

Diss. Utrecht 1972, 879

**PRODUCTION AND PERCEPTION  
OF VOWEL DURATION**

**A STUDY OF DURATIONAL PROPERTIES OF VOWELS IN DUTCH**

**S. G. NOOTEBOOM**

at







# PRODUCTION AND PERCEPTION OF VOWEL DURATION

A STUDY OF DURATIONAL PROPERTIES OF VOWELS IN DUTCH

## PROEFSCHRIFT

TER VERKRIJGING VAN DE GRAAD VAN DOCTOR  
IN DE LETTEREN AAN DE RIJKSUNIVERSITEIT TE  
UTRECHT, OP GEZAG VAN DE RECTOR  
MAGNIFICUS PROF. DR. SJ. GROENMAN, VOLGENS  
BESLUIT VAN HET COLLEGE VAN DEKANEN IN  
HET OPENBAAR TE VERDEDIGEN OP VRIJDAG  
20 OKTOBER 1972 DES NAMIDDAGS TE 3.15 UUR

DOOR

SIBOUT GOVERT NOOTEBOOM

GEBOREN OP 19 APRIL 1939 TE MAKASSAR

BIBLIOTHEEK DER  
RIJKSUNIVERSITEIT  
UTRECHT

PROMOTOR: PROF. DR. A. COHEN

RIJKSUNIVERSITEIT UTRECHT



0494 7663

## CONTENTS

1. INTRODUCTION . . . . .	1
1.1. Goals . . . . .	1
1.2. Some preliminary considerations . . . . .	2
2. ARTICULATORY MEASUREMENTS . . . . .	9
2.1. Method . . . . .	9
2.2. Measurements . . . . .	12
2.2.1. The role of vowel quantity . . . . .	13
2.2.2. The effect of stress and position . . . . .	16
2.2.3. The effect of number of syllables . . . . .	17
2.3. Accuracy in articulatory timing and reproduceability of an articulatory programme . . . . .	18
2.3.1. Accuracy in the timing of lip movements . . . . .	19
2.3.2. Accuracy in the timing of voice onset and end . . . . .	20
2.3.3. Constancy in the articulatory programme . . . . .	21
2.4. The role of vowel quantity in articulatory timing . . . . .	23
2.4.1. The vowels of Dutch . . . . .	25
2.4.2. Studying realisations of vowel quantity in nonsense words . . . . .	27
2.4.2.1. Vowel quantity and actual durations . . . . .	28
2.4.2.2. Vowel quantity as reflected in the total word durations . . . . .	29
2.4.2.3. Explaining the durational differences between vowels from the same quantity category . . . . .	35
2.4.2.4. The durational behaviour of the schwah . . . . .	38
2.4.2.5. Quantity and phenomena of contact . . . . .	39
2.4.2.6. Results for the [tVtVtVt] words . . . . .	42
2.4.2.7. On [u, y, i] also before [r] . . . . .	45
2.4.2.8. On the diphthongs [ɛi, ^y, au] . . . . .	47
2.4.3. Summary and conclusion . . . . .	48
2.5. The effect of stress and position and of number of syllables in the word on the durations of long and short vowels . . . . .	50
2.5.1. The different behaviour of long and short vowels . . . . .	50
2.5.2. Dominant versus non-dominant words . . . . .	51
2.5.2.1. Some acoustical differences between dominant and non-dominant words . . . . .	52
2.5.3. The durational build-up of the three-syllable dominant words . . . . .	55
2.5.3.1. The durational behaviour of the stressed vowel in the dominant word . . . . .	57

2.5.3.2.	The durational behaviour of the unstressed vowels in the dominant word . . . . .	58
2.5.3.3.	The consonant durations in the dominant word . . . . .	59
2.5.4.	The durational build-up of the three-syllable non-dominant word . . . . .	60
2.5.4.1.	The vowels in the dominant word . . . . .	60
2.5.4.2.	The consonants in the non-dominant word . . . . .	62
2.5.5.	The effect of the number of syllables in the word . . . . .	62
2.5.5.1.	The effect of number of syllables in words spoken in isolation on the durations of stressed vowels . . . . .	64
2.5.5.2.	The effect of number of syllables in words spoken in isolation on the durations of unstressed vowels . . . . .	68
2.5.5.3.	The effect of number of syllables in words spoken in isolation on the consonant durations . . . . .	68
2.5.5.4.	The effect of number of syllables in embedded non-dominant words . . . . .	68
2.5.6.	Towards a quantitative description . . . . .	71
2.6.	Summary . . . . .	75
3.	PERCEPTUAL TESTS OF DURATIONAL REGULARITIES . . . . .	76
3.1.	A method for testing the perceptual reality of durational regularities . . . . .	76
3.2.	The experimental set-up . . . . .	80
3.3.	Accuracy in perceptual timing . . . . .	82
3.4.	Testing for the perceptual reality of physiologically conditioned effects . . . . .	85
3.5.	Testing rules for vowel durations . . . . .	89
3.5.1.	Testing a durational rule for stressed vowels . . . . .	89
3.5.1.1.	The effect of the number of syllables following in the word on the preferred duration of a stressed vowel . . . . .	91
3.5.1.2.	The effect of position on the preferred duration of a stressed vowel in a three-syllable word . . . . .	94
3.5.1.3.	The effect of the total number of syllables on the preferred duration of a vowel in a stressed final syllable . . . . .	95
3.5.2.	Testing for the perceptual reality of some durational rules for unstressed vowels . . . . .	97
3.5.2.1.	Testing rules (2), (4) and (5) . . . . .	97
3.5.2.2.	Testing rules (3) and (4) . . . . .	100
3.6.	Summary . . . . .	101



4. SOME INTERPRETATIONS AND SPECULATIONS . . . . .	103
4.1. The place of vowel-duration rules in linguistic theory: the phonetic component . . . . .	103
4.2. Accuracy in production and perception of vowel durations and the coding of timing information in the brain . . . . .	108
4.3. On the perceptual knowledge of physiologically conditioned effects . . . . .	111
4.4. On the comparability of vowel quantity in different languages . . .	112
4.4.1. The specification of Dutch vowel quantity on the level of phonetic representation . . . . .	112
4.4.2. The quantitative relation between long and short vowels . . . .	113
4.5. On the durational rules for vowels . . . . .	115
4.5.1. The generality of the rules . . . . .	115
4.5.2. The boundary between Dutch long and short vowels . . . .	117
4.5.3. Other determinants of vowel duration . . . . .	117
4.5.4. On the universality of the rules proposed . . . . .	118
4.6. On pitch accent, stress and duration . . . . .	119
4.7. Other applications of the method of matching to internal criterion	121
4.8. Summary . . . . .	122
 APPENDIX A: Tables 1–20 with data from the articulatory measure- ments . . . . .	125
APPENDIX B: Tables 21–30 with data from adjustment tests and spec- trographic measurements . . . . .	147
 References . . . . .	153
Author index . . . . .	158
 List of phonetic symbols with key words . . . . .	161
 Summary . . . . .	162
Samenvatting . . . . .	164

## Acknowledgement

Although the actual wording of this study and the responsibility for all its defects are my own, if there is anything of value in it this is the outcome of close collaboration of many people. Professor Anthony Cohen stimulated my interest in temporal phenomena in speech, helped me through many difficulties and depressions in the course of the investigations and guided me in the writing. I profited greatly from his insight and keen interest in my work. The experiments reported upon were done as much by Iman Slis as by myself. Without his experience in speech research and tireless curiosity this work would not have been possible. In our countless discussions his unorthodox way of thinking often gained the victory over my own laziness of thought. Many improvements resulted from his reading of the manuscript. Most of the technical devices used in our experiments were designed by Lei Willems. His thorough knowledge of experimental phonetics and anticipatory understanding of our problems made that he often knew better what we wanted than we did ourselves. The technical assistance of Gerard Moonen, Chris Lammers, Henny Melotte, Henk de Vries and André van Nes is also gratefully acknowledged. Many thanks are due to Herman Muller who wrote the necessary computer programmes.

From what I said it may be apparent that this study is the product of team work. The team was formed by many people of the Institute for Perception Research. Only those most closely involved have been mentioned. I thank also all the others who helped me with advice and friendship and who created such a stimulating environment. In particular I thank the director, Prof. Dr. J. F. Schouten, who gave me plenty of time, firstly to adapt my untrained mind to the unfamiliar world of speech research, secondly to follow the capricious ways of my own curiosity and thirdly to write this study. The long time it took me may be a measure of his patience.

I am very grateful to Prof. Dr. H. Schultink of the university of Utrecht, who both in the initial stage and in the final stage of my writing closely and critically read the manuscript and suggested many improvements.

Finally I thank Ragnhild, my wife, for her help with the bibliographical work and for often forcing me to stop work during the weekends.

## 1. INTRODUCTION

### 1.1. Goals

In this study we will be concerned with some aspects of the control of segment duration, and particularly vowel duration, in Dutch. Variations in segment duration are an important cause of acoustic variability in the realisation of linguistically identical units. It has for instance been shown by a number of authors that the spectral composition of a vowel may vary according to changes in its duration, other things being equal (see Stevens and House <sup>1</sup>), Stevens, House and Paul <sup>2</sup>), Lindblom <sup>3</sup>), Peterson and Lehiste <sup>4</sup>)). Furthermore systematic differences in vowel durations are found to be important cues for perception in a number of different ways. They may signal phonemic quantity, stress, the voiceless or voiced character of the following consonant, the presence of a major syntactic break immediately following the syllable, the presence of a word boundary. In view of this it would seem to be of some importance to formulate rules from which systematic variations in vowel durations can be predicted. Such rules may help in describing the acoustic parameters relevant to the perception and decoding of speech. They may also be of value in generating high-quality synthetic speech. The search for regularities and the formulation of rules for vowel duration are goals of this study.

By studying the systematic variations in vowel durations it may be possible to reveal some aspects of the organisation of the mental structures of language, particularly aspects which are not easily studied by the conventional methods of linguistics. This study aims to discover some properties of the mental structures of language. It is hoped that the results will help to clarify the relation between the discrete linguistic specification of speech and a phonetic specification of the more continuous properties of the speech event although we know that the contribution made by this study may be only a minor one.

When we started this study we knew very little concerning the systematic variations of vowel durations in Dutch. This made the present study rather exploratory in character. The measurements we did often led to findings and questions not directly related to the main theme of the study. We have not refrained from mentioning those when they came up in the course of our writing. It is hoped that at least in some cases this may lead to interesting experiments in the future.

In the next section of this introduction we will give some preliminary considerations. In chapter 2 we will describe and discuss a number of articulatory measurements on vowel durations. In chapter 3 we will describe and discuss a number of perceptual tests, set up for testing the perceptual relevance of a few durational rules derived from the articulatory measurements in chapter 2. In chapter 4 we will put forward some interpretations and speculations concerning the results obtained.

## 1.2. Some preliminary considerations

The primary data of a linguist are mostly not the outcome of physical measurements. In order to describe the syntactic structure or the phonological structure of sentences a linguist has no need for physical measurements. In many cases he describes structures of his mother tongue and for doing so he reflects on language material which is generated by his own knowledge of the language. He observes his data and describes the linguistic regularities in a generalised form with the help of rules the form of which is provided by a general theory. From these rules he may predict new structures and test whether these predicted structures are indeed not in conflict with his knowledge of the language. All this goes on within his own mind. He evokes his primary data from his own mind and construes the descriptive and explanatory models within his own mind. He calculates the predictions from these models and this calculation is part of his mental activity just as the test procedure which essentially is a check whether or not the structure generated by his formal rules could also follow from his knowledge of the language.

In this way a linguist may find the rules which describe many of the structures underlying the form of sentences in his language. We may hypothesise that the structures obtained in this way in some non-trivial way correspond to mental structures underlying the verbal behaviour of the speakers of the language. This does not imply that a psychological reality has to be assigned to the rules themselves.

We assume then that a linguist in some way describes aspects of the implicit knowledge a language user has about his language. Not all aspects of this knowledge, however, are accessible to the linguistic method. There are many language-universal and language-particular regularities in the verbal behaviour of human speakers and listeners that reflect underlying mental structures which escape the attention of the linguist searching his own mind. Numerous examples of such regularities can be found in the domain of phonetics. Phonetics particularly studies those properties of language that determine how words and phrases actually sound. Such properties, in the domain of interaction between adjacent speech sounds, in the domain of intonation or in the domain of temporal patterning, if they are not personal, are either universal or language-particular. In both cases their description is of importance. Not only theoretically, in that it may help us to understand fundamental properties of human speech and to explain the "naturalness" of many other properties of language, but also practically: rules for the interaction between adjacent speech sounds, for intonation and for temporal patterning are indispensable for generating high-quality synthesised speech. They also may be of great value in second-language teaching.

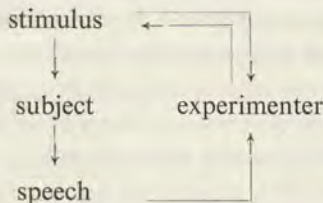
It is in phonetic research that physical measurements can contribute not only to gaining insight into the physiological and acoustic processes of speech but

also into mental structures underlying the control of speech, and in so far as these mental structures are specific for language, into those aspects of language structure that are beyond the linguist's intuition. We hope to exemplify this in this study.

The linguist's intuition seems particularly poor in the time domain. The structures generated by his rules are essentially timeless and even there where he comes closest to speech as something physically real, in his description of the phonetic form of sentences, this description is made up of a sequence of segments (speech sounds) that have no extension in time. Those segments are thought of as "articulatory states, their acoustic resultants or their neural-command antecedents, which are themselves largely 'timeless' (Abercrombie<sup>5</sup>), in that a particular segment is no more to be characterised by the time interval over which its defining physical properties are maintained than its graphical representative is by the space it may occupy on the line of print" (Lisker, ref. 6, pp. 152-153).

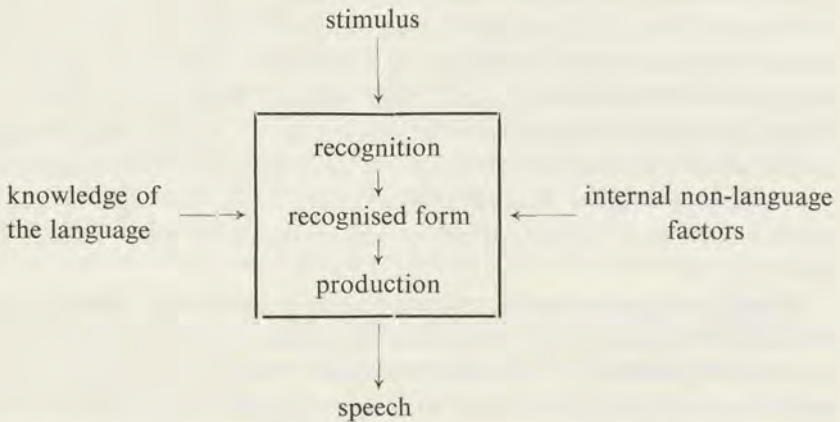
Both physical measurements on the articulatory and acoustic effects of speech-production behaviour and experiments on speech perception show that there are many regularities in the timing of speech. Those regularities are at least partly language-particular and as such reflect learned aspects of verbal behaviour. A study of such regularities may lead to the formulation of rules which model part of the knowledge the speaker has about his language. Thus the study of regularities in the verbal behaviour of language users as revealed by physical measurements to us seems a legitimate way of gaining insight into the structures of language. This requires careful experimenting under controlled laboratory conditions, because otherwise too many uncontrolled factors may affect the verbal behaviour.

A common way of studying the physically measurable aspects of speech is having a subject, speaker of the language one wishes to study, pronounce a word or phrase. This word or phrase may be visually presented to the subject. When he speaks it the required measurements may be made by registering some physiological or acoustic aspect of the resulting speech behaviour or by storing some recorded signals for later processing. By systematically varying the visual stimuli and trying to establish some relation between the stimuli and the speech behaviour properties of the mediating system may be inferred. The simplest diagram possible of the experimental situation looks as follows:



The subject is a highly complex system and many internal factors may affect his speech behaviour. He must recognise the stimulus form with the help of his knowledge of the language, and, again with the help of his knowledge of the language, translate this recognised form into actual speech.

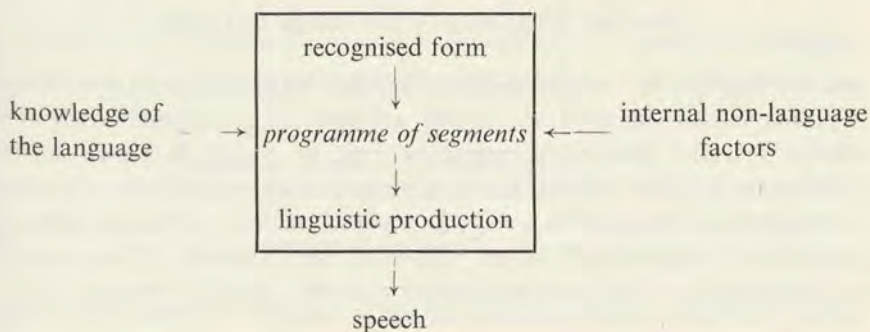
Not only may language factors affect his speech behaviour but also non-language factors. Thus we may draw the following diagram:



This diagram may remind us that in the experimental situation described we may expect to find influences from the recognition processes, from the production processes, from the knowledge of the language and from internal non-language factors. If we wish to use such an experimental situation for investigating the effect of the knowledge of the language we should try to keep all other factors constant. We may for instance optimise the chance of correct recognition by allowing only a limited number of possible stimuli in an experimental session and by an optimal stimulus presentation. We can try to keep the non-language factors constant by instructions to the subject. Many effects of linguistic production, for example those resulting from the physiology of the speech organs, will systematically vary with the form spoken. This will make it difficult, though not always impossible, to separate these effects from those stemming from the organisation of the language.

In the above diagram we have not as yet accounted for the existence of phoneme-like segments on some level of linguistic production. There are arguments, notably from errors of speech, that a speaker, knowing the form or forms which he is going to speak, constructs a programme for his speech-production system, and only after having prepared this programme turns it into actual speech by means of lower-level speech-production processes. Now presumably one may argue for the existence of such programmes on many levels of speech production, for instance the semantic level, syntactic level, phonological level, the level of motor commands.

