on two aspects of stress beat perception, the precision with which they may be determined and their absolute locations on a time scale.

If it is at all possible to perceive correctly the variation in interstress interval duration in speech, one must first be able to delimit accurately the intervals to be judged. If this is not possible then the accuracy of duration perception will suffer to a corresponding degree. To locate the stressed syllables is thus an important first step in the perception of interstress interval durations in speech. The stressed syllables can hardly be regarded as points in time, but as was reported and discussed in 2.1 and 3.4, there is evidence that the perceptual occurrence in time of a syllable is closely connected with the onset of the vowel and that other factors seem to play a less important role. The 'points' in time at which the stressed syllables seem to occur, were referred to as the 'stress beats'. The main purpose of the experiment described below is to determine how precisely these stress beats may be located.

Although there is experimental evidence that stress beat locations are closely connected with the onsets of the stressed vowels, other factors have been found to influence the perception of stress beat locations to some degree. The duration of the consonant or consonants preceding the vowel is such a variable (Rapp, 1971; Allen, 1972). Other variables that have been suggested as having an influence are consonant voicing (Lindblom, 1970; Rapp, 1971) and vowel duration (Fox and Lehiste, 1987). The influence on placement of these factors is rather small, at least if compared with the variation in

interstress interval duration one normally finds, but the effect must considered if one wants to understand the perception of speech rhythm fully. Possible effects of vowel and consonant durations were also studied.

Another question that is considered is the question of perceptual regularization. As was shown in section 3.5, there are experimental results that seem to imply that stressed syllables in speech are perceived as occurring more regularly than they do physically (Donovan and Darwin, 1979; Darwin and Donovan, 1980). But results of some other studies seem to contradict this view, at least partly (Bell and Fowler, 1984; Scott, Isard, and Boysson-Bardies, 1985). The question can, therefore, hardly be said to be resolved; more studies are needed to gain more insight. Even if the question of regularization was not a primary concern in the experiment described below, it was decided to analyse the responses with respect to regularity to see if any such tendencies could be detected.

As I have mentioned (2.1, 3.4), there is a theory challenging the idea of the close connection between stress beats and vowel onsets and arguing that the 'p-centre' is the relevant cue in speech rhythm perception. Although p-centres too seem to be linked to the stressed syllables, it has not been possible to connect them with any particular acoustical correlate and it is, thus, not possible to propose any absolute locations for p-centres. There is a serious methodological problem if one wants to test the validity of the p-centre hypothesis for continuous speech. The types of experiments on which the theory is based use isochronous readings of lists of isolated words and manipulation of inter-item distances in word lists or syllable lists to make them sound isochronous. It is unclear to me how these techniques could be modified to be applicable to continuous speech. Also, the differences in interstress interval durations depending upon whether one considers stress beat locations (vowel onsets) or p-centres are very small compared to the great variation in interstress interval duration in continuous speech. From this point of view, p-centres explain perceptual isochrony no better than stress beat locations. I have made no attempt to find an experimental method to determine p-centre locations in continuous speech. I have instead centred the study on finding the stress beat locations for a particular stimulus phrase. In the analysis of the results, however, I have tried a way of calculating relative p-centre locations, using the formula proposed by Marcus (1981) (see 3.4) to see if the interval durations obtained by using this method deviate in any interesting way from the ones based on stress beat locations.

It is far from obvious how these questions should be studied experimentally and what measures are the most appropriate to use. In the following section, I will discuss and motivate the particular choice of method used in the experiment in this study.

6.1. Some methodological considerations.

Many of the experiments that have been carried out in order to determine stress beat locations in spoken language have involved some kind of production or reproduction of rhythm. Often the stimuli have been presented repeatedly and the task of the subjects has been to tap to the beats or react in some similar manner or they have been asked to read to an outside regular stimulus rhythm. From these experiments, it has often been possible to propose rather precise locations of the stress beats. Also detailed models of stress beat placement have been made that take into account such variables as type and duration of prevocalic consonants in the stressed syllable and so on (Allen, 1972, Rapp, 1971, Fox and Lehiste, 1987).

Several of the experiments of the above mentioned type were reported and discussed in some detail in 3.4 There is no need to repeat what was said there but I will expand on some methodological points that were considered in connection with planning the experiment presented here.

One of the difficulties one encounters when using production or reproduction as the means of representing the perception of rhythm is to separate the part of the result that is due to the particular form of production from the part that depends on perception only. Several complicating factors may be involved.

If stimuli are presented repeatedly or if subjects are told to tap to several consecutive syllables there is an influence of a general tendency to behave regularly in motor activity to be considered. A small pilot study I made with myself as a subject (unpublished) several years ago comes to mind. When transcribing conversations from video tapes I noticed that some parts of the conversations sounded rhythmical. I wanted to see if the speech rhythm I thought I heard could be connected with any observable non-verbal behaviour. To establish the underlying rhythm of speech I used finger-tapping. I made 20 rounds of finger-tapping to 2 selected intervals of speech. No endless tape loop was used so there was no regularity in the presentation of the sequence listened to. Instead the tape was simply rewound between repetitions. Given the aim of the study, the results were very disappointing. When the taps from all the rounds were marked on a transcript with a time scale the individual taps seemed almost randomly scattered. It was not possible to determine cluster points with any reasonable degree of accuracy. However, the inter-tap intervals, in each round taken separately, showed remarkable constancy. My conclusion then was only that the method was unusable for my purpose. But now, in retrospect, the results are beginning to make more sense. There need not be any direct or simple correspondence between a certain behaviour and the physical properties of the stimulus that elicits the behaviour. We know that there is a strong tendency for motor behaviour to be regular. Subjects usually tap very regularly even in the absence of any external stimulus. In fact, Fraisse and others

have shown in experiments that subjects find it difficult *not* to tap regularly. (see for example Fraisse, 1956). It is, therefore, likely that subjects will tend to tap in a generally regular fashion that need have little to do with any physical properties of the stimulus. A problematic aspect if one wants to use the method of finger-tapping or any other reproduction type of technique is, thus, the tendency to regularity in precisely that type of motor behaviour.

In one of the experiments reported by Allen, the stimulus phrase was repeated regularly with approximately 4 seconds between presentations. Even if 4 seconds is near the upper limit of the 'psychological present' (according to Fraisse) it is still possible that one may establish some regularity of motor action over this time interval. When analysing his results for a possible correlation between successive taps, this was indeed what Allen found; Successive taps were correlated. Admittedly the effect was not great and there are statistical means of checking it (at least to some extent) but the fact that it is there is a hint that some other method should be looked for that does not have this disadvantage.

Another factor that may influence the results in an undesirable way is variability in motor behaviour between subjects. Reaction time and general motor ability may play a role. Some evidence that this may influence the results can also be found in Allen's results. When the results from a tapping experiment were compared to those of a click matching experiment, it turned out that inter-subject differences in beat locations were greater in the tapping experiment.

If stimuli are not presented at regular intervals or if subjects only tap to one syllable at the time then the tendencies to motor regularity may not play such an important role, but then individual reaction time becomes a factor that must be controlled carefully.

In connection with all these experiments there is also the phenomenon of rhythmic anticipation, mentioned in 1.6 and also discussed in 3.4, to be considered.

Considering all these methodological difficulties involved in using motor responses it was decided to try to eliminate, as far as possible, any motor involvement and test stress beat placement as directly as possible as a function of perception. One way of attempting this is to place some kind of 'acoustical marker' so that it perceptually coincides with a particular syllable. This can be done in two ways. One may construct an experimental apparatus which permits the subject to move the marker around until he is satisfied that it coincides with the target syllable, or the experimenter may prepare a set of stimuli with different marker locations and present them to subjects for judgement.

Both these techniques were also tried in the study by Allen mentioned above. The click placement task resulted, as predicted, in smaller inter-subject differences than when tapping was used. There was however another factor present that may have influenced the distribution of the placements. Within a given round of trials, subjects tended to move the clicks in a give direction on several consecutive trials and stop on the first click placement

that seemed acceptable. The resulting distributions were flatter than normal, possibly as a consequence of this cautious strategy adopted by the subjects.

The other type of click matching, judging whether a click copied into the phrase coincided with a given syllable or not, was also used by Allen. He did not regard the experiment as very successful though, since most subjects were unable to perform the task at all. Allen reports the responses of some of the subjects as being randomly scattered over the entire test interval (600 ms). Only data from the three best subjects was therefore usable. For these subjects, however, there were no inter-subject differences in stress beat placement. The locations measured in terms of the range of placements regarded as acceptable was quite wide (200 ms), however. Allen regarded the locations as being 'broad slurs' rather than points in time.

Now all things considered, the method that seems to be the least biased is the one using judgement of click placement. For this reason, it was decided to choose it as the one to be used in the experiment described below. The disadvantage, if Allen's results are representative, is that subjects seem to be willing to accept clicks over a wide range as acceptable. Should this generally be the case, one may question how much sense it makes to calculate stress beat placements down to the last millisecond. However, this method may still yield a more accurate description of stress beat perception than the other alternatives.

6.2. Method.

6.2.1 Subjects.

Students of speech therapy at the University of Gothenburg served as subjects in this study. They were in their second year of study and had thus had some formal training in linguistics. None of them, however, had any prior knowledge of the particular aspect of speech examined in this study. The tests were carried out as an optional part of a laboratory course. Participation was voluntary and unpaid.

6.2.2 Stimuli.

The stimulus material used in the experiment described below consisted of a set of copies of a line of poetry in which a click had been placed at, or near, the onset of one of the stressed vowels.

The poem from which the line was taken (*Atenarnes sång* by the Swedish poet Viktor Rydberg) was one of several poems that had been recorded earlier for a different purpose. The poems were read by an experienced reader. They were recorded in a soundproof studio using high quality recording equipment; Neumann KM 84 microphones and a NAGRA IV-S open reel tape recorder.

The line used as the test-phrase was read in a scanning style with clearly marked stresses. The reason for choosing a phrase with this particular quality was that it was thought desirable, in this and the duration perception experiments described in Chapter 7, that subjects should have no difficulty in deciding which syllables were stressed. This means, of course, that the results in the experiment described in this study are representative only for speech for which stresses are easily identifiable but not necessarily for speech where they are not.

To permit analysis and preparation of the test tapes, the recorded phrase was sampled into a computer (DG, Eclipse S/200) with 20 kHz sampling frequency. The stressed syllables were located and the onsets of the nuclear vowels were marked on a time scale with the help of a computer program (MIX, written by Rolf Carlson, Royal Institute of Technology, Stockholm).

Durations of the different interstress intervals (ISI) are shown in Figure 6.1. Interstress intervals will be referred to as intervals 1—6 following the numbering in the figure. The corresponding stressed syllables will be referred to as syllables 1—6. A spectrogram

	(h) 'ärligärd	'ödenärm	'odigtifr	'ämstal	'edetud	'ignar
ISI#	1	2	3	4	5	6
duration	685	733	873	560	726	622

Figure 6.1. The test phrase "Härlig är döden när modigt i främsta ledet du dignar" (Glorious is death when you succumb courageously on the front line) divided up into interstress intervals (ISI) in the traditional manner.

printout, with orthographic and phonetic transcriptions of the different interstress intervals marked on the same time scale, is shown in Figure 6.2.

The phrase is structurally quite regular. There is little variation in the number of syllables per interstress interval—four of the intervals are trisyllabic and the other two disyllabic. When measured, however, in the traditional way (vowel onset to vowel onset) the actual durations of the intervals are far from equal (see Figure 6.1). The reason for choosing a phrase with durationally unequal intervals was to be able to see if any kind of perceptual regularization could be detected in the results. If such a mechanism is involved in perception of speech, it might influence the perception of where the stress beats are and displace them perceptually in order to achieve a greater regularity in the spacing of stress beats.

The interstress intervals that deviate most markedly are the middle ones (3 and 4 in Figure 6.1). The other intervals are fairly equal in duration. If there is some kind of perceptual regularization, its effect should be most noticeable for the two most deviant intervals. This may result in a bias in the perception of stress beat locations. In addition to connecting

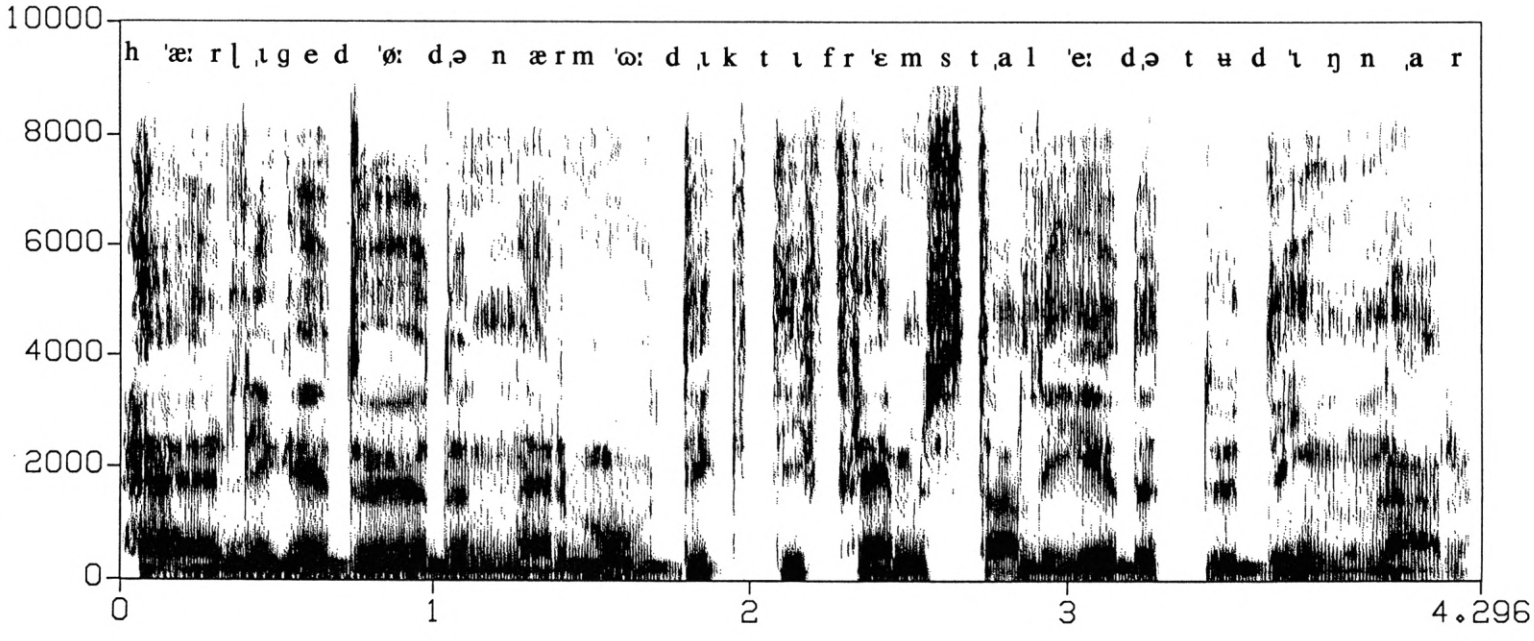

Figure 6.2. Wide band spectrogram of the phrase used as stimulus in the experiments. The divisions in the orthographic and phonetic transcriptions are on the same time scale as the spectrogram.

stress beats with the stressed syllables, subjects might prefer click placements that make the rhythm seem more regular, thus accepting 'early' clicks in position 3 more readily because the resulting pattern would be more regular.

I have to admit, however, that when this experiment was designed I was not aware of the possibility that rhythmic anticipation (see 1.6) may also play a role. Rhythmic anticipation has been found in connection with motor responses to stimulus rhythms, but it seems entirely plausible that such an effect may also influence perception in a similar manner. Thus, if one assumes that subjects establish some kind of prediction of interval durations during the first beats (which they seem to be doing in synchronization experiments) they would tend to expect the third beat to come a little earlier than it actually does. One effect of this expectancy could be a bias to accept early click placements as correct in this particular location more often than for the other syllables.

The end result of some overall perceptual regularization would thus be indistinguishable from that of an anticipation effect. To decide between these two options some other type of experiment must be performed.

The preparation of the stimuli to be used in the experiment involved placing a click in the stimulus phrase somewhere in the vicinity of one of the stressed vowel onsets. It is not easy to say what constitutes a 'good' click sound. It seems reasonable, though, that the duration of the click should be short compared to the duration of the tested syllable. It also seems desirable that the rise time of the sound should be short to make it possible to determine the location of the click with sufficient accuracy. After listening to a number of possible candidates it was decided to use a 'rim beat' from a digital drum machine (BOSS, DR-220A). The sound had a total duration of only about 35 ms, the major part of its energy concentrated in the first 10 ms. The rise time was approximately 3 ms. The click sound used in the experiment was sampled into a different file. It could be copied into the test phrase with the aid of a specially made computer program (written by myself using the BLOD library of programming routines, developed by Peter Branderud, Department of Linguistics, University of Stockholm).

In the experiment by Allen described above, the intervals within which the clicks were placed were [+300 ms, –300 ms] relative to the vowel onsets. While preparing the material, used in this experiment, I made several small pilot studies with myself as a subject to determine a suitable test interval size. The pilot studies indicated that an interval 600 ms wide, as in Allen's experiments, was much greater than necessary. 'Correct' judgements were only very rarely more than 75 ms away from the vowel onset. From the results using myself as a subject it thus seemed as if an interval of 100 ms on each side of the vowel onset would be quite sufficient. It is desirable to have the different click placements not too far apart and also not to have too many different stimuli, in order not to exhaust the subjects. Based on the results of the pilot studies, an interval 200 ms wide with 9 different click locations was considered as a reasonable compromise.

Nine different stimuli for each stressed syllable were prepared. The clicks that were copied into the phrase occurred at one of the locations k*25 ms, k = ±0, ±1, ±2, ±3, ±4, relative to the vowel onset. A test tape was made for each of the stressed syllables. On the tape, each of the 9 different stimuli was presented 4 times. The stimuli were presented in random order with a 5 second silent interval between each stimulus. Each test was preceded by a trial round consisting of 5 randomly selected stimuli.

6.2.3 Procedure.

The tests were made with a group of students of speech therapy as a part of their laboratory work. This meant certain restrictions on how and when the tests could be administered. Although each test took only about 20 minutes it was, for example, not possible to carry out all the tests on one occasion. It is doubtful, however, if it would have meant any real advantage to do all the tests in one session. The tests demanded a high degree of mental effort and the results may, therefore, have been adversely affected by the subjects' gradual loss of concentration had all tests been given on one occasion. The tests were, therefore, carried out on 4 different occasions, the disadvantage being that not all subjects were able to participate in every test. The different syllables were not tested in the order they come in the phrase, but in a randomly decided order. The first syllable, which was going to be tested in the very last session was never tested, because of a time table change. This means that only syllables 2—5 were tested. The test group size varied between 8 and 13, with 6 subjects participating in all 5 test sessions.

The stimuli were presented over loudspeakers in a lecture room. The subjects were instructed to listen carefully to a test phrase with a click copied into it and try to determine whether or not the click coincided in time with a specific syllable. No information was given as to what it would mean for the click and the syllable to occur simultaneously. Subjects had to make up their own subjective criteria. The answers were to be given on a pre-printed form where the subjects were to tick off 'right' or 'wrong' depending upon whether they regarded the click and the syllable to be simultaneous or not. Each session was begun with a trial round of 5 stimuli to make the task understood by the subjects. The answers from these trial rounds were discarded from the final results. The test was carried out as a 'forced choice' test, that is only 'right' or 'wrong' were permitted as answers, and the subjects were forced to give an answer for every stimulus whether they felt certain or not.

6.3 Results.

6.3.1 Stress beats; locations and distributions.

The results from all five experiments are summarized, in a graphical form, in Figure 6.3. An observation one may make is that the interval [–100ms,+100ms] seems a little too

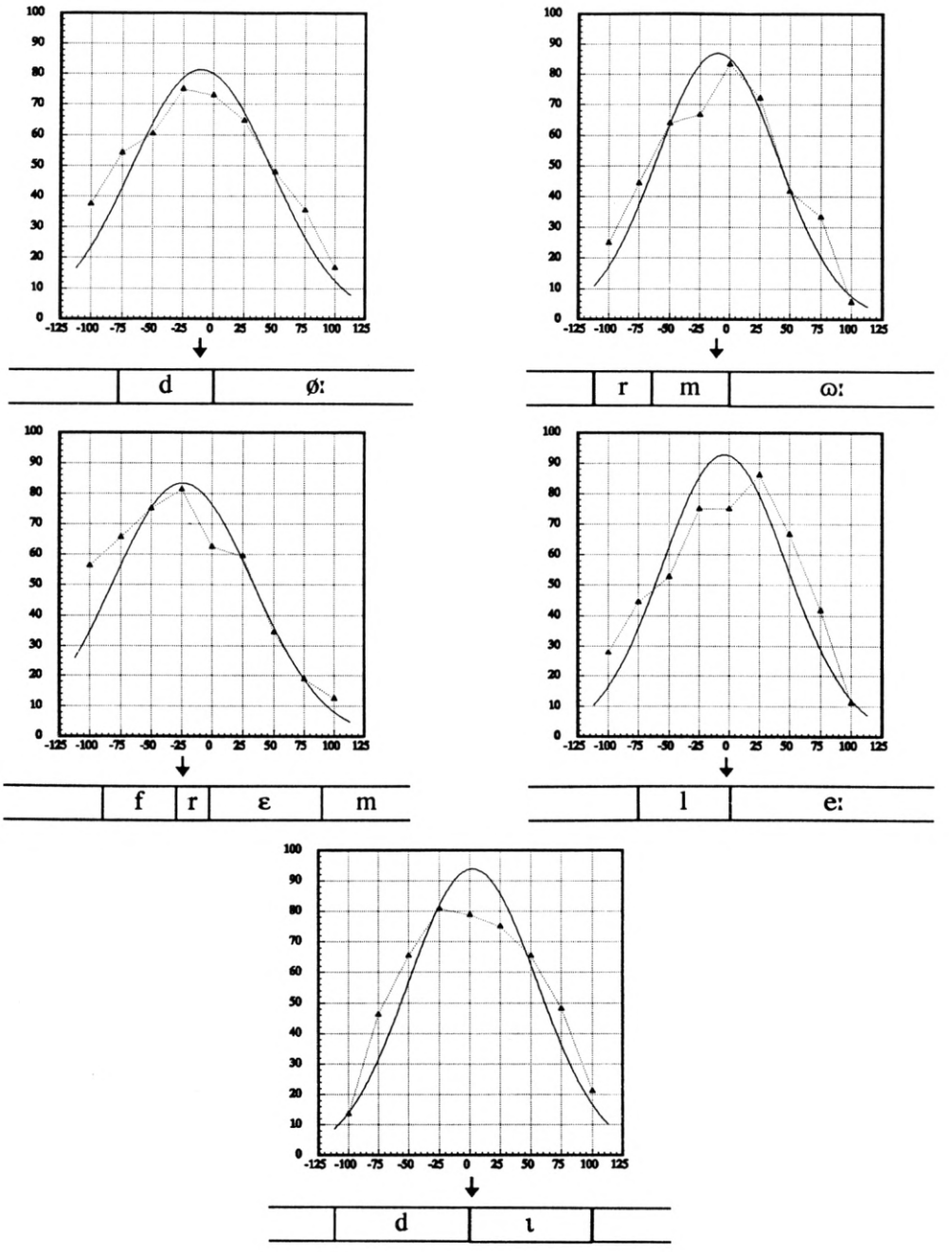

Figure 6.3 The distribution of answers meaning that a click was perceived as coinciding with a given syllable. Normal distribution curves with the same means and standard deviations as the response distributions have been overlaid for reference. The vertical divisions in the transcriptions below each diagram indicate the beginnings and ends of the phonemes using the same time scale as that of the abscissas in the diagrams. Mean click placements are indicated with arrows.

narrow since the number of 'right'-answers does not fall to zero at the end points of the interval. A closer inspection of the individual results revealed a great variation in the ability to decide if the clicks and syllables were simultaneous or not. For some of the subjects, an interval of 200 ms was simply not wide enough. For these subjects, the 'right'-answers were also concentrated around the onset of the stressed vowel but they were a lot more uncertain and judged some of the clicks at the end points of the interval to be 'O.K.' too. There are different ways of dealing with this problem. One would be simply to exclude the results from the subjects for whom the interval seemed too small. A separate calculation using only the results of the 'worst' performers, however, showed that including their results introduced no systematic bias with respect to click placement. The only effect was to increase the standard deviation. Since one aim of this study was to be able to say something about the precision with which subjects in general are able to place the stress beats (e.g. some function of the standard deviation) it was decided to include all the results in the study. With the results at hand, it thus seems as if the prediction, based on the results of the pilot studies using myself as a subject that an interval of 200 ms around the vowel onset would be quite sufficient was a bit over optimistic. For many of the subjects the interval was large enough, but not for all of them.

The fact that some subjects would have required a larger interval than the one actually used, if one wants the scores to fall to zero at the end points, implies some extra caution in treating the results. If there is no systematic bias then the weighed averages should give a fairly accurate picture of the mean click placements. The standard deviations, however, may turn out to be underestimated using this method, because values at the end points are 'missing'. To take care of this potential problem, I have used a different method to calculate means and standard deviations. Assuming the distribution of 'right' judgements to be normal, the cumulative frequency distribution should follow a normal ogive. If the frequencies are transformed to z-scores, the result, assuming normality, is a linear function of the click placements. If the fit between a regression line and the z-scores is good, this also provides a rough check of the normality assumption. Mean click placements and standard deviations calculated following the procedure just described are shown in Table 6.1. The regression coefficients show that the fit between data and the regression line is quite good. To provide a check of the method, calculations of the averages and standard deviations presented in Table 6.1 and all subsequent presentations have also been made using a simple weighed means procedure. A comparison of the results using different methods supports the prediction that means should be approximately equal while standard deviations may differ. A paired t-test of the differences between the calculated click placements using the two methods shows that the mean difference in click placement is only .1 ms and the difference is, of course, far from significant. Standard deviations, however, do differ in the predicted way. The method of weighed averages results in standard deviations that, on the average, are 5.6 ms lower. The difference is significant

Table 6.1. The number of times, for each click placement, that the click was perceived as coinciding with the stressed syllable. Each stimulus was presented four times. Between 9 and 13 subjects (N) participated in each experiment. The results in the table are thus based on 32—52 judgements per click placement. M is the mean click placement and SD the standard deviation.

	% 'right' answers for different click placements												
Syll. #	−100	−75	−50	−25	0	+25	+50	+75	+100	N	M	SD	r
2	38	54	60	75	73	65	47	35	17	12	−10.3	57.0	.999
3	25	44	64	67	83	72	42	33	6	9	−10.2	48.7	.996
4	56	66	75	81	63	59	34	19	13	8	−24.2	57.4	.999
5	28	44	53	75	75	86	67	42	11	9	− 4.0	51.7	.995
6	13	46	65	81	79	75	65	48	21	13	+2.4	51.3	.996

(paired t-test, $t = 11.15$, $df = 50$, $P < .001$). All results presented in this section have been calculated using the method of regression of z-scores on click placement.

The results confirm the results of other experiments showing that subjects place stress beats at, or very near, the onsets of the nuclear vowels. An analysis of individual mean click placements reveals no significant inter-subject differences (ANOVA, $P = .123$). There are differences, however, and on a larger data base they may turn out to be significant, but in this context they are not. The groups are not quite equal for all syllables. A subgroup of the six subjects who took part in all five sessions will be singled out and treated separately. If the groups tested here are considered as equally representative, one may test for significance the differences in click placement for the different syllables using the Tukey-B test for multiple comparisons. Using this test, the only two differences that turn out to be significant ($P < .05$) are syllable 4 vs. syllable 5, and syllable 4 vs. syllable 6. To summarize the results of this first analysis, one may say as a first approximation that subjects do not seem to differ significantly in their perception of where the stress beats are. The beats are located very near the vowel onsets and placements relative to the onsets differ significantly in only two comparisons out of ten possible ones.

Using the standard deviation as a measure of the precision with which the beats may be located, the situation is somewhat the opposite of that concerning beat locations. There are no significant differences between the standard deviations for the different syllables. There are, however, highly significant differences between subjects. These figures tell us something about the ability to locate stress beats as a function of the individual subject and will, therefore, be analysed in more detail. The results are summarized in Table 6.2. To facilitate comparison, the subjects are ranked in ascending order by their standard deviations.

Two important observations may be made. First, the range of values is very wide; 35.1 ms to 71.0 ms. The mean standard deviation for the 'best' subject is only half of that of the least able one. Another interesting observation is that the variation in standard deviation for a specific subject is generally very low. What this means, if these results are repre-

Table 6.2. The mean standard deviations of click placements for the individual subjects (SD) and the standard deviations of the distributions of standard deviations (sd). 'N' is the number of tests that an individual subject participated in.

Subject	SD	sd	N
7	35.1	5.5	5
6	41.0	4.0	5
5	44.0	2.5	3
1	44.5	5.0	3
2	45.4	7.9	2
3	46.5	1.6	5
13	52.9	10.3	3
4	54.0	9.8	4
12	59.4	8.6	5
8	60.0	10.5	5
9	65.5	4.6	4
11	66.6	5.6	5
10	71.0	4.7	2
Mean	52.5	12.3	51

sentative, is that the ability to determine the location of stress beats is a reasonably stable function for an individual subject. Using a rather conservative measure of difference like the Tukey-B procedure, differences of 18 ms or more are generally significant ($P < .05$). In this case it means that about one third of all possible inter-subject differences are significant.

6.3.2 Correlation between stress beat locations and the durations of vowels and prevocalic consonants.

The analysis in the previous section was made with the results from test groups which were not identical. In the following sections, I will analyse the results with respect to the relations between absolute stress beat locations and vowel and consonant durations and also look at the results from the point of view of regularization and p-centre locations. To eliminate the possible source of error it means to compare unequal groups, I have singled out the results of the six subjects who took part in all five tests for these analyses. I will begin the analysis by giving a brief description of this subgroup.

The mean click placements for the subjects in the subgroup and all syllables are presented in Table 6.3. As might be expected, the results do not differ in any dramatic way from those obtained by pooling the results from all subjects. It can be seen that click placements rank exactly the same with respect to the different syllables and if 'click placement' is tested using the Tukey-B test for multiple comparisons, again only two differences turn out to be

Table 6.3. Mean click placements, in milliseconds, relative to the vowel onsets in the stressed syllables (reference) for those subjects who took part in all the tests.

Syllable	DÖ	MO	FRÄ	LE	DI
Reference	0.0	0.0	0.0	0.0	0.0
Subject					
3	–21.9	–28.0	–34.4	–11.5	+2.6
6	–27.7	–3.4	–41.1	–29.4	–22.6
7	–47.2	–12.5	–72.0	+36.7	+16.4
8	–35.8	–14.2	–25.0	–0.6	–3.0
11	–0.0	+14.3	–18.5	–5.2	+8.9
12	–6.3	–24.3	–32.7	–0.6	+28.1
Subgroup	–19.9	–15.1	–33.6	–7.4	+3.6
Whole group	–10.3	–10.2	–24.1	–4.0	+2.4

significant. As with the whole group, the differences in click placement between syllable 4 {FRÄ} and syllables 5 {LE} and 6 {DI} are significant at the .05 level. There are no significant inter-subject differences in mean click placement. There are, however, significant differences in standard deviations, with subjects 3, 6, and 7 forming a group which is significantly 'better' than the other three.

It has been proposed that there is a correlation between the durations of the consonant(s) preceding the stressed vowel and stress beat location (Rapp, 1971; Allen, 1972). Vowel duration has also been suggested as having an influence (Fox and Lehiste, 1987). Both these possibilities were considered with respect to the results obtained above. To see whether click placement could be correlated with consonant duration in this experiment, durations of prevocalic consonants were tested against click placement. Two cases were tested, depending upon whether one regards only the consonant immediately preceding the vowel or the whole cluster in the case of more than one consonant. Table 6.4 shows these relations for the two possible cases.

Allen (1972) found a positive correlation between mean click precession and the duration of the initial consonants. The longer the duration of the initial consonant, the earlier the click seemed to be placed. Allen concluded that *"initial consonant length is a good predictor of mean click precession for the stressed syllables in these utterances"* (p. 99).

The correlation coefficients (Spearman) were not high, however; .51, and .60 with .50 being the limit of significance for $P < .05$. For the tapping experiment the correlations were even lower. Given those results, I would hesitate to regard consonant duration as a *good* predictor, even if there seems to be a weakly significant correlation.

As for the results in this study, the rank correlations (Spearman) in the two cases examined above are –.7 for case 1 and –.3 for case 2. None of these correlations are significant at the

Table 6.4. The durations of the stressed vowels and the consonants or consonant clusters preceding them. Durations are given in milliseconds. 'Click. place.' is the time in ms by which the average click placements precede the onsets of the vowels.

Case 1.					
Consonant(s)	[d]øː	[m]ɷː	[r]ɛ	[l]eː	[d]ɪ
Duration	79	63	26	75	110
Case 2.					
Consonant(s)	[d]øː	[rm]ɷː	[fr]ɛ	[l]eː	[d]ɪ
Duration	79	109	86	75	110
Vowel dur.	236	214	91	240	100
Click prec.	20	15	34	7	–4

.05 level, because of the small number of cases. The correlation in case 1 (single consonants) is nearly significant, however, but the correlation is in the opposite direction of that suggested in the studies by Allen and by Rapp.

A regression analysis, using the same data again gives a very low correlation for case 2 (Pearson, r = –.41, P = .25), but for case 1 the correlation is significant (r = –.92, P = .01) indicating that the shorter the consonant the earlier the perceived beat. This is exactly the opposite of Allen's results. A graphic representation of mean click precession as a function of consonant duration may be found in Figure 6.4.

The slope of the regression line is –.43. I have made the same calculations using Allen's data and found corresponding slopes of .36 and .46 for the data in the click placement experiment. For Rapp's data, the coefficient is approximately .42 (as judged from a diagram). That is, the slopes have almost exactly the same magnitude but in the opposite direction. I can offer no explanation for this difference other than that the dependency on consonant duration, if it is a reality at all, may be outweighed entirely by other factors. But this question needs some further investigation. It should also be pointed out that the differences in click placement that we are considering here are very small and in most cases, as I pointed out above, not significant.

Fox and Lehiste (1987) found click placement to be correlated with vowel duration. They found that the longer the duration of the vowel, the later the click tended to be placed. No such correlation can be found in the data presented here. A rank order test (Spearman) gives a rank order coefficient of only –.3 and a regression analysis yields a regression coefficient (Pearson) of –.09. These correlations are, of course, not significant.

Figure 6.5 summarizes, in a graphical form, the relation between mean click placements and segment onsets.

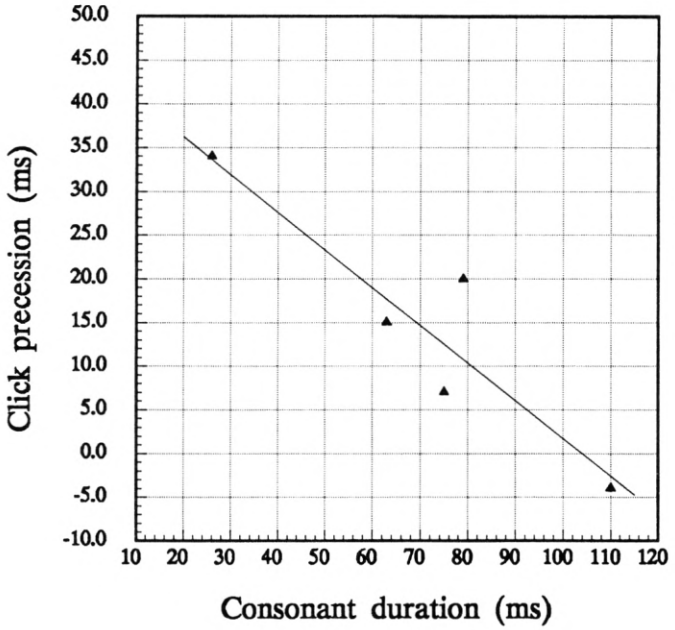

Figure 6.4. Mean click precession as a function of the duration of the prevocalic consonant (r = –.92).

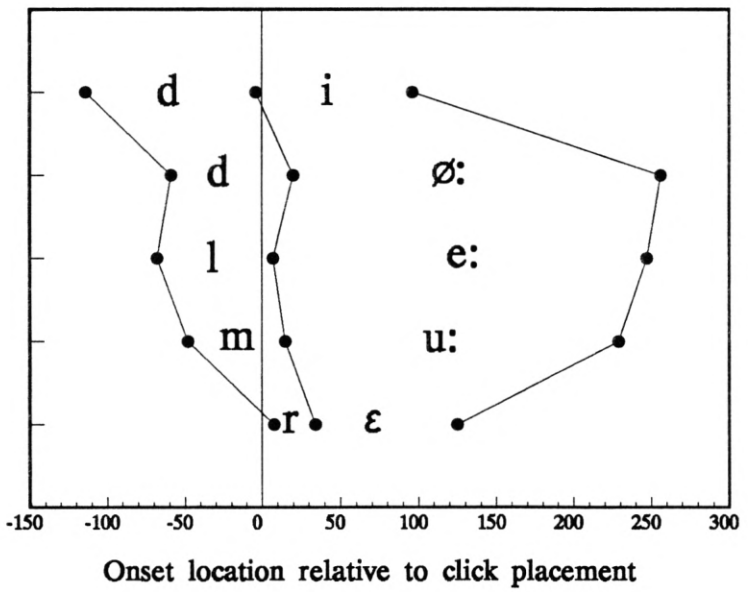

Figure 6.5. Locations of segment boundaries relative to mean click placements for the different stressed syllables. The syllables are ordered in descending order according to consonant durations.

6.3.3 Perceptual regularization.

Mean click placements can be used to calculate subjective interstress intervals. This has been done in Table 6.5. It was predicted that if any regularization in perception occurs, it would first of all have the effect of decreasing the longest interstress interval and increasing the shortest one. Looking at the values in Table 6.5, the impression is that some regularization has taken place. The calculated duration of the longest interval is shorter than the measured duration (855 ms vs. 873 ms) for all subjects, and the shortest interval is longer than the measured one (586 ms vs. 560 ms) for all subjects.

The impression is, thus, that the results lend some support to the idea that the rhythm is regularized. The difference is very small, however, and far from enough to result in isochrony. The difference is only 18 ms for the longest interval and 26 ms for the shortest one. If the average interval duration is used as the target value of perfect regularization, then a shortening of the longest interval by some 160 ms would have been required. From this point of view, the increased regularity is totally insignificant.

Two test-variables that have been proposed to compare and test regularity were described in 3.7 (Donovan and Darwin, 1979; Scott, Isard, and Boysson-Bardies, 1985). One of these variables, that by Scott *et al.*, was:

$$t = \sum_{1 \le i < j \le n} \left| \ln \frac{d_i}{d_j} \right|$$

The distribution of this variable is not known and it is not easy to interpret the results of a test using it. But assuming normal distribution of the scores they differ significantly from the score obtained from the measured durations ($P < .05$). The results may, thus, be interpreted as supporting the idea that the perceived stress beat locations form a more

Table 6.5. Interstress interval durations based on the subjective stress beat placements. The last column is the test scores for regularity based on the formula proposed by Scott *et al.*

Subject					
3	727	867	583	741	1.2096
6	757	836	572	733	1.1707
7	767	814	669	706	.6714
8	754	863	585	724	1.2070
11	747	841	573	741	1.1592
12	708	865	592	724	1.1600
All	737	855	586	737	1.1333
Measured	733	873	560	726	1.3416
Difference	+4	−18	+26	+11	

regular sequence than the vowel onsets do. Two things must be stressed, however. The change in regularity, even if it may be significant, is very small and nowhere near what would be required to achieve isochrony. Also, in view of the limited material involved in the test, only one phrase, it is entirely possible that the effect may be caused by other factors that just happen to work in the direction of greater regularity in this particular case. But more importantly, the precise significance of the test variable has not been examined thoroughly enough for any very definite conclusions to be drawn.

6.3.4 P-centres.

The theory of p-centres is based on data from experiments with lists of isolated words or syllables. The formula proposed by Marcus (1981) to account for relative p-centre locations is explicitly meant to be about isolated words. In a discussion following the presentation of the results, Marcus claims, however, that "*the simple acoustic model ... is of far greater generality*" (p. 253). He proposes that the differential version of his equation should hold also for continuous speech (see 3.4). It is not possible to test the validity of this claim on the material in his study. The formula used for isolated syllables could be applied, though, if one regards the stressed syllables as the 'isolated words' in Marcus' formula. Using the parameter values reported by Marcus, the formula would be:

$$P = -0.35*C + 0.25*V + k$$

where 'C' is the duration of the consonants preceding the stressed vowel, 'V' is the duration of the vowel and the following consonants and 'k' is an arbitrary constant. Since the absolute locations of the p-centres are not important but only the relative ones, in connection with the question of regularity, one may assume the constant 'k' to be 0, to facilitate calculations. The p-centre adjustments relative to the vowel onsets can now be calculated. For the five stressed syllables they are, 46, 42, 67, 49 and 50 ms respectively. The first observation one makes is that the relative adjustments are very small. The changes in interstress interval durations they will introduce are in the order of 10—20 ms, which is far too little to make any important change in regularity. Table 6.6 shows the durations of the intervals in the phrase calculated from three different measures; vowel to vowel onsets, stress beat locations and relative p-centre locations.

Table 6.6. Interstress interval durations, in milliseconds, calculated from vowel-to-vowel onsets, stress beat locations, and p-centres using the formula put forward by Marcus.

Onset-to-onset	733	873	560	726
Stress beats	737	855	586	737
P-centres	729	898	542	727

It is clear from the values in Table 6.6 that using the p-centre values does not produce a more isochronous sequence. Applying the regularity measure used above, the p-centre locations turn out to be the least regular. It is not obvious, of course, that the method I used is the relevant one, but other alternatives produce similar results. If one uses a morphological definition of syllables and treats the stressed syllables as isolated syllables, the resulting intervals turn out as slightly more regular than stress beat locations. If, on the other hand, one regards the words, of which the stressed syllables form a part, as isolated then the p-centre locations form a sequence which is slightly less regular than the stress beats. It must be said, however, that the differences are all the time very marginal. The greatest deviation from measured values in any interval under any of the above mentioned alternative procedures is only 42 ms.

The hypothesis of isochrony of p-centres obviously finds little support in this experiment. It must be pointed out however that the theory has not yet been developed for the analysis of continuous speech.

6.4. Discussion and conclusions.

Stress beats were studied from two points of view; the precision with which they may be determined, based on the standard deviations of click placements, and their absolute locations.

With respect to the question of with what kind of precision it is possible for a listener to pin-point stress beats, using perception only, the following conclusions may now be drawn. If we use the 'difference limen' (DL = .67*SD) as the measure, then the DLs for the subjects in this study range between 23.7 ms and 47.9 ms with a mean of 35.4 ms. This means that larger deviations than that have a better than 50 % chance of being detected. For the six 'best' subjects the average DL is only 28.8 ms. If this is compared with typical variation in interstress interval duration, it seems clear that perceptual isochrony, if such a thing indeed exists, cannot be due to an inability to delimit the intervals accurately enough. Compared to the average interstress interval duration (700 ms) for the material used in this experiment, the uncertainty in determining the end points of an interval is only about 4% of the duration of an interval. Also compared to the variation in interstress interval duration (SD = 107 ms, Range = 313 ms, cf. also Chapter 2), the precision in determining interstress interval end points is high. If these results are representative it means that an impression of regularity or isochrony is not likely to be due to an inability to delimit the intervals to be judged with great enough precision but must have some other cause.

The inter-individual variation, with respect to the ability to place stress beats accurately, reported by other investigators is confirmed. With respect to stress beat locations, on the other hand, there were no significant inter-individual differences. This agrees with the results from one of Allen's experiments where he used the same method. The standard

deviations are somewhat greater than those in Rapp's study (35 ms to 71 ms vs. 20 to 45 ms), and also greater than those in Allen's tapping experiment (28 ms to 42 ms, as calculated from the reported figures). This lends support to my suggestion that the precision in some of the production experiments may be a slightly exaggerated measure of perceptual precision. The great uncertainty displayed by many of the subjects in Allen's click matching experiment was not confirmed, however. For his subjects, an interval of 600 ms seemed insufficient. There is no evidence in this study that this should be the case. The interval of 200 ms used here was too small for some of the subjects in the sense that there were some responses also at the end points of the interval. But even for these subjects, the responses were by no means randomly scattered, but clearly centred around some mean near the vowel onset.

The correlations between stress beat locations and the durations of vowels and prevocalic consonants reported by others, found very little support in this study. The only significant correlation found was that between single consonants and click placements. And it must be noted that the dependency is in the opposite direction of the results reported by others. Considering these results and the generally low correlations and great variations found in this and other studies, one must conclude that the claims of precise dependencies between click placement and vowel or consonant durations seem to rest on somewhat shaky ground. More studies are needed to clarify this point.

The results found in this study may be interpreted as some support, although weak, for the regularization effect reported in other types of studies (Donovan and Darwin, 1979; Darwin and Donovan, 1980). The effect is there in the sense that the intervals calculated from subjects' stress beat locations are more regular than the measured onset to onset intervals. If regularity is tested using the test variable proposed by Scott, Isard, and Boysson-Bardies (1985), subjectively determined intervals are found to be more regular than the physical ones. But there are several reservations one must make. First of all, the adjustment is very small compared to the adjustment needed to come anywhere near isochrony. From this point of view, it is totally insignificant. Secondly, there are other possible explanations for the result. If one assumes that the results reported in 1.6 on rhythmic anticipation may be present also when the subjects are not themselves producing the 'beats', but only judging them, then the result obtained above would be expected since an expectation based on the first two intervals would predict the third beat as coming earlier than it does. If such an effect is present it would, of course, have as a result that the sequence of acoustic events would be regularized in some sense. But the mechanism behind it would be of a rather different nature from the one usually proposed and it should not be restricted to speech stimuli. More experiments need to be done to resolve this question.

Applying the formula for p-centres to the material used here did not produce a more regular sequence than did vowel to vowel onsets or stress beat locations. It should be pointed out, however, that there is no developed formalism for p-centres in continuous speech.

The general conclusion I would like to draw from this experiment is that the hypothesis that stress beats are closely connected with vowel onsets in the stressed syllables finds rather strong support. The precision with which subjects manage to place stress beats perceptually is also quite good. The other results mean slight modifications in beat placement but it should be noted that first of all, the modifications are very small compared to ISI durations, and secondly, the causes of the modifications are far from obvious. Click placement was significantly correlated with consonant duration. Stress beats were somewhat more regularly spaced than vowel onsets. But although the correlations were significant in this study and in others where similar correlations have been observed, I still feel far from convinced that the correlations must necessarily be interpreted as causal relationships. I think it is best, awaiting further results, to keep open the possibility that these finer details are no more than artefacts of less than well designed and understood experiments.

Chapter 7

Perceptual estimation of interstress interval durations.

7.1 Introduction.

In this chapter, I will report the results of a series of 8 experiments designed to gain some insight into the precision with which our perceptual system permits us to judge the durations of interstress intervals in speech. In the previous chapter, the precision with which interstress intervals can be delimited was studied. It was shown that this precision is quite good. For the six 'best' subjects, the difference limen was just under 30 ms. This difference limen is quite small compared to typical interstress interval durations. This does not necessarily mean that their durations may be perceived with any corresponding accuracy. In fact some authors (e.g. Lehiste, 1977) have suggested that the inability to perceive the differences in interstress interval duration accurately is at least one of the reasons why some languages (e.g. English) have been claimed to be isochronous. In her study subjects were to judge a series of four intervals and say which was the longest or shortest. They managed to do so in many cases but there were also many errors and Lehiste concluded that the results were closer to chance than to perfect discrimination. This question needs to be studied further, however, before any more definite conclusions may be drawn. In particular, one must try to determine in quantitative terms the 'just noticeable differences' for duration perception at this level. The study presented in this chapter is an attempt to do just that.

As a measure of the just noticeable difference, standard deviations and differential fractions will be used. The relevance of these measures was discussed in 5.1.3. These

measures seem to me the most natural to use in the technical description of the experiments (7.2—7.10). In the discussion of the results in 7.11.5, I will use another concept which is related to the standard deviation, the difference limen. Since the discussion, particularly in connection with isochrony, is often about whether the difference is perceptible or not, the difference limen, expressing the border-line between chance responses and actually perceived differences, will be used in the concluding discussion. The technical procedure used to calculate the differential fractions will be described in 7.4.2.

The series of experiments grew out of an original study comprising the first three experiments described below. The experiments were of three kinds. In one type of experiment, a phrase was presented repeatedly and the task was to rank order the interstress interval durations. This type of task is of course a very difficult one. In addition to the difficulty of correctly perceiving durations in continuous speech, the task also presents a memory difficulty. It is necessary to keep all the different intervals in memory long enough to be able to compare their durations. It was originally thought that this type of task would simply be too difficult and that subjects' performances would be more or less random. As it turned out, however, this did not seem to be the case, even though the results clearly indicated that the task was very difficult. In a second type of experiment, the interstress intervals were 'cut out' from the phrase and presented pairwise in a duration discrimination task. This type of task presents a lower processing load and tells us more about the duration perception of isolated interstress intervals as such. One idea behind the use of this technique was to be able to say something about how much the increased memory load presented by the first type of task would influence the possibility of correctly judging the durations. Another reason for doing this type of experiment was to be able to relate the perception of duration in speech to results from other duration perception experiments done with noise and empty intervals as stimuli. To this end, a control experiment was also designed. The experiment was identical to the discrimination experiment using speech fragments, except that the interstress intervals from the phrase had been replaced by noise pulses with the same durations as the interstress intervals. This provides a check on the experimental technique, including the range of stimuli used and the procedures, and makes it easier to compare the results from this test and the speech discrimination tests with the results on duration perception obtained by others using similar techniques.

The original three experiments were followed by additional experiments to add more data in order to make the analysis more reliable. Rather than vary the material used and the techniques, the same material and techniques were used but with different groups. The group that participated in the first speech discrimination experiment was used in a second experiment with the noise stimuli. It was planned to re-test the first 'noise group' with speech as well but this turned out to be impossible for practical reasons. Instead a group of participants in a seminar at the phonetics department at the University of Stockholm were used as subjects in two additional discrimination tests—one using speech and the other noise.

In addition to the perception of interstress interval durations, two more questions were studied. One of these questions was the problem of time-order error discussed in 5.1.5. This question was studied for two reasons. One reason for including this variable was methodological. It is necessary to bring this variable under control in experiments of this kind. But there are also results that indicate that this effect may be different for speech stimuli than for other types of stimuli. In the experiment by Lehiste (1976), reported in 5.1.6, where she carried out a duration discrimination test using vowel like stimuli, there was a considerable effect of TOE. For equal durations, the bias was as great as 68.7% against 31.3% in favour of the first presented stimulus in a pair. One implication of her result is that there may be a dramatic increase in TOE when speech material is used compared to the much lower values found for noise stimuli. In non-speech material, TOE is not always present at all (e.g. Small and Campbell, 1962) and when it is, it is not usually of that order. Stott (1935) for example obtained TOEs of around 56—44% in experiments with comparable durations, which is considerably lower than the TOE in Lehiste's experiment. One reason for the difference may be a stronger expectation of final lengthening when speech material is involved. It was, therefore, thought that by studying TOE in the experiments carried out here, some light might be thrown on this question.