For the moment we will focus attention on the level at which phoneme-like segments are readied to be spoken. This is expressed in the following diagram:



The form of this programme may be manipulated by the stimulus form. By studying regularities in the speech output we may infer properties of the segments if we are sufficiently sure that we have kept all other factors constant, or know their effects. We assume that the segments are direct representations of the speech sounds as generated by the knowledge of the language. Thus this segmental programme, being the lowest level of speech production where the form of a phrase which is about to be spoken is specified in a sequence of non-overlapping segments corresponding to speech sounds, may be thought of as something like the level of phonetic representation in generative phonology.

That a level (programme) of non-overlapping speech segments exists in speech performance may be argued from the study of speech lapses, in which single speech segments take each other's places, and, in fact, within certain constraints, move around like the letters of a compositor (Cohen <sup>7</sup>), Nootboom <sup>8</sup>), Mackay <sup>9</sup>), Fromkin <sup>10</sup>), Boomer and Laver <sup>11</sup>)). Examples of such lapses are the following, taken from Fromkin:

keep a tape → teep a cape,  
 fish and tackle → fash and tickle.

Such speech errors lead us to include a segmental programme in our performance diagram. There are reasons to believe, however, that not all observable aspects of speech should be derived from segments or from the features of which segments are made up. The study of speech errors reveals also that both stress patterns and intonation patterns are programmed independently of the actual segments. This may be argued from errors in which speech segments, when taking each other's places, leave the prosodic structure as it was, that is segments do not take the prosodic features with them to their new places. This

presumably holds for all errors of speech but in some cases it becomes particularly clear. Boomer and Laver <sup>11)</sup> cite the following slip:

how *bád* things were → how things *bad* were,

and they mention the “interesting possibility that the physiological determinants of primary stress, whatever that may be, are programmed independently of the other articulatory features of utterances” (ref. 11, p. 55). In Dutch we have found examples from which it becomes probable that not only the articulatory determinants of tonic stress (= pitch accent), but also of lexical stress are programmed independently of the individual speech sounds. These examples are comparable to the following English example, cited by Fromkin <sup>10)</sup>:

Wang’s bibliography → Wing’s *babliography*.

In such errors to our experience the realisation of the transposed vowels is definitely changed due to their new positions. This means that whereas the vowels themselves change places the stress levels do not. We therefore assume that the stress pattern of a word is programmed independently of the segmental composition of the word. Let us now consider the following example also cited by Fromkin:

$\begin{matrix} 2 & & 3 & & 1 & & & & 2 & & 3 & & 1 \end{matrix}$   
 a computer in our own laboratory → a laboratory in our own computer.

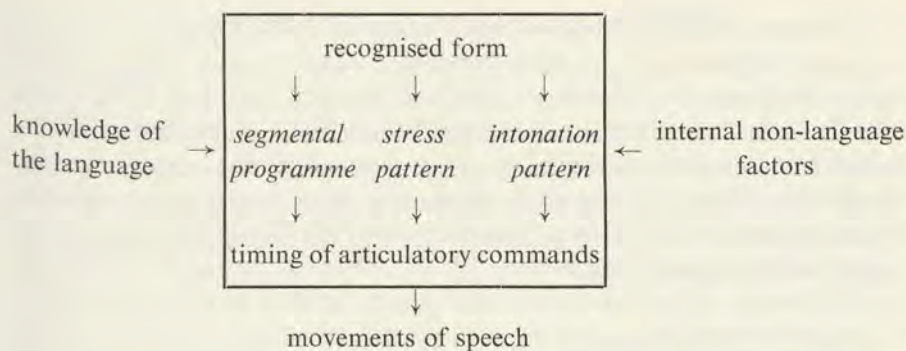
This error shows that “while the word position of primary stress in the phrase is not transposed, the stressed syllable of the word in isolation is the syllable which receives sentence stress” and this leads Fromkin to suggest that “the word stress is stored as part of the articulatory specifications of the stored unit ‘word’, but that the sentence or phrase stress and over-all intonation contour is generated separately” (ref. 10, p. 43).

Thus the study of speech lapses suggests that the place of the major stresses in a sentence or phrase and the intonation contour are generated separately with respect to the actual words chosen and that the lexical stress pattern of a word is specified independently of the segmental composition of the word. We may conceive of the segments as specifying targets, whereas the stress pattern on the one hand and the intonation contour on the other hand are factors determining the amount of deviation from these targets in actual production.

Although on some level of programming the sequence of segments, the stress pattern and the intonation pattern may have their separate form, they must in some way be synchronised on the level of actual timing of articulatory commands.

Some major factors affecting the timing of articulatory commands in our experimental situation may now be schematically represented as follows:





This rather simplistic diagram of the performance factors which may affect the measurable aspects of speech may provide a convenient starting point for our experimental approach.

By systematically changing the “recognised form” and studying the accompanying differences in the speech output we may recover some properties of the mediating system. This, essentially, has been the way of experimenting described in chapter 2 of this study. In order not to vary too many parameters at a time, we have restricted our measurements to nonsense words with a rather simple phonemic make-up. This also enabled us to choose our stimulus material in such a way that we could easily and accurately measure the durational build-up of the resulting productions.

If one wishes to study aspects of the knowledge of the language by studying the output of the speech-production system one runs the risk that the other factors displayed in the above diagram interfere so much with the actual production of speech that the structures of language as contained in the language user’s knowledge are difficult to describe. It would be nice to do away with all the interference of the physical systems of speech production and the limitations they impose on the output. To give an example: it is known that an open vowel has a longer duration than a close one. This may be attributed to mechanical limitations of the speech organs. The mouth has to open further for the open vowel than for the close one and thus the movement of the lower jaw and the lips covers a longer distance for the open vowels than for the close ones. This takes more time and explains the greater duration for the open vowels. There is no a priori reason to assume that the longer duration of the open vowels is part of the mental structures underlying speech. We may assume that open and close vowels would be equally long, other things being equal, but for the mechanical effect of openness. One would like to have a method of bypassing the effects of the mechanics of speech in studying the mental structures underlying it. As one cannot very well open the subject’s skull and see what is inside one should design a task in which the subject externalises the internal representation of the form concerned, or some aspect of it, without actually speaking.

For instance this may be done in a perceptual task in which the subject is asked for some judgment on an auditorily presented word. A special form of such a task is the method of matching to internal criterion, i.e. a task in which the subject is asked to adjust some perceptual characteristics of a repeatedly presented stimulus with a control, according to some internal criterion. Experiments of this kind will be described in chapter 3. They were set up to test some durational rules, formulated in chapter 2, as to the perceptual reality of the regularities described by them.

## 2. ARTICULATORY MEASUREMENTS

### 2.1. Method

Being primarily interested in the factors which underlie the control of articulatory timing in speech we thought it natural to measure durations of speech segments in articulation. In order to do this we used lip contacts for measuring the moments of opening and closing of the lips, tongue contacts in an artificial palate for measuring the moments of beginning and end of palatal tongue closure, and a throat microphone for measuring the moments of beginning and end of phonation. Most of the special-purpose equipment used has been developed in the Institute for Perception Research. It was devised by Willems<sup>12)</sup> and used in earlier phonetic work by Slis<sup>13)</sup> who initiated this type of articulatory measurements in our Institute and who took part in the experiments reported upon in this study.

The lip contacts and the tongue contacts were made by H. E. M. Melotte. The time-measuring device DONDERS was designed by Schouten and Domburg<sup>14)</sup> and built by G. J. J. Moonen. The necessary computer programme for processing the data has been written by H. F. Muller. The main advantage of the measuring equipment is that it makes possible to process a large amount of rather precise articulatory measurements in a much shorter time than more-conventional ways of measuring would require.

A block diagram of the measuring equipment is given in fig. 2.1. The subject was provided with lip contacts as shown in fig. 2.2, or with the artificial palate, shown in fig. 2.3, or, in some cases, with both. He was also provided with a throat microphone. Both lip contacts were silver strips, approximately 0.3 mm thick, 6 mm broad and 4 cm long. They were smoothly bent around the lower

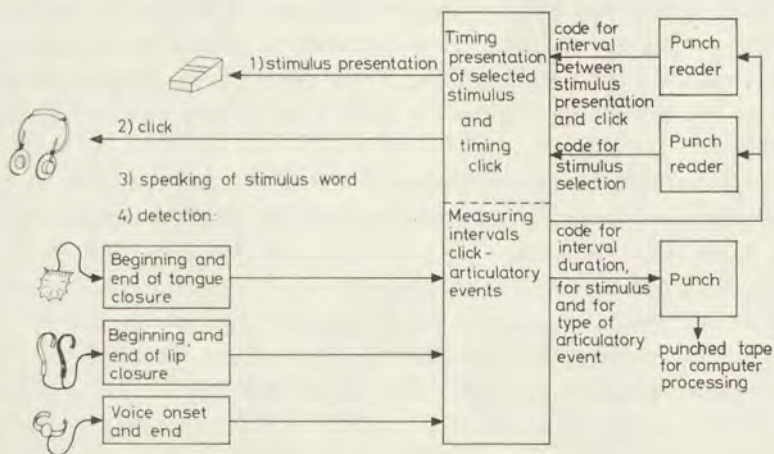


Fig. 2.1. Block diagram of the equipment used in measurements on articulatory timing.



Fig. 2.2. Schematic drawing of the pick-ups used for measuring lip-closure durations.

lip in such a way that they fitted closely to the lip. They do not interfere with normal articulation. The strip in the middle was covered with insulating material, except for a blank thin silver thread which was soldered lengthwise on top of the silver strip, in such a way that it made contact with the upper lip when the mouth was closed. It was approximately 0.5 mm in diameter. The other, non-insulated, silver strip functioned as the second electrode. When the mouth closed the electrical resistance between the two electrodes diminished sharply and discontinuously. When the mouth opened the resistance sharply and discontinuously increased. These sudden changes in resistance were measured by means of an alternating voltage of 0.4 V and the frequency of the signal might e.g. be 8000 Hz. The discontinuous changes in resistance were translated into electrical pulses which were fed to DONDERS.

Two artificial palates were made for two subjects. These artificial palates were of the kind normally used in dentistry. In the surface of the palates a number of pairs of small gold contacts were connected to a corresponding number of measuring circuits outside the mouth by means of thin insulated wires. By using a different frequency for each circuit it is possible to measure the electrical resistance between each gold contact and the blank lip contact separately. In this way the moments of beginning and ending of contact between the tongue and the artificial palate could be measured. The frequencies of the circuits are 5000, 6500, 8000 and 9500 Hz. How the frequencies can be separated is described by Willems<sup>12</sup>). One of the circuits was used for the lip contacts.

In the measurements reported upon in this study only the contacts in the alveolar region of the artificial palate were used. In effect the contact that gave the most reliable results for [t] productions was chosen. Reliability was established in preliminary measurements in which the output of the artificial palate was compared with the output of a throat microphone and the audio signal. These measurements were done for each subject separately.

The subject was also provided with a throat microphone. The signal from this microphone was fed into an electronic device which may be called a "voice onset and end detector". This device gave a pulse when the voice signal rose above a certain threshold and also when the signal fell below this threshold.



Fig. 2.3. Artificial palate with contacts for measuring tongue-closure durations.

These pulses, just as the pulses from the lip and the tongue contacts, were fed to DONDERS.

A small box was placed in front of the subject. It had a window on which six different stimulus words were faintly legible. Behind each stimulus form was a light bulb. Each measuring cycle began with the lighting up of these lamps, making one of the stimulus forms clearly legible. This was the sign for the subject to prepare himself to speak that particular stimulus word. The selection of the light to be lit was controlled by a closed punched tape loop, which was fed into the DONDERS, making the lamp to light up for a short period. The six lamps, and thus the six possible stimuli, were coded in random order in the closed punched tape in such a way that no particular sequence of stimuli was repeated within 120 stimulus presentations. Once the illumination period of a stimulus was ended, DONDERS produced a click which was fed to the subject via headphones. The click was the sign for the subject to speak the stimulus he had prepared himself for. The interval of time between the end of the illumination period and the click varied randomly between 1 and  $2\frac{1}{2}$  seconds. The duration of this interval was controlled by a second closed punched tape loop which was fed to the DONDERS.

The moment the click was produced was also when DONDERS began measuring in ms the moments of voice onset and end and the moments of lip and/or tongue opening and closure. DONDERS stopped measuring after 2 seconds. The click was, so to speak, a general reference point in time for all our time measurements. DONDERS was originally designed as a "reaction-time meter", and was able to measure intervals of time between a stimulus moment (in our case the click) and a large number of independent reactions. After each measuring cycle DONDERS produced the measured intervals in a predetermined order in punched tape. The punched tape of each experimental run was processed further by a general-purpose computer of the Technical University in Eindhoven.

The processing was as follows. The interval between the clock and the first closure was subtracted from all other intervals. Thus the first closure moment was taken as a specific reference point in time. Then the computer calculated all desired intervals between moments of opening and closure and moments of voice onset and end. The standard deviation over all occurrences of such an interval in an experimental run was calculated. Which intervals were actually taken will become apparent from the results.

## 2.2. Measurements

Our measurements were to a large extent exploratory in nature. This gave some freedom in the choice of the stimulus material, perhaps more freedom than one would wish. We have, however, rather narrowly limited our stimulus material in order not to vary too many parameters at once. The choice of the

material was limited by the following considerations and questions. It is known that the articulatory timing in the realisation of a speech sound is much influenced by the adjacent speech sounds. Thus the duration of a vowel before a voiceless consonant is shorter than before a voiced consonant, and before a plosive shorter than before a fricative. This was found for many languages (e.g. Elert, ref. 15, pp. 132–133) and also for Dutch (Slis and Cohen, ref. 16, p. 89). The effect of adjacent speech sounds can be rather dramatic in consonant clusters (Slis, ref. 17, p. 74). In this study we have chosen to concentrate primarily on those aspects of durational control in speech that do not stem from this kind of interaction between adjacent speech sounds. As this kind of interaction is presumably always present we have tried to keep its effects as constant as possible. This was done by using nonsense words with a simple structure such as [pVpVpVp], [tVtVtVt] or [mVmVmVm], in which V stands for a vowel. The vowels were in most cases identical in all syllables of the word.

Due to limitations in our experimental set-up the number of differing stimuli per experimental run could not exceed 6. Thus we made lists of 6 or less stimuli. Each stimulus was presented at least 20 times during an experimental run. Due to errors both in the productions of the subject and in the apparatus the total of productions per stimulus included in the results was often somewhat less than that of stimulus presentations.

The main questions which guided us in making the lists of stimuli were the following:

- (1) What is the role of vowel quantity in the programming of articulatory timing?
- (2) What is the effect of stress and syllable position on articulatory timing in three-syllable nonsense words?
- (3) What is the effect of the number of syllables in the word on articulatory timing in polysyllabic nonsense words?

### 2.2.1. *The role of vowel quantity*

In languages where quantity differences exist they play a role in the programming of timing in articulation. In Dutch there is a quantity opposition for vowels. Quantity is thought of here as a phonological or phonetic feature controlling differences in duration which cannot be attributed to other causes such as interaction of the vowel quality with adjacent consonants or stress or tempo. See further sec. 2.4. In order to explore the durational effects of vowel quantity systematically we used stimuli of the form [pVpVpVp] as e.g. [pa:pa:pa:p]. To begin with this was done for 15 Dutch vowels, including 3 diphthongs (the Dutch vowel system will be more fully dealt with in the discussion of the results). We constructed three lists of stimuli, together covering the 15 vowels. The vowel [a:] was included in each list for purposes of comparison. As it turned out that data from different experimental runs are not

completely comparable, because of intersession differences, the three lists did not make comparison of all possible vowel pairs possible. Therefore a fourth list was constructed for comparing vowels having the same degree of "height" but differing in frontness or backness and lip rounding. One more list was added for comparing a 16th, neutral, vowel [ə], occurring in unstressed syllables only, with long and short vowels. The 5 lists obtained in this way were the following:

list 1	list 2	list 3	list 4	list 5
[pa:pa:pa:p]	[pa:pa:pa:p]	[pa:pa:pa:p]	[pe:pe:pe:p]	[pa:pa:pa:p]
[pe:pe:pe:p]	[pø:pø:pø:p]	[po:po:po:p]	[pø:pø:pø:p]	[pəpa:pəp]
[peipeipeip]	[pʌypʌypʌyp]	[paupaupaup]	[po:po:po:p]	[paɾa:paɾa]
[pepepep]	[pæpæpæp]	[paɾaɾaɾa]	[pɪpɪpɪp]	[pəpəpəp]
[pɪpɪpɪp]	[pyypyyp]	[pɔpɔpɔp]	[pæpæpæp]	[pɪpɪpɪp]
[pipipip]		[pupupup]	[pɔpɔpɔp]	[pəpɪpəp]

Lists 1-4 were used in experimental runs with 3 subjects, one of them being the present author, the other two being colleagues of his. List 5 was used in experimental runs with SN and JtH only. Subject JtH was not involved in any other way in the measurements than as a subject. It needs some justification that non-naive subjects were used in these measurements. In preliminary experiments with lip contacts it was found that there was no obvious difference in the behaviour of naive and non-naive subjects. There were great practical advantages in using non-naive subjects because the subjects had to get used to the experimental situation, the experimental sessions took much time and were rather annoying for the subjects, which made a high motivation level desirable. Furthermore there were no artificial palates available for naive subjects. Despite these considerations the use of non-naive subjects would not have been justified if the results were not confirmed in some independent way. Such confirmation is provided in the perceptual experiments with naive subjects described in chapter 3.

The subjects were instructed to keep the intonation pattern as constant as possible throughout an experimental run. Each stimulus was presented at least 20 times within one experimental run, which took approximately 15 minutes. The results are presented in full in tables 1.A-5.A and 1.B-5.B in appendix A. In the tables 1.A-5.A the first columns give the stimuli, the second columns (n) the number of times the stimulus was spoken by the subject and measured. The columns p<sub>1</sub>-p<sub>4</sub> give the durations of mouth closure for the successive [p]s. The columns V<sub>1</sub>-V<sub>3</sub> give the durations of periods the mouth was open, thus including noise burst and, if present, aspiration. Columns tot. give the interval from first lip opening to last closure. Columns sd give the standard deviations of the preceding intervals.



In the tables 1.B-5.B, which are organised in the same way as tables 1.A-5.A, the columns  $p_1V_1-p_3V_3$  give the intervals between the moment of lip opening and the moment of voice onset, thus presenting the duration of noise burst plus aspiration. The columns  $V_1p_2-V_3p_4$  give the intervals between the moments of lip closure and the voice end, thus presenting the durations of the voice "tails".