Another question, which will be touched upon indirectly is the question of speech mode perception. As was discussed in section 3.5, some authors (e.g. Darwin and Donovan, 1980) have suggested that the 'illusion of isochrony' is a phenomenon restricted to speech and that it may be the case that we perceive speech in a different way from other 'sounds'. But others (Bell and Fowler, 1984; Scott, Isard, and Boysson-Bardies, 1985) have questioned this view and suggested that it may be the difference in complexity between non-speech stimuli (often noise) and speech stimuli which is responsible for the difference. To approach questions of this kind, one would like to be able to compare speech stimuli with non-speech stimuli of comparable complexity. Now, speech sounds are complex at many different levels. Not only are they acoustically complex. They also contain semantic and syntactic information, not present in non-speech stimuli. An interesting question is, therefore, how it would affect our perception of durations in speech if these particular aspects were removed or obscured, but without affecting the level of acoustic complexity. To do so, reiterant speech or sequences of monosyllables are used in many experiments aimed at studying prosodic properties of speech. In most of the experiments on p-centres described above, this type of material was used. Other possibilities are to use non-speech analogues (Darwin and Donovan, 1980; Bell and Fowler, 1984) or to 'degrade' recorded speech in some way so that it becomes unintelligible (Scott, Isard, and Boysson-Bardies, 1985). The idea behind using acoustically distorted speech is that the full acoustical complexity of the speech is preserved. Still another possibility that seems useful, at least for some types of prosodic studies, would be to use real speech, but from a language not understood by the subjects. Given that the experimental task is such that what is to be judged is not supposed to be a function of the listeners' native language this technique could be used. It seems

reasonable to assume that duration discrimination is a rather general ability that should not be dependent upon a subject's native language. One way of testing duration perception of speech-like stimuli would, therefore, be to use stimuli from a language not understood by the subjects.

During a visit to Holland, I had the opportunity to give two of the tests to two groups of Dutch students. None of the students had any knowledge of Swedish. It was thought that this would provide an opportunity to study the perception of duration in the material used in the previous experiments in identical tests, with all the acoustic properties preserved but with semantic information removed. This would provide some insight into whether the semantic information, which Swedish subjects will inevitably perceive and probably process to some extent, degrades the duration perception. It was predicted that the Dutch subjects, for whom the stimuli were more like 'speech like noise' than actual speech, might have an easier task in processing the durational properties and thus perform somewhat better.

The different experiments will be described in the order they were performed. In connection with the presentation of each experiment, the result from the experiment will be presented and analysed. The first three experiments, which comprised the original study, will be analysed in more detail. In connection with the description of these experiments, the preparation of stimuli as well as the experimental procedures will be described. The following experiments (4—8) which were carried out to obtain more data, but with the same material and using the same procedures, will be reported in a more summary form, emphasizing the presentation of results. Some parts of the analysis which involve the comparison of results from different experimental groups and experimental conditions will be postponed until all experiments have been presented. In section 7.11 the results will be summarized and further analysis will be carried out.

7.2 A description of the stimulus material used in the experiments.

In the two types of experiments where speech was used, stimuli were taken from the phrase used in the stress beat experiment described in Chapter 6. Figure 7.1 shows the syllabic and durational structure of the phrase. As can be seen, 4 of the 6 interstress interval are trisyllabic and the remaining two disyllabic. There is, thus, slightly less variation than one normally finds in speech. That is one of the reasons why this particular phrase was chosen. When listened to, it gives an impression of temporal regularity and it was, therefore, thought to be a suitable test case. If the temporal irregularities can be detected in this fairly regular phrase they should also be possible to detect in other types of speech which are often more irregular than this phrase.

In one type of experiment, the whole phrase was used as stimulus. In a second type, the different interstress intervals were cut out 'electronically' from the phrase and presented

ISI #	Structure			Tot. duration
1	(h) **ärl**	**ig**	**ed**	
	338	137	211	685
2	**öd**	**en**	**ärm**	
	295	227	210	733
3	**od**	**ikt**	**ifr**	
	322	298	253	873
4	**ämst**	**al**		
	388	173		560
5	**ed**	**et**	**ud**	
	300	225	200	726
6	**ign**	**ar**		
	355	266		622

Figure 7.1. The syllable structure and temporal structure of the test phrase, "Härlig är döden när modigt i främsta ledet du dignar", used in the experiments. The figures indicate durations of intervals and syllables in milliseconds.

pairwise. How this was done will be described in more detail in connection with experiment 2. In a third type of experiment, noise was used. The noise used was white noise that had been filtered to obtain approximately the same average frequency spectrum as that of the test phrase. For details the reader is referred to the description of experiment 3.

The interstress intervals differed in a number of ways, of course, except for their durations. One variable that may have an influence on the perception of duration is loudness. It is not obvious how this should be measured to say something about duration perception. In experiments done on duration perception where loudness has been the variable, the level throughout a given stimulus has always been constant. In speech stimuli this is not the case. There is a constant fluctuation in intensity level. The figures given below refer to the relative average RMS-levels for each stimulus. The figures are normalized, assuming a 60 dB average level for the phrase.

Table 7.1. Loudness of the speech fragments used in the discrimination tests. The loudness values are the mean RMS values for the whole stimulus normalized to an assumed mean level of 60 dB for the whole phrase.

ISI	1	2	3	4	5	6
Loudness	59	71	62	65	54	49

7.3 Experiment 1.

In this experiment, the subjects heard the entire test phrase. The task was to rank the interstress intervals of the phrase according to their durations.

7.3.1 Method.

Subjects

The subjects were students of speech therapy at the University of Gothenburg. The experiment was carried out as a part of a laboratory course. The students were in their second year of study. They thus had some general knowledge of linguistics. None of them, however, had any knowledge of the particular aspects of speech tested here. Participation was voluntary and unpaid.

Stimuli

The phrase described above served as stimulus for the experiment. It was available in a digitized version that had been sampled into a computer in connection with the stress beat experiment described in Chapter 6 (20 kHz sampling frequency). The phrase was re-synthesized and copied back onto an open reel tape (Revox B 77). On the final test tape, the phrase was repeated, first 4 times with 3 seconds between repetitions and then 20 times at 20 second intervals, the idea being that subjects should use the first 4 repetitions to get acquainted with the phrase and make a preliminary ranking of the interstress interval durations and then use the following repetitions to check and correct their rankings if necessary. The total duration of the test tape was 8 minutes. (The preparation of the test tapes used in this and all subsequent experiments was made on a DG, Eclipse S/200 computer using the MIX program written by Rolf Carlson, Royal Institute of Technology.)

Procedure

Before the test, the subjects were informed about the purpose of the experiment. They were told that the aim was to see if it were possible to rank correctly the interstress intervals in a given phrase according to their physical durations. The test was carried out as a 'forced choice' test. That is the subjects had to rank the intervals even if they regarded them as being more or less equal. Each subject was given an answer form on which the test phrase was printed. The stressed syllables had been clearly marked. The subjects were told to mark the ranks on the answer form below the respective interstress intervals using the numbers 1 to 6. Number 1 was to be used to mark the longest and 6 the shortest interval.

The test was carried out in a soundproof perception lab where the stimuli were presented via headphones. Several pilot studies and demonstrations led by the author had indicated that it is next to impossible to set a level of loudness in a headphone presentation that is regarded as optimal by all subjects. There are two ways to deal with this problem. One way

is to decide upon a level on some 'objective' ground and force subjects to accept this level whether they like it or not. This would have the advantage of standardizing the presentation level, but the risk is of course that subjects are seriously disturbed by a level felt to be either too high or to low. The other possibility is to let subjects decide the level individually. A sort of compromise between the two approaches was chosen and used in this and all subsequent headphone presentation. The level was adjusted to an initial level of approximately 70 dB for all phones. Each phone had an individual level control by which the listening level could be adjusted, up or down, by approximately 5 dB. Subjects were told to use these controls to adjust the level to a level they felt to be optimal, if they were not satisfied with the original setting. Recorded speech was used in this adjustment process. Some, but not all, of the subjects used this possibility.

For practical reasons a time limit of 8 minutes was set for the test. The subjects were given the instruction that they had completed the test as soon as they had decided upon a certain rank ordering and did not feel that they could improve it any further.

7.3.2 Results.

The first observation was that subjects decided quite rapidly how they wanted to rank the interstress intervals. None of them felt that they needed to hear all 24 repetitions of the phrase. They had all made up their minds after the initial 4 repetitions and another 3 or 4. Table 7.2 shows how they ranked the interstress intervals.

Table 7.2. The subjective rankings of interstress interval durations. For ease of comparison the intervals are ordered according to their physical durations; ISI #3 being the longest.

	ISI					
	3	2	5	1	6	4
Subject						
1	3	2	1	4	5	6
2	1	4	5	3	6	2
3	1	5	4	2	3	6
4	2	1	5	4	6	3
5	5	2	4	3	1	6
6	3	1	2	4	6	5
7	1	2	4	5	6	3
8	4	1	5	3	6	2
Mean rank	2.50	2.25	3.75	3.50	4.88	4.13

An inspection of the results indicates that the task was indeed a difficult one. The results are very varied. But they are by no means random as might have been expected. Two statistical tests were used to analyse the results. The Kendall Coefficient of Concordance (W) was used to test inter-subject agreement and the T_c-test to test agreement between responses and a given criterion—the criterion in this case being the ranks obtained by using the measured durations. (Both tests are described in Siegel and Castellan, 1988.)

The concordance test applied to the results of the whole group gives the following result:

$$W = .280 \qquad \chi^2 = 11.214 \qquad 5 \text{ df} \qquad P = .0473 \qquad \text{ave}(r_S) = .1776$$

The meaning of 'ave(r_S)' here and in the subsequent analyses is the average value of the Spearman rank-order coefficients between all possible pairs of rankings. There is fair agreement between subjects ($P < .05$) but the average of the correlations is rather low, indicating that individual subjects show a great deal of uncertainty. (An inspection of the individual correlations shows the performances of subjects 5 and 8 to be nearly random. With these subjects excluded, the significance level falls to .023 and the ave(r_S) increases to .322)

A test of agreement between the subjects' rankings and the 'criterion' (= the physical durations) gives a higher degree of agreement.

$$T_C = .333 \qquad z = 2.590 \qquad P = .0048$$

T_C is the average of the Kendall rank-order correlation coefficients between each ranker and the criterion ranking. As can be seen, the agreement with the criterion is significant at the .5 % level.

The results of this first experiment confirm the suspicion that a task of this kind is difficult but also that it is not impossible. There is significant agreement between subjects and also between the subjective rankings and the physical durations. The conclusion must be that it is indeed possible for judges, particularly the better ones, to judge fairly correctly the durations of interstress intervals in a phrase of this type. It should also be pointed out that this is a far more difficult task than merely realizing that the intervals are not equal; something that would be sufficient to question the concept of perceptual isochrony. No subject expressed any doubt about whether the intervals were unequal.

7.4 Experiment 2.

In this experiment, the subjects were confronted with the different interstress intervals from the phrase used in experiment 1, but presented in a discrimination task. The intervals had been cut out of the phrase and were presented pairwise in all possible combinations. In order to counterbalance a possible time-order effect in the total result, all possible combinations of interstress intervals were presented in both orders.

7.4.1 Method.

Subjects

The subjects in this experiment were students from a course in experimental methodology at the department of linguistics at the University of Gothenburg. Six of the subjects were graduate students and two were staff members. Although the general level of linguistic knowledge was high, no one had any prior knowledge of the particular aspects of perception tested in this experiment. Participation was voluntary and unpaid.

Stimuli

Stimuli in the experiment were the interstress intervals from the test phrase used above that had been cut out and were now presented pairwise with 1 second between the durations in a pair and 5 seconds between stimulus pairs. Onsets of the stressed vowels were marked in the time-wave and the phrase was cut up into interstress intervals which were then placed in separate files. Care was taken to make 'cuts' at zero-crossings to ensure that there should be no disturbing transient noises in the resulting stimuli.

The order of presentation of the 30 different combinations of intervals was decided with the help of a random number table. There may exist better ways of deciding the order of elements in lists of this type. Ross (1934) and Wherry (1938) have suggested criteria for determining optimum presentation orders and also suggested lists in agreement with their criteria. The criteria are based on even spacing between items and the avoidance of proximity of like items. The lists used in this study are not optimal from this point of view (as analysed the way proposed by Wherry) but the deviations are not serious. The papers cited were not known to the author at the time the lists were constructed. Had they been known the lists may have been constructed in a different way.

The files were read onto a tape in the given order with 1 second between durations and 5 seconds between pairs. An additional list of 7 randomly chosen pairs preceded the actual test sequence. These 7 stimuli acted as a pre-test trial sequence. The sequence of stimuli in the test was divided up into groups of 12 stimuli with a 1 kHz tone marker between groups to facilitate for the subjects to keep track of stimuli while ticking off their answers on the answer forms. The total duration of the test tape was 8 minutes.

Procedure

Subjects were instructed to compare the durations of the two fragments of speech presented in each pair and tell which was the longer of the two. The test was carried out as a forced choice test. Subjects were given an answer sheet where each stimulus pair was represented by a number and two boxes representing the two stimuli in the pair. They were told to mark with a tick in one of the boxes which of the two stimuli they perceived as the longer one. The experiment was carried out in the same perception lab that was used in experiment 1. The stimuli were presented via headphones. To give the subjects a chance to become familiar with the experimental procedure, the test was begun with a trial round of 7 stimuli after which there was a short pause in order to give the subjects a chance to tell whether they had any problems in understanding the task. Then the test proper was given.

7.4.2 Results.

Estimation of durations

As an indication of how well subjects succeeded in solving the task, their performance score will be used. The score is simply the percentage of correct judgements. The scores will be used in a later section to compare different experimental groups and stimulus types. In this experiment the average score was 74.2 % (SD = 10.2).

By comparing, interval by interval, how the subjects have judged the relative durations in each pair, it is possible to construct a ranking of the subjective durations of the 6 intervals for each of the subjects. If these subjective rankings are compared with the measured durations they can be used as a measure of how well the subjects have succeeded in accurately perceiving the different durations. The rankings for each subject as well as the mean ranking for the group as a whole are given in Table 7.3.

Table 7.3. The subjective rankings of interstress interval durations.

	ISI					
	3	2	5	1	6	4
Subject						
1	1	2.5	2.5	4	6	5
2	1.5	3	4	5	1.5	6
3	1	3	4	2	5	6
4	1	3	4	5	2	6
5	1	3	4	2	5	6
6	1	4	2.5	2.5	6	5
7	1	2.5	4.5	2.5	4.5	6
8	1	2	3	4	6	5
Mean rank	1.06	2.88	3.56	3.38	4.50	5.63

The *W*-test shows a high degree of agreement between subjects.

$$W = .691, \qquad \chi^2 = 27.618, \qquad 5\ df, \qquad P < .001, \qquad ave(r_S) = .646$$

The *W*-test reveals a markedly higher degree of inter-subject agreement than the corresponding test in experiment 1. The concordance is significant at the .001 level. Also agreement with the criterion is highly significant.

$$T_C = .650, \qquad z = 5.114, \qquad P < .001$$

This is not an unexpected result. The most important factor contributing to this result is, of course, that the task is considerably simpler than the ranking task in experiment 1.

It is obvious that the success or failure to discriminate between two stimuli must depend upon the durational contrast between the two. In the following analysis I will try to analyse how much of the discrimination that can be explained by duration.

The difference between the durations of two stimuli can be described in two ways. The difference can be described in absolute terms, as the number of milliseconds by which one duration exceeds the other, or in relative terms, for instance as a percentage using one of the two as the standard of comparison. In Table 7.4, I have compared 3 sets of rankings. I have ranked the different combinations of durations with respect to how many times they were correctly judged and compared that ranking with rankings based upon absolute and relative durational differences between the stimuli.

Table 7.4. Ranking of 'correct discrimination' compared to the rankings of absolute and relative durational differences between the stimuli.

Durations	% correct	Rank	Diff.abs.(ms)	Diff.rel.(%)
873—560	100	1	313 (1)	56 (1)
726—560	97	2	166 (5)	30 (4)
733—560	91	3.5	173 (4)	31 (3)
685—560	91	3.5	125 (8)	22 (6)
873—622	84	5	251 (2)	40 (2)
873—733	78	6.5	140 (7)	19 (8)
733—622	78	6.5	111 (9)	18 (9)
873—726	75	8	147 (6)	20 (7)
873—685	72	9.5	188 (3)	27 (5)
685—622	72	9.5	63 (11)	10 (12)
733—685	69	11	48 (13)	7 (13)
733—726	63	12.5	7 (15)	1 (15)
622—560	63	12.5	62 (12)	11 (11)
726—622	59	14	104 (10)	17 (10)
726—685	44	15	41 (14)	6 (14)

It is now possible to test for correlations between 'correct discrimination' and the two difference measures. Using the Spearman rank correlation test on the respective ranks one obtains the following correlations:

Absolute difference—% correct	r = .78	P = .0034
Relative difference—% correct	r = .85	P = .0015

There seems to be a fairly high correlation between both measures of difference and discrimination, but with a slightly higher correlation for relative difference.

I have also tested for a possible linear relationship between 'correct discrimination' vs. absolute (ms) and relative (%) differences.

Absolute difference—% correct	r = 0.74
Relative difference—% correct	r = 0.83

Again there seems to be a fair degree of correlation. The conclusion that can be drawn from these results is that the ability to discriminate between the durations of two intervals, presented one after the other, depends upon the size of the difference to a fairly high degree, most strongly on the relative difference. It also seems as if a linear dependency would be a reasonable first approximation, particularly for the relative differences. Relative durational contrast seems to explain more than 70 % of the behaviour ($r^2 = .72$). It should be noted, however, that even if as high a proportion as 70% of the behaviour seems to be accounted for by the differences in durations, there is also a fairly high proportion that must have some other explanation.

In the following, and for all the subsequent discrimination experiments, I will compute standard deviations and differential fractions ($\Delta T/T$) for the discriminability of the different stimuli to give a more precise picture of the minimal size of perceptible durational contrasts. If the standard deviation is used as 'ΔT', the differential fraction defines an interval around the base duration, outside of which the probability of correct discrimination is 68% or better, expressed as a fraction of the base (or standard) duration. Since the differential fraction is strongly connected with Weber's law, it is often also referred to as the 'Weber fraction' or 'Weber number'. The connection between Weber's law and the correlations presented above should perhaps be mentioned. It was shown above that the relative differences in duration explain most of the discriminability between stimuli. But this is precisely what Weber's law says, but in a slightly different way. It maintains that the just noticeable difference is a constant proportion of the durations judged, and as was shown above, the relative difference did indeed seem to be the strongest determining factor.

The procedure used to compute the differential fractions from the experimental results is the following. Assuming each one of the six durations in the experiment, one at a time, to be the standard (T) against which the other durations (comparisons) are compared, one may form tables showing how many times the comparison was judged to be longer (CL)

than the standard. These numbers expressed as a probability ($0 \leq P \leq 1$) form a psychometric function whose theoretical distribution is thought to be the same as the cumulative normal distribution (the normal ogive). Therefore, if the probabilities are converted to z-scores and these scores are plotted against the comparison durations, the result will be a straight line. Making these assumptions, linear regression can be used to calculate means and standard deviations (SD) for the discrimination function for each of the base durations. A more detailed description and a theoretical background may be found in Luce and Galanter (1963) and Baird and Noma (1978).

The proportions of answers meaning that the 'comparison' is judged as the longer are presented in Table 7.5. If these figures are regarded as representative, they can be interpreted as the probabilities that the comparison is judged longer in a given context.

Table 7.5. The probabilities that the comparison duration is judged 'longer' (CL).

Ref.	Test	CL	Ref.	Test	CL	Ref.	Test	CL
560	622	.625	622	560	.375	685	560	.094
	685	.906		685	.719		622	.281
	726	.969		726	.594		726	.469
	733	.906		733	.781		733	.688
	873	1.000		873	.844		873	.719
726	560	.031	733	560	.094	873	560	.000
	622	.406		622	.219		622	.156
	685	.531		685	.313		685	.281
	733	.625		726	.375		726	.250
	873	.750		873	.781		733	.218

From these proportions, the differential fractions can be calculated. These are presented in Table 7.6.

Table 7.6. SD and differential fractions ($\Delta T/T$) as a function of the reference durations.

Ref.	560	622	685	726	733	873
SD	88	243	166	142	155	422
$\Delta T/T$	.16	.39	.24	.20	.21	.48

The regression coefficients for the linear regression equations of the z-scores on base durations can be used as a measure of goodness of fit of the data to a normal distribution. The correlation coefficients are presented in Table 7.7. The fit is quite satisfactory for all durations except the longest one (873 ms). For this duration, the regression coefficient is only .654 which means that the results should be interpreted with a certain amount of caution.

Table 7.7. Regression coefficients showing the correlations between the z-scores and the test durations.

r(560) = .897	r(622) = .892	r(685) = .917
r(726) = .852	r(733) = .995	r(873) = .654

Time-order errors

As was pointed out in the introduction, there are methodological reasons for controlling for a possible time-order error. But also, the results obtained by Lehiste (1976) indicate that TOEs may be greater when speech material is used than they normally are with non-speech material. TOEs will, therefore, be given some consideration in the analysis of the experiments in this study. In this experiment (2) and the following (3), TOE will be analysed in some detail. But for the rest of the experiments (4—8), I will only use the time-order variable proposed by Allan and Kristofferson (1974) to analyse TOE (see section 5.1.5 and below).

A detailed account of how all the different combinations were judged is given in Table 7.8. The theoretical distribution, assuming perfect duration discrimination, would be 16—0 (or 0—16) in each combination and 240—240 for the test as a whole. The distribution for the whole test (255—225) is thus slightly biased i favour of the first stimulus. Expressed as percentages, however, the effect is very modest (53%—47%).

This and all subsequent experiments have also been analysed using the 'Allan and Kristofferson variable'.

$$TOE = P(R_{10} \mid S_1S_0) - P(R_{01} \mid S_0S_1)$$

This variable bears a direct relation to the distribution presented above, of course. It can be calculated by simply dividing the difference between 'first answers' and 'second answers' by the total number of answers. The use of it has the advantage, however, of assigning each distribution a number which can be used to test the significance of the differences between different experiments. This will be done in the summary in section 7.11.4. For this experiment, the TOE was found to be +.067 with a standard deviation of .150.

For individual comparisons, there is great variation in TOE. But TOE does not seem to depend on the durational contrast in any systematic way other than in the trivial sense that when discrimination is perfect, as in the case of the greatest contrast (873—560), there is no TOE. For lower contrasts, TOE varies around a mean of +.65. TOE seems to drop in absolute magnitude when discrimination rises above some 80 %. In a general sense this agrees with the results obtained by Stott (1935) for comparable durations.

Table 7.8. The number of times a particular interstress interval has been judged to be the longer in all possible combinations and presentation orders. 'Score' is the number of times each of the two intervals in a pair has received a 'longer' judgement. The order in the 'Durations' column corresponds to the presentation order.

ISI	Durations	Score	ISI	Durations	Score
1—2	685—733	7— 9	2—6	733—622	12— 4
2—1	733—685	13— 3	6—2	622—733	3—13
		20—12			15—17
1—3	685—873	6—10	3—4	873—560	16— 0
3—1	873—685	13— 3	4—3	560—873	0—16
		19—13			16—16
1—4	685—560	16— 0	3—5	873—726	13— 3
4—1	560—685	3—13	5—3	726—873	5—11
		19—13			18—14
1—5	685—726	12— 4	3—6	873—622	14— 2
5—1	726—685	11— 5	6—3	622—873	3—13
		23— 9			17—15
1—6	685—622	10— 6	4—5	560—726	1—15
6—1	622—685	3—13	5—4	726—560	16— 0
		13—19			17—15
2—3	733—873	4—12	4—6	560—622	4—12
3—2	873—733	13— 3	6—4	622—560	8— 8
		17—15			12—20
2—4	733—560	14— 2	5—6	726—622	12— 4
4—2	560—733	1—15	6—5	622—726	9— 7
		15—17			21—11
2—5	733—726	9— 7			
5—2	726—733	5—11		Total: 256—224	
		14—18			

7.5 Experiment 3.

This experiment was the first of the 'noise experiments', designed as a control experiment to be able to compare duration discrimination of speech stimuli with discrimination of noise in otherwise identical tests.

7.5.1 Method.

Subjects

Subjects in this experiment were a group of students in an undergraduate course given at the department of linguistics at the University of Gothenburg. They were of approximately the same age as the subjects in experiment 1 and had a comparable background in linguistics. Participation was voluntary and unpaid.

Stimuli

Stimuli in this experiment were noise pulses. The pulses were made of white noise that had been filtered in order to get the same long-time frequency spectrum as the speech material used in experiment 2 in order to preserve some of the spectral characteristics. The intensities were equal for all stimuli. The durations of the pulses were made exactly the same as the durations of the interstress intervals. In order to prevent transient 'clicks' at the onsets or offsets these were modified using a ramp corresponding to a rise-time (and decay) of approximately 5 ms. The resulting stimuli had no audible distortions connected with the onsets or offsets. A test tape was prepared which was identical to that used in experiment 2, except for the different content of the durations.

Procedure

The experiment was carried out in the same way, and with the same type of instructions as experiment 2.

7.5.2 Results.

Estimation of durations

The performance score in this experiment was 88.1 (SD = 6.1). This is significantly better than the score in the speech discrimination test described above.

Table 7.9 gives the results for the rankings computed for each subject.

Table 7.9. The subjective rankings of the noise pulses according to duration.

	ISI					
	3	2	5	1	6	4
Subject						
1	1	3	2	4	5	6
2	1	2	4	3	5	6
3	1	4	2.5	2.5	5	6
4	1	2	3	4	5	6
5	1	3	2	4	5	6
6	1	3	2	4	5.5	5.5
7	1	2	3	4	5	6
8	1	3	2	4	5	6
Mean rank	1.00	2.75	2.56	3.69	5.06	5.94

The *W*-test shows a high degree of agreement between subjects.

$$W = .927, \qquad \chi^2 = 37.086, \qquad 5\ \text{df}, \qquad P < .001, \qquad \text{ave}(r_S) = .917$$

The *W*-factor is almost 1 and the agreement is highly significant. Also agreement with the criterion is highly significant and considerably higher than in the speech experiment.

$$T_C = .850, \qquad z = 6.709, \qquad P < .001$$

As with the speech experiment, I will analyse the performance as a function of the durational contrasts between stimuli. Table 7.10 shows the rankings of correct discrimination vs. relative and absolute differences.

Table 7.10. Ranking of 'correct discrimination' compared to the rankings of absolute and relative durational differences between the stimuli.

Durations	**% correct**	**Rank**	**Diff.abs.**	**Diff.rel.**
873—560	100	2	313 (1)	56 (1)
873—622	100	2	251 (2)	40 (2)
733—560	100	2	173 (4)	31 (3)
873—726	97	5	147 (6)	20 (7)
726—560	97	5	166 (5)	30 (4)
733—622	97	5	111 (9)	18 (9)
873—733	94	8	140 (7)	19 (8)
873—685	94	8	188 (3)	27 (5)
685—560	94	8	125 (8)	22 (6)
726—622	91	10	104 (10)	17 (10)
685—622	88	11	63 (11)	10 (12)
733—685	78	12	48 (13)	7 (13)
726—685	75	13	41 (14)	6 (14)
622—560	72	14	62 (12)	11 (11)
733—726	47	15	7 (15)	1 (15)

A Spearman rank correlation test using the values in Table 7.10 gives the following result:

Absolute difference—% correct	r = .90 P = .0007
Relative difference—% correct	r = .91 P = .0006

The agreement is very good and, as with the speech data, the relative difference seems to be the best predictor although the difference is now very small. The fit of a regression line is less good which is explained by the high number of like values.

Absolute difference—% correct	r = 0.73
Relative difference—% correct	r = 0.76

The conclusion from these results is the same as that from the corresponding analysis of the speech data; Relative durational differences seem to explain most of the ability to discriminate between durations and this is even more true in the case of noise stimuli. Correct discrimination is almost entirely a function of durational contrast.

The analysis of differential fractions is found in the following three tables. As can be seen in Table 7.12, the differential fractions are considerably lower than in the speech experiment. The fit of the regression line is also closer. The reason why it is not meaningful to make a regression analysis for the longest duration is apparent by looking at the data in Table 7.11. There is too great a contrast between the 'standard' and the comparisons. Almost all subjects were able to discriminate correctly between the standard and comparison in all contexts so there is too little variation in the data. For the rest of the durations, regression lines can be found and the fit is better than for the speech data.

Table 7.11. The probabilities that the comparison duration is judged 'longer' (CL).

Ref.	Test	CL	Ref.	Test	CL	Ref.	Test	CL
560	622	.719	622	560	.281	685	560	.063
	685	.938		685	.875		622	.125
	726	.969		726	.906		726	.750
	733	1.000		733	1.000		733	.781
	873	1.000		873	1.000		873	.938
726	560	.031	733	560	.000	873	560	.000
	622	.094		622	.000		622	.000
	685	.250		685	.219		685	.063
	733	.469		726	.531		726	.031
	873	.969		873	.938		733	.063

Table 7.12. SD and ΔT/T as a function of the reference durations.

Ref.	560	622	685	726	733	873
SD	80	83	94	83	86	—
ΔT/T	.14	.13	.14	.11	.12	—

Table 7.13. Regression coefficients showing the correlations between the z-scores and the test durations.

r(560) = .988	r(622) = .987	r(685) = .962
r(726) = .995	r(733) = .986	r(873) = —

Time-order errors

Table 7.14. The number of times a particular noise pulse has been judged to be the longer in all possible combinations and presentation orders. 'Score' is the number of times each of the two noise pulses in a pair has received a 'longer' judgement. The order in the 'Durations' column corresponds to the presentation order.

Pair	Durations	Score	Pair	Durations	Score
1—2	685—733	3—13	2—6	733—622	16— 0
2—1	733—685	12— 4	6—2	622—733	0—16
		15—17			15—17
1—3	685—873	1—15	3—4	873—560	16— 0
3—1	873—685	15— 1	4—3	560—873	0—16
		16—16			16—16
1—4	685—560	16— 0	3—5	873—726	15— 1
4—1	560—685	2—14	5—3	726—873	0—16
		18—14			15—17
1—5	685—726	5—11	3—6	873—622	16— 0
5—1	726—685	13— 3	6—3	622—873	0—16
		18—14			16—16
1—6	685—622	15— 1	4—5	560—726	1—15
6—1	622—685	3—13	5—4	726—560	16— 0
		18—14			17—15
2—3	733—873	0—16	4—6	560—622	7—9
3—2	873—733	14— 2	6—4	622—560	14— 2
		18—14			21—11
2—4	733—560	16— 0	5—6	726—622	15— 1
4—2	560—733	0—16	6—5	622—726	2—14
		16—16			17—15
2—5	733—726	9— 7			
5—2	726—733	10— 6		Total: 249—231	
		19—13			

The time-order error seems to be on the same order as for the speech experiment. The bias is in favour of the first presented stimulus (52%—48%). Average TOE in this experiment is +.038 with a standard deviation of .139. There is a downward trend in TOE from about .1 for the lowest contrasts falling to 0 when the relative durational contrast reaches 30%.

7.6 Experiment 4. *(noise)*

This was the first of the additional experiments carried out to obtain more data.

7.6.1 Method.

Subjects

Subjects were the same subjects as those who took part experiment 2. Participation was voluntary and unpaid.

Stimuli

The stimulus material was the same noise discrimination list that was used in experiment 3.

Procedure

The test procedure was identical to that described for experiments 2 and 3. The test was given in the same soundproof perception lab. The stimulus material was presented via headphones.

7.6.2 Results.

The average performance score was 85.0 % (SD = 5.2) correct responses. The scores do not differ significantly from those obtained by the group tested in experiment 3.

Agreement between subjects was found to be significant.

$$W = .933, \qquad \chi^2 = 37.330, \qquad 5 \text{ df}, \quad P < .001, \qquad \text{ave}(r_s) = .924$$

Also agreement with the criterion was highly significant.

$$T_c = .900, \qquad z = 7.107, \qquad P < .001$$

The results of the W and T_c are almost identical to those in experiment 3 described above.

The mean time-order error was +.092 with a standard deviation of .157. The TOE does not differ significantly from the TOE for the speech group described in experiment 3.

The fit of the regression line is quite good for all base durations except 873 ms. The differential fractions differ somewhat from those in experiment 3, but a discussion of the different results and conclusions will be postponed till section 7.11.

Table 7.15. The subjective rankings of interstress interval durations.

	ISI					
	3	2	5	1	6	4
Subject						
1	1	2	3.5	3.5	5	6
2	2	1	4	3	5	6
3	1	2	3	4	5	6
4	1	2	3	4	5	6
5	1	3	2	4	5	6
6	1	3	2	4	5	6
7	1	3	2	4	5	6
8	1	2	3	4	5	6
Mean rank	1.13	2.25	2.81	3.81	5.00	6.00

Differential fractions.

Table 7.16. The probabilities that the comparison duration is judged 'longer' (CL).

Ref.	**Test**	**CL**	**Ref.**	**Test**	**CL**	**Ref.**	**Test**	**CL**
560	622	.844	622	560	.156	685	560	.094
	685	.906		685	.875		622	.125
	726	1.000		726	.875		726	.500
	733	.969		733	.875		733	.781
	873	1.000		873	.938		873	.969
726	560	.000	733	560	.031	873	560	.000
	622	.125		622	.125		622	.063
	685	.500		685	.219		685	.031
	733	.500		726	.500		726	.094
	873	.906		873	.813		733	.188

Table 7.17. SD and ΔT/T as a function of the reference durations.

Ref.	560	622	685	726	733	873
SD	133	127	92	109	114	214
ΔT/T	.24	.20	.14	.15	.16	.25

Table 7.18. Regression coefficients showing the correlations between the z-scores and the test durations.

r(560) = .972	r(622) = .863	r(685) = .972
r(726) = .970	r(733) = .986	r(873) = .583

7.7 Experiment 5. *(speech)*

This experiment and the following one was carried out in connection with a seminar held at the phonetics department at the University of Stockholm. The experimental conditions may not have been optimal, but since the general performance of the subjects used here did not differ significantly from that of the groups used in the discrimination experiments described above, the conditions seem to have been acceptable and the results have been included in the study.

7.7.1 Method.

Subjects

The subjects were students and teachers at the department of phonetics at the University of Stockholm and the Speech Transmission Laboratory of the Royal Institute of Technology. Participation was voluntary and unpaid.

Stimuli

The stimulus material was the speech discrimination list used in experiment 2.

Procedure

The test was given in a lecture room. The stimulus material was presented via loudspeakers.

7.7.2 Results.

The average performance score was 75.8 % (SD = 7.2). In this respect the group is comparable to the first tested speech group.

Table 7.19. The subjective rankings of interstress interval durations.

	ISI					
	3	2	5	1	6	4
Subject						
1	1	2	3	4	6	5
2	1	2	3	5	4	6
3	4	1.5	5	3	1.5	6
4	1	2	6	3.5	3.5	5
5	3.5	1	5	3.5	2	6
6	1	2.5	2.5	4	5	6
7	3	1	4	2	5	6
8	2.5	1	2.5	4.5	4.5	6
9	1	3	3	3	6	5
10	4	1.5	5	3	1.5	6
11	1	2	3.5	3.5	5	6
Mean rank	2.09	1.77	3.86	3.55	4.00	5.73

Agreement between subjects was found to be significant.

$$W = .608, \qquad \chi^2 = 33.458, \qquad 5\ \text{df}, \qquad P < .001, \qquad \text{ave}(r_S) = .569$$

Also agreement with the criterion was significant.

$$T_C = .539, \qquad z = 4.985, \qquad P < .001$$

The mean time-order error was +.061 with a standard deviation of .198. The TOE does not differ significantly from the TOE for the speech group described in experiment 2.

Differential fractions.

Table 7.20. The probabilities that the comparison duration is judged 'longer' (CL).

Ref.	Test	CL	Ref.	Test	CL	Ref.	Test	CL
560	622	.841	622	560	.159	685	560	.136
	685	.864		685	.591		622	.409
	726	.932		726	.523		726	.341
	733	1.000		733	.750		733	.796
	873	.932		873	.750		873	.614
726	560	.068	733	560	.000	873	560	.068
	622	.477		622	.250		622	.250
	685	.659		685	.205		685	.386
	733	.750		726	.250		726	.346
	873	.659		873	.500		733	.500

Table 7.21. SD and ΔT/T as a function of the reference durations.

Ref.	560	622	685	726	733	873
SD	504	189	239	189	301	143
ΔT/T	(.90)	.30	.35	.26	.41	.16

Table 7.22. Regression coefficients showing the correlations between the z-scores and the test durations.

r(560) = .818	r(622) = .867	r(685) = .689
r(726) = .725	r(733) = .892	r(873) = .931

Differential fractions are higher and goodness of fit scores are somewhat lower than the corresponding values for the other speech group. This may be an indication that the less favourable listening conditions may have had an adverse effect on performance. But the differences are too small for any definite conclusions to be drawn.

7.8 Experiment 6. *(noise)*

This is the second experiment with the 'seminar group' at the University of Stockholm. The experiment was carried out approximately 20 minutes after the completion of the speech test.

7.8.1 Method.

Subjects

The subjects were students and teachers at the department of phonetics at the University of Stockholm and the speech transmission laboratory of the Royal Institute of Technology. The group was almost identical to that of the preceding experiment (exp. 5.) 10 of the 12 subjects were the same. Participation was voluntary and unpaid.

Stimuli

The stimulus material was the same noise discrimination list as that used in experiments 3 and 4.

Procedure

The test was given in a lecture room. The stimulus material was presented via loudspeakers.

7.8.2 Results.

The average performance score was 88.2 % (SD = 5.2). The result does not differ significantly from those of experiments 3 and 4.

Table 7.23. The subjective rankings of interstress interval durations.

	ISI					
	3	2	5	1	6	4
Subject						
1	1	3	2	4	5	6
2	1	3	2	4	5	6
3	1	2.5	2.5	4	5	6
4	1	2	3	4	5	6
5	1	2	4	3	5	6
6	1	2	3	4	5	6
7	1	2	3	4	5	6
8	1	4	3	2	5	6
9	1	4	2	3	5	6
10	1	2	3.5	3.5	5	6
12	1	2.5	2.5	4	5	6
13	1	2.5	4	2.5	5	6
Mean rank	1.00	2.63	2.88	3.50	5.00	6.00

As with the other noise experiments, agreement between subjects is very high.

$$W = .918, \qquad \chi^2 = 55.060, \qquad 5\ \text{df}, \qquad P < .001, \qquad \text{ave}(r_s) = .910$$

Also agreement with the criterion was highly significant.

$$T_c = .856, \qquad z = 8.298, \qquad P < .001$$

The mean time-order error was +.058 with a standard deviation of .133. The TOE does not differ significantly from the TOE for the speech group described in experiment 2.

Differential fractions.

Table 7.24. The probabilities that the comparison duration is judged 'longer' (CL).

Ref.	Test	CL	Ref.	Test	CL	Ref.	Test	CL
560	622	.854	622	560	.146	685	560	.021
	685	.979		685	.875		622	.125
	726	.979		726	.958		726	.583
	733	.979		733	.958		733	.729
	873	1.000		873	.979		873	.979
726	560	.021	733	560	.021	873	560	.000
	622	.042		622	.042		622	.021
	685	.417		685	.271		685	.021
	733	.500		726	.500		726	.063
	873	.938		873	.938		733	.063

Table 7.25. SD and ΔT/T as a function of the reference durations.

Ref.	560	622	685	726	733	873
SD	114	101	76	84	83	204
ΔT/T	.20	.16	.11	.12	.11	.23

Table 7.26. Regression coefficients showing the correlations between the z-scores and the test durations.

r(560) = .910	r(622) = .885	r(685) = .995
r(726) = .981	r(733) = .991	r(873) = .861

7.9 Experiment 7. *(whole phrase)*

This and the following experiment were carried out with Dutch subjects. The idea behind using subjects with no knowledge of Swedish was that the sounds of an unknown language may be perceived as 'speech-like noise' rather than speech in experiments on duration perception. The prediction was that this would reduce processing load, particularly memory load, and thereby improve duration perception performance. Two of the experimental conditions were used—rank ordering of the interstress intervals when the whole phrase was presented and duration discrimination of speech fragments.

The first experiment was rank ordering of the interstress intervals in the whole phrase. The experiment was identical to experiment 1, described in section 7.3.

7.9.1 Method.

Subjects

The subjects were undergraduate students at the department of Taal en Minderheden (speech and minorities) at the University of Tilburg, in the Netherlands. Participation was voluntary and unpaid.

Stimuli

The stimulus material was the same as that used in the first ranking experiment (exp. 1).

Procedure

The test was given in a lecture room. The stimulus material was presented via loudspeakers. The instructions were given in English, but were supplemented by instructions in Dutch to clarify some points. The answer form was the same one that was used in the Swedish experiment. That meant that the written version of the phrase was in Swedish. No translation of the phrase was given nor any hint of its origin or type.

7.9.2 Results.

The supposedly simpler stimulus properties were counteracted to some extent by subjects' ability to correlate the auditory stimulus with the written version of the phrase. This meant for example that they wanted to hear more repetitions of the phrase than the Swedish subjects in the corresponding experiment. Not being able to understand the meaning of the phrase presents a memory problem of course. The subjects were told to concentrate on the stressed syllables only, which they could easily identify, and try to rank the intervals between the stresses according to duration. After a number of repetitions and a few further clarifications all subjects seemed to understand the task fully. It must be pointed out,

however, that the fact that the phrase was not understood turned out to be a greater methodological problem than anticipated.

Table 7.27. The subjective rankings of interstress interval durations.

	ISI					
	3	2	5	1	6	4
Subject						
1	3	2	6	4	5	1
2	3	2	4	1	6	5
3	1	2	6	3	5	4
4	5	2	3	1	4	6
5	3	4	6	2	5	1
6	1	3	4	2	5	6
7	2	1	6	3	4	5
8	2	3	4	1	5	6
9	1	2	3	4	5	6
10	6	2	5	3	4	1
Mean rank	2.70	2.30	4.70	2.40	4.80	4.10

The *W*-test gave the following results:

$$W = .387, \qquad \chi^2 = 19.371, \qquad 5 \text{ df}, \qquad P = .0016, \qquad \text{ave}(r_S) = .319$$

If compared to the results of experiment, 1 it can be seen that the agreement between subjects is higher and so is the level of significance. This is an indication that the performance of the Dutch subjects was indeed somewhat better.

Agreement with the criterion was also highly significant and the T_C score higher than for the Swedish subjects.

$$T_C = .307, \qquad z = 2.673, \qquad P = .0038$$

A more detailed comparison will be given later but the figures presented here tend to lend support to the idea that discrimination is somewhat easier when semantic information is removed.

7.10 Experiment 8. *(speech)*

The experiment presented here is identical to the speech discrimination experiments presented above.

7.10.1 Method.

Subjects

The subjects were again undergraduate students at the department of 'Speech and minorities' at the University of Tilburg. It was not possible to test the same group as the one in experiment 7, but 6 of the 10 subjects were the same.

Stimuli

The stimulus material was the same speech discrimination list that was used in experiments 2 and 5.

Procedure

The test was given in a lecture room. The stimulus material was presented via loudspeakers. The answer forms were the same as those used in the Swedish experiments but instructions were translated to English. Oral instructions were given in English and supplemented in Dutch. In this experiment subjects had, of course, no added difficulty in understanding the task.

7.10.2 Results.

The average performance score was 82.3 % (SD = 7.2). The score is approximately 7 % higher that the corresponding results for Swedish subjects but the difference is not statistically significant (ANOVA, P = .105).

Table 7.28. The subjective rankings of interstress interval durations.

	ISI					
	3	2	5	1	6	4
Subject						
1	1	3	4	2	5	6
2	1	2	4	3	5	6
3	1	2	3	4	5.5	5.5
4	2.5	1	4	2.5	6	5
5	1	2	3	4	5	6
6	2.5	1	2.5	4	5.5	5.5
7	1	2	4	3	5	6
8	1	2	4	3	6	5
9	1	2.5	2.5	4	5	6
10	2.5	2.2	4	1	5	6
Mean rank	1.45	2.00	3.50	3.05	5.30	5.70

Agreement between subjects was found to be very high. The W score is the highest for all speech groups. The significance level is also higher than for any of the other speech groups

$$W = .857, \qquad \chi^2 = 42.834, \qquad 5\ \text{df}, \qquad P < .001, \qquad \text{ave}(r_s) = .841$$

Also agreement with the criterion is the highest for any speech group and so is the level of significance.

$$T_c = .747, \qquad z = 6.594, \qquad P < .001$$

The mean time-order error was –.033 with a standard deviation of .121. This group is the only one who produced a negative TOE. However, the TOE does not differ significantly from the TOE for the other speech groups.

Differential fractions.

Table 7.29. The probabilities that the comparison duration is judged 'longer' (CL).

Ref.	Test	CL	Ref.	Test	CL	Ref.	Test	CL
560	622	.700	622	560	.300	685	560	.025
	685	.975		685	.975		622	.025
	726	.925		726	.900		726	.225
	733	.975		733	.975		733	.650
	873	1.000		873	1.000		873	.800
726	560	.075	733	560	.025	873	560	.000
	622	.100		622	.025		622	.000
	685	.775		685	.350		685	.200
	733	.750		726	.250		726	.200
	873	.800		873	.700		733	.300

Table 7.30. SD and ΔT/T as a function of the reference durations.

Ref.	560	622	685	726	733	873
SD	95	76	100	125	118	231
ΔT/T	.17	.12	.15	.17	.16	.27

The goodness of fit is quite good for all durations except the longest one indicating that this value should be interpreted with a certain caution.

Table 7.31 Regression coefficients showing the correlations between the z-scores and the test durations.

r(560) = .791	r(622) = .901	r(685) = .927
r(726) = .822	r(733) = .936	r(873) = .612

7.11 Summary of experimental results and discussion.

In the following sections, I will further analyse the results of the eight experiments described above and compare the different groups and experimental conditions.

7.11.1 General performance, W and T_C scores.

As performance score, the percentage of correct responses was used. A summary of the scores for the discrimination experiments is presented in Table 7.32. The performance in terms of the number of correct responses can be seen as a measure of the difficulty in duration discrimination for a given condition.

Table 7.32. A summary of performance scores for the 6 discrimination experiments.

Stimulus type		Noise			Speech	
Experiment	3	4	6	2	5	8 (Dutch)
Mean	88.1	85.0	88.2	74.2	75.8	82.3
SD	6.1	7.7	5.2	10.2	7.2	7.2

An ANOVA test of scores by experiments reveals significant differences between the two types of stimuli (noise vs. speech) ($P < .001$). The groups are not completely independent. A subgroup of subjects participated under both conditions (noise vs. speech). But if the results of this subgroup are compared in a paired t-test the scores are again found to differ significantly between the two conditions ($P = .001$). The impression given by the figures in the table that the speech test is more difficult than the noise test is thus found to be statistically significant. This is a first (rather crude) measure of the increased difficulty it means in duration discrimination if speech is used instead of noise. The scores of the Dutch group is higher than the scores for the two Swedish groups under the speech condition but lower than the scores for noise, supporting the hypothesis that part of the increased difficulty in judging the durations of speech stimuli is due to the increased processing load when semantic information is present in addition to the increased acoustic complexity compared to noise.

An observation worth underlining is that the scores are as high as 75% or better for all groups. This means that, for the durations used here, subjects managed to judge the difference in duration between two stimuli correctly 75 times out of 100. No single subject in any of the groups performed at chance level.

The differences between the different conditions can be further analysed by looking at the results of the W and T_C tests. The results from these tests are summarized in Table 7.33. It can be seen that inter-subject agreement (W) and agreement between subjects' rankings and that based on the objective durations (T_C) is a direct function of the experimental

Table 7.33. A summary of the results from W and T_c tests on the results from the 8 experiments.

Exp.	k	W	χ^2	P	ave(r_s)	T_c	z	P
SPEECH:								
Rank ordering of the whole phrase								
1	8	.280	11.214	.0473	.178	.333	2.590	.0048
7 (Dutch)	10	.387	19.371	.0016	.319	.307	2.673	.0038
Discrimination								
2	8	.691	27.618	.0000	.646	.650	5.114	.0000
5	11	.608	33.458	.0000	.569	.539	4.985	.0000
8 (Dutch)	10	.857	42.834	.0000	.841	.747	6.594	.0000
NOISE:								
3	8	.927	37.086	.0000	.917	.850	6.709	.0000
4	8	.933	37.330	.0000	.924	.900	7.107	.0000
6	12	.918	55.060	.0000	.910	.856	8.298	.0000

condition. Although agreement is significant for all experimental conditions, the degree of agreement varies considerably. These differences can be seen as reflecting the difficulty of the task. In the noise discrimination task, agreement is almost perfect. When speech is used the agreement decreases markedly and with the whole phrase to judge, agreement scores are down to about one third compared to the noise condition. As was shown in the analysis of experiments 2 and 3, the correlation between relative durational differences and ability to judge these differences correctly is higher when noise is used than with speech. Another way of saying the same thing is that less of the duration discrimination performance is explained by durational differences when speech is used. The most likely explanation for this is that when speech is used, some attention and processing is directed towards the added information content of the stimuli (e.g. spectral, semantic etc.) and that this increased load on processing distracts the subject from duration processing to some degree.

7.11.2 Differential fractions.

One of the main aims of this study is to establish some kind of just noticeable difference for duration perception of interstress intervals. As was discussed in Chapter 5, the differential fraction is such a measure. This section will be devoted to a summary and comparison of the differential fractions found in the different experiments. Differential fractions were presented in detail in connection with the presentations of the different experiments above. Rather than repeat these figures, I will present them in a different form

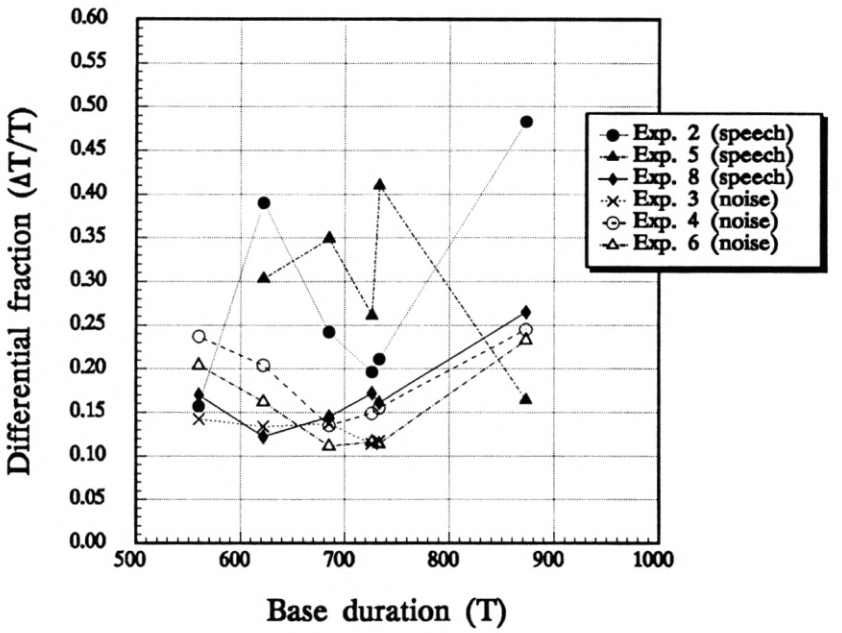

Figure 7.2. Differential fractions for all the discrimination experiments. Two data points have been excluded because data was too uncertain. 'Exp. 8' is the Dutch group.

which makes comparison easier. In Figure 7.2, a graphic representation of the differential fractions for all the groups is shown. There is a great variability in the values for speech data from the Swedish groups. But the values are generally considerably higher than for noise. The values for the Dutch group, on the other hand, are comparable with those for Swedish noise data. This supports the prediction that Dutch subjects have judged stimuli as complex noise rather than speech.