In the above lists of stimuli a bilabial consonant was chosen because the durational interaction due to coarticulation between vowel and adjacent consonants will probably be less for bilabial consonants than for consonants formed with the tongue. In the case of tongue consonants the primary articulator for consonants and vowels is the same and therefore interference seems more probable. In order to see whether such expected interference effects indeed show up and to what extent they may disturb the realisation of vowel quantity three more lists of stimuli were made, identical with the first 3 lists of stimuli above, except that all [p]s are replaced by [t]s. These lists are:

list 6	list 7	list 8
[ta:ta:ta:t]	[ta:ta:ta:t]	[ta:ta:ta:t]
[te:te:te:t]	[tø:tø:tø:t]	[to:to:to:t]
[tɛitɛitɛit]	[t^yt^yt^yt]	[tautautaut]
[tɛtɛtɛt]	[tætætæt]	[tatatat]
[tItItIt]	[tytytyt]	[tøtøtøt]
[tititit]		[tututut]

These lists were used for subjects IS and SN only, because no artificial palates were available for other subjects. The results are presented in tables 6-8 in appendix A, which are organised in the same way as tables 1.A-5.A. In these measurements the moments of voice onset and end were not measured due to technical difficulties.

One problem, having to do with vowel quantity in Dutch, cannot be studied in the above measurements. That is the problem of the effect of a following [r] on the quantity of the vowel. It is claimed in the literature that the vowels [u], [y] and [i] are short except before [r] where they become long (e.g. Moulton<sup>18</sup>). With our equipment it is not possible to measure the beginning of a postvocalic [r], as this [r] behaves very much as a diphthong-like change in the vowel, or as a velar fricative. We can, however, measure the duration of vowel plus [r]. As a long vowel plus [r] in Dutch distributionally cannot be followed by a [p], we chose [t] as the consonant following [r]. As the prevocalic consonant we chose [p] because in this way we obtained very natural-sounding Dutch phoneme sequences. We this time restricted the measurements to monosyllables, to avoid unnecessarily complicated nonsense words. Thus we used the following lists:

list 9	list 10
[pa:rt]	[pø:rt]
[po:rt]	[pe:rt]
[purt]	[pyrt]
[part]	[pirt]
[pɛrt]	[pœrt]
[pɔrt]	[pɪrt]

The results for IS and SN are presented in tables 9–10.

### 2.2.2. *The effect of stress and position*

It is known that stressed syllables tend to have a longer duration than unstressed syllables. In order to find out how the difference between articulatory timing in stressed and unstressed syllables depends on the position within the word we made the following list of stimuli:

list 11
[pa:pa:pa:p]
[pa:pa:pa:p]
[pa:pa:pa:p]
[pa:pa:pa:p]
[pa:pa:pa:p]
[pa:pa:pa:p]

This list enabled us to study the effect of stress and position on the realisation of both a long vowel and a short one. We chose the vowels [a:] and [a] because these form the most open long/short pair, and thus the opening and closing movements will be faster than for other vowels which may lead to better-defined moments of lip opening and closing.

The list was used for two subjects, IS and SN, who spoke the words embedded in a carrier phrase as follows:

*de uiting [pVpVpVp] is onzin*  
*(the utterance [pVpVpVp] is nonsense).*

The stressed syllable of the nonsense word received a pitch accent. The results are given in table 11 in appendix A.

In normal speech many polysyllabic words do not receive a pitch accent on the syllable bearing the lexical stress. In order to study the effect of stress and position in such non-accentuated words the same nonsense words were used in a different way. This time the stimulus words were spoken in a carrier phrase, as follows:

*de uiting [pVpVpVp] is onzin*  
*(the utterance [pVpVpVp] is nonsense).*

Thus a 12th list of stimuli was obtained. The subjects were instructed to produce an emphatic stress on the word "onzin". In doing this no other pitch accents were produced in the sentence. To demonstrate this, pitch curves and amplitude envelopes were recorded of individual cases of both ways of producing the stimulus words. These recordings will be presented and dealt with at the discussion of the results. The results of the durational measurements are presented in table 12.

The data on the moments of voice onset and end for the measurements concerning the effect of stress and position are not presented at all because they did not show anything new compared to the measurements as to the role of vowel quantity.

### 2.2.3. *The effect of number of syllables*

It has often been stated that the number of syllables in a word affects the duration of the individual segments of that word (e.g. Lehiste, ref. 19, pp. 40-41). In order to study this effect we made the following lists of stimulus words:

list 13	list 14	list 15	list 16
[ma:m]	[mam]	[ma:ma:m]	[mamam]
[ma:ma:m]	[mamam]	[ma:ma:ma:m]	[mamamam]
[ma:ma:ma:m]	[mamamam]	[ma:ma:ma:ma:m]	[mamamamam]
[ma:ma:ma:ma:m]	[mamamamam]	[ma:ma:ma:m]	[mamamam]
[ma:ma:m]	[mamam[	[ma:ma:ma:ma:m]	[mamamamam]
[ma:ma:ma:m[	[mamamam]		

In these lists we used the consonant [m] instead of [p] for two reasons. Firstly the preceding measurements had shown that the moments of voice onset and end are completely dependent on the moments of lip opening and closure respectively and thus not much information is gained by using a voiceless consonant. Secondly it was found in preliminary measurements that words of up to four syllables lead more easily to production errors when all consonants are [p] than when all consonants are [m].

For this first exploration of the effect of number of syllables on the articulatory timing within a word we did not think it necessary to take into account all possible stress placements for the words of one up to four syllables. In lists 13 and 14 we can study the effects of number of syllables both following an initial stressed syllable and before a final stressed syllable. In lists 15 and 16 we can study the effect of adding unstressed syllables preceding the stressed syllable on the rest of the word. The data can be found in tables 15-16. The items in lists 13-16 were spoken by one subject, SN, in isolation and with a pitch accent on the stressed syllable. In order to see how the effect of the number of syllables comes about in words which do not play an important role

in the phrase and do not receive a pitch accent, the same nonsense words were used in a carrier phrase as follows:

*de uiting . . . is onzin*  
(*the utterance . . . is nonsense*).

Thus 4 more lists of stimuli, 17, 18, 19 and 20 were obtained. These phrases were spoken by the same subject with emphatic stress on the word "onzin" and no pitch accent on the stressed syllable of the nonsense word.

The results are presented in tables 17–20. The data of our articulatory measurements may be of interest to other workers in phonetics and therefore they are presented fairly completely in appendix A. For those who are interested in the acoustic durations of the vowels it is worth noting that for tables 1–5 they can be calculated by subtracting the voice onset times given in tables 1.B–5.B from the [V] durations given in tables 1.A–5.B. Furthermore one may note that, due to the rounding off in the mean segment durations, the sum of the mean segment durations of a word may differ slightly from the total word duration given in the tot. columns.

### 2.3. Accuracy in articulatory timing and reproduceability of an articulatory programme

In the perception of speech we are sensitive to amazingly small differences in duration. Well-known examples are the effect of the closure duration on the voiced/voiceless distinction (Lisker<sup>20</sup>, Slis and Cohen<sup>16</sup>) and the effect of voice onset time on the perception of the voiced/voiceless distinction and of aspiration (Lisker and Abramson<sup>21</sup>); differences below 20 ms may cause a change in the perceptual judgements. The importance of durations in the perception of speech also is clearly demonstrated by the fact that it is possible to synthesise understandable and even acceptably sounding speech from spectrally homogeneous segments by carefully shaping the time function of the amplitude envelopes of the segments (Cohen, Schouten and 't Hart<sup>24</sup>).

Reviewing the literature on psychophysical measurements of the just-noticeable differences in duration, Ilse Lehiste (ref. 19, p. 13) comes to the conclusion that "it appears that in the range of the durations of speech sounds — usually from 30 to about 300 ms — the just-noticeable differences in duration are between 10 and 40 ms". The amazing accuracy in perceiving durational differences is not only present when a subject has to discriminate between phoneme categories. Huggins<sup>25,26</sup> has shown that a listener can detect changes in the duration of a segment within a spoken sentence when these changes are as small as 10–20 ms and no discrimination task is involved. It seems reasonable to expect that a speaker is able to control his articulatory timing with the same degree of accuracy. And indeed one has found that a speaker, when repeating the same phrase over and over again, may reach standard deviations for the

duration of syllables of 10–15 ms (Kozhevnikov and Chistovich, ref. 27, p. 98). We may assume that in repeating the same phrase the same articulatory programme is realised over and over again and that the fluctuations found in the measurable durations are due to inaccuracies in the measurements, fluctuations in the production processes and perhaps minor fluctuations in the stored programme. We may study the amount of variation caused by these factors in our measurements by looking at the standard deviations for segment durations in repeating the same nonsense words within one experimental run. On the other hand we may see something of the amount of freedom there exists in constructing such a timing programme by comparing the durational build-up of the same nonsense word as spoken by different subjects, or as by the same subject in different experimental runs.

### 2.3.1. Accuracy in the timing of lip movements

Figure 2.3.1 gives an idea about the accuracy which may be reached in articulatory timing. There the standard deviations are given for the successive segments of the word [papapap] as spoken 17 times by JtH (table 2.A in appendix A). One may notice that 4 of the standard deviations are below 5 ms. This shows that the control of segmental duration, in this case of the intervals between opening and closing of the lips, can be even more precise than suggested by the evidence from perceptual experiments. Now this is an exceptional case in our measurements.

In fig. 2.3.2 the mean standard deviations for the successive segments of all [pVpVpVp] words in tables 1.A, 3.A and 4.A are given for the long-vowel

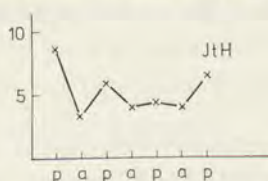


Fig. 2.3.1. Standard deviations in ms of successive articulatory segment durations as defined by the moments of lip closing and opening in the nonsense word [papapap] as spoken 17 times by subject JtH.

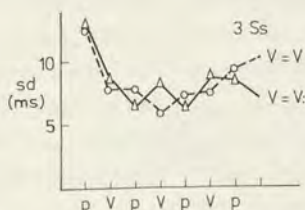


Fig. 2.3.2. Standard deviations in ms of successive articulatory segment durations as defined by the moments of lip closing and opening in long-vowel and short-vowel nonsense words of the form [pVpVpVp]. The standard deviations, taken from tables 1.A, 3.A and 4.A, were averaged over different words and over three subjects.

words and the short-vowel words separately. The data are averaged over the three subjects. We had intended to average over all the tables with data from three subjects, but we excluded table 2.A because of the very high standard deviations in the words with [y] for two subjects. These high standard deviations occur both in the vowel segments and in the consonant segments. This suggests that they are due to an error in the measurements. Because the lips were protruded a great deal and did not open very much in the words with the high rounded vowel [y], presumably the lip contacts did not function well.

From fig. 2.3.2 it is apparent that, except for the first [p] closure, all standard deviations are well below 10 ms. Thus the accuracy in articulatory timing seems to be at least as good as is suggested by the evidence from perception. This high accuracy in particular shows that our subjects are indeed able to keep their articulatory programme for a particular nonsense word constant during an experimental run. From fig. 2.3.2 it is also apparent that the standard deviations in the long-vowel words are somewhat different from those in the short-vowel words. The stressed short vowel in particular has a lower standard deviation than the stressed long vowel. This may have to do with the durational difference between the two. Also the [p]-closure durations in the long-vowel words seem to have lower standard deviations than those in the short-vowel words. This can hardly be explained by durational differences and suggests that somehow the accuracy in the articulatory timing of the closure is affected by the preceding and/or following vowel. A tentative explanation of these phenomena has been suggested by Slis<sup>28)</sup> who assumes that in the case of the long vowel the syllable boundary falls before the consonantal closure and in the case of the short vowel within the consonantal closure.

### 2.3.2. Accuracy in the timing of voice onset and end

We have measured the voice onset time relative to the consonant release and the voice "tail" defined as the interval between the moment of consonantal closure and voice end. Voice onset time was measured for all three syllables of the words with the form [pVpVpVp], voice tail was measured for the first two syllables only. The data from tables 1.B, 3.B and 4.B are summarised below:

	voice onset time						voice tail			
	first		second		third		first		second	
	syll.		syll.		syll.		syll.		syll.	
	ms	sd	ms	sd	ms	sd	ms	sd	ms	sd
IS	22	9.0	13	3.8	13	4.4	23	4.8	27	4.0
JtH	16	3.8	10	2.9	11	2.3	39	3.6	37	3.5
SN	10	3.3	10	2.7	10	3.0	30	4.0	29	3.9

The voice onset time in the first syllable is somewhat longer than in the other syllables for subjects IS and JtH. It is plausible, however, that this is due to a systematic error in measurement. The moment of voice onset was defined as the moment the amplitude of the throat-microphone signal rose above a certain threshold value. In the short unstressed first syllable the slope of the amplitude increase of this signal may be less steep than in the other syllables, thus causing a delay in voice onset time.

One may notice that the voice onset time differs very little for the three subjects and is but short. The standard deviations are also very small. Actually in 27 cases the standard deviations were below 2 ms. No further systematic difference was found in voice onset time for the different vowels. Thus it seems that the moment of voice onset is not programmed independently of the moment of consonant release. In particular the low standard deviations suggest that the moment of voice onset is directly dependent on the moment of consonant release. This dependence may be aerodynamic, i.e. it may be the case that the vocal cords are brought into position just before the consonant release and automatically begin to vibrate when the air stream increases due to escape of the air. A similar reasoning applies to the voice end. We would suggest that the moment of voice end is in a way controlled by the moment of consonantal closure, such that when the moment of consonantal closure shifts in time the moment of voice end automatically follows.

This high dependence of both voice onset and voice end on the moments of opening and closing of the lips means that little additional information on articulatory timing in these nonsense words comes from studying the data on voice onset and end. In studying articulatory timing we will henceforth concentrate on the moments of oral opening and closure only.

### 2.3.3. *Constancy in the articulatory programme*

We have seen above that a subject is able to keep the timing programme for a nonsense word constant during an experimental run with an amazing degree of accuracy. It would be interesting to know, however, what the amount of freedom is in the construction of such a programme. In how far is a subject free to time his articulation and in how far is his timing governed by the internal organisation of speech behaviour? Our measurements were not designed to find the upper limits of this freedom but we get some idea about the fluctuations in the timing programmes by comparing the durational build-up of the nonsense word as spoken by different subjects or as spoken by the same subject in different experimental runs.

In fig. 2.3.3 one sees a graphical representation of the durations of articulatory segments of the words [pa:pa:pa:p] and [papapap] as spoken many times by three different subjects in one experimental run for each subject. The vowel

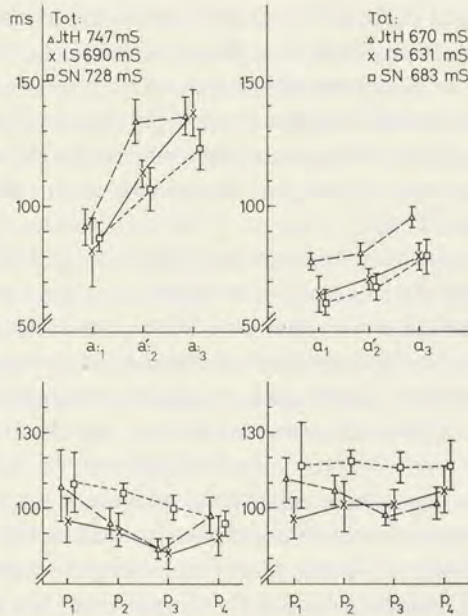


Fig. 2.3.3. Durations in ms of articulatory segments of the words [pa:pa:pa:p] and [papa:papa:p] spoken many times by three different subjects. Top: the durations of the successive vowels, bottom: the durations of the successive [p] closures. For each duration the standard deviation is indicated on both sides by a vertical line.

durations are given in the upper half, the [p]-closure durations in the lower half of the figure. For each duration, the standard deviation is indicated on both sides by a vertical line.

Although the durational values per word are significantly different for the three subjects, the differences are surprisingly small. The subjects seem to adhere to very similar patterns, of which the main characteristics seem to be a short vowel duration in the first syllable, a rather long vowel duration in the second, stressed syllable, and a vowel duration which is at least as long as that in the stressed syllable, in the third, unstressed syllable. This pattern holds good for the long-vowel word and also, as it were on a decreased scale, for the short-vowel word. The closure durations do not show a very pronounced pattern.

In fig. 2.3.4 a graphical representation is given of the durations of articulatory segments of the word [pa:pa:pa:p] as spoken by one subject in 4 different experimental runs. Again two times the standard deviation is indicated for each duration. It is clear that the fluctuations in the timing programme for one subject are not less than those between subjects. But again these fluctuations are not very great and the subject seems to stick rather well to the same overall timing pattern.

These results suggest that there are fairly fixed preferred patterns for the control of articulatory timing in polysyllabic nonsense words.



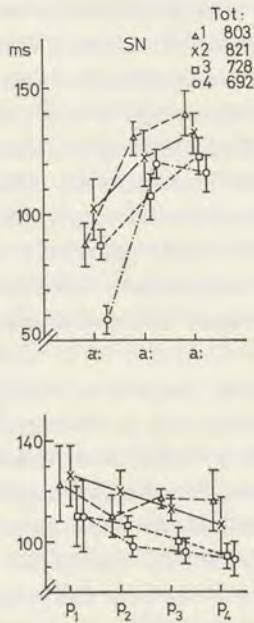


Fig. 2.3.4. Durations in ms of articulatory segments in the word [pa:pa:pa:p] as spoken many times by one subject in four different experimental runs. Top: the durations of the vowel segments; bottom: the durations of the [p] closures.

In terms of our simplistic diagram for the control of articulatory timing in speech as given in sec. 1.2, the above results make probable that the contribution of the non-language processes does not lead to widely differing timing programmes. The fluctuations which do exist in the timing programmes possibly arise from non-language factors. The effect of the language processes seems to be overriding. We intend to investigate the contribution of some of the factors underlying the control of articulatory timing in polysyllabic nonsense words and provide rules for describing the main characteristics of the patterns in articulatory timing.

#### 2.4. The role of vowel quantity in articulatory timing

In many languages some vowel phonemes are characterised by longer durations than other vowel phonemes. We assume here that such differences in characteristic vowel duration are controlled by a phonetic feature of vowel quantity. Phonetic feature is taken here in the Chomsky and Halle sense of a partly independently controllable aspect of speech (Chomsky and Halle, ref. 29, p. 297). The phonetic feature of vowel quantity thus controls differences in vowel duration that cannot be attributed to other causes such as interaction of the vowel quality with the adjacent consonants or stress or tempo. Mani-

festations of this feature can best be studied by comparing supposed long and supposed short vowels in identical conditions. In such studies of vowel quantity this comparison has mostly been restricted to vowels in stressed syllables. Exceptions are Delattre and Hohenberg<sup>30)</sup> and Lehiste (ref. 19, pp. 138–139). Delattre and Hohenberg studied duration as a cue to the tense/lax distinction in German unstressed vowels. Their data are difficult to interpret because of their definition of vowel duration which was restricted to the steady-state portions of the vowels. Lehiste's study concerns vowel quantity in word and utterance in Estonian. Her basic finding was that the quantity ratio defined as  $V/V'$  was the same for stressed and unstressed syllables, but that absolute durations differed.

We have undertaken to study the role of vowel quantity in Dutch in articulatory timing, both in stressed and in unstressed syllables. The assumption that a feature of vowel quantity should be included in the list of phonetic features may be argued as follows. In some languages there are pairs of long and short vowels which have essentially the same targets, i.e. ideal vocal-tract configurations (e.g. Norwegian, Finnish, Czech). In other languages the vowel targets of paired long and short vowels are different (German, Swedish, Dutch). Whether the vowel targets of long and short vowels are the same or different in a specific language may be found by asking native speakers to produce these vowels continuously, which many people are very well able to do. In languages where paired long and short vowels have different vowel targets the continuous versions of long vowels and their short counterparts are markedly different; in languages where long and short vowels have essentially the same targets, these sustained versions give the same perceptual results. In the latter case the only difference between the members of the long/short pairs is the difference in duration, and from this we conclude that duration may be controlled independently in human speech.

It may be noted that in discussions of the tense/lax opposition in English vowels, differences in duration are often considered to be secondary to other differences (Hockett, ref. 31, p. 31; Chomsky and Halle, ref. 29, pp. 324–325; Perkell, ref. 32, p. 64). These discussions also seem relevant to research on Dutch vowels, because the Dutch long vowels are considered to be tense and the short vowels to be lax (Cohen et al., ref. 33, pp. 12–18). We uphold the view that even though there is a tense/lax opposition the differences in vowel duration are controlled independently of other differences. We essentially agree with Delattre (ref. 34, p. 1143), who stated "that any implication that lax/tense might be the cause of short/long is badly misleading".

This view is in accordance with the results of an experimental phonetic study by Lindblom<sup>3)</sup> who showed that the deviation of a vowel from its target position may for a given consonantal environment be calculated solely from the vowel duration.

Whenever we use the terms long, short and quantity alone we will refer to the abstract feature of vowel quantity. When we intend to refer to the measurable durations we will speak of long and short durations. It is part of the theory of phonology that in the phonetic representation features are specified by positive integers that indicate the degrees to which a feature is present in the segment concerned (Chomsky and Halle, ref. 29, pp. 65, 297). We will refer to this as the value or specification of the feature. We will try to recover from measurable articulatory durations some conditions which have to be imposed on the specification of the feature of quantity in the phonetic representation of Dutch vowels. In the following brief discussion of the vowels of Dutch we will clarify the specific questions we intend to seek an answer to.

### 2.4.1. *The vowels of Dutch*

In Dutch as spoken by our subjects there are, diphthongs included, 15 vowels apart from the schwa occurring only in unstressed positions and three vowels [ɛ:, œ:, ɔ:] occurring only in loan words. In the Dutch of other people there may be one more vowel: besides [ɔ] those people also have [o]. This will not be taken into account here. Also the vowels occurring only in loan words will not be studied because of the difficulty of pronouncing these vowels in nonsense words. Thus we will restrict our investigation to the 16 vowels in fig. 2.4.1. Figure 2.4.1 may be considered a stylised version of an acoustical

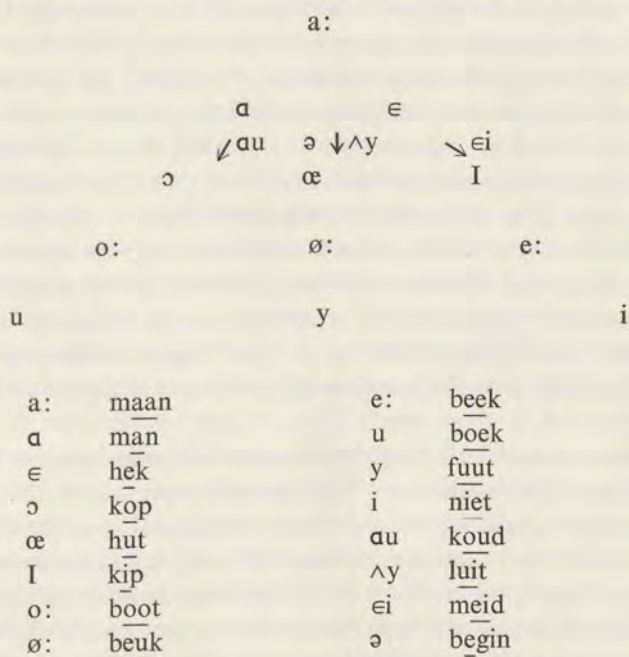


Fig. 2.4.1. Stylised vowel triangle of Dutch vowels, with key words.

vowel triangle. It may also be given an articulatory interpretation in that the vertical axis mainly corresponds to degree of height, [a:] being the lowest vowel, and the horizontal axis mainly corresponds to frontness versus backness, [i] being the most-front vowel.

There are four long/short pairs, viz. [a:, ɑ], [o:, ɔ], [ø:, œ] and [e:, I], in which a difference in duration, alongside with other differences, is supposed to be important for keeping the vowels apart. These other differences may be vowel quality, diphthongisation of the long vowels, and, at least in isolated vowels, less-abrupt decay of amplitude for the long vowels (Cohen et al. <sup>35</sup>). None of these other differences, however, can explain the differences in duration in a way that would make them follow automatically from these other differences. The suggestion that the main difference is one of tenseness (Cohen et al., ref. 33, pp. 12–18) has not been substantiated. No good arguments have been found for the idea that the primary difference is a perceptual one, the long vowels plus [u, y, i] being "clear" (Dutch: *helder*) the short vowels being "dull" (Dutch: *dof*) (De Groot <sup>36</sup>). It has also been suggested that in Dutch there is no quantity but rather an opposition in "contact" or "Silbenschnitt". For a discussion see sec. 2.4.2.5.

We assume that each vowel can be assigned a value for the quantity feature. A question which we want to answer is whether all long vowels have the same value and all short vowels have the same value. This is not necessarily the case. It has been found for Swedish, for instance, that the durational ratio for the pair [u, u:] is less than for the other vowel pairs (Elert, ref. 15, p. 113). Elert explained this by the different articulation: "the short [u], phonetically [θ], has a considerably more open and back articulation and hence, greater intrinsic duration (...) than the long allophone". Hadding-Koch and Abramson <sup>37</sup>), however, having found in a perceptual experiment that the cue value of duration in this pair is less than in other pairs, suggested that the cue value has shifted to quality in this pair and that thus the constraint upon the speakers to maintain a clear durational difference lessened (Lehiste, ref. 19, p. 33). This could mean that on the level of phonetic representation the values for the quantity feature would be different for the [u, u:] pair than for the other pairs. We will try to establish from our measurements whether in Dutch all long vowels form one class and all short vowels do likewise.

Apart from the four long/short pairs mentioned there are four vowels, viz. [ɛ, u, y, i] that do not participate in a quantity opposition. The [ɛ] is considered to be short by Dutch phoneticians and phonologists. The [u, y, i] are generally considered to be short except before [r]. We raise the question whether they indeed are as short as the genuine short vowels [a, ɔ, œ, I] or whether they perhaps should be assigned some in-between value for the quantity feature. Furthermore we would like to know whether indeed [u, y, i] behave in the same way as the long vowels before [r].

There are three diphthongs in Dutch [ɛi, ʌy, au]. Measurements of Zwaardemaker and Eykman (ref. 38, p. 298) suggest that these have a longer duration than the long vowels. This could be explained by assuming that Dutch long vowels constitute one segment on the level of phonetic representation and the diphthongs two segments. If, however, the durational behaviour of long vowels and diphthongs is not systematically different, this could be interpreted as evidence for assigning the same number of underlying segments to them and assigning to these segments identical values of quantity.

A final question concerns the neutral vowel [ə] (schwah). This vowel does not participate in a quantity opposition and occurs in unstressed syllables only. We should like to know whether in those unstressed syllables the [ə] should be assigned a quantity value different from that of the short vowels.

The questions posed here together boil down to the more general question as to which restrictions have to be imposed on the specification of the quantity feature in the phonetic representation of Dutch vowels. We will try to answer this question and at the same time study the way in which quantity is realised in both stressed and unstressed syllables.

#### 2.4.2. *Studying realisations of vowel quantity in nonsense words*

The specification of the quantity feature concerns the level of phonetic representation. This is a level within transformational grammar and as such is supposed to be descriptive of the knowledge an ideal speaker/hearer has of his language, in this case of the sound structure of his language. This knowledge of the language is only one of the factors entering into the timing of articulation. If we wish to study the effect of some aspect of the knowledge of the language on the timing in articulation we should keep all other factors as constant as possible. We have tried to do this by making the conditions on the realisations of quantity differences as rigorous as possible. The use of phonologically possible nonsense words diminishes undesired semantic effects on the timing. The use of polysyllabic words with the same stress pattern increases the possibility for the subject to stick to the same rhythmical pattern. This was also helped by the instruction to keep the intonation pattern as constant as possible throughout an experimental run.

By using three-syllable words with the stress on the second syllable we were enabled to study realisations of quantity in three rather different types of syllable, in the unstressed short first syllable, in the stressed second syllable and in the unstressed but stretched third syllable.

In the words with the form [pVpVpVp] the effect of the interaction between consonants and vowels due to coarticulation was diminished because the main articulator for consonant and vowel was not the same.

By using three-syllable words with three times the same vowel in a number

of measurements we were able to use the differences in total word duration as a measure for the effect of quantity differences.

By comparing the [pVpVpVp] words with the [tVtVtVt] words we may get some idea about the effect of coarticulation on the realisation of quantity.

The special behaviour of [u, y, i] before [r] can be studied by comparing the durations of the [Vr] part in words of the form [pVrt] with a long vowel, a short vowel or one of those doubtful cases. We will now discuss our results on vowel quantity.

#### 2.4.2.1. Vowel quantity and actual durations

Figures 2.4.2 and 2.4.3 exemplify the kind of results we get on the effects of vowel quantity and vowel quality. The figures give values averaged over the three subjects. The vowel durations are represented in the top part of the figures, the consonant durations in the bottom part. From these figures we may

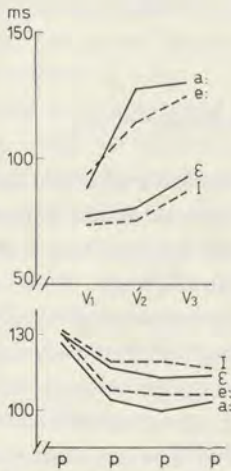


Fig. 2.4.2. Top: durations in ms of long and short vowels in three different positions in words of the form [pVpVpVp]. Bottom: the corresponding [p]-closure durations.

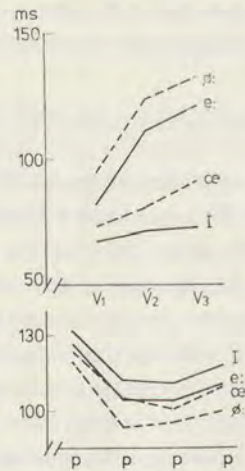


Fig. 2.4.3. As fig. 2.4.2.

see that the long vowels have a longer duration than the short vowels in all three syllables. This difference is not the same, however, for each position. In the stressed syllable and in the unstressed final syllable the difference is considerable, in the unstressed initial syllable only slight (for the first, second and third syllable the differences are approximately 15, 40 and 50 ms respectively, whereas the V/V: ratios are approximately 0.85, 0.65 and 0.70). Clearly the actual vowel durations depend among other things on the vowel quantity, on position within the word and on stress. This holds good not only for the abso-

lute differences in duration of long and short vowels, but also for the V/V: ratio. Thus neither the absolute difference nor the V/V: ratio is constant.

Sometimes Dutch is considered a language in which stressed vowels alone carry information on quantity (e.g. Elert, ref. 15, p. 110). Our data show that unstressed vowels may also carry information on quantity, be it not to the same degree for all unstressed syllables.

It seems significant to us that a durational difference is also kept up in the short unstressed first syllable. Here this difference is sometimes so small that it cannot be expected to have much perceptual relevance (e.g. 8 ms for [e:] vs [I] in list 4.A, subject IS). This probably indicates that articulatory timing is not only controlled by perceptual necessities but also by fixed underlying properties which are systematically present even when their effect is attenuated to the extent that they become perceptually irrelevant.

Thus our data demonstrate that the quantity feature is systematically at work in all positions in an attempt to keep the long and short vowels apart. The extent to which it succeeds depends on non-segmental factors such as stress and position.

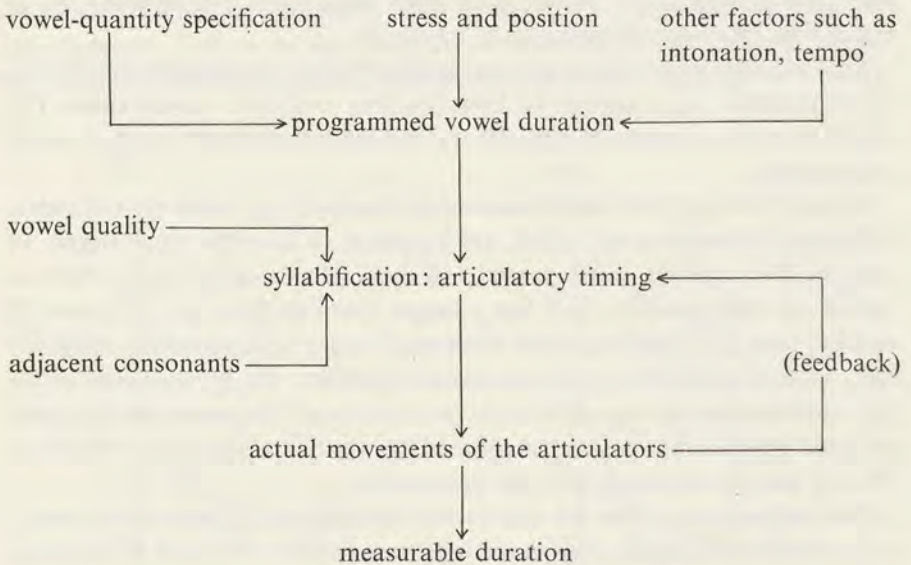
Figures 2.4.2 and 2.4.3 also demonstrate, however, that there are systematic differences between vowels which are supposed to have the same degree of quantity. For example, in the second and third syllable [a:] is longer than [e] and in all three positions [ø:] has a longer duration than [e:], [ɛ] than [I] and [œ] than [I]. Note that these differences within vowel-quantity categories seem to be compensated in the consonant durations: the [p] durations in the [a:] words are shorter than those in the [e:] words, etc. This means that probably the total word durations are not affected by the vowel-quantity differences. We will use this observation in the next section.

One may also note that the [p]-closure durations are systematically longer in the short-vowel words than in the long-vowel words. This will be discussed in sec. 2.4.2.5 in relation to the question of close and loose contact.

#### 2.4.2.2. Vowel quantity as reflected in the total word durations

It is well known that the measurable duration of a vowel depends among other things on the adjacent consonants and particularly on the following consonant. Eli Fischer-Jørgensen (ref. 39, p. 207) found that "the duration of the vowel depends (under otherwise equal conditions) on the extent of the movement of the speech organs required in order to come from the vowel position to the position of the following consonant". Thus, for instance, the vowel [u] had a longer duration in [ud] than in [ug] and the vowel [i] had a longer duration in [ig] than in [id]. These and similar findings led Fischer-Jørgensen to hypothesise that "the motor command for the timing is the same irrespective of the quantity of the vowel, but the execution of the command may be delayed owing to the movements to be made". This, in our opinion, means that on

some level of the programming of speech the vowels [u] and [i] and other vowels having the same degree of quantity, are assigned the same programmed duration, but in the actual production the measurable durations may come about differently due to the limitations of the speech-production system. This puts us in a difficult position, because we intend to find out whether or not some categories of vowels have the same programmed duration by studying measurable durations. These measurable durations result only partly from some abstract programmed duration, and may be thought to be derived from the combined effects of a quantity value, stress and position and perhaps intonation. Possible effects on measurable vowel duration may be represented schematically in the following diagram:



In our measurements we have tried to reduce the effect of stress and position for each vowel within the nonsense words by keeping these factors as constant as possible. For the moment we assume that differences in these effects for the different vowels are negligible. The effect of adjacent consonants, however, is certainly not negligible. Although we only compare vowels in identical environments it is not to be thought that the effect of the environment is identical for the different vowels. Fischer-Jørgensen's <sup>39)</sup> measurements show that the effect of the same consonant may be different for different vowels. From our diagram it may be postulated that this effect can come about in at least three ways: it can be preprogrammed in the syllabification processes. For instance, we find that an [a:] has a longer duration in the environment [pVp] than an [e:]. We may think that this is caused by the syllabification processes which automatically, due to some subroutine, lead to a greater interval of time between the



commands for lip opening and closing for [a:] than for [e:]. This would imply that the larger opening of [a:] compared to [e:] is anticipated on the planning stage. Thus the movement for reaching the target position will take more time for [a:] than for [e:]. We may also think, however, that at the planning stage of syllabification the arrangement of opening and closing commands on some internal time axis is identical for [a:] and [e:]. We may then account for the durational difference by assuming that in the execution of the movement the realisation of the [a:] movement takes longer than that of the [e:] movement and that the closing movement is somewhat delayed by feedback from the speech apparatus. At present we are not able to decide between these two possibilities. They both explain the durational difference by a greater interval of time between the neural command for the opening movement and the neural command for the closing movement. There is still another explanation. It may be possible that the time interval between the neural commands is identical in the two cases and that the durational difference is caused by the actual articulatory movements only. In that case we would expect the durational difference between the words with [a:] and the words with [e:] not to affect the total word duration at all. The timing of the neural commands would not depend on the vowel quantity and the durational differences in measurable vowel duration would be completely compensated in the consonant durations. This idea can be checked in our data, and if this really did seem to be the case, then the degree of vowel quantity should be reflected in the total word durations of those [pVpVpVp] words which have three identical vowels, in a much more precise way than in the measurable intervals between lip opening and lip closing. Thus the total word durations would be a measure for the quantity of the vowels. If we find that the total word durations of [pa:pa:pa:p] and [pe:pe:pe:p] within one experimental run are identical this can be explained by assuming that these vowels have the same degree of quantity and per syllable have the same programmed duration, which leads to an identical spacing of lip-opening and lip-closing commands on the time axis. This naturally holds also for other vowel pairs.