The results shown in Figure 7.2 are the results from several different groups of subjects. To make a closer comparison between the noise and speech conditions possible the two equal subgroups that took part in both types of tests were singled out and treated separately. The result is presented in Figure 7.3. The picture is essentially the same as that in 7.2. Beginning with the noise data it seems as if discrimination was best for the middle range of durations. However, this may to some extent be an artefact of the choice of durations that were compared. When the shortest and longest durations are compared they form the end-points of the scale, which means that the durations at the other end of the range will be judged as longer (or shorter) in almost 100 % of the cases which is, to some extent, in contradiction with the normality assumption. They are also considerably longer (or shorter) than the nearest comparison. This may have caused the effect rather than differences in duration perception as a function of base duration. If this is true, then the values for the

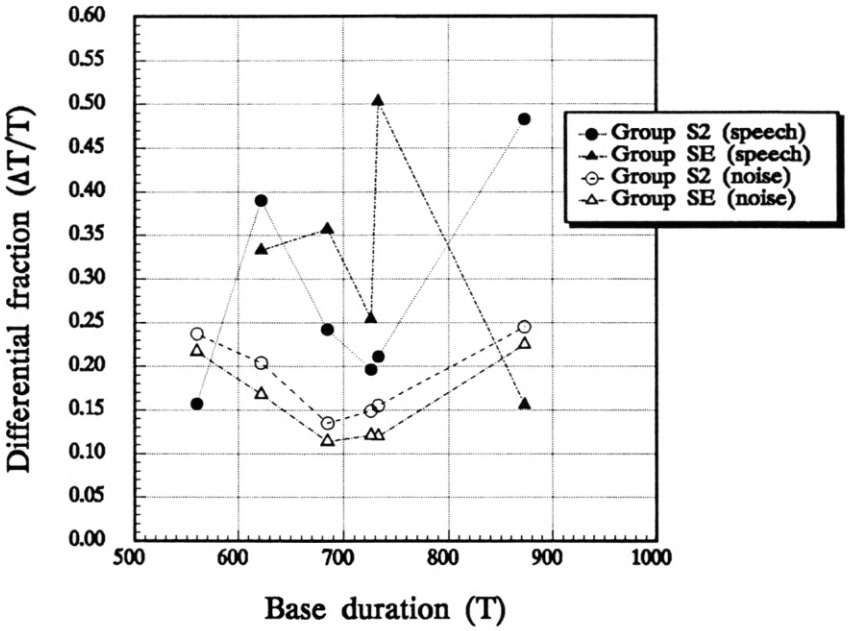

Figure 7.3. Differential fractions for the two subgroups that participated in both speech and noise discrimination experiments. One data point has been excluded because data was too uncertain. S2 is the group that took part in experiments 2 and 4. SE is the subgroup of subjects who took part in both experiments 5 and 6.

middle durations (685—733 ms) are the most representative. The uncertainty in the peripheral durations is also present in the data from the speech discrimination experiments. But here there is a great deal of variability in the other durations as well. To a great extent, this reflects the increased difficulty of the task, meaning that other factors than just duration have played a role.

An inspection of the data showed that the subjects who had the lowest scores are responsible for almost all the variability. If the best performers are treated separately, this variability vanishes almost completely. In Figure 7.4, differential fractions are plotted for the 11 Swedish subjects who scored 80% correct responses or better. For comparison, the results from the Dutch subjects and the mean values for the noise tests are plotted in the same diagram. Here the differences between the different conditions and groups are greatly reduced. If the 'middle durations' are considered the groups are ranked according to the prediction that noise is best discriminated with differential fractions around .12. Speech discrimination appears to get differential fractions around .20 and for the Dutch group, to whom it was assumed that the stimuli resembled complex noise, the differential fractions are somewhere in between. But the variability is too great for these observations to be claimed with certainty. More experiments must be done to clarify the question. The

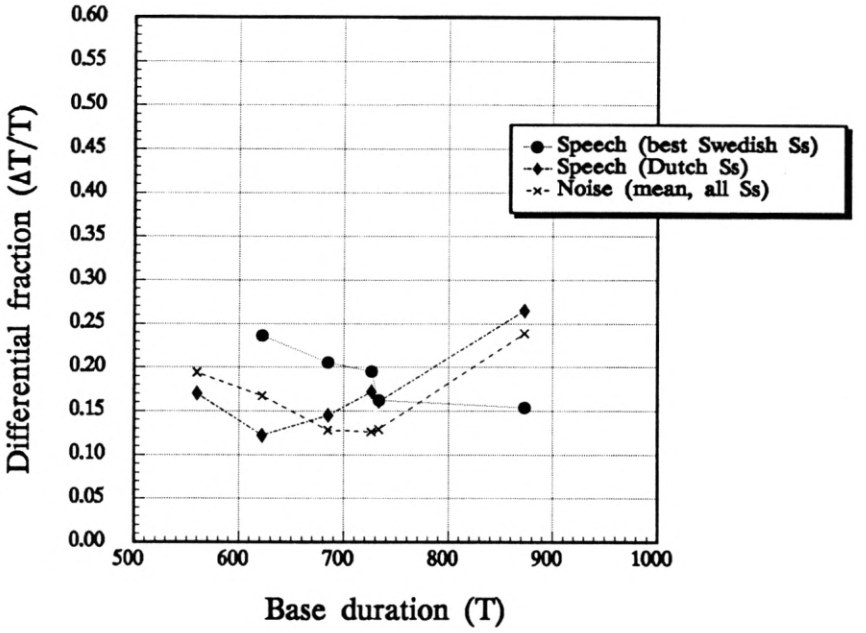

Figure 7.4. Differential fractions for speech data for those Swedish subjects who performed 80% or better on the two tests.

tendencies are clear enough for one to want to suggest, as a prediction for the outcome in future experiments, that these relations between different stimulus types would be preserved. The orders of magnitude of the differential fractions are probably also reasonably representative.

7.11.3 Subjective durations.

From the regression lines of z-scores on base durations used to compute differential fractions it is also possible to compute subjective durations for the different stimuli. In comparable studies, these durations are often referred to as the 'PSEs' (Point of Subjective Equality). I will use the same terminology here. The first diagram (Figure 7.5) is again a summary of results from all experiments.

The picture is a rather scattered one. If, however, one disregards the variation at the end points of the scale and looks at the middle values, there is an obvious pattern. The PSEs for noise are almost exactly the same as the objective durations but the PSEs for speech vary for all groups in a non-random way. The base durations 685 ms and 733 ms seem to be overestimated.

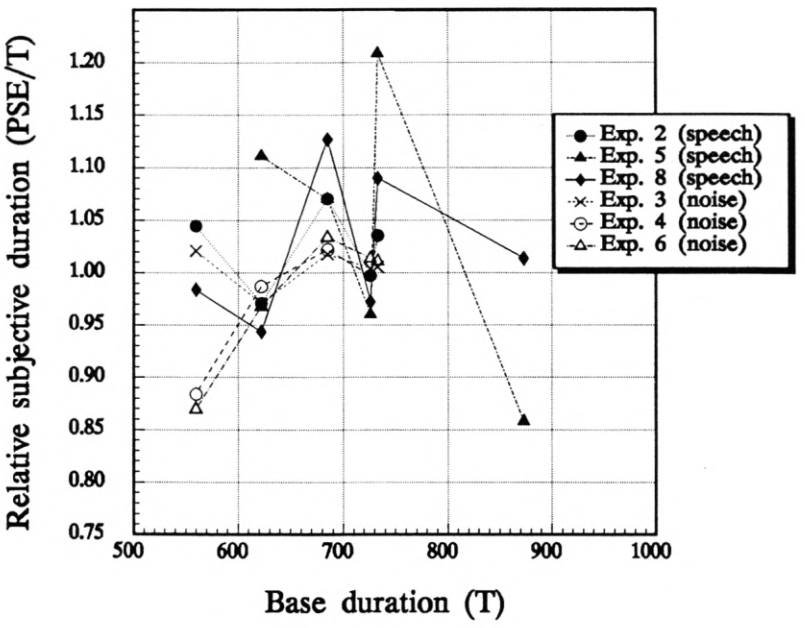

Figure 7.5. Subjective durations computed from regression equations for all groups.

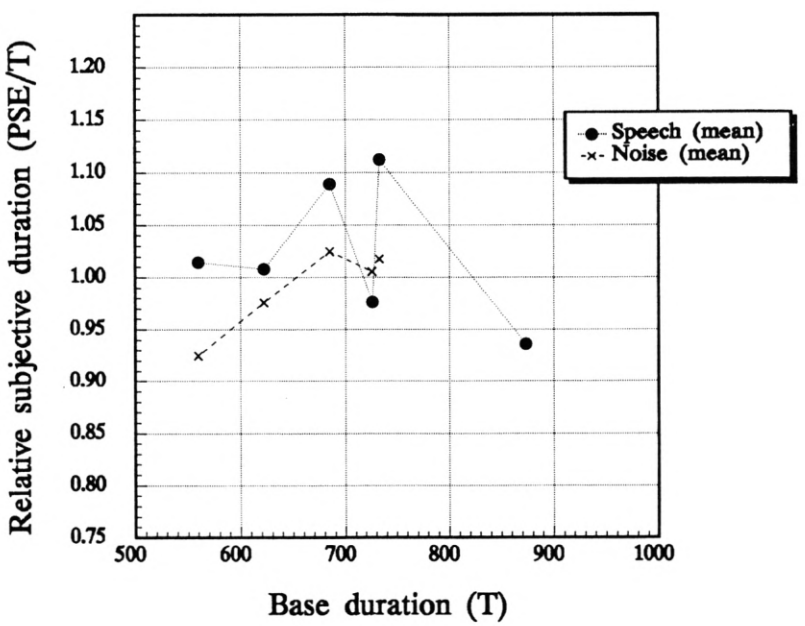

Figure 7.6. Subjective durations based on average values of data from Swedish speech and noise data.

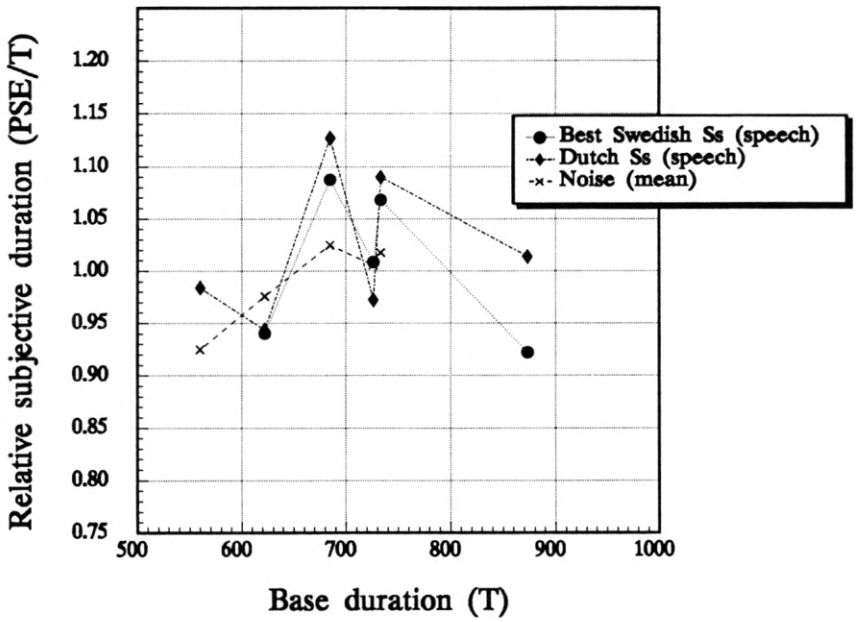

Figure 7.7. Subjective durations based on average values of data from the best Swedish speech group, the Dutch speech data, and the noise data.

If the results from speech and noise experiments are pooled and the mean values are considered, the picture becomes clearer. These data are shown in Figure 7.6 (Swedish data only).

In contrast to what seemed to be the case with differential fractions, these deviations do not seem to be restricted to the 'poorer' subjects. The same relations are true for the group of best performing subjects as well as for the group of Dutch subjects. Figure 7.7 shows these data.

It has been found in many studies that stimulus loudness affects duration perception. To explore the possibility that loudness could also explain the deviations in PSEs, the correlations between loudness and perceived duration was investigated. In Figure 7.8 the PSEs are plotted against loudness. As can be seen in the diagram there seems to be some correlation between loudness and relative subjective duration, but the correlation is rather weak ($r = .417$, $r^2 = .174$) indicating that other factors as well play important roles.

Some interesting observations may be made, however, that indicate that loudness may be an important variable when considering duration perception in speech. If one considers the interstress intervals with durations 726 ms and 733 ms, one would predict that they would be judged as equal. The distribution of 'longer' answers should be 50%—50% for

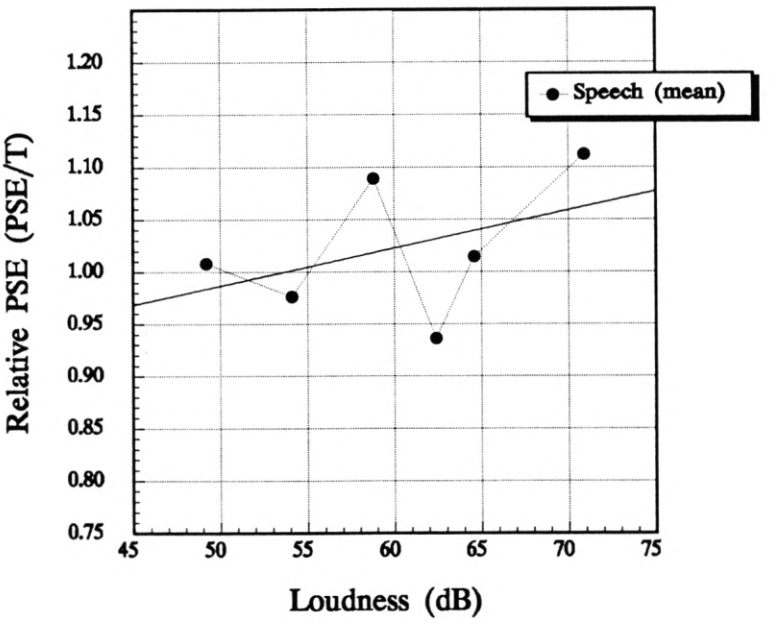

Figure 7.8. Subjective durations based on average values as a function of stimulus loudness. The regression line is also plotted in the diagram.

comparisons of the two. This is indeed the case if one looks at the noise discrimination results in Table 7.14. The distribution of 'longer' responses is 15—17. For the speech test (Table 7.8) the result is quite different. Here the distribution is 20—12 in favour of the 733 ms duration. This agrees well with the prediction one would make based on their respective loudness values (71 dB vs. 54 dB). If one considers the results from the other groups (not presented in detail above), the picture is exactly the same. For the noise tests there is no bias at all (16—16 and 24—24, respectively) whereas there are strong biases in the speech tests (33—11 for the Swedish group and 30—10 for the Dutch one).

It is difficult to assess the bias for other durations, but if one considers the other 'middle' duration (685 ms) one may see that it is louder than the 726 ms interval (59 dB vs. 54 db) and less loud than the 733 ms interval (71 db). In comparisons of the two durations 733 ms and 726 ms against the 685 ms duration the scores for the 733 ms duration should be markedly higher in the speech tests. This is indeed also the case for all groups.

It must be pointed out, however, that in view of the limited material, these figures can hardly be seen as proof of anything. But they could perhaps be seen as an indication that loudness may play a role and that the influence of this variable is worth exploring.

7.11.4 Time-order errors.

Time-order errors were studied for two reasons. As was pointed out in 7.1, one of the reasons is methodological. How much does TOE bias the result of a pairwise comparison of speech stimuli? But a more important question was the possibility, suggested by the results in Lehiste's (1976) study, that this type of bias is considerably greater for speech than for non-speech.

A summary of TOEs in the 6 experiments involving duration discrimination is given in Table 7.34. As can be seen in the table, the effect of TOE seems to be rather small. The largest bias for any of the experiments is a 55—45% bias in favour of the first presented stimuli. TOEs in the other experiments are smaller. A significance test of the means (ANOVA) reveals no significant differences, either between the experiments taken together or between those using noise stimuli and those using speech. As with the scores test, the groups are not independent. If, however, the subgroup of subjects who took part in both conditions are compared, again no significant differences are found. The conclusion must be that the TOE effect is the same under both conditions.

The results also confirm the finding in many other investigations that time-order errors for durations in this range are positive. The result for the Dutch group deviates from this pattern but the TOE is very small and the deviation is not significant. The TOEs found here are of the same order as those found by Stott (1935) using non-speech stimuli. The ones found here are slightly smaller but the difference would probably not be significant. The very considerable TOE found by Lehiste, using synthetic vowels, was not confirmed. Now, synthetic vowels and the stimuli used here are of course very different in character. But a conclusion one may draw is that large TOEs do not seem to be a characteristic of speech stimulus discrimination in general.

Table 7.34. A summary of time-order error data for the 6 discrimination experiments.

Stimulus type		Noise			Speech	
Experiment	3	4	6	2	5	8
Mean	+.038	+.092	+.058	+.067	+.061	–.033
SD	.139	.157	.133	.150	.198	.121
Bias	249–231	262–218	381–339	256–224	350–310	290–310
	52%–48%	55%–45%	53%–47%	53%–47%	53%–47%	48%–52%

7.11.5 Discussion.

The main aim of this study was to determine 'just noticeable differences' for speech stimuli in duration discrimination. In conformity with the treatment of the results in Chapter 6, I will use relative difference limens (DL/T = .67*ΔT/T) rather than the differential fractions to describe the results with respect to duration discrimination.

The relative difference limen based on the mean value of the differential fractions for the two Swedish speech groups is .19. Given the variation (SD = .07) the exact figure should perhaps be interpreted with a certain amount of caution, but the order of magnitude is probably quite representative. Compared to typical interstress interval durations in the order of 500 ms, contrasts of more than 95 ms should thus have a better than 50 % chance of being detected. As was seen in Chapter 4, typical standard deviations in the material used in the production study are of the same order and typical ranges 3 to 5 times greater. And if mean values for interstress interval durations are used then relative durational contrasts between adjacent intervals are greater than .19 in 9 cases out of 15.

The difference limens have been found in a pairwise discrimination task. But the results from experiments involving duration discrimination in a series of durations imply that discrimination in that type of task is at least as good (e.g. Michon, 1964; Halpern and Darwin, 1982). The implication of the results obtained here must, therefore, be that the durational differences in interstress interval duration that one finds in speech should often be well within the capability of the perceptual system to detect.

The mean DL/T value for the experiments where noise stimuli were used is .11 and the mean value for the Dutch group is .12. The size of the fractions are comparable to those obtained by others using noise and tones of similar durations (Abel, 1972a, DL/T = .10; Small and Campbell, 1962, DL/T = .18) These results indicate that the semantic content as well as the acoustic complexity of the stimulus may affect duration discrimination. The prediction that the removal (at least to some extent) of semantic information should make the task easier seems to be met. In fact the results for the Dutch speech group are comparable to the results for non-speech stimuli with the Swedish groups. Unfortunately, it was not possible for practical reasons to test the Dutch group with non-speech stimuli. It is, therefore, not possible to say with any certainty how they would have performed. It cannot be ruled out that they would have performed better than the Swedish groups also in the non-speech tests. It is unlikely, however, that a difference in duration discrimination ability is the sole explanation for the markedly better result in this group. Particularly since the two Dutch groups were not identical and both performed better. The most likely conclusion is instead that the semantic information which is present for the Swedish subjects and will be processed to some extent, whether the subjects want to or not, is the cause of at least part of the difference.

At a first glance, the considerably poorer results from the experiments involving judgment of interstress interval durations in the whole phrase may seem to contradict the conclusions drawn above to some extent. But it must be pointed out that the task in these experiments was considerably more difficult than merely deciding whether intervals are equal or not. If that had been the task, the success would probably have been 100%. No subject had any doubts that the intervals were unequal and the relative success with which they managed to rank the interval durations shows that their intuitions had a clear correspondence with the actual durations. From this point of view, the results from experiments 1 and 7 also clearly speak in favour of the idea that differences in interstress interval durations are indeed perceptible. In this particular sense, Lehiste's assumption that we hear interstress intervals as fairly equal because we are unable to tell the difference is not borne out. But it must be pointed out, of course, that the results from the experiments described above only tell us something about what we *can* do, and not necessarily much about what it is that we *normally* do when listening to speech.

The considerable time-order bias found by Lehiste was not found in the experiments carried out here. TOE was of the same order as has been found in typical experiments using non-speech stimuli. No support was, thus, found for the idea that speech is processed any differently from non-speech in this particular respect.

Some evidence was also found that the variation in loudness between different stimuli may have influenced duration perception. If the results found here are representative it means that accurate duration perception may be made more difficult by the constantly varying intensity in normal speech. It seems unlikely, however, that this effect would cause speech to sound any more or less regular.

PART IV

Discussion and suggestions for further research.

Chapter 8

Summary of the results, discussion, and suggestions for further research.

Aspects of speech rhythm were studied experimentally in both production and perception. In addition, some theoretical and methodological questions were discussed. I will now try to summarize the results from the experimental study as well as some of the theoretical issues. The discussion following the summary will be centred around the question of regularity. On what grounds can we claim that there are tendencies to regularity in speech production or speech perception? In the concluding sections, I will make some suggestions for further research.

8.1 Summary of theoretical results.

In Chapter 3, a number of theoretical and methodological issues were raised. The two main questions were: how can interstress intervals be modelled? and does a linear increase in interstress interval duration put any restrictions on the possible temporal compensation mechanisms at syllable level within the intervals?

8.1.1 The linear model.

Using data published in other studies, it was shown that, as far as it is possible to tell from available data, a linear model for interstress interval duration as a function of the number

of syllables seems to fit data very well (see section 3.2). Moreover, languages seem to fall into two fairly distinct groups using a linear model. Interstress intervals in all languages seem to increase by around 100 ms per added syllable, but the constant terms in the linear equation, assumed to reflect the added duration in stressed syllables, differ between languages. For those languages conventionally called 'syllable-timed' the constant term was found to be approximately 100 ms and for the stress-timed languages around 200 ms (perhaps somewhat more; 232 ms was found for the regression line calculated on the average values for a number of 'stress-timed' languages). If these results hold true in general, the implication seems to be that the difference between the languages called 'syllable-timed' and those called 'stress-timed' lies not so much in the way interval durations grow as a function of the number of syllables, but more in the relative prominence of stressed syllables, assumed to be reflected in the increase in their durations compared to unstressed ones.

8.1.2 Compression of syllables.

If interstress intervals grow at a constant rate per added syllable this may seem to imply that the added unstressed syllables should also be of constant duration. But in 3.3 it was shown that this is not necessarily the case. Even if intervals increase by a constant amount per added syllable, there are still possibilities for compensatory changes. Stressed syllables may be gradually compressed while unstressed ones are unchanged, or both stressed and unstressed syllables may be compressed. It was shown, in fact, that it is even possible for compression in both stressed and unstressed syllables when the increase in interval duration is accelerated.

It was mentioned in 3.3 that if one assumes the existence of a final lengthening effect also in interstress intervals which are not phrase-final, then additional possibilities for variation in syllable duration within an interval arise. In 4.7, an example of how this may work was shown, and proposed as a model to describe the data analysed in Chapter 4. One result of such an effect is that, if mean syllable durations for unstressed syllables are computed, it will seem as if syllable duration is successively decreasing with increasing interval length, suggesting a gradual compression. But the model used in 4.7 shows that there need not be any such compression effect. What looks like compression, if mean durations are used, is only an artefact of the way the means are computed.

The implication of these findings is, of course, that it is necessary to study the internal composition of interstress intervals, as well as their total durations, very closely before any conclusions about compression can be reached. These findings also have implications for the study of languages from a classificatory point of view. Even if languages should appear as identical when total durations of interstress intervals are considered, they may differ widely with respect to interval-internal processes.

8.2 Summary of the production study.

In the production study (Chapter 4) the emphasis was on regularity and variation in interstress interval duration. The focus was on three main questions: 1) What is a typical mean value for interstress interval duration, and how much variation is there? 2) Does the linear model for interval duration as a function of the number of syllables in the interval, proposed in 3.2, hold for the empirical data used in this study? 3) Are there any compression tendencies as a function of interval length indicating a tendency for greater interval regularity? In addition, two methodological questions were given consideration: do the variables studied here, 'interstress interval duration' and 'syllable duration', seem to depend sex or age? and, what constitutes a reasonable group size in a study of this kind if one wants to be sure that the results are fairly representative?

8.2.1 Variation in interstress interval duration.

Whenever interstress intervals have been studied by measurements in the speech signal, their durations have been found to be highly variable (see 2.3.1). The results in Part II of this study constitute no exception to that rule. The range of durations was found to be half, or more than half, of that of the mean interstress interval duration for an individual speaker. This means that if mean duration is in the order of 500 ms, the range of durations may be expected to be 250 ms or more. The range is not a measure of the contrast between adjacent intervals but an inspection of the material shows that maximum differences between adjacent intervals are almost of the same order as the range, the most extreme case being an interval of 829 ms followed by one of 493 ms. What this means, of course, is that little or no support may be found in the data presented and analysed here to claim that interstress interval durations 'tend to be more or less equal' or the like.

There is no corresponding variability in mean durations of interstress intervals between subjects, and no significant differences were found between the three groups. The mean values found also agree well with mean values found in other studies of Swedish (e.g. Fant and Kruckenberg, 1989), and other languages (e.g. Dauer, 1983). With respect to language universals, it is interesting to see that interstress intervals seem to cluster around a mean of some 500 ms for all languages, at least if read prose is concerned. There is considerable variation around these means, with respect to the realization of individual interstress intervals, but mean values seem to converge rather systematically towards a value near 500 ms.

8.2.2 The linear model.

The variation found in interstress interval duration was by no means random. It was found that interval duration could be described as a linear function of the number of syllables in

the interval to a very close approximation. And the number of phonemic segments in the interval was found to be an even better predictor of interval duration, this relationship too being approximately linear. The increase in duration per added syllable was found to be around 110 ms. The corresponding increase per added segment was 53 ms on the average. From this point of view, adding more syllables or segments to the interval appears to be a more or less concatenative process.

Whereas individual variation was considerable, no significant differences were found between groups, indicating that the values found are fairly stable characteristics of the language itself when averaged over a reasonably large group of subjects.

Again, there seem to be certain values that are more or less the same for all languages. The increment in interstress interval duration per added syllable (110 ms) is typical of values found in similar studies (e.g. Dauer, 1983. see 3.2). And, judging from the size of the constant term (281 ms) in the regression equation, Swedish seems to place itself among the other languages with marked stress (e.g. English), although the value found here is somewhat higher than the 232 ms found as an average for the 'stress-timed' languages presented in Figure 3.1.

There is certainly a strong suggestion in these findings that there may be fairly stable universal, as well as language specific, properties with respect to the dependency of interstress interval duration on the number of syllables or the number of phonemic segments. This seems particularly true for the 'rate'. There seems to be more room for variation in the constant terms.

8.2.3 Compression of syllables.

Superficially, it would seem as if interstress intervals were made up by adding stressed and unstressed syllables of constant but unequal durations, a process suggested as a possibility by Faure, Hirst, and Chafcouloff (1980). But an analysis of syllable duration showed that this was too simple a model. As was shown in 3.3, it is perfectly possible for various compression processes to operate within interstress intervals without destroying the linear increase in total duration as a function of the number of syllables. However, no conclusive evidence for any compression tendencies was found in this study. But it was shown that, primarily due to the number of phonemes in the target material in different syllables, syllable duration varied as a function of position and the number of syllables in the interval. One of the effects of this variation was that unstressed syllables were considerably longer than the 110 ms suggested by the increase in total interval duration per added syllable. Syllable duration for unstressed syllables in interval final position was also found to be significantly longer than in interval-medial position. The result was that mean syllable durations for stressed (always interval-initial), unstressed interval-medial, and unstressed interval-final syllables were 274 ms, 143 ms, and 202 ms respectively, in rather striking contrast to what one would expect by looking at total interval durations only.

8.2.4 Methodological issues.

Although considerable inter-individual variation was observed, no comparable differences seemed to be present when group results were compared. Mean values for interstress interval durations and syllable durations did not differ significantly between groups. Neither did the linear equations expressing interstress interval duration as a function of the number of syllables or phonemic segments. Neither sex nor age seem to be crucial variables. What seems to be the important factor in a study of this type is that the size of the group is not too small. In the study presented here, a group size of 10 subjects seems to have been sufficient. It must be noted, however, that the material used here was such that it did not inspire too much variation. In studies of other types of material, for example spontaneous speech, the conditions may be different. A theoretical implication of this finding is that such variables as interstress interval duration and syllable duration seem to be functions of the language structure and that values for these parameters converge towards typical means when averaged over reasonably large groups.

8.3 Summary of the perception study.

In the perception study (Chapters 6 and 7), two 'key' questions were addressed: 1) With what kind of precision is it possible to determine the locations of stressed syllables? and 2) What is the just noticeable difference for interstress interval duration? An attempt was also made to compare duration discrimination in linguistic stimuli with discrimination of noise, and with linguistic stimuli where some of the semantic content was 'removed'.

8.3.1 Stress beat perception.

The main aim of the 'stress beat study' was to determine how accurately subjects are able to locate stressed syllables on a time scale. Using the concept of the difference limen, it was found that DLs for different subjects varied between 23.7 ms and 47.9 ms. The six best subjects had DLs below 30 ms. There were no significant differences between subjects with respect to how they located the stress beat for a particular syllable. The difference was only in the accuracy with which they performed.

Some correlation was found between the duration of the consonant preceding the vowel and stress beat location, but the correlation was weak and the precise interpretation of the result is not absolutely clear. The same may be said for a small regularization effect that was found. It was found to be significant, but the reason for it is unclear. It may be interpreted as regularization, but also as some kind of perceptual anticipation. But most importantly, both these types of displacements relative to the vowel onsets are very small compared to the interstress intervals in the test-phrase used. To accomplish complete

regularization, a displacement of some 160 ms would have been required. Compared to a displacement on that scale, the one found (20 ms) seems totally insignificant.

8.3.2 Duration perception.

If the variation in interstress interval duration is seen in the light of the results of the duration perception experiments, it must be concluded that the irregularities are well within the range of detectable deviations.

In one of the experiments described in Chapter 7, the task was to rank order interstress intervals according to duration. Although this is a considerably more difficult task than simply deciding whether they are equal or not, subjects performed at significantly better than chance level.

Differential fractions based on the duration discrimination experiments were found to be .15—.20 for the better subjects and slightly higher for the group as a whole. If the concept of the difference limen is used as the just noticeable difference, one may conclude that this difference is in the order of 10 to 15 % of the comparison duration. These results were achieved in discrimination tasks but there are results from other experiments (e.g. Michon, 1964; Hirsh, Monahan, Grant, and Singh, 1990; Monahan and Hirsh, 1990) which suggest that duration discrimination in a series of events is at least as good. If each interstress interval is regarded as the comparison against which the following interval is judged, it is found, in the material studied here, that contrasts of 10—15% or more are very frequent and that contrasts as high as 35—45% are not uncommon. If these values are representative of speech in general, the inevitable conclusion must be that the temporal irregularities found in speech should be detectable in perception.

8.4 Discussion of regularity in production.

In this study, durations were studied at both syllable and interstress interval level. An attempt was made to describe the variation as well as regularities in the durations of these intervals. One may say that both a great deal of variation and a great deal of regularity was found depending upon from which angle the results are viewed. If one looks at total durations of syllables and interstress intervals, there is variation within wide ranges. Syllable durations, for example, may vary between some 30 ms and more than 300 ms. Variation on almost the same scale is found in interstress interval duration. But if one tries to look into the causes behind the variation, a different pattern emerges. Durations were shown to depend in a very systematic way on such factors as the number of phonemic segments in an interval, stress, syllable position etc. From this point of view, the behaviour of both syllable and interstress interval durations is highly regular. It was shown, for example, that mean durations of interstress intervals were almost perfectly predictable, assuming duration to be a linear function of the number of syllables or phonemic segments

in the interval. The same type of relations hold true for syllables, although here one must also include such factors as stress and syllable position as variables to obtain good predictions. Given results like the ones just mentioned, one must ask what status such concepts as 'syllable-timing' and 'stress-timing' really have.

The very words 'stress-timing' and 'syllable-timing' imply that there are timing mechanisms operating directly on interstress intervals and syllables respectively. What reasons do we have to believe in the existence of such mechanisms? Let us look at two hypotheses that have been proposed as explanations for syllable timing. On the basis of studies of various motor activities, Lenneberg (1967) has hypothesized a frequency of 6 ± 1 Hz as an underlying frequency that governs the production of syllables. This frequency would correspond to durations of 140 to 200 ms. Lenneberg's conjecture is neurologically based and he cites, as one piece of evidence, findings of an EEG rhythm of approximately 7 Hz in neurological research. If Lenneberg is right then one would be justified in speaking about syllable timing in this particular sense. But an underlying neurological frequency is not the only possible explanation for a certain regularity at syllable level. Brodda (1979) has proposed an alternative view explaining typical syllable durations as a purely mechanical consequence of jaw movement. Based on a pilot study with himself as a subject he found an eigenfrequency of about 6 Hz (corresponding to a cycle time of 160—170 ms) for the jaw considered as a vibrating mass. This should determine normal (unstressed) syllable duration to a considerable extent, but wider gestures, for example connected with stress, would require more time and syllable duration would, as a consequence, be longer. The qualitative difference between the two hypotheses should be noted. Lenneberg's hypothesis describes a mechanism which is supposed to be actively involved in timing syllable durations whereas Brodda's hypothesis may be seen as a constraint on timing rather than an active process. It should also be pointed out that the two views are not contradictory. Both mechanisms could very well coexist.

Superficially, one may say that the above hypotheses agree with the results of the experimental study presented in Chapter 4. Syllables, particularly unstressed ones (mean duration 177 ms), are indeed often in the 160 ± 20 ms range as Lenneberg's and Brodda's hypotheses predict. But this is not enough to constitute evidence for syllable-timing. All it means is that the hypotheses are compatible with the results. What one must show is that there are phenomena which can only (or at least most likely) be explained by hypothesizing a special timing constraint for syllables. But this is where the hypotheses run into trouble. As was shown in the study in Chapter 4, syllable duration as well as interstress interval duration is a function of the number of phonemic segments to a very close approximation, and for syllables, stress and position may explain much of the remaining variation. It is true, however, that there is additional variation in syllable duration that cannot be explained by the number of phonemes, stress or position. If one regards, say, unstressed syllables with two segments, the variation is typically 125—325 ms, that is a factor of 2.5 approximately, for an individual subject, and occasionally more. It is likely that much of

this variation may be explained by the types of phonemic segments in the syllable, but this factor was not studied here. There is, thus, not regularity but additional variation that needs to be further explained. That being the case, it is difficult to see where syllable-timing fits in.

One may also put it in a slightly more negative way: who needs syllable-timing? What is it that syllable-timing must explain? If syllables were equal, or at least more equal than the factors we already know to condition syllable duration would lead us to believe, we might be looking for a mechanism that could explain 'unexpected regularity'. But this is not the case. On the contrary, syllable duration is highly variable. Most of the variation may be explained by factors like the number of phonemic segments, stress and so on, and what is left to explain is certainly not any added regularity but rather the fact that durations vary even more. There is additional *variation* which is not explained by those factors that we already have fairly well under control. So, again, who needs a theory that predicts *less* variation?

With respect to stress-timing the situation is similar to what has been said above about syllable-timing. The variation in interstress interval duration amounts to 50 % or more of the mean interval duration and very large durational contrasts may occur. As is the case with syllables, interstress interval duration is primarily a function of the number of segments in the interval. In the study presented in Chapter 4, no other factor was significantly correlated with interval duration. And it seems possible to explain interstress interval duration as a simple sum of its component syllable durations through what basically looks like a concatenative process.

The only serious objection to such a view is that, in some studies, what seems like a compression effect in longer intervals has been found. As was pointed out in the discussion in 3.3, the effect has only been found in some studies, usually of the 'target word in a carrier phrase' type. In the study presented above, no significant effect was found and other studies have produced the same negative result (e.g. Lehiste, 1990), or a weak or ambiguous tendency (e.g. Strangert, 1985).

But let us, for the sake of argument, assume that a compression effect exists, only it is too small to be always significant. It must then be said that we are talking about an effect which is in the order of 5% or less of normal interstress interval duration. The range of durations for stressed vowels in Fowler's (1977) study was 25 ms (mean value) for 1 to 3-syllable intervals, and the effect found for syllables by Fourakis and Monahan (1988) was in the same order. If it should be meaningful to speak about a stress-timing mechanism one must assume that it is independent of the syllable production to some degree and not only an automatic consequence of syllable timing and segmental timing. One must now ask what could reasonably be the purpose and function of such a mechanism. The alternation of stressed and unstressed syllables means imposing a hierarchical structure on speech. This is most certainly an important characteristic. Expecting stresses to recur

regularly in this sense may very well be important in perception. When Lehiste (1977) proposes that *"The listener expects isochrony—expects the stresses to follow each other at approximately equal intervals."* (p. 262) she may have had something like that in mind. And if there were a mechanism that managed successfully to achieve isochrony of interstress intervals, this might also facilitate perception, but it is highly unlikely that the small adjustments we are talking about in connection with possible syllable compression effects in longer intervals (25 ms or less!) may have this function. As was shown in Chapter 7, differences in duration in the order of 25 ms should not even be detectable. On these grounds alone one might question the existence of such a mechanism. Why would there be a mechanism involved in speech production which hardly ever achieves its purpose? In most cases, there is no significant effect and when there is, it is in the order of 25 ms which may not even make a detectable difference! If isochrony were really an important characteristic of speech, would one not, on the contrary, have every reason to expect the mechanism behind it to be highly successful?

Perceptual arguments could also be invoked against the stress-timing view. Lehiste's argument, that we hear speech as regular for two reasons, first of all because we are unable to detect most of the differences and secondly because we expect isochrony, could, in fact, be reversed. If it were really the case that we expect isochrony, why then are we so insensitive to the realization of that which we are said to expect. Would it not be more reasonable to expect us to be highly sensitive to deviations from isochrony if it were really an important quality. Huggins (1972a) found that disrupting the rhythm, in the sense of changing segment durations, as little as 2—3 % of an interstress interval duration created a detectable effect. This contrasts with the relative insensitivity to differences in total durations of intervals, found here and by Lehiste herself. At least some 10 % seems to be necessary for intervals to differ, if the difference should be perceptible. Why this difference in sensitivity between the two levels? Should one not expect a corresponding sensitivity to disruptions of isochrony if isochrony was really expected? This difference in perceptual sensitivity indirectly supports the concatenative view. It is important to realize segment durations accurately (at least relative durations given a certain articulation rate). But this requirement is very difficult to combine with any kind of isosyllabicity or isochrony. Equal timing at levels higher than segments would require a constantly, and very rapidly, changing articulation rate if relative durations of adjacent segments are to be even approximately preserved. Now, if one assumes that isochrony of interstress intervals would facilitate perception, constantly changing articulation rate would be the prize to pay, and this might also make segment perception more difficult, perhaps to a corresponding degree. So one must ask what, if anything, would be the advantage of such an organization. Now, even if such an organization seems unlikely on these somewhat speculative grounds, it cannot be dismissed a priori. But the combination of its relative unlikelihood and the fact that hardly any empirical support for it may be found speaks very strongly against even tendencies to isochrony and stress-timing.

A conclusion which seems valid based upon the discussion above is that there seems to be very little need to postulate any syllable-timing or stress-timing mechanisms. The results obtained in various experimental studies may be explained and modelled without the help of any hypotheses of this kind. This does not exclude the possibility that such mechanisms may exist, although they seem to leave no convincing traces, but in the absence of any obvious need for them one might as well follow the old rule: *"Frustra fit per plura quod potest fieri per pauciora"*.

8.5 Discussion of regularity in perception.

It is often claimed that although speech is not isochronous as far as any events in the speech signal are concerned, it is nevertheless perceived as regular. But a weak point here is that this claim has been very little studied. We do not really know to what extent this is true, and precisely under what conditions this illusion occurs. In this section, I will consider what conclusions one may draw based upon what is known at present.

The results from what seems to be the only study that has approached the question of stress-timing and syllable-timing from a perceptual point of view, that by Miller (1984) discussed in section 2.3.3, were ambiguous to say the least.

Other studies have approached the question in a more indirect way. Lehiste (1977) has proposed that we hear speech as regular because we are unable to perceive the differences in duration between intervals. And Darwin and Donovan (Darwin and Donovan, 1980; Donovan and Darwin, 1979) have suggested that we hear speech in a different way (speech mode) and perceive speech as regular on other grounds than merely its physical properties.

To be able to evaluate Lehiste's proposal that we hear speech as regular because we are unable to detect the differences, we must have a clear view of what the limitations of the perceptual system are in this respect. The studies presented in Chapters 6 and 7 were attempts to approach this question. Two factors were recognized as crucial for accurate interval duration perception; the delimiting of intervals and the perception of duration. With respect to the question of how accurately the intervals may be delimited it was shown that this may be done with an uncertainty in the order of some 30 ms. And duration perception of interstress intervals was shown to be accurate to about 10—15 % of the interval duration. Since interstress interval durations vary a lot more than that in normal speech, the conclusion was that the variation should also be detectable in many, if not most, cases. And when subjects were told to rank order the intervals in a phrase, they were indeed able to do so with a significantly better success than mere chance. These results seem to speak against Lehiste's assumption. Now, the reservation one may raise against such a conclusion is, of course, that the results only tell us something about what the perceptual system is capable of under optimal, or near optimal, conditions and perhaps not very much about normal listening.

The other hypothesis, that we listen to speech in 'speech mode' and doing so perceive speech as more regular than it is physically, was discussed in detail in 3.5 and there is no reason to repeat that discussion here. Let us only remind ourselves that those who have questioned that view (Bell and Fowler, 1984; Scott, Isard, and Boysson-Bardies, 1985) did so on principally two grounds; the same regularization effects were obtained by using other types of complex stimuli and the effect was only found with some, but not all, subjects.

To summarize what we know so far, one may say that: 1) Subjects are perfectly able to detect durational differences between interstress intervals (at least under laboratory listening conditions) 2) The perceptual regularization effects may also occur with complex non-speech sounds, which implies that the effect may simply be the result of regular responses when the task becomes too difficult, and 3) Not all subjects regularize, some may reproduce the intervals accurately, indicating that they are able to perceive the rhythm veridically.

Looking at these results, there certainly seems to be little basis for claiming that speech is perceived as regular. Must we not conclude then that Marcus, Lehiste and others who think that we perceive speech as regular, at least under certain circumstances, are simply wrong? Well perhaps, but then again, perhaps not. In the following, I will suggest an alternative view which at least permits the acceptance of both types of results.

The very fact that not all subjects behaved in the same way in the regularization experiments may provide a clue to what it is that speech mode perception is all about. Speech exhibits many types of regularities. Not so much in the sense of some physical events occurring at very regular intervals in time, but in a number of other ways. Stresses are regularly recurring in the sense that under normal circumstances there is often a stressed syllable every two or three syllables or so. And they are, for good reasons, the syllables we should pay most attention to. They are the elements which most of the time are richest in information—they often carry most semantic weight. Stress often helps in disambiguating otherwise ambiguous sentences and stressed syllables are involved when there is focus on something, when matters are topicalized and so on. There are, thus, a number of reasons to pay particular attention to stresses. In doing so, we are often helped by a number of acoustic properties that make the stressed syllables more salient. They may be longer, higher in pitch, louder etc.

If instead of seeing the speech stream primarily as a flow of acoustic events on a physical time scale, we regard it as a flow of information, there are definitely regularities that we must pay close attention to. It is, therefore, crucial, if one wants to understand fully the perception of rhythm, to know a lot more about 'attention' in speech perception.

My experiments and those of Bell and Fowler (1984) have certainly demonstrated that subjects are able to judge, veridically, the rhythm and durations of speech if that is what they specifically put their attention to. At least that is true for many of them. But it goes without saying that this is not how we normally listen to speech. Now, when people report

that they perceive speech as physically regular, all we may be fairly sure of is that that is what they *think* they do. That is how they interpret their perceptions. But, (at least for some of them), I would like to propose the following alternative explanation: They perceive regularly occurring events, often connected with stresses ⇒ they interpret the structure of what they hear as regular, because in the particular sense just described that is precisely what it is ⇒ they tap a regular rhythm, or express the same thing by some other means, having a general feeling that what they just heard was quite regular.

Now, I will not try to deny the speculative element in what I have just said, but I will maintain that it has a lot to speak for it. If it were true, the results from some of the experiments which inspired this discussion in the first place, are exactly what one would expect, rather than seem contradictory as may otherwise be the case. There is nothing wrong with our perceptual system. If we manage to put our attention to the physical properties of the speech signal and specifically listen for its durational structure, we (= some of us) are perfectly able to detect many (if not most) of the irregularities. (The case of the Dutch subjects in my study is another strong point in case here, helped as they were in concentrating on the physical side of the sounds because the semantic side was 'incomprehensible' to them.) Now as with other abilities, some people are able and some are less able. Those who are able to control their attention to such a degree that they may 'filter out', to a greater or lesser extent, information which is normally highly relevant but in this particular situation constitutes 'noise', will also succeed in detecting the 'real' durational properties, and those who are less able will not. Thus the two seemingly contradictory types of behaviour. Now, how to test this hypothesis, if that is what it may be regarded as, is of course, an entirely different question.

8.6 Some suggestions for further research.

As was shown in 2.3.3, the study of speech rhythm has been carried out in a rather unsystematic way. And most studies have attempted to test whether a particular language fits into one or the other of the two categories 'stress-timing' and 'syllable-timing' rather than take an unbiased look at the actual durations of syllables and interstress intervals and how they interact. This means, for example, that very little data exists upon which to make comparisons of the languages conventionally classified as stress-timed, in the case of which normally only interstress interval durations have been studied, and those believed to be syllable-timed, usually only studied with respect to syllable durations. This 'cart before the horse' type of approach should be abandoned, it serves no useful purpose.

What one would like to see is some kind of programme of basic questions with respect to speech rhythm, which could be systematically studied with a comparative perspective in mind. In this programme it is important also to include studies of the perception of speech rhythm, an area which, for some reason, seems to have been almost entirely neglected. In

the following sections I will try to suggest a few questions that may be worth giving some consideration in such a programme.

8.6.1 Interstress interval duration as a function of the number of syllables or phonemic segments.

Data in studies by Faure, Hirst and Chafcouloff (1980), Nakatani, O'Connor and Aston (1981) and Dauer (1983), and the one presented here suggest that, as a first approximation, interstress interval durations grow linearly as a function of the number of syllables. Moreover, in all studies, the 'growth rates' found seem to be approximately the same; usually around 110 ms per added unstressed syllable. The constant in the linear function, that is the added duration due to the stressed syllable, however, seems to be language dependent. The data suggest that the constant term is either around 100 ms or a little more than 200 ms. These values mean a slight oversimplification with respect to data, but are a very good first approximation.

The data available at present thus suggest that it may be possible to base a classification on the linear function expressing interstress interval duration as a function of the number of syllables. Even if the simple relation that seems to hold for the limited set of data available does not hold when more languages are included, it is still possible that 'growth functions' based on linear regression analysis could be used as a basis for classification. What the data available so far suggest is at least that the possibility is worth exploring.

The results in Chapter 4 indicated that the number of phonemic segments was an even better predictor of interval duration than the number of syllables. If the results found here and by Fant and Kruckenberg (1989) are representative, it also seems as if the 'growth rate' is extremely constant between experiments (approximately 53 ms). How would this compare with corresponding rates for other languages than Swedish? Here is another possible candidate for a universal.

8.6.2 The internal structure of interstress intervals—syllable duration.

Even if it turns out to be true that interstress interval durations grow linearly as a function of the number of syllables and segments, this does not imply that syllable durations must be independent of the number of syllables in the interstress interval. As I have shown (see 3.3), constant syllable duration is only one of several possibilities. There may also be compression (or stretching) processes affecting syllables (stressed, unstressed or both) as the interval size grows. This did not turn out to be the case in the material studied above but different linguistic material may yield different results. Again, this may form a basis for typological classification. Different types of compression, stretching, or constancy as a function of the number of syllables may be operating within interstress intervals in

different languages, even if interval duration growth is constant. The timing constraints at interstress interval level may be different enough to constitute a basis for typological classifications.

It should also be pointed out that even negative evidence with respect to this question must be regarded as an important result. If it should be the case that there are no differences between languages in this respect, it would mean that these particular timing mechanisms are language independent. This would, of course constitute an equally important piece of knowledge. A study of this kind is thus worth doing regardless of what expectations one may have about the outcome.

It was suggested in 3.2, that the difference between languages lies primarily in differences in the status (and as a consequence the duration) of stressed syllables. Thus the durational contrast between stressed and unstressed syllables seems to be an important factor with respect to speech rhythm. Now, this contrast may be accomplished in various ways. It may, for example, be the case that stressed syllables are simply shorter in the 'syllable-timed' group but that everything else is much the same. There are some indications, however, that this may be too simple an assumption. If Delattre's (1966) results are representative then the differences in contrast are reflected in both stressed and unstressed syllables. It seems as if a low contrast, like in Spanish (typically 1.30:1), is the result of shorter stressed syllables as well as longer unstressed ones. In fact, this trend seems to be present for all languages in the material presented by Delattre, with French forming one end point of the scale and Spanish the other. Results of this kind are, of course, highly relevant with respect to the possibility of finding grounds for a classification and should be studied much more closely and extensively.

8.6.3 The internal structure of interstress intervals—syllable structure.

In connection with the study of syllable durations in different positions within the interstress intervals, syllable structure must also be studied. As was shown in 4.6.4 the number of segments in a syllable varied considerably with position, medial syllables being the shortest. Since syllable duration was found to be a function of the number of segments, the variation in the number of segments will, of course, have immediate effects on duration. The question of syllable structure must, therefore, be studied further.

Mean values for interstress interval durations, averaged over a reasonable amount of data, seem to be around 500 ms with rather small deviations between studies. Mean values in Dauer's (1983) study were just under 500 ms for all languages. Faure, Hirst, and Chafcouloff (1980) found mean interval duration for two speakers to be 476 ms. In the study by Bolinger (1965) referred to in 2.3.1, the mean is again approximately 500 ms (calculated by myself). And in the study presented in Chapter 4, a value of 580 ms was found. Now, if as has been shown to be the case in the material studied here, interval duration is a function

of the number of syllables and segments in the interval, then the explanation for the consistency in mean values, within and between languages, may have its explanation in the syllable structure. This question should be studied further.

8.6.4 Perception of rhythmic structure in speech.

Although the perception of speech rhythm is as important as its production, surprisingly little research has been done in this area. The major part of the studies in this field have been concerned with the perception of stress beats or p-centres. Hardly anything has been done on the perception of structure and regularity, and the relation between the two. One question one might ask, as I mentioned in Chapter 1, is whether regularity of structure or regularity of timing is the most important aspect of the perception of regularity. The fact that structurally regular poems, say iambic verse, are often perceived as highly regular, although interstress interval durations may actually vary as much as in other types of speech, suggests that structure may be even more important than other characteristics. And Manrique and Signorini (1983) in their study of speech rhythm and durations in Argentine Spanish suggest that it may be the syllable structure of Spanish that is responsible for the its perceived rhythm rather than syllable or interstress interval durations. The correlations between structural aspects and perception should be further explored in experimental studies.

8.6.5 Typology based on speech rhythm perception.

As was mentioned in 2.3.3, one may hypothesize that, even if the classification of languages as stress-timed or syllable-timed should find little or no support in studies of speech production, it may still be the case that a classification could be constructed on perceptual grounds. This is a hypothesis well worth exploring and one would expect to find quite a few studies of this kind. But for some reason this possibility has not been given much consideration. The study by Miller (1984) discussed in 2.3.3 is to my knowledge the only one exploring this possibility. In Miller's study, seven languages were studied—Arabic, Spanish, Japanese, Yoruba, Polish and Finnish. For those who believe that the languages of the world can be classified as stress-timed or syllable-timed the outcome of this experiment must be rather discouraging. Only Arabic, conventionally classified as stress-timed, met the prediction. If Miller's results are representative it means that the classification of languages as stress-timed or syllable-timed on perceptual grounds is untenable. But more studies must be done to resolve this problem definitely.