In figs 2.4.4–2.4.7 the total word durations are graphically represented per experimental run. The small horizontal lines give the arithmetic means and the vertical lines the estimated standard error defined as  $m \pm s t/\sqrt{n}$ , in which  $m$  is the mean,  $s$  the standard deviation,  $t$  the  $t$ -factor for small samples and  $n$  the number of items in the sample;  $t$  was chosen for a confidence level of  $p < 0.05$ . This representation gives only a rough visual indication whether total word durations are significantly different or not. A more precise calculation may be made by applying the  $t$ -test for independent means on each pair of total word durations within the same experimental run (McNemar, ref. 40, pp. 102–103). Doing this we find that within each experimental run the words with vowels belonging to the category [a:, o:, ø:, e:, au, ay, ei] have total

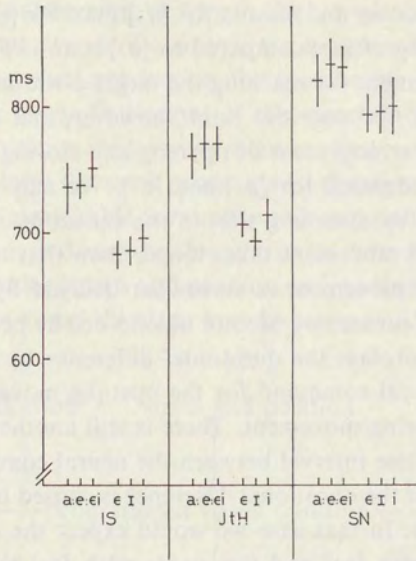


Fig. 2.4.4. Total word durations per experimental run for [pVpVpVp] words. Arithmetic means and estimated standard errors indicated.

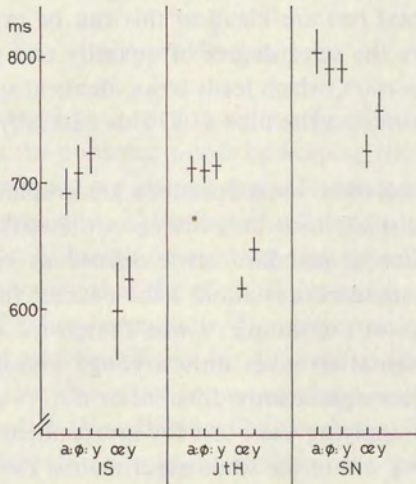


Fig. 2.4.5. As fig. 2.4.4.

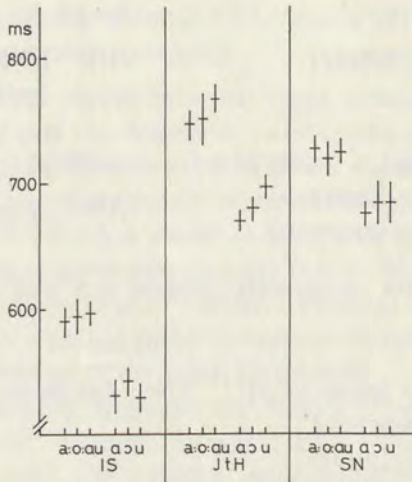


Fig. 2.4.6. As fig. 2.4.4.

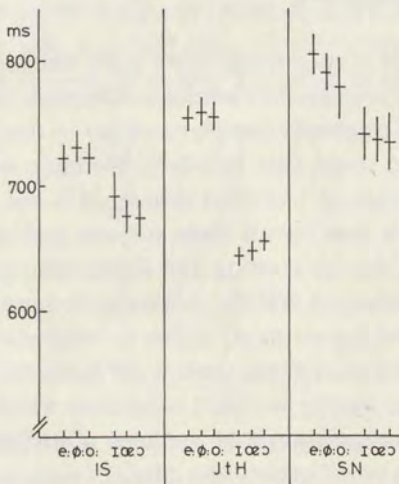


Fig. 2.4.7. As fig. 2.4.4.

word durations which are significantly different from those of the words with vowels of the category [ɑ, ɔ, ɛ, I, ε, u, y, i]. This is not very surprising. The programmed durations of long vowels and diphthongs are apparently longer than those of short vowels plus [u, y, i] although the difference is smaller than one would have expected on the basis of the vowel durations alone. This is due to the fact that the words with short vowels have somewhat longer consonant durations than the words with long vowels. We will return to this later.

Applying the *t*-test to word pairs with vowels of the same category shows the following pairs to be significantly different at a level of  $p < 0.02$ :

for list 1

JtH: [pipipip] > [pIpIpIp]

for list 2

IS: [p^y p^y p^y p^y] > [pa:pa:pa:p]

JtH: [pæpæpæp] > [pyypyyp]

SN: [pæpæpæp] > [pyypyyp]

for list 3

JtH: [pupupup] > [papapap]

[pupupup] > [pɔpɔpɔp]

for list 4

JtH: [pɔpɔpɔp] > [pIpIpIp]

The following pairs are significantly different at a level of  $p < 0.1$ :

for list 1

IS: [pɛipɛipɛip] > [pe:pe:pe:p]

[pipipip] > [pɛpɛpɛp]

JtH: [pɛpɛpɛp] > [pIpIpIp]

for list 4

SN: [pe:pe:pe:p] > [pɔ:pɔ:pɔ:p]

for list 3

IS: [pɔpɔpɔp] > [pupupup]

In all there are 12 out of 66 possible word pairs which reach a level of significance. In the majority of cases the durational difference between vowels within the same category is completely compensated for in the consonant durations. A good example is the vowel pair [a:]–[e:]. We have seen that the duration of [a:] is longer than that of [e:]. This difference is not reflected in the total word durations. This is true for all three subjects and may indicate that the articulatory timing of the lip opening and lip closing on the level of neural control has been the same and that the difference in durational build-up of the words [pa:pa:pa:p] and [pe:pɛ:pe:p] is due to mechanical effects only.

That there are 12 word pairs which show word durations that are significantly different, in spite of our having provided conditions which were as identical as possible, means that reorganisation at the level of articulatory timing due to the interaction between vowel quality and adjacent consonants is not abnormal. Our data show, however, that by using bilabial consonants and strict experimental conditions, the effects of reorganisation can be minimised such that indeed the total word durations may be taken to represent the degrees of vowel quantity.

Our data may be given the following interpretation. Underlying the measurable vowel durations in our nonsense words there are only two degrees of vowel quantity, one for the vowels [a:, ɔ:, ø:, e:, au, ^y, ɛi] and one for the vowels [a, ɔ, œ, I, ɛ, u, y, i]. Thus the diphthongs behave like long vowels and the vowels [u, y, i] like short vowels. We may think of these degrees of vowel quantity as targets as it were in the time dimension, which may or may not be reached in the actual production.

### 2.4.2.3. Explaining the durational differences between vowels from the same quantity category

In figs 2.4.8–2.4.11 our results on actual vowel durations are summarised. For our hypothesis of only two degrees of vowel quantity to be acceptable we should be able to explain the differences in actual durations between vowels of the same quantity from the workings of the speech apparatus. The best way to do this would be to provide a model of the speech apparatus from which the durational differences would automatically follow. We have no such model, however. Instead we will try to show that the differences found are not in conflict with predictions of a partial model concerning degree of opening and with common-sense observations of the vowel production.

It is known that in many languages the degrees of vowel height or vowel opening affects the vowel duration (for a discussion of the relevant literature see Elert, ref. 15, p. 122). For a bilabial environment this phenomenon has been explained by Lindblom<sup>41)</sup> in terms of a quantitative model of lip mandible coordination. This model, which accounts quite accurately for Lindblom's measurements, attributes the longer duration of more-open vowels largely to the sluggishness of the jaw. It simply seems to take more time to open the jaw (and lips) further than to open them less. The effect of vowel height can partly

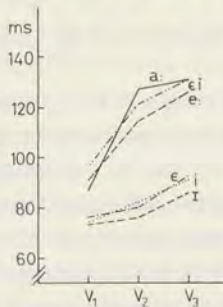


Fig. 2.4.8. Durations in ms of long and short vowels in [pVpVpVp] words. Averaged over three subjects.

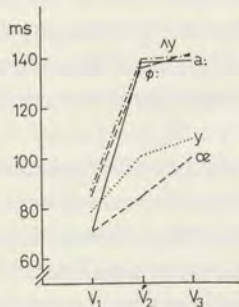


Fig. 2.4.9. As fig. 2.4.8.

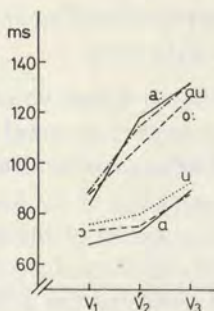


Fig. 2.4.10. As fig. 2.4.8.

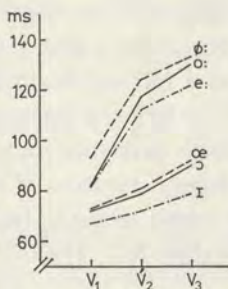


Fig. 2.4.11. As fig. 2.4.8.

explain our data. In most positions the lowest vowel [a:] has a duration which is longer than or at least as long as the duration of the other long vowels. Note that this does not hold for the [a:] duration in the short unstressed first syllable. There the [a:] is always shorter than the other long vowels. We assume that in such short syllables the realisation of [a:] which demands a much greater articulatory movement to reach its target than the other vowels do, is more affected than the realisation of other vowels. Tendencies of this kind could be responsible for such well-known effects of vowel neutralisation in words like Dutch *banaan*: [ba:na:n] → [bana:n] or [bəna:n]. Our data suggest that such neutralisation may be more frequent for [a:] than for other vowels.

We also see in our data that the [ɛi] has consistently longer durations than [e:]. This too can be explained by the effect of vowel height. Among the short vowels we see that [ɛ] has longer durations than [I] and this also can be due to vowel height.

There are a number of vowel-pair differences which are not consistent with the effect of vowel height. We assume that in these cases the effect of vowel height is counteracted by other effects. We notice that [i] has longer durations than [I] and in the stressed position also than [ɛ], although [i] is the highest of these three vowels. Evidently in this case vowel duration is not directly dependent on the extent of movement of the lower jaw. We assume that the

spreading of the lips, necessary to pronounce an optimal [i], has interfered with the effect of jaw opening and changed the moments of opening and closing of the lips as measured by a lip contact.

In fig. 2.4.11 we may study the systematic effects of lip rounding and frontness vs backness. We see that round [ø:] and [œ] are considerably longer than their non-round counterparts [e:] and [I] respectively.

This effect of lip rounding can be understood as follows. The lips protrude somewhat in rounding and form more or less a circle. They keep this protruded form also during the complete closure interval of the [p], thus being more compressed during [p] closure adjacent to round vowels than adjacent to other vowels. In the opening phase of the vowel movement this will lead to a somewhat earlier moment of lip separation as measured in the middle of the lips. In the closing phase it will lead to a somewhat later moment of lip contact in the middle of the lips. Thus the relatively long duration of round vowels in bilabial environment can be explained as a result of coarticulation between lip closure and lip rounding.

In fig. 2.4.11 one also sees that the front round vowels [ø:] and [œ] are longer than the back round vowels [o:] and [ɔ] respectively. This presumably is due to the fact that in the front round vowels the lips are more protruded than in the back round vowels.

The data in fig. 2.4.10 have to be explained by the combined effect of vowel height and lip rounding. In the short vowels lip rounding seems to prevail. The effect seems to become stronger as the vowels become higher. This is consistent with the observation that lip protrusion in round vowels is stronger for high vowels than for lower vowels.

Again, in fig. 2.4.9 the combined effect of lip rounding and vowel height seems to be present. For the long vowels these effects seem to be balanced, whereas for the short vowels the effect of extreme lip protrusion in the vowel [y] which is round, front and high has increased the duration considerably, in this case leading to extensive reorganisation of the articulatory timing of opening and closing of the lips.

The above interpretation of our results suggests that vowels having the same quantity specification would have had the same duration when in the same syllable position, but for the coarticulation with the surrounding consonants. Difference in quality thus leads to difference in duration by the effects of coarticulation which in our view takes place on a lower level than where quantity is specified. Our data show that the durational effects of coarticulation of vowel quality with bilabial closure are not in all cases purely mechanical in nature. At least in some cases there is considerable reorganisation of the timing of opening and closing movements. The present data offer no way to find out whether this reorganisation results from feedback mechanisms in the ongoing control of speech or is preprogrammed on some programming level of motor

control. The existence of this reorganisation in our data, although interesting in itself, means that we did not completely succeed in making the underlying specifications of quantity transparent in the measurable articulatory durations. The degree to which we succeeded is sufficient, however, to lend support to the hypothesis that there are only two possible specifications of vowel quantity on a level of mental programming which would correspond to the linguistic level of phonetic representation.

#### 2.4.2.4. The durational behaviour of the schwah

In fig. 2.4.12 our data concerning the schwah are summarised. The values are averaged over two subjects. These data show that the articulatory timing of the opening and closing movements in the stressed syllable is affected by the choice of the vowel in the unstressed syllables. All three vowels [a:], [a]

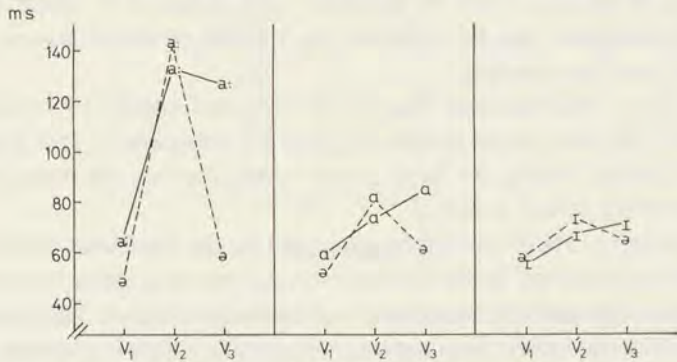


Fig. 2.4.12. Durations in ms of the vowels in [pa:pɑ:pɑ:p] and [pəpɑ:pəp]; [pɑpɑpɑp] and [pəpɑpəp]; [pɪpɪpɪp] and [pəpɪpəp]. Averaged over two subjects.

and [ɪ] have a longer duration when the unstressed syllables contain schwachs than when they contain vowels which are identical to the stressed one. Thus the units in which reorganisation of articulatory timing due to interaction of adjacent segments can take place are larger than a syllable. This is confirmed once again by the durations of the schwachs: when the stressed vowel is an [a:] these durations are shorter than when the stressed vowel is an [ɑ] and when the stressed vowel is an [ɑ] these durations are shorter than when the stressed vowel is an [ɪ]. Thus, here also, reorganisation of articulatory timing seems to take place over at least 2 syllables.

When we now compare the durations of the schwah with those of the short vowels [ɑ] and [ɪ] we see that in nearly all cases the schwah is shorter. Although these differences are not reflected in the total word durations, it cannot be assumed that these differences are purely mechanical and do not involve reorganisation of timing. As we have seen above the differences between the schwah and the other vowels in the same position are partly compensated in the stressed



vowel and this presupposes extensive reorganisation. This suggests that the programmed duration of the schwa is shorter than the programmed duration of the other vowels.

We cannot be sure, however, that this is due to a difference in the specification of quantity and not to other factors. The reorganisation found may be the result of the fact that in the words with [ə] the configuration of the vocal tract has to be changed extensively from the first to the second and from the second to the third syllable, whereas in the other words this configuration need not be changed except for the bilabial closure.

Thus, whereas these data on the durational behaviour of the schwa demonstrate an interesting case of reorganisation of articulatory timing, they do not show whether in a linguistic description the [ə] should be assigned the same or a different specification for quantity as the short vowels. This also leaves open the question whether or not the [ə] should be considered a separate phoneme.

It has been claimed that the [ə] and the [œ] have the same features and thus should be considered to constitute one phoneme (De Groot, ref. 42, pp. 173–174). By others it has been claimed that [ə] and [œ] are different phonemes (Cohen et al., ref. 33, pp. 18 and 55). De Groot argues that, where there is a difference between the two vowels, this should be attributed to a difference in stress level. Cohen et al. argue that as this pair of vowels constitutes the only case in which such a difference in stress levels is needed in order to keep their realisations apart it seems more elegant to consider these vowels to constitute separate phonemes. Our data seem to support this idea. On the other hand we have not compared [ə] and [œ] in identical conditions. In order to do so we should have used e.g. a pair of words as [rœpœpœp] and [rəpœpəp]. It seems, however, intuitively difficult and rather artificial to pronounce these two words in a different way: the subject feels embarrassed when asked to do this. This leaves unsettled the problem of whether or not [ə] and [œ] should be assigned the same features.

#### 2.4.2.5. Quantity and phenomena of contact

Earlier we noted that the [p]-closure durations in the words with short vowels are somewhat longer than in the words with long vowels. The data on this are summarised in fig. 2.4.13. The values given there are averaged over three subjects and 4 experimental runs for each. They are derived from tables 1.A–4.A. The difference between the initial [p] before short and long vowel is about 5 ms. The difference between the other [p]s, including the final one, is about 10 ms for the long- and the short-vowel words. We conclude that the effect of the quantity difference on the postvocalic consonant is more important than that on the prevocalic consonant. We will for the moment concentrate on this effect on the postvocalic consonant.

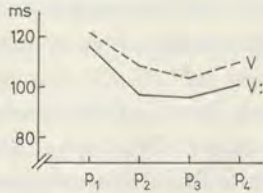


Fig. 2.4.13. Durations of [p] closures in ms in words with long vowels as opposed to words with short vowels. The data are averaged over three subjects and four experimental runs for each.

This inevitably leads to a discussion of the problem of “fester und loser Anschluss” (close and loose contact), or “scharfer und schwacher Silbenschnitt”, or “checked vs free vowels”. These are different names for the same phenomenon, viz. that there seems to be a difference between the transition from the short vowel to the following consonant and the transition from the long vowel to the following consonant.

This phenomenon has been described by Sievers (see Fischer-Jørgensen, ref. 39, pp. 138–139) in relation to the “Silbengipfel” (syllable crest). The consonant following the short vowel comes immediately after this syllable crest and the consonant following the long vowel comes when the vowel has become somewhat weaker.

Jespersen (ref. 43, pp. 202–206) gave the following description of this phenomenon: “Kommt er (der Konsonant) schnell und bricht den Vokal in dem Augenblick ab, wo dieser am kräftigsten gesprochen wird, so haben wir ‘festen Anschluss’ . . . wenn er dagegen erst eine Zeit nach der kräftigsten Aussprache des Vokals kommt, wenn der Vokalklang also schon vor Eintritt des Konsonanten geschwächt ist, so haben wir ‘losen Anschluss’ ”.

It has been suggested that in Dutch there is no quantity opposition but rather an opposition in “contact” or “Silbenschnitt”. (Van Wijk, ref. 44, p. 39; Trubetzkoy, ref. 45, p. 234). The main source of this hypothesis seems to be Jakobson’s statement “Die monotonsche Tonstufenkorrelation kann nicht mit der Quantitätskorrelation der Vokale im selben phonologischen Plan eines Sprachsystems koexistieren” (ref. 46, p. 135). In languages with free word stress, such as Dutch, the stressed phonemes are lengthened, according to Jakobson. This apparently led him to the idea that in those languages quantity cannot be used as a phonological opposition.

Van Wijk has also had the idea that the real phonological opposition was one in intonation and with intonation he meant the “movement of sound in the syllable” (ref. 47, pp. 9–10 and *passim*).

In view of the experimental evidence to be discussed below there seems to be no solid reason for still accepting Jakobson’s statement.

Fliflet <sup>48</sup>) showed that the perception of “contact” in a number of languages

depends on durational characteristics. By shortening long vowels "loose contact" became "close contact". By shortening postvocalic consonants "close contact" became "loose contact".

An experimental approach to the problem of contact in German can be found in Fischer-Jørgensen<sup>49</sup>). She did extensive measurements on differences in air stream, subglottal pressure, labial pressure, duration of vowel and consonant segments and auditory perception between cases of "fester" and "loser Anschluss". It may be worth while to quote her conclusion in full: "In dem untersuchten Material deutscher Wörter haben die Konsonanten nach kurzem Vokal im allgemeinen einen stärkeren Luftstrom bei der Implosion, einen höheren intraoralen Luftdruck, einen stärkeren Organdruck, und eine längere Dauer. Der stärkere Luftstrom hängt wahrscheinlich mit der Silbendauer zusammen, und der Luftdruckunterschied ist sehr klein. Am wichtigsten ist demnach der Unterschied in bezug auf Organdruck und Dauer, der auf eine grössere Konsonantenintensität nach Kurzvokal deutet. . . . Für die Perzeption des Anschlusses scheint die Vokaldauer entscheidend zu sein, die Konsonantendauer und der, auch akustisch sehr unregelmässige, Intensitätsverlauf des Vokals scheinen keine Rolle zu spielen. Die Bedeutung der Konsonantintensität bedarf nähere Untersuchungen" (ref. 49, p. 163).

Thus, according to Fischer-Jørgensen's results the vowel length seems to be the most relevant cue in perception. It seems reasonable, then, to suppose that the accompanying phenomena in the transition of vowel to consonant result from the implementation of vowel quantity in the production. We postulate that the higher intensity and longer duration of consonant articulation after short vowels is the result of the speaker's effort in making the vowel short without producing the vowel in a sloppy way. A characteristic of short vowels may be that they are stopped abruptly.

This idea is confirmed by the outcome of a perceptual experiment of Cohen, Slis and 't Hart and by some electromyographic results obtained by Slis.

Cohen et al.<sup>35</sup>) had subjects adjust the duration of maximum amplitude and the duration of the decay time of the amplitude envelope of synthetic vowels. It was found that not so much the overall duration of the vowels seemed to be relevant, but rather the duration of the decay time. The short vowel had to be heard as stopped abruptly.

Slis<sup>50</sup>) found a higher electromyographic activity in the lip-closing movements in [p] and [b] following a short vowel than in [p] and [b] following a long vowel. Thus a stronger closing command seems to follow the short vowels, which accounts for the increased air stream, subglottal pressure and labial pressure found by Fischer-Jørgensen. Furthermore Slis found that a higher emg activity is correlated with an earlier peak in the emg activity. This indicates that a stronger command is advanced in time compared to a less strong command. This accounts for the longer durations of consonants following short

vowels. The articulatory closure comes earlier in time and so increases the duration of the closure.

The experimental evidence concerning the phenomenon of contact indicates that all phenomena which are traditionally described under the heading of “fester” and “loser Anschluss” or “scharfer” and “schwacher Silbenschnitt” or “checked” and “free” vowels result from the implementation of the feature of quantity. We explicitly assume that short vowels are programmed to be stopped more abruptly than tense vowels in order to make them sound short.

#### 2.4.2.6. Results for the [tVtVtVt] words

One of our reasons for working with vowels in a bilabial environment was that the durational effects of coarticulation in such an environment would supposedly be less than e.g. in a dental environment. It may be of interest to see whether reorganisation in a dental environment is indeed more extensive than in a bilabial environment and to what extent this interferes with the realisation of vowel quantity. To do this we will restrict our considerations to the total word durations of the [tVtVtVt] words. The differences between the two subjects and other inconsistencies in the results concerning segment durations have led us to doubt whether the palatal contacts, which should indicate beginning and end of [t] closure, have worked reliably in these measurements. The relative effect of systematic errors in measuring the moments of opening and closing of the tongue is great for the segment durations but only slight for the total word durations.

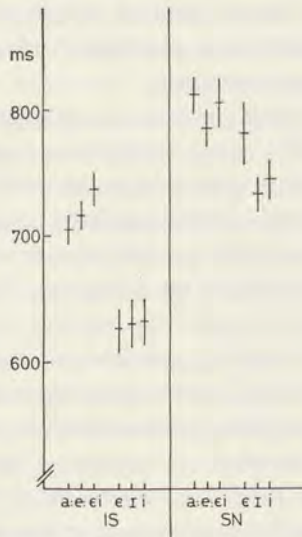


Fig. 2.4.14. Total word durations per experimental run for [tVtVtVt] words. Arithmetic means and estimated standard errors indicated.

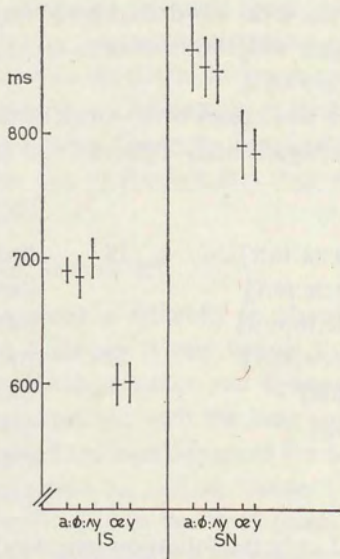


Fig. 2.4.15. As fig. 2.4.14.

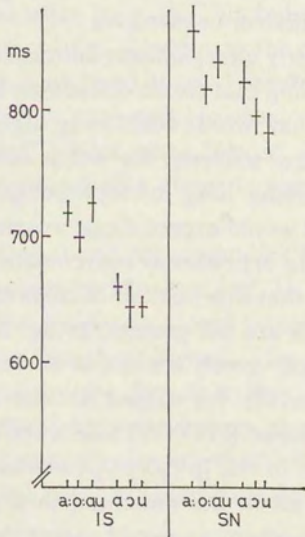


Fig. 2.4.16. As fig. 2.4.14.

In figs 2.4.14–2.4.16 the total word durations are graphically represented per experimental run, again with the arithmetic mean and the estimated standard error defined as  $m \pm s t/\sqrt{n}$ .

If we apply the *t*-test to word pairs with vowels of the same category we find the following pairs to be significantly different at a level of  $p < 0.02$ :

	for list 6		for list 8
IS:	[teiteiteit] > [ta:ta:ta:t]	IS:	[ta:ta:ta:t] > [to:to:to:t]
	[teiteiteit] > [te:te:te:t]		[tautautaut] > [to:to:to:t]
	[ta:ta:ta:t] > [te:te:te:t]		[tatatat] > [tututut]
SN:	[teiteiteit] > [te:te:te:t]	SN:	[ta:ta:ta:t] > [to:to:to:t]
	[tetetet] > [tftftft]		[ta:ta:ta:t] > [tautautaut]
	[tetetet] > [tititit]		[tautautaut] > [to:to:to:t]
			[tatatat] > [tototot]
			[tatatat] > [tututut]

At the level of  $p < 0.1$  only the following pair was found to be significantly different:

	for list 8
IS:	[tatatat] > [tututut]

In all, there are 15 out of 32 possible pairs that reach a level of significance. Earlier we have seen that for [pVpVpVp] words there were 12 out of 66 possible pairs reaching a level of significance. Thus indeed it seems to be so that the effect of the coarticulation on articulatory timing is stronger for the dental consonants than for the bilabial consonants.

One may notice that nearly all significant differences in total word durations can be explained by assuming that words containing low (or open) vowels have a tendency to be longer than words containing high (or close) vowels. There seems to be a second weaker tendency for words containing diphthongs to be longer than words containing long monophthongs. Both tendencies are in accordance with what one would expect if one assumes that covering a greater distance takes more time in articulatory movements.

It is surprising, though, that in a number of cases these lengthening effects of diphthongs and low vowels are not present. In fig. 2.4.14 we see that for subject IS the words with open vowels are not at all longer than the close-vowel words. In fig. 2.4.15 we see that for subject SN the diphthong does not lead to increased duration. In these [tVtVtVt] words the articulatory timing seems to be less predictable than in the [pVpVpVp] words and, although the differences which do exist are all in the same and in a predictable direction, the absence of such differences where we would expect them seems to indicate that the natural tendencies may be counteracted by other tendencies, e.g. a desire

for keeping total word durations constant. Such tendencies may lead to a reorganisation of articulatory timing which counteracts the reorganisation due to coarticulation. On the whole the dental environment seems to be less suitable for studying the effect of quantity differences, at least with the present way of measuring articulatory durations. It may be the case that the inconsistent behaviour of our subjects was due to the fact that they were not sufficiently used to the artificial palate.

#### 2.4.2.7. On [u, y, i] also before [r]

Traditionally there has been a difficulty in classifying the vowels [u, y, i] for Dutch phonologists. Although it was known that these are phonetically short except before [r] (Zwaardemaker and Eykman, ref. 38, pp. 298–299), phonologically they were classified with the long vowels (e.g. De Groot<sup>36</sup>). Several phonetic arguments have been advanced for doing this. Thus De Groot described the long vowels plus [u, y, i] as “helder” (clear) against the other short vowels which were thought to be “dof” (dull). As Moulton (ref. 18, p. 299) pointed out, this was merely finding impressionistic acoustic labels for an intuition and not a phonetic argument. Van Wijk (ref. 44, p. 39) considered the long vowels plus [u, y, i] to be “zwak gesneden” (schwach geschnitten, having loose contact) and the other short vowels “scherp gesneden” (scharf geschnitten, having close contact). Van Haeringen (ref. 51, p. 160) already noted that this hypothesis does not seem to apply, as the contact between the short vowels and the following consonant is not found to be different from the contact between [u, y, i] and the following consonant. Our own data show that the longer duration of the [p] closure after short vowels compared to long vowels is equally present after [u, y, i]. This indicates that according to the phonetic criterion of contact [u, y, i] belong to the class of short vowels. Cohen et al. (ref. 33, pp. 12–18) considered [u, y, i] together with the long vowels to be tense, whereas the other short vowels are considered lax. Just as De Groot’s criteria “helder” and “dof”, these terms “tense” and “lax” are simply labels for an intuition as no demonstrable phonetic meaning has been attached to them.

The difficulties in finding sound arguments for the traditional classification of [u, y, i] have been resolved by Moulton (ref. 18, pp. 311–312), who explains the phonologist’s intuition by distributional criteria. Distributionally [u, y, i] behave like long and not like short vowels. The distributional behaviour of [u, y, i] can be explained by the fact that they have developed from long vowels, which as a result of historic changes at a given stage had no short correlates and thus could easily develop short allophones in most environments. Actually it seems to be the case that they have done so in all positions except before [r].

In order to see whether indeed the vowels [u, y, i] before [r] have the same

effect on articulatory timing as the long vowels, we decided to measure articulatory durations in words containing long vowels, short vowels and [u, y, i] before [r]. As the [r] in Dutch in the postvocalic position in the speech of our subjects is either some diphthong-like change in the vowel or something which approaches a velar or uvular fricative, we are not able to determine the beginning or end of the [r] itself in our articulatory measurements. Therefore we decided to measure the duration of the whole complex of vowel plus [r] before a [t]. We constructed words of the form [pVrt]. The results are summarised in figs 2.4.17 and 2.4.18. It may be seen that [u, y, i] do not differ in their

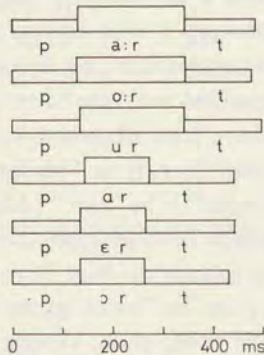


Fig. 2.4.17. Durational build-up of words of the form [pVrt]. Averaged over two subjects.

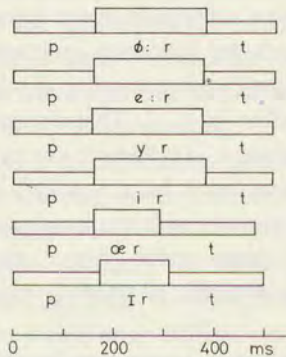


Fig. 2.4.18. As fig. 2.4.17.

effect on articulatory timing from the long vowels, whereas the short vowels have a very different effect on timing. Note also that the [t] closures after short vowel plus [r] are about 40 ms longer than those after long vowel plus [r].

This may mean that the effect of vowel contact is present also for the sound groups [Vr] and [V:r]. Perhaps these sound groups may be considered units in the programming of duration.

In these measurements we again find that according to the programming of



duration, Dutch vowels fall into two groups, but this time [u, y, i] belong to the group of long vowels. Thus in this environment these vowels should be assigned the same degree of quantity as the long vowels.

#### 2.4.2.8. On the diphthongs [ɛi, ʌy, ɑu]

Dutch phonologists do not agree on the question whether the three diphthongs [ɛi, ʌy, ɑu] are to be described as one or as two phonemes. So Van Wijk (ref. 44, pp. 41–42) and Van den Berg (ref. 52, pp. 22–23) describe them as single phonemes, this being most in line with the native speaker's intuition. Moulton (ref. 18, p. 297) and Cohen et al. (ref. 33, pp. 27–30) prefer a biphonemic description. One may notice that a biphonemic solution is in agreement with the bisegmental description of diphthongs and tense vowels in generative phonology.

In a recent article Cohen (ref. 53, p. 288), reviewing the available evidence, stated that the biphonemic solution is unsatisfactory "since:

- (a) it does not seem 'natural' to naive native speakers,
- (b) it introduces a class feature distinction between first and second components which cannot be supported by phonetic evidence,
- (c) it does not explain why e.g. in errors of speech two segments together are always involved, whereas in the case of other close knit units, such as consonant clusters, individual members of these groups are found to play a part".

Argument (b) refers to phonetic evidence of several kinds. In acoustic measurements it was found that in diphthongs the change in formant frequencies, particularly the  $F_1$ , takes place throughout the duration of the sound, and in a well-defined direction. In the long vowels [o:, ø:, e:] the change in the frequencies of the first two formants takes place only towards the end of the vowels (Mol <sup>54</sup>), Koopmans-Van Beinum <sup>55</sup>). In perceptual experiments with synthetically generated speech-like sounds it was found that the direction of change in the  $F_1$ - $F_2$  field is more limited for the diphthongs than for the diphthongised long vowels. On the other hand the formant frequencies of the initial component show much less variability for the long vowels than for the diphthongs (Slis and Van Katwijk <sup>56</sup>). In both the acoustic measurements and the perceptual experiments it was found that the direction of the diphthong movement is much more important than the end point of the movement. These results are in line with those of Gay <sup>57-59</sup>) who from both acoustic measurements and perceptual experiments concluded that [a<sup>l</sup>, a<sup>u</sup>] in American English are "unit phonemes in that the gliding movements of each are not compatible with those of a vowel plus a vowel or vowel plus semivowel sequence in terms of either frequency course or duration".

Cohen's argument (c) is derived from studies of errors of speech such as Cohen <sup>7</sup>), Nooteboom <sup>8</sup>), Fromkin <sup>10</sup>). In those studies it was found that

single phonemic segments may take each other's places as in the following example: *leest niet* → *liest neet* ([le:st nit] → [list ne:t]). Numerous examples have been found in which diphthongs are interchanged with other vowels, such as e.g. *aan en uit* → *uin en aat* ([a:n en ^yt] → [^yn en a:t]). None have been found in which components of diphthongs are interchanged with other vowels. This suggests that diphthongs have to be considered single units.

The description of the diphthongs as single units, which nevertheless seem to be phonetically different from the long vowels, is scarcely possible within the framework of present-day phonetic theory. This has led Cohen<sup>53</sup>) to suggest that a special feature should be incorporated in phonetic theory to account for the idiosyncratic behaviour of diphthongs: "The type of command to be postulated in the case of diphthongs, ei, ui, au, would be to the effect that a certain trajectory constitutes the single target".

This hypothesis concerning the underlying nature of this kind of diphthongs is supported by the results of our durational measurements. We have seen that the diphthongs adapt to the Dutch vowels as far as their effect on articulatory timing is concerned.

If we assume that underlying the production of a long vowel there is only one linguistic segment, we may equally assume that underlying the production of a diphthong there is only one linguistic segment. One may perhaps argue that it is equally well possible to describe both the diphthong and the long vowel as made up of two segments. Cohen's arguments, however, rather convincingly show that such a solution does not account for the typical differences between diphthongs and long vowels.