It must also be stressed that the study of speech rhythm should not only be carried out within the framework of syllable-timing and stress-timing. Studies of musical rhythm which seem highly relevant also in the context of speech rhythm have been mentioned. The technique used by Gabrielsson (1973a) asking subjects to rate different musical stimuli

using adjectives like 'simple', 'varied', 'wild', 'pulsating', 'aggressive' etc. to characterize different rhythmical impressions may very well be used in speech rhythm research. Using factor analysis, it may be possible to find new and more interesting dimensions of speech rhythm than those used hitherto. There is also a possibility that different languages may fall into different categories that can be constructed on the basis of the results of such experiments. This opens the possibility of finding new, and perhaps more interesting, ways of constructing a language typology for rhythm based on perception.

Bibliography

ABEL, SHARON M. 1972a. Duration discrimination of noise and tone bursts. *Journal of the Acoustical Society of America*, **51**, 1219—1223.

ABEL, SHARON M. 1972b. Discrimination of temporal gaps. *Journal of the Acoustical Society of America*, **52**, 519—524.

ABERCROMBIE, DAVID. 1965. A phoneticians view of verse structure. In David Abercrombie, *Studies in Phonetics and Linguistics*, 16—25. London: Oxford University Press.

ABERCROMBIE, DAVID. 1967. *Elements of General Phonetics*. Edinburgh: Edinburgh University Press.

ADAMS, CORINNE. 1979. *English Speech Rhythm and the Foreign Learner*. The Hague: Mouton.

ALLAN, LORRAINE G. 1979. The perception of time. *Perception & Psychophysics*, **26**, 340—354.

ALLAN, LORRAINE G. & ALFRED B. KRISTOFFERSON. 1974a. Psychophysical theories of duration discrimination. *Perception & Psychophysics*, **16**, 26—34.

ALLAN, LORRAINE G. & ALFRED B. KRISTOFFERSON. 1974b. Judgments about the duration of brief stimuli. *Perception & Psychophysics*, **15**, 434—440.

ALLAN, LORRAINE G. & ALFRED B. KRISTOFFERSON. 1974c. Successiveness discrimination: Two models. *Perception & Psychophysics*, **15**, 37—46.

ALLAN, LORRAINE G., ALFRED B. KRISTOFFERSON & MARNIE E. RICE. 1974. Some aspects of perceptual coding of duration in visual duration discrimination. *Perception & Psychophysics*, **15**, 83—88.

ALLAN, LORRAINE G., ALFRED B. KRISTOFFERSON & E. W. WIENS. 1971. Duration discrimination of brief light flashes. *Perception & Psychophysics*, **9**, 327—334.

ALLEN, GEORGE D. 1972. The location of rhythmic stress beats in English: An experimental study I & II. *Language and Speech*, **15**, 72—100, and 179—195.

ALLEN, GEORGE D. 1973. Segmental timing control in speech production. *Journal of Phonetics*, **1**, 219—237.

ALLEN, GEORGE D. 1975. Speech rhythm: Its relation to performance universals and articulatory timing. *Journal of Phonetics*, **5**, 75—86.

ALLEN, GEORGE D., & SARAH HAWKINS. 1980. Phonological rhythm: Definition and development. In Grace H. Yeni-Komshian, James F. Kavanagh and Charles A. Ferguson (Eds.), *Child Phonology*, 227—256. N.Y.: Academic Press.

ARISTOTLE. *On the Art of Poetry*. Translated by Ingram Bywater. Oxford: University of Oxford Press.

ARISTOTLE. *The "Art" of Rhetoric*. Translated by John Henry Freese. (The Loeb classical library) London: Heinemann.

AVANT, LLOYD L. & PAUL J. LYMAN. 1975. Stimulus familiarity modifies perceived duration in prerecognition visual processing. *Journal of Experimental Psychology: Human Perception and Performance*, **1**, 205—213.

AVANT, LLOYD L., PAUL J. LYMAN & JAMES R. ANTES. 1975. Effects of stimulus familiarity upon judged visual duration. *Perception & Psychophysics*, **17**, 253—262.

BAIRD, JOHN C. & ELLIOT NOMA. 1978. *Fundamentals of Scaling and Psychophysics*. New York: John Wiley & Sons.

BECKMAN, MARY. 1982. Segment duration and the 'mora' in Japanese. *Phonetica*, **39**, 113—135.

BELL, ALAN & CAROL A. FOWLER. 1984. Perception of the rhythm of English and of nonspeech analogues. *Paper presented to the 108th meeting of the Acoustical Society of America, Minneapolis, October 1984*.

BENGTSSON, INGEMAR & ALF GABRIELSSON. 1983. Analysis and synthesis of musical rhythm. In Johan Sundberg (Ed.), *Studies of Music Performance, Publications issued by the Royal Swedish Academy of Music*, **39**, 15—24.

BENGUEREL, ANDRÉ-PIERRE. 1986. Comments on "Perceptual isochrony in English and in French" (*Journal of Phonetics*, 13, 155—162). *Journal of Phonetics*, **14**, 331—332.

BENGUEREL, ANDRÉ-PIERRE & JANET D'ARCY. 1986. Time-warping and the perception of rhythm in speech. *Journal of Phonetics*, **14**, 231—246.

BERGLUND, BIRGITTA, ULF BERGLUND, GÖSTA EKMAN & MARIANNE FRANKENHAEUSER. 1969. The influence of auditory stimulus intensity on apparent duration. *Scandinavian Journal of Psychology*, **10**, 21—26.

BERNOULLI, DANIEL. (1738). Exposition of a new theory on the measurement of risk (translated by L. Sommer), *Econometrica*, **22**, 23—36.

BHARUCHA, JAMSHED JAY & JOHN H. PRYOR. 1986. Disrupting the isochrony underlying rhythm: An asymmetry in discrimination. *Perception & Psychophysics*, **40**, 137—141.

BLANKENSHIP, DONALD A. & NORMAN H. ANDERSON. 1976. Subjective duration: A functional analysis. *Perception & Psychophysics*, **20**, 168—172.

BOCHNER, JOSEPH H., KAREN B. SNELL & DOUGLAS J. MACKENSIE. 1988. Duration discrimination of speech and tonal complex stimuli by normally hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America*, **84**, 493—500.

BOLINGER, DWIGHT. L. 1965. Pitch accent and sentence rhythm. In Isamu Abe and Tetsuya Kanekiyo (Eds.), *Forms of English: Accent, Morpheme, Order*, 139—180. Cambridge, Mass.: Harvard University Press.

BOLTON, THADDEUS L. 1894. Rhythm. *American Journal of Psychology*, **6**, 145—238.

BOND, Z.S. & JOANN FOKES. 1985. Non-native patterns of English syllable timing. *Journal of Phonetics*, **13**, 407—420.

BORCHGREVINK, HANS M. 1982. Prosody and musical rhythm are controlled by the speech hemisphere. In Manfred Clynes (Ed.), *Music, Mind, and Brain*, 151—157. New York: Plenum Press.

BORING, EDWIN G. 1952. *Sensation and Perception in the History of Experimental Psychology*. N. Y.: Appleton-Century-Crofts.

BROADBENT, DONALD E. & PETER LADEFOGED. 1959. Auditory perception of temporal order. *Journal of the Acoustical Society of America*, **31**, 1539.

BRODDA, BENNY. 1979. Något om de svenska ordens fonotax och morfotax. *Papers from the Institute of Linguistics, University of Stockholm*, **38**.

BRUCE, GÖSTA. 1983. On rhythmic alternation. *Working Papers*, **25**, 35—52. Lund: Linguistics-Phonetics, Lund University.

BRUCE, GÖSTA. 1984. Rhythmic alternation in Swedish. In Claes-Christian Elert, Irene Johansson and Eva Strangert (Eds.), *Nordic Prosody III*, 31—41. Stockholm: Almquist & Wiksell International.

BURGHARDT, H. 1973a. Die subjektive Dauer schmalbandiger Schalle bei verschiedenen Frequenzlagen. *Acustica*, **28**, 278—284.

BURGHARDT, H. 1973b. Über die subjektive Dauer von Schallimpulsen und Schallpausen. *Acustica*, **28**, 284—290.

BUXTON, HILARY. 1983. Temporal predictability in the perception of English speech. In Anne Cutler and D. R. Ladd (Eds.), *Prosody: Models and Measurement*, 111—121. Berlin: Springer Verlag.

CARBOTTE, RAMONA M. 1973. Retention of time information in forced choice duration discrimination. *Perception & Psychophysics*, **14**, 440—444.

CARLSON, ROLF & BJÖRN GRANSTRÖM. 1975. Perception of segmental duration. In Antonie Cohen and Sieb G. Nooteboom (Eds.), *Structure and Process in Speech Perception*, 90—106.

CARLSON, ROLF, BJÖRN GRANSTRÖM & DENNIS H. KLATT. 1979. Some notes on the perception of temporal patterns in speech. *Proceedings of the ninth international congress of phonetic sciences, Copenhagen 1979, Vol II*, 260—267.

CARLSON, ROLF, ANDERS FRYDEN, BJÖRN GRANSTRÖM & JOHAN SUNDBERG. 1987. Speech and music performance: Parallels and contrasts. *STL-QPSR 4/1987.*

CARLSON, V. R. & I. FEINBERG. 1968. Individual variations in time judgment and the concept of an internal clock. *Journal of Experimental Psychology*, **77**, 631—640.

CARLSON, V. R. & I. FEINBERG. 1970. Time judgment as a function of method, practice, and sex. *Journal of Experimental Psychology*, **85**, 171—180.

CATFORD, JOHN C. 1977. *Fundamental Problems in Phonetics*. Edinburgh: Edinburgh University Press.

CLASSE, ANDRÉ. 1939. *The Rhythm of English Prose*. Oxford: Basil Blackwell.

CLYNES, MANFRED & JANICE WALKER. 1982. Neurobiological functions of rhythm, time, and pulse in music. In Manfred Clynes (Ed.), *Music, Mind, and Brain*, 171—216. New York: Plenum Press.

COHEN, LOUIS & MICHAEL HOLLIDAY. 1982. *Statistics for Social Scientists*. London: Harper & Row.

COOPER, WILLIAM E., JEANNE M. PACCIA & STEVEN G. LAPOINTE. 1978. Hierarchical coding in speech timing. *Cognitive Psychology*, **10**, 154—177.

CORDER, S. PIT. 1973. *Introducing Applied Linguistics*. Harmondsworth, Middlesex: Penguin.

CRAIG, JAMES C. 1973. A constant error in the perception of brief temporal intervals. *Perception & Psychophysics*, **13**, 99—104.

CREELMAN, C. DOUGLAS. 1962. Human discrimination of auditory duration. *Journal of the Acoustical Society of America*, **34**, 582—593.

CRYSTAL, THOMAS H. & ARTHUR S. HOUSE.1982. Segmental durations in connected speech signals: Preliminary results. *Journal of the Acoustical Society of America*, **72**, 705—716.

CRYSTAL, THOMAS H. & ARTHUR S. HOUSE. 1988a. Segmental durations in connected speech signals: Current results. *Journal of the Acoustical Society of America*, **83**, 1553—1573.

CRYSTAL, THOMAS H. & ARTHUR S. HOUSE. 1988b. Segmental durations in connected speech signals: Syllabic stress. *Journal of the Acoustical Society of America*, **83**, 1574—1585.

CRYSTAL, THOMAS H. & ARTHUR S. HOUSE. 1988c. A note on the variability of timing control. *Journal of Speech and Hearing Research*, **31**, 497—502.

CURTIS, DWIGHT W. & STANLEY J. RULE. 1977. Judgement of duration relations: Simultaneous and sequential presentations. *Perception & Psychophysics*, **22**, 578—584.

CUTLER, ANNE. 1980. Syllable omission errors and isochrony. In Hans W. Dechert and Manfred Raupach (Eds.), *Temporal Variables in Speech*, 183—190. The Hague: Janua Linguarum Series Maior, 86, Mouton Publishers.

DARWIN, CHRISTOPHER J. & ANDREW DONOVAN. 1980. Perceptual studies of speech rhythm: Isochrony and intonation. In J. C. Simon (Ed.), *Spoken Language Generation and Understanding*, 77—85. Dordrecht, Holland: D. Riedel Publishing Company.

DAUER, REBECCA M. 1983. Stress–timing and syllable–timing reanalyzed. *Journal of Phonetics*, **11**, 51— 62.

DECHERT, HANS W. & MANFRED RAUPACH, (Eds.). 1980. *Temporal Variables in Speech*. The Hague: Mouton Publishers.

DELATTRE, PIERRE. 1966. A comparison of syllable length conditioning among languages. *International Review of Applied Linguistics*, **4**, 183—198.

DEVANE, J. R. 1974. Word characteristics and judged duration for two response sequences. *Perception and Motor Skills*, **38**, 525—526.

DIEHL, RANDY L. 1987. On segments and segment boundaries. *Journal of Phonetics*, **15**, 289—290.

DIVENYI, PIERRE L. & WILLIAM F. DANNER. 1977. Discrimination of time intervals marked by brief acoustic pulses of various intensities and spectra. *Perception & Psychophysics*, **21**, 125—142.

DONOVAN, ANDREW & CHRISTOPHER J. DARWIN. 1979. The perceived rhythm of speech. *Proceedings of the ninth international congress of phonetic sciences, Copenhagen 1979, Vol. II*, 268—274.

DUEZ, DANIELLE & YUKIHORO NISHINUMA. 1985. Some evidence on rhythmic patterns of spoken French. *Experiments in Speech Processes, Report IV*, 30—40. Stockholm: University of Stockholm, Institute of linguistics.

EEFTING, W. & A. C. M. RIETVELD. 1989. Just noticeable differences of articulation rate at sentence level. *Speech Communication*, **8**, 355—361.

EHRLICH, STÉPHANE. 1958. Le mécanisme de la synchronisation sensori-motrice. *L'Année Psychologique*, **58**, 7—23.

EHRLICH, STÉPHANE. 1960. Rhythm: Recent French contributions. *Acta Psychologica*, **17**, 155—176.

EHRLICH, STÉPHANE, GENEVIEVE OLÉRON & PAUL FRAISSE. 1956. La structuration tonale des rythmes. *L'Année Psychologique*, **56**, 27—45.

EISLER, HANNES. 1975. Subjective duration and psychophysics. *Psychological Review*, **82**, 429—450.

EISLER, HANNES. 1976. Experiments on subjective duration 1868—1975: A collection of power function exponents. *Psychological Bulletin*, **83**, 1154—1171.

EKMAN, GÖSTA. 1959. Weber's law and related functions. *Journal of Psychology*, **47**, 343—352.

ELING, PAUL A., JOHN C. MARSHALL & GERARD P. VAN GALEN. 1980. Perceptual centres for Dutch digits. *Acta Psychologica*, **46**, 95—102.

FAURE, GEORGES, D. J. HIRST & MICHEL CHAFCOULOFF. 1980. Rhythm in English: Isochronism, pitch, and perceived stress. In Linda R. Waugh and C. H. van Schoonevald (Eds.), *The Melody of Language*, 71—79. Baltimore: University Park Press.

FANT, GUNNAR & ANITA KRUCKENBERG. 1989. Preliminaries to the study of Swedish prose reading and reading style. *Speech Transmission Laboratory Quarterly Progress and Status Report*, **2/1989**, 1—80. Stockholm: Royal Institute of Technology (KTH).

FANT, GUNNAR, ANITA KRUCKENBERG & LENNART NORD. 1989. Stress patterns, pauses, and timing in prose reading. *Speech Transmission Laboratory Quarterly Progress and Status Report*, **1/1989**, 7—12. Stockholm: Royal Institute of Technology (KTH).

FECHNER, GUSTAV THEODOR. 1860. *Elemente der Psychophysik*. Leipzig: Breitkopf & Härterl.

FLEGE, J. E. & W. S. BROWN, JR. 1982. Effects of utterance position on English speech timing. *Phonetica*, **39**, 337—357.

FONAGY, IVAN. 1979. Artistic vocal communication at the prosodic level. In Harry and Patricia Hollien (Eds.), *Current issues in the phonetic sciences, Proceedings of the IPS–77 congress, Florida*, 245—260. Amsterdam: John Benjamins B.V.

FOURAKIS, MARIOS & CAROLINE B. MONAHAN. 1988. Effects of metrical foot structure on syllable timing. *Language and Speech*, **31**, 283—306.

FOWLER, CAROL A. 1977. *Timing Control in Speech Production*. Bloomington, Indiana: Indiana University Linguistics Club.

FOWLER, CAROL A. 1979. "Perceptual centers" in speech production and speech perception. *Perception and Psychophysics*, **25**, 375—388.

FOWLER, CAROL A. 1983. Converging sources of evidence on spoken and perceived rhythms of speech: Cyclic production of vowels in monosyllabic stress feet. *Journal of Experimental Psychology: General*, **112**, 386—412.

FOWLER, CAROL A. 1984. Segmentation of coarticulated speech in perception. *Perception & Psychophysics*, **36**, 359—68.

FOWLER, CAROL A. 1987. Consonant-vowel cohesiveness in speech production as revealed by initial consonant exchanges. *Speech Communication*, **6**, 231—44.

FOWLER, CAROL A. & LOUIS G. TASSINARY. 1981. Natural measurement criteria for speech: The anisochrony illusion. In John Long and Alan Baddeley (Eds.), *Attention and Performance, IX*, 521—535. Hillsdale, N.J.: Earlbaum.

FOWLER, CAROL A., MARY R. SMITH & LOUIS G. TASSINARY. 1986. Perception of syllable timing by prebabbling infants. *Journal of the Acoustical Society of America*, **79**, 814—825.

FOX, ROBERT ALLEN & ILSE LEHISTE. 1987. The effect of vowel quality variations on stress-beat location. *Journal of Phonetics*, **15**, 1—13.

FRAISSE, PAUL. 1948a. Les erreurs constantes dans la reproduction de courts intervalles temporels. *Archives de Psychologie*, **32**, 161—176.

FRAISSE, PAUL. 1948b. Rythmes auditifs et rythmes visuels. *L'Année Psychologique*, **49**, 21—42.

FRAISSE, PAUL. 1952. La perception de la durée comme organisation du successif. *L'Année Psychologique*, **52**, 39—46.

FRAISSE, PAUL. 1956. *Les structures rythmiques*. Paris: Érasme.

FRAISSE, PAUL. 1961. Influence de la durée et de la frequency des changements sur l'estimation du temps. *L'Année Psychologique*, **61**, 325—339.

FRAISSE, PAUL. 1964. *The Psychology of Time*. London: Eyre & Spottiswoode.

FRAISSE, PAUL. 1966. L'anticipation de stimulus rythmiques. Vitesse d'établissement et précision de la synchronisation. *L'Année Psychologique*, **66**, 15—36.

FRAISSE, PAUL. 1971. L'apprentissage de l'estimation de la durée et ses repères. *L'Année Psychologique*, **71**, 371—379.

FRAISSE, PAUL. 1978. Time and rhythm perception. In Edward C. Carterette and Morton P. Friedman (Eds.), *Handbook of Perception, Vol. VIII*, 203—254. New York: Academic Press.

FRAISSE, PAUL. 1982. Rhythm and tempo. In Diana Deutsch (Ed.), *The Psychology of Music*, 149—180. New York: Academic Press.

FRAISSE, PAUL & STÉPHANE EHRLICH. 1955. Note sur la possibilité de syncoper en fonction du tempo d'une cadence. *L'Année Psychologique*, **55**, 61—65.

FRAISSE, PAUL & GENEVIEVE OLÉRON. 1950. La perception de la durée d'un son d'intensité croissante. *L'Année Psychologique*, **50**, 327—343.

FRAISSE, PAUL & GENEVIEVE OLÉRON. 1954. La structuration intensive des rythmes. *L'Année Psychologique*, **54**, 35—52.

FRAISSE, PAUL & CLAUDE VOILLAUME. 1971. Les repères du sujet dans la synchronisation et dans la pseudo-synchronisation. *L'Année Psychologique*, **71**, 359—369.

FRY, D. B. 1958. Experiments in the perception of stress. *Language and Speech*, **21**, 126—152.

FRY, D. B. 1971. Time constants in speech. In L. L. Hammerich, Roman Jacobson and Eberhart Zwirner (Eds.), *Form & Substance*, 171—180. Copenhagen: Akademisk Forlag.

FUJIMURA, OSAMU. 1987. A linear model of speech timing. In Robert Channon and Linda Shockey (Eds.), *In honour of Ilse Lehiste*, 109—123. Dordrecht: Foris Publications.

FUJISAKI, HIROYA, KIMIE NAKAMURA & TOSHIAKI IMOTO. 1975. Auditory perception of duration of speech and non-speech stimuli. In Gunnar Fant and M. A. A. Tatham (Eds.), *Auditory Analysis and Perception of Speech*, 197—220. London: Academic Press.

GABRIELSSON, ALF. 1969. Empirisk rytmforskning. *Svensk tidskrift för musikforskning*, **51**, 49—118.

GABRIELSSON, ALF. 1973a. Adjective ratings and dimension analysis of auditory rhythm patterns. *Scandinavian Journal of Psychology*, **14**, 244—260.

GABRIELSSON, ALF. 1973b and 1973c. Similarity ratings and dimension analyses of auditory rhythm patterns, I and II. *Scandinavian Journal of Psychology*, **14**, 138—160 and 161—176.

GABRIELSSON, ALF. 1974. Performance of rhythm patterns. *Scandinavian Journal of Psychology*, **15**, 63—72.

GABRIELSSON, ALF. 1979. Experimental research on rhythm. *The Humanities Association Review*, **30**, 69—92.

GABRIELSSON, ALF. 1982. Perception and performance of musical rhythm. In Manfred Clynes (Ed.), *Music, Mind and Brain*, 159—169. New York: Plenum Press.

GABRIELSSON, ALF. 1985. Interplay between analysis and synthesis in studies of music performance and music experience. *Music Perception*, **3**, 59—86.

GAY, THOMAS. 1978. Physiological and acoustic correlates of perceived stress. *Language and Speech*, **21**, 347—353.

GILI Y GAYA SAMUEL. 1940. La cantidad silábica en la frase. *Castilla (Valladolid)*, **I**, 287—298.

GESCHEIDER, GEORGE A. 1985. *Psychophysics: Method, Theory, and Application.* Hillsdale, N. J.: Lawrence Earlbaum.

GETTY, DAVID J. 1975. Discrimination of short temporal Intervals: A comparison of two models. *Perception & Psychophysics*, **18**, 1—8.

GETTY, DAVID J. 1976. Counting processes in human timing. *Perception & Psychophysics*, **20**, 191—197.

GROSJEAN, FRANÇOIS & HARLAN LANE. 1981. Temporal variables in the perception and production of spoken sign language. In P. P. Eimas and J. L. Miller (Eds.), *Perspectives on the Study of Speech*, 207—237. Lawrence Earlbaum Associates.

GUILFORD, JOY P. 1954. *Psychometric Methods.* New York: McGraw Hill.

GUSTAFSON, KJELL. 1988. The graphical representation of rhythm. *Progress reports from Oxford phonetics*, 6—26. Oxford: University of Oxford, Phonetics Laboratory.

GÅRDING, EVA. 1975. The influence of tempo on rhythmic and tonal patterns in three Swedish dialects. *Working Papers*, **12**, 71—83. Lund: Phonetics Laboratory, University of Lund.

HAGGARD, MARK. 1973. Correlates between successive segment durations: Values in clusters. *Journal of Phonetics*, **1**, 111—116.

HALLIDAY, MICHAEL A. K. 1967. *Intonation and Grammar in British English.* The Hague: Mouton.

HALLIDAY, MICHAEL A. K. 1970. *A Course in Spoken English: Intonation.* London: Oxford University Press.

HALPERN, ANDREA R. & CHRISTOPHER J. DARWIN. 1982. Duration discrimination in a series of rhythmic events. *Perception & Psychophysics*, **31**, 86—89.

HANDEL, STEPHEN & PEYTON TODD. 1981. Segmentation of sequential patterns. *Journal of Experimental Psychology: Human Perception and Performance*, **7**, 41—55.

HAWKINS, SARAH. 1979. Processes in the development of speech timing control. *Current Issues in the Phonetic Sciences, Proceedings of the IPS–77 Congress, Florida*, 267—278. Amsterdam: John Benjamins B.V.

HELSON, HARRY. 1964. *Adaptation-Level Theory: An Experimental and Systematic Approach to Behaviour*. New York: Harper and Row.

HENRY, FRANKLIN M. 1948. Discrimination of the duration of a sound. *Journal of Experimental Psychology*, **38**, 734—743.

HELLSTRÖM, ÅKE. 1977. Time errors are perceptual. *Psychological Research*, **39**, 345—388.

HILL, D. R., WIKTOR JASSEM & IAN H. WITTEN. 1979. A Statistical approach to the problem of isochrony in spoken British English. *Current issues in the phonetic sciences, Proceedings of the IPS–77 Congress, Florida*, 286—294. Amsterdam: John Benjamins B.V.

HIRSH, IRA J. 1959. Auditory perception of temporal order. *Journal of the Acoustical Society of America*, **31**, 759—767.

HIRSH, IRA J., CAROLINE B. MONAHAN, KEN W. GRANT & PUNITA G. SING. 1990. Studies in auditory timing: 1. Simple patterns. *Perception & Psychophysics*, **47**, 215—226.

HIRST, DANIEL. 1983. Structures and categories in prosodic representations. In Anne Cutler and D. R. Ladd (Eds.), *Prosody: Models and measurements*, 93—109. Berlin: Springer Verlag.

HOEQVIST, JR., CHARLES. 1983a. Durational correlates of linguistic rhythm categories. *Phonetica*, **40**, 19—31.

HOEQVIST, JR., CHARLES. 1983b. Syllable duration in stress-, syllable- and mora-timed languages. *Phonetica*, **40**, 203—237.

HOEQVIST, JR., CHARLES. 1983c. The perceptual center and rhythm categories. *Language and Speech*, **26**, 367—376.

HUGGINS, A. W. F. 1972a. Just noticeable differences for segment duration in natural speech. *Journal of the Acoustical Society of America*, **51**, 1270—1278.

HUGGINS, A. W. F. 1972b. On the perception of temporal phenomena in speech. *Journal of the Acoustical Society of America*, **51**, 1279—1290.