One may note that Cohen's suggestion of a trajectory-like command underlying a diphthong is difficult to interpret in terms of a single phonetic feature of the Chomsky and Halle kind. In the phonetic representation more than one feature seems to be needed for describing a segment as a diphthong and at the same time defining the trajectory of the diphthong. There seems to be no self-evident choice of these features. If we think of the mental representations of vowels as points or areas in some perceptual vowel space made up of two or three dimensions for defining vowel quality and one dimension being essentially an internal representation of the time axis, then we think of diphthongs as trajectories through this space with well-defined directions and we may think of long vowels as having some longer extension on the time axis than the short vowels while the vowel quality stays essentially the same. The diphthongised long vowels may be somewhat tilted with regard to one or more of the quality dimensions.

#### 2.4.3. *Summary and conclusion*

The above discussion has shown that vowel quantity is an important determinant of articulatory timing, both in stressed and unstressed syllables. Further-

more it has shown that only two degrees of quantity suffice for explaining the actual articulatory timing in vowel production provided we know the effect of stress and position and of vowel quality and coarticulation with adjacent consonants. Specifically the results have shown that those vowels which do not participate in a phonological-quantity opposition phonetically adapt to either of the categories of vowels which do so. Some doubt remains as to the [ə].

As an additional result we have found that the effect of coarticulation of vowel quality and bilabial closure sometimes does and sometimes does not lead to detectable reorganisation of articulatory timing. In as far it does we may account for this in a model of speech production in at least two ways. This reorganisation may be preprogrammed on the level of programming where motor commands stemming from different adjacent phonemes are integrated within a syllable gesture, or it may result from a feedback from the actual movement of the articulators to this syllabification level. In many cases the effect of such reorganisation is not detectable at all. Those cases particularly enabled us to recover from the total word durations that the vowels and diphthongs of Dutch fall into two categories with respect to their inherent effect on articulatory timing. In those cases there still are differences in the measurable durations of the vowels. These can be explained in terms of the mechanical behaviour of the speech organs given the same timing of commands for lip opening and closing for different vowel qualities.

The effect of reorganisation of articulatory timing due to coarticulation of vowel quality and consonantal closure seems to be much more extensive in the case of [t] closure than in the case of [p] closure. This seems reasonable in view of the fact that the primary articulator in the former case is identical for consonant and vowel, whereas in the latter it is not.

An interesting finding was that reorganisation of articulatory timing due to the choice of segments may in some cases extend over at least two syllables. This throws some doubt on the theory that the syllable is the unit for the integration of subsequent phonemes into complex motor-command patterns (cf. Kozhevnikov and Chistovich, ref. 27, pp. 119-163).

Our finding that [p] closures following short vowels are somewhat longer than [p] closures following long vowels is in agreement with evidence from the literature that consonants following short vowels are pronounced with more force of articulation than consonants following long vowels. This suggests that short vowels due to some implementation rule are programmed to be stopped more abruptly than long vowels. It is also in the [p] closures that [u, y, i], sometimes considered to be "schwach geschnitten", betray themselves as being programmed in the same way as the other short vowels, i.e. as "scharf geschnitten". Before [r], however, these vowels fully adapt to the long vowels.

That the diphthongs fully adapt in their durational behaviour to the long vowels seems in agreement with Cohen's suggestion that both long vowels and

diphthongs should be described as constituting only one segment (this argument loses force, however, in view of the discussion in sec. 4.1). Underlying such a segment one may assume some trajectory through a perceptual vowel space.

## 2.5. The effect of stress and position and of number of syllables in the word on the durations of long and short vowels

In the preceding sections we have shown that as far as the effect of the phonetic feature of vowel quantity is concerned, Dutch vowels fall into two categories, long and short vowels. From the data presented there it also became evident that measurable vowel durations are affected by a number of other factors. In the following sections we will try to reveal some major characteristics of the underlying patterns which control vowel durations in Dutch in monomorphemic polysyllabic words. We assume these underlying patterns to result from an interplay of vowel quantities, stress and position within the word.

### 2.5.1. *The different behaviour of long and short vowels*

There are a number of indications that long and short vowels react differently to variations in stress and position. Stetson wrote about English vowels: "The 'short' vowel is one that cannot be prolonged if the syllable is lengthened; the prolongation must occur in the arresting consonant of the intersyllabic interval". The long vowel, however, can be prolonged (Stetson, ref. 60, p. 104). Statistically the different behaviour of long and short vowels is apparent in the "much greater scatter of values in long syllable nuclei compared to short syllable nuclei" (Lehiste, ref. 19, p. 36). Lehiste found this for Czech. The same is observed for German by Georg Heike (ref. 61, p. 44). Lehiste relates this finding to Trubetzkoy's idea that in a quantity opposition, the short member corresponds to a point in time, while the long member has a length dimension and is stretchable at will (*undehnbar* vs *dehnungsfähig*). About the Dutch long-short opposition it is also traditionally said that long vowels can be prolonged and short vowels cannot (Zwaardemaker and Eykman, ref. 38, p. 296). It may be noted that such a statement as "short vowels cannot be prolonged" should not be taken too literally. It probably must be interpreted as "short vowels generally vary less in duration than long vowels". Short vowels may undergo expressive lengthening, as Moulton noticed (Moulton, ref. 18, p. 298). Furthermore in non-expressive pronunciation short vowels are not always of exactly the same duration in all conditions, as has been shown in sec. 2.4. Thus short vowels, just as long vowels, undergo changes in duration according to stress and position, but the extent of these changes seems to differ.

The above considerations show that, if we want to study the effect of stress and position on vowel duration we must do so for long and short vowels

separately. On the other hand, the measurements in sec. 2.4 make it probable that all Dutch short vowels will behave similarly and all Dutch long vowels, including the diphthongs, will behave similarly. This makes it reasonable to restrict our further measurements for the time being to one long vowel and one short vowel. We have chosen the vowel pair [a:]-[ɑ], because these vowels interfere less with the bilabial closure of surrounding bilabial consonants as measured by means of lip contacts than other vowel phonemes.

### 2.5.2. *Dominant versus non-dominant words*

In the description of Dutch intonation patterns Cohen and 't Hart (ref. 62, p. 184) have felt the need to distinguish between dominant and non-dominant words. Dominant words are in some way felt as important by the speaker and therefore intonationally characterised. Non-dominant words are not so characterised. Typically, words referring to something which has not been mentioned before in the conversation are intonationally characterised, whereas words referring to something mentioned before in the conversation, or which is thought to belong to the knowledge speaker and hearer share, are not intonationally characterised.

A word may be intonationally characterised by means of a pitch accent on the syllable which bears the lexical stress. The pitch accent may be realised as a rise, a fall or a rise plus a fall on the stressed syllable. The stressed syllable of the non-dominant word receives no pitch accent. The intonation contour continues through the non-dominant word without any prominence-lending pitch movement.

The question naturally arises whether the temporal pattern of dominant words is different from that of non-dominant words. Specifically one would like to know whether lexical stress in a non-dominant word is a determining factor for segment duration or not. It might also be the case that lexical stress in non-dominant words is not realised at all. Other aspects of the temporal pattern may also differ between dominant and non-dominant words. To study the temporal patterns of both dominant and non-dominant words one must be able to vary the "dominance" of the word. Although we are not able as yet to predict the dominance from the syntactic and/or semantic structure according to explicit rules, there seems to be no difficulty in using the concept in an intuitive way.

In our measurements we used the following phrase: *de uiting pVpVpVp is onzin* (the utterance *pVpVpVp is nonsense*). One of the syllables of the nonsense word received lexical stress. In the normal pronunciation of this phrase the nonsense word is felt as dominant by the speaker and its stressed syllable receives a pitch accent. This can easily be changed by placing a contrastive accent on the word "onzin". Then automatically all other words in the phrase

become non-dominant and the stressed syllable of the nonsense word loses its pitch accent. In this way we have varied the dominance of the nonsense word.

Below we will first show some acoustical differences between dominant and non-dominant words (2.5.2.1). We will then discuss the durational build-up of three-syllable dominant nonsense words (2.5.3) and ditto non-dominant words (2.5.4). Then we will discuss the effect of the number of syllables in the word (2.5.5). In 2.5.6 we will give a quantitative description of the major characteristics of the underlying patterns which control vowel durations in Dutch.

#### 2.5.2.1. Some acoustical differences between dominant and non-dominant words

To demonstrate some acoustical differences between dominant and non-dominant words a recording was made of the fundamental frequency contour and the amplitude envelope of 12 spoken phrases, viz. six phrases containing the six different nonsense words in dominant position and six phrases containing these nonsense words in non-dominant position. The 12 phrases were selected by taking the first version occurring of each type in the two experimental runs in which the speech signal was picked up by microphone and recorded on tape. The fundamental frequency contours were measured by means of the Trans Pitchmeter of Frøkjær-Jensen.

The output of the Trans Pitchmeter plus the audio signal were recorded by means of a Visicorder, i.e. a UV oscillograph. The pitch curves and the amplitude envelope of the audio signal were retraced by hand. These retracings are shown in figs 2.5.1–2.5.6, each figure showing the curves for one type of nonsense word in both dominant and non-dominant position.

It may be seen that in the stressed syllable of the dominant word the fundamental frequency makes a jump with respect to the preceding syllable. In 4 out of 5 cases the frequency definitely rises during the vowel of this stressed syllable. Where this is not the case the frequency has risen markedly in the unvoiced segment preceding the vowel. It may be noted that the abrupt lowering of the frequency contour which in Dutch immediately follows voiceless plosives interferes with the realisation of the pitch accent.

In the stressed syllable of the non-dominant word the frequency contour in most cases does not differ markedly from the frequency contour in the unstressed syllables. In all syllables of all non-dominant nonsense words the lowering of the fundamental frequency after the voiceless [p] is strongly present. In some cases, e.g. in fig. 2.5.5, the frequency change in the stressed syllable of the non-dominant word is more pronounced than in the unstressed syllables. In these cases the effect of stress on the fundamental frequency is still much less in the non-dominant word than in the dominant word. Perhaps the slight effect of stress in the non-dominant word reflects some inhibited tendency towards pitch accent.

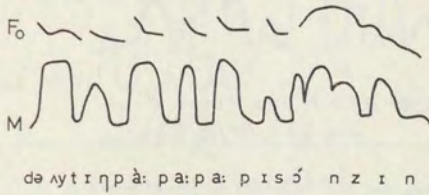
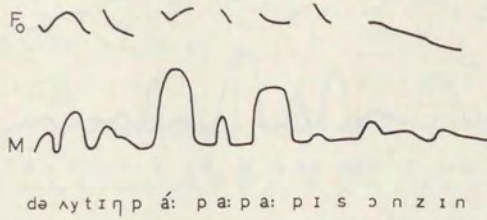


Fig. 2.5.1. Retracings by hand of recordings of spoken phrases containing three-syllable nonsense words in embedded position.  
Top: with a pitch accent on the stressed syllable of the nonsense word.  
Bottom: without a pitch accent on the stressed syllable of the nonsense word.  
F<sub>0</sub>: fundamental frequency as measured with a Frøkjaer-Jensen Trans Pitchmeter; M: amplitude envelope of the microphone signal.

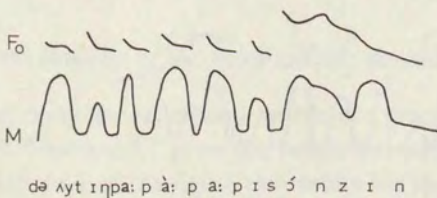
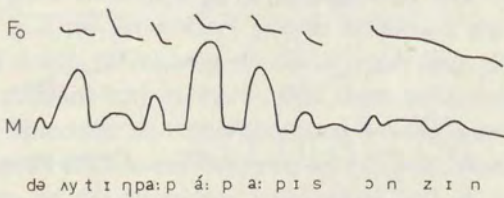


Fig. 2.5.2. As fig. 2.5.1.

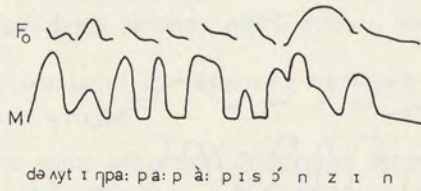
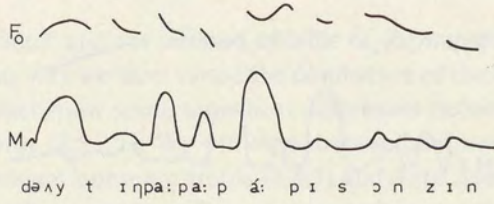


Fig. 2.5.3. As fig. 2.5.1.

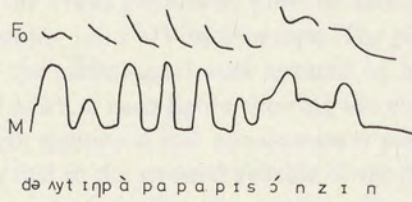
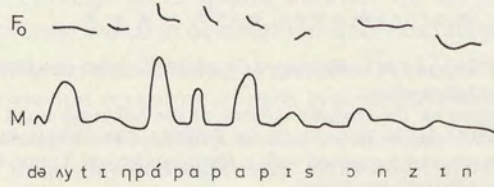


Fig. 2.5.4. As fig. 2.5.1.

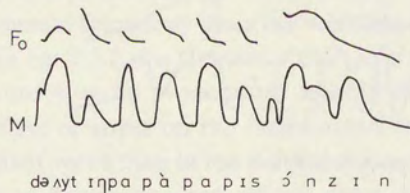
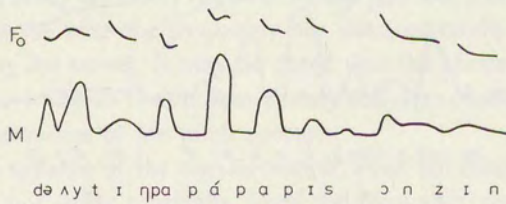


Fig. 2.5.5. As fig. 2.5.1.



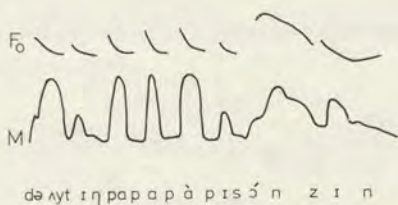
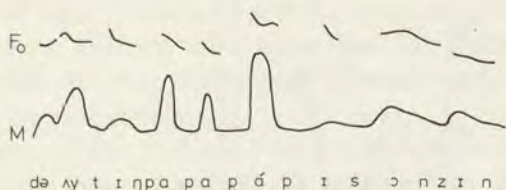


Fig. 2.5.6. As fig. 2.5.1.

That the realisation of a dominant word is markedly different from the realisation of a non-dominant word is not only shown by the fundamental-frequency contours, but also by the amplitude envelopes.

In the dominant words the peak amplitudes of the three vowels differ markedly. In the non-dominant words the peak amplitudes of the three vowels may be considered identical. This holds good for all six pairs of phrases. In the dominant words the stressed syllable has a markedly higher peak amplitude than the unstressed syllables. We may assume that in the realisation of the pitch accent in these cases subglottal pressure was increased. The second syllable shows a very low peak amplitude if unstressed, whereas the first syllable if unstressed shows a very low peak amplitude only if immediately followed by a stressed syllable. It seems as if the articulatory system anticipates the great effort for producing a pitch accent by relaxing somewhat more immediately before.

The above observations may exemplify that there are definite differences in the realisation of dominant and non-dominant words, both with regard to the fundamental-frequency contour and with respect to amplitude envelopes. Below we will concentrate on the durational build-up of both dominant and non-dominant words.

### 2.5.3. *The durational build-up of the three-syllable dominant words*

The results concerning the dominant words are graphically represented in figs 2.5.7 and 2.5.8. Figure 2.5.7 gives the results for subject SN, fig. 2.5.8 for subject IS. In the left part of the figure the results for the words with the long vowel [a:] are given, in the right part the results for the words with the short vowel [a].

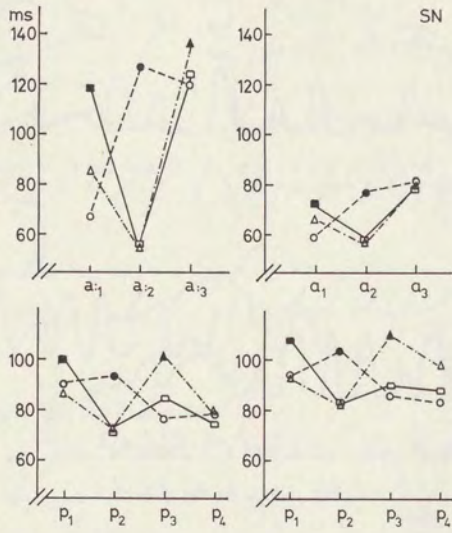


Fig. 2.5.7. Durations of vowel and consonant segments in embedded three-syllable nonsense words with varying stress placement and a pitch accent on the stressed syllable. Top: the vowel durations. The black markings indicate the stressed, and the open ones the unstressed vowels. Vowels belonging to the same word are connected by lines. Bottom: the consonant durations. The black markings indicate the consonant durations preceding the stressed vowels, the open ones the other consonant durations.

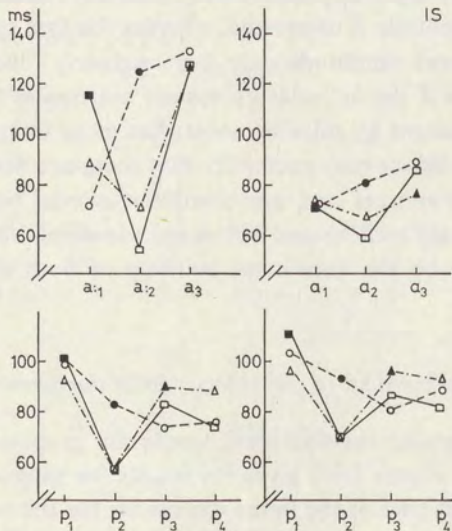


Fig. 2.5.8. As fig. 2.5.7.

The top parts of figs 2.5.7 and 2.5.8 give the vowel durations for the nonsense words with the three different stress placements. Plots taken from the same word are connected with lines. Black markings stand for stressed vowels, open markings stand for unstressed vowels.

The bottom parts of figs 2.5.7 and 2.5.8 give the consonant durations. Here also plots taken from the same word are connected by lines. The black markings stand for [p] durations immediately before stressed vowels, the open markings stand for all other [p] durations.