JAMIESON, DONALD G. 1977. Two presentation order effects. *Canadian Journal of Psychology*, **31**, 184—194.

JAMIESON, DONALD G. & WILLIAM M. PETRUSIC. 1975. Presentation order effects in duration discrimination. *Perception & Psychophysics*, **17**, 197—202.

JASSEM, WIKTOR, D. R. HILL & IAN H. WITTEN. 1984. Isochrony in English speech: Its statistical validity and linguistic relevance. In Dafydd Gibbon and Helmut Richter (Eds.), *Intonation, accent and rhythm*, 203—225. Berlin: Walter de Gruyter.

JONES, DANIEL. 1918, (1962 ninth ed.). *An Outline of English Phonetics*. Cambridge: W. Heffer & Sons.

JONES, MARI RIESS. 1978. Auditory patterns: Studies in the perception of structure. In Edward C. Carterette and Morton P. Friedman (Eds.), *Handbook of Perception, Vol. VIII*. New York: Academic Press.

JONES, MARI RIESS, GARY KIDD & ROBIN WETZEL. 1981. Evidence for rhythmic attention. *Journal of Experimental Psychology: Human Perception and Performance*, **7**, 1059—1073.

KILLEEN, PETER R. & NEIL A. WEISS. 1987. Optimal timing and the Weber function. *Psychological Review*, **94**, 455—468.

KLATT, DENNIS H. 1975. Vowel lengthening is syntactically determined in connected discourse. *Journal of Phonetics*, **3**, 129—140.

KLATT, DENNIS H. 1976. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, **59**, 1208—1221.

KLATT, DENNIS H. 1979. Synthesis by rule of segmental durations in English sentences. In Björn Lindblom and Sven Öhman (Eds.), *Frontiers in Speech Communication Research*, 287—299. London: Academic Press.

KLATT, DENNIS H. & WILLIAM E. COOPER. 1975. Perception of segment duration in sentence contexts. In Antonie Cohen and Sieb G. Nooteboom (Eds.), *Structure and Process in Speech Perception*, 69—89.

KING, H. E. 1962. Anticipatory behaviour: Temporal matching by normal and psychotic subjects. *The Journal of Psychology*, **53**, 425—440.

KORSAN-BENGTSEN, MARGARETA. 1973. Distorted speech audiometry. *Acta Oto-Laryngologica*, Supplement 310.

KRISTOFFERSON, ALFRED B. 1973. Psychological timing mechanisms. A paper presented at the Fourth Annual Meeting of the Lake Ontario Vision Establishment, Niagara Falls, Ontario.

KRISTOFFERSON, ALFRED B. 1977. A real time criterion theory of duration discrimination. *Perception & Psychophysics*, **21**, 105—117.

LADEFOGED, PETER. 1967. *Three Areas of Experimental Phonetics*. London: Oxford University Press.

LADEFOGED, PETER. 1975. *A Course in Phonetics*. New York: Harcourt Brace Jovanovich.

LEHISTE, ILSE. 1971. Temporal organization of spoken language. In L. L. Hammerich, Roman Jacobson & Eberhart Zwirner, *Form & Substance*, 159—169. Copenhagen: Akademisk Forlag.

LEHISTE, ILSE. 1973. Rhythmic units and syntactic units in production and perception. *Journal of the Acoustical Society of America*, **54**, 1228—1234.

LEHISTE, ILSE. 1976. Influence of fundamental frequency pattern on the perception of duration. *Journal of Phonetics*, **4**, 113—117.

LEHISTE, ILSE. 1977. Isochrony reconsidered. *Journal of Phonetics*, **5**, 253—263.

LEHISTE, ILSE. 1979a. The perception of duration within sequences of four intervals. *Journal of Phonetics*, **7**, 313—316.

LEHISTE, ILSE. 1979b. Perception of sentence and paragraph boundaries. In Björn Lindblom and Sven Öhman (Eds.), *Frontiers in Speech Communication Research*, 287—299. London: Academic Press.

LEHISTE, ILSE. 1980. Interaction between test word duration and length of utterance. In Linda R. Waugh & C. H. van Schoonevald, *The Melody of Language*, 169—176. Baltimore: University Park Press.

LEHISTE, ILSE. 1985. Rhythm of poetry, rhythm of prose. In Victoria A. Fromkin (Ed.), *Phonetic linguistics*, 145—155. Orlando: Academic Press.

LEHISTE, ILSE. 1990. Some aspects of the phonetics of metrics. In Kalevi Wiik and Ilkka Raimo (Eds.), *Nordic Prosody V, Papers from a symposium*, 206—218. Turku: University of Turku, Phonetics.

LEHISTE, ILSE, JOSEPH P. OLIVE & LYNN A. STREETER. 1976. Role of duration in disambiguating syntactically ambiguous sentences. *Journal of the Acoustical Society of America*, **60**, 1199—1202.

LENNEBERG, ERIC H. 1967. *Biological Foundations of Language*. New York: John Wiley & Sons, Inc.

LERDAHL, FRED & RAY JACKENDOFF. 1981. On the theory of grouping and meter. *Musical Quarterly*, Vol LXVII, No. 4, 479—506.

LIEBERMAN, MARK & ALAN PRINCE. 1977. On stress and linguistic rhythm. *Linguistic Inquiry*, **8**, 249—336.

LIEBERMAN, MARK & LYNN A. STREETER. 1978. Use of nonsense syllable mimicry in the study of prosodic phenomena. *Journal of the Acoustical Society of America*, **63**, 231—233.

LINDBLOM, BJÖRN & KARIN RAPP. 1973. Some temporal regularities of spoken Swedish. *Papers from the Institute of Linguistics University of Stockholm*, **21**, 1—59. Stockholm: University of Stockholm.

LINDBLOM, BJÖRN, BERTIL LYBERG and KARIN HOLMGREN. 1981. *Durational Patterns of Swedish Phonology: Do they reflect short-term motor memory processes?* Bloomington, Indiana: Indiana University Linguistics Club.

LLOYD JAMES, ARTHUR. 1940. *Speech Signals in Telephony*. London:

LONGUET-HIGGINS, H. C. 1976. Perception of melodies. *Nature*, **263**, 646—653.

LOOTS, MARIJKE E. 1979. *Metrical Myths: An Experimental-Phonetic Investigation into the Production and Perception of Metrical Speech*. Utrecht: Drukkerij Elinkwijk BV.

LUCE, R. DUNCAN & EUGENE GALANTER. 1963. Discrimination. In R. Duncan Luce, Robert R. Bush and Eugene Galanter (Eds.), *Handbook of Mathematical Psychology*, Vol. 1, 191—243. New York: John Wiley and Sons, Inc.

LUNNEY, H. W. M. 1974. Time as heard in speech and music. *Nature*, **249**, 592.

LYBERG, BERTIL. 1977. Some observations on the timing of Swedish utterances. *Journal of Phonetics*, **5**, 49—59.

LYBERG, BERTIL. 1979. Final lengthening—partly a consequence of restrictions on the speed of fundamental frequency change? *Journal of Phonetics*, **7**, 187—196.

LYBERG, BERTIL. 1981a. Some consequences of a model for segment duration based on F_0-dependence. *Journal of Phonetics*, **9**, 97—103.

LYBERG, BERTIL. 1981b. Some observations on the vowel duration and the fundamental frequency contour in Swedish utterances. *Journal of Phonetics*, **9**, 261—272.

LYBERG, BERTIL. 1984. Some fundamental frequency perturbations in a sentence context. *Journal of Phonetics*, **12**, 307—317.

MAJOR, ROY C. 1981. Stress-timing in Brazilian Portuguese. *Journal of Phonetics*, **9**, 343—351.

MANRIQUE, ANA MARIA BORZONE DE & ANGELA SIGNORINI. 1983. Segmental duration and rhythm in Spanish. *Journal of Phonetics*, **11**, 117—128.

MARCUS, STEPHEN M. 1979. Perceptual centers. (P-centers). Paper presented at the ninth international congress of phonetic sciences, Copenhagen 1979. Abstract of paper in the Proceedings, Vol I, 238—239. Copenhagen: Institute of Phonetics, University of Copenhagen.

MARCUS, STEPHEN M. 1981. Acoustic determinants of perceptual center (P-center) location. *Perception and Psychophysics*, **30**, 247—256.

MARTIN, JAMES G. 1972. Rhythmic (hierarchical) versus serial structure in speech and other behaviour. *Psychological Review*, **79**, 487—509.

MARTIN, JAMES G. 1979. Rhythmic and segmental perception are not independent. *Journal of the Acoustical Society of America*, **65**, 1286—1297.

MICHON, JOHN A. 1964. Studies on subjective duration. *Acta Psychologica*, **22**, 441—450.

MICHON, JOHN A. 1967a. *Timing in Temporal Tracking*. Soesterberg: Institute for perception.

MICHON, JOHN A. 1967b. Magnitude scaling of short durations with closely spaced stimuli. *Psychonomic Science*, **9**, 359—360.

MILLER, JOANNE L. & FRANÇOIS GROSJEAN. 1981. How the components of speaking rate influence perception of phonetic segments. *Journal of Experimental Psychology: Human Perception and Performance*, **7**, 208—215.

MILLER, G. A. 1947. Sensitivity to changes in the intensity of white noise and its relation to masking and loudness. *Journal of the Acoustical Society of America*, **19**, 609—619.

MILLER, M. 1984. On the perception of rhythm. *Journal of Phonetics*, **12**, 75—83.

MISHIMA, JIRO. 1951–1952. Fundamental research on the constancy of "mental tempo". *Japanese Journal of Psychology*, **22**, 27—28.

MISHIMA, JIRO. 1956. On the factors of the mental tempo. *Japanese Psychological Research*, **4**, 27—38.

MISHIMA, JIRO. 1965. *Introduction to the Morphology of Human Behaviour. The Experimental Study of Mental Tempo*. Tokyo: Tokyo, Publishing.

MIYAKE, ISHIRO. 1902. Researches on rhythmic action. In E. W. Scripture (Ed.), *Studies from the Yale Psychological Laboratory*, **10**, 1—48.

MONAHAN, CAROLINE B. & IRA J. HIRSH. 1990. Studies in auditory timing: 2. Rhythm patterns. *Perception & Psychophysics*, **47**, 227—242.

MOORE, BRIAN C. J., BRIAN R. GLASBERG, C. J. PLACK & A. K. BISWAS. 1988. The shape of the ear's temporal window. *Journal of the Acoustical Society of America*, **83**, 1102—1116.

MORTON, JOHN, STEVE MARCUS & CLIVE FRANKISH. 1976. Perceptual centers (P-centers). *Psychological Review*, **83**, 405—408.

NAKATANI, LLOYD H., KATHLEEN D. O'CONNOR & CARLETTA H. ASTON. 1981. Prosodic aspects of American English speech rhythm. *Phonetica*, **38**, 84—106.

NAVARRO TOMÁS, TOMÁS. 1916. Cantidad de las vocales acentuadas. *Revista de Filología Española*, **III**, 387—408.

NAVARRO TOMÁS, TOMÁS. 1917. Cantidad de las vocales inacentuadas. *Revista de Filología Española*, **IV**, 371—388.

NAVARRO TOMÁS, TOMÁS. 1918. Diferencias de duratión entre las consonantes españolas. *Revista de Filología Española*, **V**, 367—393.

NAVARRO TOMÁS, TOMÁS. 1922. La cantidad silábica en unos versos de Rubén Darío. *Revista de Filología Española*, **IX**, 1—29.

NISHINUMA, YUKIHIRO. 1984. Prediction of phoneme duration by a distinctive feature matrix. *Journal of Phonetics*, **12**, 169—173.

NOOTEBOOM, SIEB G. 1973. The perceptual reality of some prosodic durations. *Journal of Phonetics*, **1**, 25—45.

O'CONNOR, J. D. 1973. *Phonetics*. Harmondsworth, Middlesex: Penguin.

OHALA, JOHN J. 1975. The temporal regulation of speech. In Gunnar Fant and M. A. A. Tatham (Eds.), *Auditory Analysis and Perception of Speech*, 431—453. London: Academic Press.

OHALA, JOHN J. & BERTIL LYBERG. 1976. Comments on "Temporal interactions within a phrase and sentence context". *Journal of the Acoustical Society of America*, **59**, 990—992.

OLÉRON, GENEVIEVE. 1952. Influence de l'intensité d'un son sur l'estimation de la durée apparente. *Année Psychologique*, **52**, 383—392.

OLIVER, DOUGLAS. 1984. Voicing patterns as one key to the pace of poetry. *Journal of Phonetics*, **12**, 115—132.

OLLER, KIMBROUGH D. 1973. The effect of position in utterance on speech segment duration in English. *Journal of the Acoustical Society of America*, **54**, 1235—1247.

OLLER, KIMBROUGH D. 1979. Syllable timing i Spanish, English and Finnish. *Current Issues in the Phonetic Sciences, Proceedings of the IPS–77 Congress, Florida*, 331—343. Amsterdam: John Benjamins B.V.

OLSEN, C. L. 1972. Rhythmical patterns and syllabic features of the Spanish sense-group. *Proceedings of the Seventh International Congress of Phonetic Sciences, Montreal 1971*. Mouton.

PIKE, KENNETH L. 1945. *The Intonation of American English.* Ann Arbor: University of Michigan Press.

PLATO. *The Laws.* Tranlslated by Trevor J. Saunders. London: Penguin.

POINTON, GRAHAM E. 1980. Is Spanish really syllable timed? *Journal of Phonetics*, **8**, 293—304.

POMPINO-MARSHALL, BERND, HANS-GEORG PIROTH, KLAUS TILK, PHILIP HOOLE & HANS G. TILLMANN. 1982. Does the closed syllable determine the perception of 'momentary tempo'? *Phonetica*, **39**, 358—367.

PORT, ROBERT F., JONATHAN DALBY & MICHAEL O'DELL. 1987. Evidence for mora timing in Japanese. *Journal of the Acoustical Society of America*, **81**, 1574—1585.

POVEL, DIRK-JAN. 1977. Temporal structure of performed music: Some preliminary observations. *Acta Psychologica*, **41**, 309—320.

POVEL, DIRK-JAN. 1981. Internal representations of simple temporal patterns. *Journal of Experimental Psychology: Human Perception and Performance*, **7**, 3—18.

RAPP, KARIN. 1971. A study of syllable timing. *Papers from the Institute of Linguistics University of Stockholm*, **8**, 14—19. Stockholm: University of Stockholm.

RASCH, RUDOLF. 1978. The perception of simultaneous notes such as in polyphonic music. *Acustica*, **40**, 22—33.

RASCH, RUDOLF. 1979. Synchronization in performed ensemble music. *Acustica*, **42**, 121—131.

RASCH, RUDOLF. 1981. *Aspects of the Perception and Performance of Polyphonic Music.* Utrecht: Drukkerij Elinkwijk BV.

RIETVELD, A. C. M. & C. GUSSENHOVEN. 1987. Perceived speech rate and intonation. *Journal of Phonetics*, **15**, 273—285.

ROACH, PETER. 1982. On the distinction between "stress-timed" and "syllable-timed" languages. In David Crystal (Ed.), *Linguistic Controversies*, 73—79.

ROSS, ROBERT T. 1934. Optimum orders for the presentation of pairs in the method of paired comparisons. *Journal of Educational Psychology*, **25**, 375—382.

ROUSSEAU, ROBERT & ALFRED B. KRISTOFFERSON. 1973. The discrimination of bimodal temporal gaps. *Bulletin of the Psychonomic Society*, **1**, 115—116.

RUHM, H. B., E. O. MENCKE, B. MILBURN, W. A. COOPER, JR & D. E. ROSE. 1966. Differential sensitivity to duration of acoustic signals. *Journal of Speech and Hearing Research*, **9**, 371—384.

SCOTT, DONIA R. 1982. Duration as a cue to the perception of a phrase boundary. *Journal of the Acoustical Society of America*, **71**, 996—1007.

SCOTT, DONIA R., S. D. ISARD & BÉNÉDICTE DE BOYSSON-BARDIES. 1985. Perceptual isochrony in English and in French. *Journal of Phonetics*, **13**, 155—162.

SCOTT, DONIA R., S. D. ISARD & BÉNÉDICTE DE BOYSSON-BARDIES. 1986. On the measurement of rhythmic irregularity: a reply to Benguerel. *Journal of Phonetics*, **14**, 327—330.

SELKIRK, ELISABETH O. 1980. On Prosodic structure and its relation to syntactic structure. In Thorstein Fretheim (Ed.), *Nordic Prosody II: Papers from a symposium*, 111—140. Trondheim: Tapir.

SHAFFER, L. HENRY. 1982. Rhythm and timing in skill. *Psychological Review*, **89**, 109—122.

SHAFFER, L. HENRY, ERIC F. CLARKE & NEIL P. TODD. 1985. Metre and rhythm in piano playing. *Cognition*, **20**, 61—77.

SHEN, YAO & GILES G. PETERSON. 1962. Isochronism in English. *University of Buffalo Studies in Linguistics, Occasional Papers*, **9**, 1—36.

SHIFFMAN, H. R. & DOUGLAS J. BOBKO. 1974. Effects of stimulus complexity on the perception of brief temporal intervals. *Journal of Experimental Psychology*, **103**, 156—159.

SHIFFMAN, H. R. & DOUGLAS J. BOBKO. 1977. The role of number and familiarity of stimuli in the perception of brief temporal intervals. *American Journal of Psychology*, **90**, 85—93.

SIEGEL, SIDNEY & N. JOHN CASTELLAN, JR. 1988. *Nonparametric Statistics for the Behavioural Sciences*. Second edition. New York: McGraw-Hill.

SIGURD, BENGT. 1973. Maximum rate and minimal duration of repeated syllables. *Language and Speech*, **16**, 373—395.

SMALL, ARNOLD M. JR. & RICHARD M. CAMPBELL. 1962. Temporal differential sensitivity for auditory stimuli. *American Journal of Psychology*, **75**, 401—410.

SMITH, BRUCE L. 1978. Temporal aspects of English speech production: A developmental perspective. *Journal of Phonetics*, **6**, 37—67.

STEINER, S. 1968. Apparent duration of auditory stimuli. *Journal of Auditory Research*, **8**, 195—205.

STETSON, RAYMOND H. 1951. *Motor Phonetics.* Amsterdam: North Holland Publishing Company. (First edition published in 1928 by Archives Néerlandaises de Phonétique Experimentale)

STEVENS, S. S. 1957. On the psychophysical law. *Psychological Review*, **64**, 153—181.

STEVENS, S. S. & EUGENE H. GALANTER. 1957. Ratio scales and category scales for a dozen perceptual continua. *Journal of Experimental Psychology*, **54**, 377—411.

STONE, MAUREEN. 1981. Evidence for a rhythm pattern in speech production: Observations of jaw movement. *Journal of Phonetics*, **9**, 109—120.

STOTT, LELAND H. 1935. Time-order errors in the discrimination of short tonal durations. *Journal of Experimental Psychology*, **18**, 741—766.

STRANGERT, EVA. 1981. *Rhythmic patterns of Swedish in a cross-linguistic perspective.* Department of Phonetics, Publication 19, 1—29. Umeå: Umeå University.

STRANGERT, EVA. 1984. Temporal characteristics of rhythmic units in Swedish. In Claes-Christian Elert, Irene Johansson and Eva Strangert (Eds.), *Nordic Prosody III*, 201—213. Stockholm: Almqvist & Wiksell International.

STRANGERT, EVA 1985. *Swedish Speech Rhythm in a Cross-Language Perspective.* Umeå: Umeå Studies in the Humanities 69.

STRANGERT, EVA 1988. Syllable durations obtained from the KTH speech data base. *Speech Transmission Laboratory Quarterly Progress and Status Report,* **4/1988**, 51—57. Stockholm: Royal Institute of Technology (KTH).

SUMMERFIELD, QUENTIN. 1981. Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance*, **7**, 1074—95.

SUNDBERG, JOHAN. 1989. Synthesis of singing by rule. In Max V. Mathews and J. R. Pierce (Eds.), *Current Directions in Computer Music Research*, 45—55 and 401. Cambridge, Mass.: MIT Press.

THOMAS, EWART A. C. & IRVIN BROWN, JR. 1974. Time perception and the filled-duration illusion. *Perception & Psychophysics*, **16**, 449—458.

THOMAS, EWART A. C. & WANDA B. WEAVER. 1975. Cognitive processing and time perception. *Perception & Psychophysics*, **17**, 363—367.

THOMPSON, JACK G., H. RICHARD SHIFFMAN & DOUGLAS J. BOBKO. 1976. The discrimination of brief temporal intervals. *Acta Psychologica*, **40**, 489—493.

THURSTONE, LOUIS LEON. 1927. A law of comparative judgment. *Psychological Review*, **34**, 273—286.

TORGERSON, WARREN S. 1958. *Theory and Methods of Scaling*. New York: John Wiley & Sons.

TREISMAN, MICHEL. 1963. Temporal discrimination and the indifference interval: Implications for a model of the "internal clock". *Psychological Monographs*, **77.**

TULLER, BETTY & CAROL A. FOWLER. 1980. Some articulatory correlates of perceptual isochrony. *Perception & Psychophysics*, **27**, 277—283.

TULLER, BETTY & CAROL A. FOWLER. 1981. The contribution of amplitude to the perception of isochrony. *Haskins Laboratories Status Reports on Speech Research, SR-65*, 245—250.

TULLER, BETTY, KATHERINE S. HARRIS & J. A. SCOTT KELSO. 1981. Articulatory motor events as a function of rate and stress. *Haskins Laboratories Status Reports on Speech Research, SR-65*, 33—62.

TULLER, BETTY, J. A. SCOTT KELSO & KATHERINE S. HARRIS. 1982. Interarticulator phasing as an index of temporal regularity in speech. *Journal of Experimental Psychology: Human Perception and Performance*, **8**, 460—472.

ULDALL, ELISABETH T. 1971. Isochronous stresses in R.P. In L. L. Hammerich, Roman Jacobson and Eberhart Zwirner (Eds.), *Form & Substance*, 205—210. Copenhagen: Akademisk Forlag.

UMEDA, NORIKO. 1975. Vowel duration in American English. *Journal of the Acoustical Society of America*, **58**, 434—445.

UMEDA, NORIKO. 1977. Consonant duration in American English. *Journal of the Acoustical Society of America*, **61**, 846—858.

UMEDA, NORIKO & ANN MARIE S. QUINN. 1981. Word duration as an acoustic measure of boundary perception. *Journal of Phonetics*, **9**, 19—28.

VAANE, EVELINE. 1982. Subjective estimation of speech rate. *Phonetica*, **39**, 136—149.

VENTSOV, ANATOLY V. 1983. What is the reference that sound durations are compared with in speech perception? *Phonetica*, **40**, 135—144.

VOS, JOOS & RUDOLF RASCH. 1982. The perceptual onset of musical tones. In Manfred Clynes (Ed.), *Music, Mind, and Brain*, 299—319. New York: Plenum Press.

VOILLAUME, CLAUDE. 1971. Modèles pour l'étude de la régulation des mouvements cadencés. *Année Psychologique*, **71**, 347—358.

WARM, JOEL S., LEWIS F. GREENBERG & C. STUART DUBE. 1964. Stimulus and motivational determinants in temporal perception. *The Journal of Psychology*, **58**, 243—248.

WARM, JOEL S. & RONALD E. MCCRAY. 1969. Influence of word frequency and length on the apparent duration of tachistoscopic presentations. *Journal of Experimental Psychology*, **79**, 56—58.

WARNER, REBECCA M. 1979. Periodic rhythms in conversational speech. *Language and Speech*, **22**, 381—396.

WENK, BRIAN J. & FRANÇOIS WIOLAND. 1982. Is French really syllable timed? *Journal of Phonetics*, **10**, 193—216.

WHERRY, ROBERT J. 1938. Orders for the presentation of pairs in the method of paired comparisons. *Journal of Experimental Psychology*, **23**, 651—660.

WITTEN, IAN H. 1977. A flexible scheme for assigning timing and pitch to synthetic speech. *Language and Speech*, **20**, 240—260.

WOODROW, HERBERT. 1911. The rôle of pitch in rhythm. *Psychological Review*, **18**, 54—77.

WOODROW, HERBERT. 1930. The reproduction of temporal intervals. *Journal of Experimental Psychology*, **13**, 473—499.

WOODROW, HERBERT. 1935. The effect of practice upon time- order errors in the comparison of temporal intervals. *Psychological Review*, **42**, 127—152.

WOODROW, HERBERT. 1951. Time perception. In S. S. Stevens (Ed.), *Handbook of Experimental Psychology*, 1224—1236, New York: Wiley.

YILMAZ, HÜSEYIN. 1967. Perceptual invariance and the psychophysical law. *Perception & Psychophysics*, **2**, 533—538.

ZELKIND, IRVING. 1973. Factors in time estimation and a case for the internal clock. *Journal of General Psychology*, **88**, 295—301.

GOTHENBURG MONOGRAPHS IN LINGUISTICS ISSN 0346-6248
University of Gothenburg
Department of Linguistics
S-412 98 Gothenburg, Sweden

1. Lars-Gunnar Andersson: *Form and Function of Subordinate Clauses.* 1975.

2. Jens Allwood: *Linguistic Communication as Action and Cooperation.* 1976.

3. Pierre G. Javanaud: *The Vowel System of Lemosin: A Phonological Study.* 1981.

4. Sven Strömqvist: *Make-Believe Through Words: A Linguistic Study of Children's Play with a Dolls' House.* 1984.

5. Elisabeth Ahlsén: *Discourse Patterns in Aphasia.* 1985.

6. Sally Boyd: *Language Survival: A Study of Language Contact, Language Shift, and Language Choice in Sweden.* 1985.

7. Richard Hirsch: *Argumentation, Information, and Interaction.* 1989.

8. Sören Sjöström: *Spatial Relations: Towards a Theory of Spatial Verbs, Prepositions, and Pronominal Adverbs in Swedish.* 1990.

9. Anders Eriksson: *Aspects of Swedish Speech Rhythm.* 1991.

Printed in Sweden
Kompendietryckeriet – Kållered
1992