It should be noted that the horizontal axis does not represent a parameter and that the lines are not curves.

### 2.5.3.1. The durational behaviour of the stressed vowel in the dominant word

As already mentioned, the stressed vowel in the dominant word is characterised by a pitch accent. Typically it is perceptually prominent. From figs 2.5.7 and 2.5.8 it may be seen that the durations of the stressed vowels are somewhat dependent on syllable position. The stressed [a:] in the first syllable is somewhat shorter than in the third syllable. For subject SN this is 18 ms, for subject IS 11 ms. The stressed [a:] in the second syllable seems to take an intermediate position.

The difference between stressed [a] in the first and in the third syllables is 8 ms and 6 ms for SN and IS respectively. For IS the [a] in the second syllable has a slightly longer duration than the [a] in the third syllable. This is only 4 ms. Overall it seems to be the case that the duration of a stressed vowel depends on the number of syllables that come later in the word. The more syllables follow the stressed vowel, the shorter the duration of this vowel.

Lindblom (ref. 63, p. 3) reports for Swedish a considerable increase in duration of the vowel of the final syllable of the word when this syllable is stressed (ca. 50 ms). This increase, he says, is not bound to phrase final position and he suggests that it may be an attribute of the word. We do not find such a considerable increase for the vowel of the stressed final syllable in Dutch. The lengthening of the vowel of the final syllable perhaps is a language-particular property of Swedish, and/or the fact that the vowel of the stressed final syllable of a word is not considerably lengthened may reflect a language-particular property of Dutch. As we will see, the vowel of the unstressed final syllable is considerably lengthened in Dutch, if it is a long vowel.

The difference in duration between stressed long vowel [a:] and stressed short vowel [a], if taken in the same position in the word, is very near to constant. It varies between 46 and 56 ms for SN and between 44 and 50 ms for IS. It seems reasonable to assume that the difference between long and short vowels is optimised in stress conditions.

### 2.5.3.2. The durational behaviour of the unstressed vowels in the dominant word

As we saw above, the duration of the stressed vowels does not depend very much on the position within the word. For the unstressed vowels the situation is quite different. Looking first at the long vowel [a:] it may be seen from figs 2.5.7 and 2.5.8 that the duration varies considerably over the three positions. The most remarkable effect is that unstressed [a:] in the final syllable is very much longer than unstressed [a]. The difference between unstressed [a:] in the initial syllable and in the final syllable is roughly 40–50 ms. The difference between unstressed [a:] in the medial syllable and in the final syllable is roughly 50–70 ms. We tentatively suggest that lengthening of the vowel of the unstressed final syllable is an attribute of the word in Dutch. This would help the listener in dividing an utterance into words.

If we now look at the unstressed short vowel we see that both in the initial syllable and in the final syllable there is not much difference between unstressed and stressed vowel. Only in the medial syllable is there a difference of about 20 ms. It is remarkable that for subject IS unstressed [a] in the final syllable has a longer duration than stressed [a]. This difference is about 10 ms. We may assume that for the short vowel just as for the long vowel there is a tendency to increasing the duration in the unstressed final syllable. This tendency may be counteracted by a demand for keeping a short vowel short. It seems natural that the demand for keeping a short vowel short is stronger in the stressed than in the unstressed position. This would explain why the unstressed vowel may have a longer duration than the stressed vowel. It may be noted, however, that for the same subject a similar situation exists for the long vowel. The [a:] of the unstressed final syllable in the word with stress on the medial syllable is somewhat longer than the vowel of the stressed final syllable. This may be explained by assuming that there is not a general tendency to give a long vowel a long duration, but rather a tendency of giving a long vowel an optimal fixed duration. In unstressed position there are a number of conflicting tendencies which may either lead to a duration that is shorter or to one that is longer than the optimal one. This is in accordance with the idea that vowel duration is controlled by a pair of abstract target values, one for the short and one for the long vowels.

One may also note that there is no fixed difference between stressed and unstressed vowels in our measurements. For example the difference between stressed and unstressed [a:] ranged from 0 to about 70 ms. This is in contrast with the findings of Lindblom for Swedish. He reports a fixed difference between stressed and unstressed vowels (Lindblom, ref. 63, p. 3). It is rather difficult, however, to compare the situation in Swedish with that in Dutch, because in Swedish the phonemically long vowels occur only in stressed posi-

tion. Thus the difference between phonemically long and short vowels is neutralised in unstressed position, whereas in Dutch this is not the case.

In discussing the stressed vowels we saw that the stressed vowel in the initial syllable is somewhat shorter than that in the final syllable, with the stressed vowel in the medial syllable in most cases taking an intermediate position. For the unstressed vowels the picture is different.

The vowel of the unstressed medial syllable is shorter than that of the unstressed initial syllable. Thus for the vowels in unstressed syllables there seems to be an overall pattern which may roughly be described as short, very short, long, for initial, medial and final syllable respectively. This may well reflect some underlying temporal pattern for polysyllabic words.

Looking at the data for initial syllables we see that unstressed vowels directly preceding stressed syllables tend to be shorter than those directly preceding unstressed syllables. This suggests some mechanism in which a stressed syllable presses the preceding one somewhat together. Such a mechanism should be general, however. Note that the effect is not present in the second syllables. There the unstressed vowel preceding the stressed syllable is somewhat longer than the one preceding the unstressed syllable. Looking at the data in a different way, we may note that the unstressed vowels in the words with stress on the third syllable tend to be longer compared to the other unstressed vowels in the same positions. This may reflect a tendency towards equalisation of the word duration, as the words with stress on the third vowels have only one lengthened vowel, whereas the other words have two lengthened vowels, viz. the stressed one and the one in the final syllable. This tendency towards equalisation may either be a result of the particular task in the experiment in which the different words were read in as similar a way as possible, or reflect some more-general tendency in speech. The results of the perceptual experiments to be discussed in chapter 3 of this study, indicate that the latter is the case.

#### 2.5.3.3. The consonant durations in the dominant word

The consonants preceding the stressed vowels, whether long or short, have a markedly longer duration than the consonants preceding unstressed vowels in the same position in the word. This difference varies between about 10 and 25 ms.

It has been suggested (Slis, ref. 50, pp. 181-182) that this increase in consonant duration before stressed vowels results from an increase in articulatory effort accompanying the realisation of stress. Slis found that greater articulatory effort as reflected in greater emg activity is accompanied by an advancement in time of articulatory commands. In this way he explained the increased duration of the consonant preceding the stressed vowel, apparently assuming that the closing command is more affected by the wave of articulatory effort than the opening command. The longer consonant duration may also be explained in a

different way, however. Recently it has been suggested by several investigators that an increase in the duration of a prevocalic consonant may increase the perceptual stress on the vowel (Lehiste, personal communication; Klatt <sup>64</sup>); Huggins <sup>65</sup>). This suggests that the increase of the duration of the prevocalic consonant may rather be part of a perceptual pattern serving the signalling of syllable stress than a lower-level implementation rule of articulatory effort as suggested by Slis. The two explanations are not in mutual conflict, however. The origin of the effect may well lay in some lower-level implementation rule, whereas the acoustical result, i.e. the increased consonant duration, may have taken on the function of signalling syllable stress. Perceptual experiments may throw more light on this function of the consonant duration in Dutch.

#### 2.5.4. *The durational build-up of the three-syllable non-dominant word*

The results concerning the non-dominant words are graphically represented in figs 2.5.9 and 2.5.10. Figure 2.5.9 gives the results for subject SN and fig. 2.5.10 for subject IS. The figures are organised in the same way as figs 2.5.7 and 2.5.8.

##### 2.4.5.1. The vowels in the non-dominant word

As observed earlier, the stressed vowel in a non-dominant word is not affected by a pitch accent, and in this way is different from the stressed vowel in a dominant word. From comparing figs 2.5.9 and 2.5.10 with figs 2.5.7 and 2.5.8 it may be seen that for both subjects the durational behaviour of the stressed

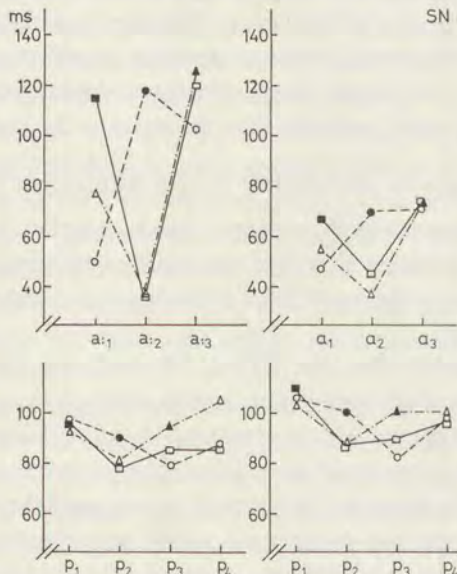


Fig. 2.5.9. As fig. 2.5.7, but without a pitch accent on the stressed syllable.

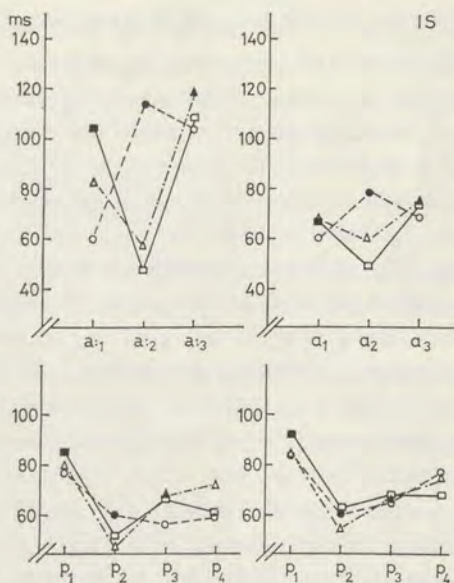


Fig. 2.5.10. As fig. 2.5.9.

vowels is the same for non-dominant words as it is for dominant words. Thus the effect of stress on duration is largely independent of the pitch accent.

In non-dominant words as in dominant words there is a tendency for the stressed vowel to be shorter as there are more syllables following in the word. The difference in duration between stressed long and short vowels seems to be identical for each position.

Under stress conditions the difference between the long and short vowels is optimised rather than that stress leads to an increase in duration, as is normally said.

The durational behaviour of the unstressed vowels is also very much the same in dominant and non-dominant words. The great differences between the unstressed vowels in different positions are no less present in non-dominant words than in dominant words. The unstressed medial vowel is, relative to the other vowels, even shorter in non-dominant than in dominant words. This reflects a difference in tempo.

The relatively long duration of the unstressed final vowel is present in both dominant and non-dominant words.

A comparison of the durational behaviour of the vowels in dominant and non-dominant words thus shows that the temporal patterns of dominant and non-dominant words are very much alike and largely independent of the pitch accent.

#### 2.5.4.2. The consonants in the non-dominant word

Whereas the vowels showed the same durational behaviour in dominant and non-dominant words, the consonants differ somewhat more. In the dominant words the consonants preceding stressed vowels show a considerable increase in duration. In non-dominant words this is much less the case. There still seems to be a tendency in the data for SN towards an increased duration in prestress position, but this is so slight that it cannot have much effect on perception and for IS is even absent. This difference between dominant and non-dominant words suggests that the lengthening of the prestress consonant is not of the same origin as the lengthening of the vowel itself. It particularly indicates that the lengthening of the prestress consonant may be linked up with pitch accent.

From figs 2.5.9 and 2.5.10 it may also be seen that final [p] may have a markedly longer duration when the preceding vowel is stressed than when this is not the case in the words with the long vowel. The increased duration of a consonant following a stressed vowel at the end of the word is somewhat inconsistent. For subject SN in dominant words the final [p] following stressed [a:] had no increased duration, but final [p] following stressed [ɑ] had. In non-dominant words SN's final [p] following [a:] had a markedly increased duration, but final [p] following [ɑ] had not. Subject IS showed an increased [p]-closure duration after stressed [a:] but not after stressed [ɑ] in both dominant and non-dominant words. The effect seems not to be necessary for the temporal pattern of the word, but may perhaps best be considered as an optional way to increase the perceptual prominence of the stressed syllable, if the syllable is immediately followed by a word boundary.

The above discussion of our measurements concerning both dominant and non-dominant three-syllable nonsense words shows that the durational build-up of dominant and non-dominant words is very much alike. We feel justified in assuming the same temporal pattern for dominant and non-dominant words given the same phonemic structure and the same speech tempo. This means that the dominant word is distinguished from the non-dominant word by the superposition of the pitch accent only, and perhaps by a different speech tempo. The pitch accent itself does not change the temporal pattern of the word, although it may have some minor durational side effects, such as an increased duration of the consonant preceding the accentuated vowel. The major characteristics of the temporal patterns will be summarised in sec. 2.5.6 together with results concerning the effect of the number of syllables in the word.

#### 2.5.5. *The effect of the number of syllables in the word*

It is often mentioned in the literature that the duration of a segment depends on the number of syllables in the word to which this segment belongs. The following data, given by Roudet (ref. 66, p. 237), may exemplify this. The



numbers give the duration in centiseconds of the syllable [pa] in words of increasing length.

pâte	27
pâté	20
pâtisserie	14
pâtisserie St. Germain	12

Similar data have been presented for a number of languages by Meyer <sup>67</sup>), Collinder <sup>68</sup>), Malmberg (ref. 69, pp. 10-11), Jones <sup>70</sup>), Lindblom <sup>62</sup>), Lehiste <sup>19</sup>).

It seems reasonable to assume that this effect found for words, in most cases spoken in isolation, is related to or identical with similar effects in longer stretches of speech such as reported by Fónagy and Magdics <sup>71</sup>) and Gaitenby <sup>72</sup>).

A typical interpretation of the data concerning one-word utterances is that given by Lehiste (ref. 19, p. 40): "It appears that in some languages the word as a whole has a certain duration that tends to remain constant, and if the word contains a greater number of segmental sounds, the duration of the segmental sounds decreases as their number in the word increases". With this interpretation in mind it may be interesting to look at the Hungarian example she cites in full from Tarnóczy <sup>73</sup>):

word	1st long	2nd short	3rd short	4th long	5th short
ta:t	210				
ta:tog	180	145			
ta:togot	140	95	115		
ta:togoto:k	120	85	105	130	
ta:togoto:knok	110	80	90	110	140

In this example we see that the duration of the vowel [a:] in the initial syllable is indeed strongly affected by the number of syllables in the word, as expected from Lehiste's interpretation. Note, however, that in the five-syllable word the [a:] in the initial syllable is shorter than the phonologically short [o] in the fifth syllable. This makes one suppose that the vowel duration in the final syllable is not affected or much less so by the number of syllables in the word. This supposition is confirmed by data presented by Lindblom and Rapp <sup>74</sup>).

Lindblom and Rapp found that compensatory adjustment of vowel duration in the stressed syllables of Swedish nonsense words occurs as a function of

word length and position within the word. This compensation was found to be related primarily to the size of the unit that remains to be produced at the beginning of that syllable. The effect of the number of syllables in the word on the vowels which come late in the word was found to be much smaller.

Thus, according to their findings, the major part of the effect of the number of syllables in the word on segment duration is due to the effect of the number of syllables that has not yet been produced. Let us now see whether this effect can be observed in Dutch.

### 2.5.5.1. The effect of number of syllables in words spoken in isolation on the duration of the stressed vowels

If in Dutch the major part of the effect of number of syllables in the word also stems from the part of the word yet to be produced we must have met this effect while discussing the temporal patterns of three-syllable nonsense words. Indeed in sec. 2.5.3.1 we observed that the more syllables follow the stressed vowel, the shorter the duration of that vowel is. This is visualised once again in fig. 2.5.11. The effect is not very strongly present there, however. This may be due to the fact that these words were embedded in a carrier phrase and the vowel durations were rather short to begin with. In the results of Lindblom and Rapp, obtained with stressed long vowels in nonsense words spoken in isolation, the vowel durations ranged from about 350 ms in mono-syllables to about 200 ms when three syllables followed. Such long vowel durations are abnormal in Dutch, also in isolated words. In fig. 2.5.12 an example is provided of the effect of the number of syllables following in a word on the duration of stressed [a:] and [ɑ] in words spoken in isolation. The data were

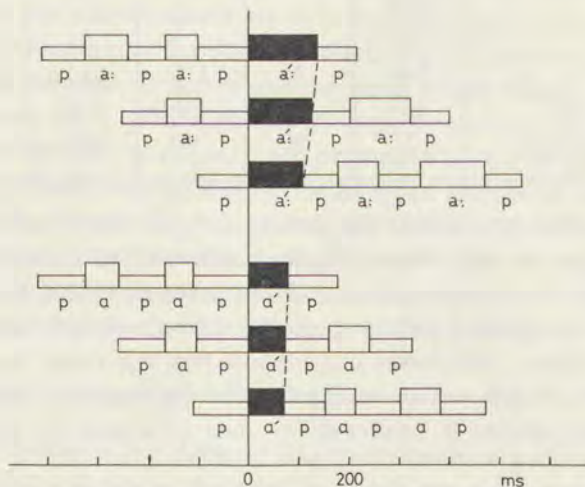


Fig. 2.5.11. Duration of the stressed vowel as a function of its position in a three-syllable nonsense word, spoken in embedded position and with a pitch accent.

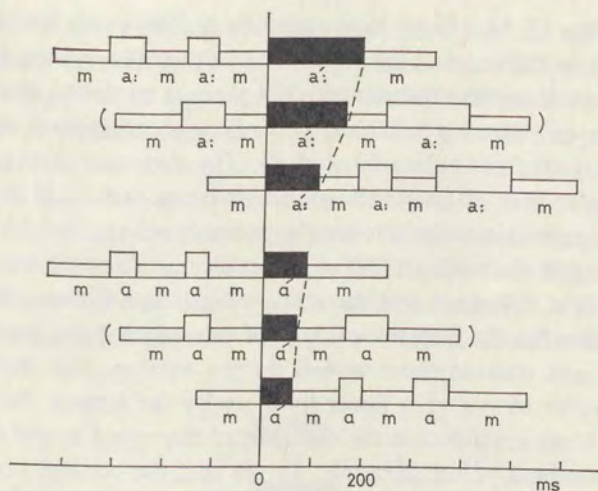


Fig. 2.5.12. Duration of a stressed vowel as a function of its position in a three-syllable non-sense word, spoken in isolation with a pitch accent. The words with stress on the second syllable in (..) were not spoken in the same experimental run as the other two.

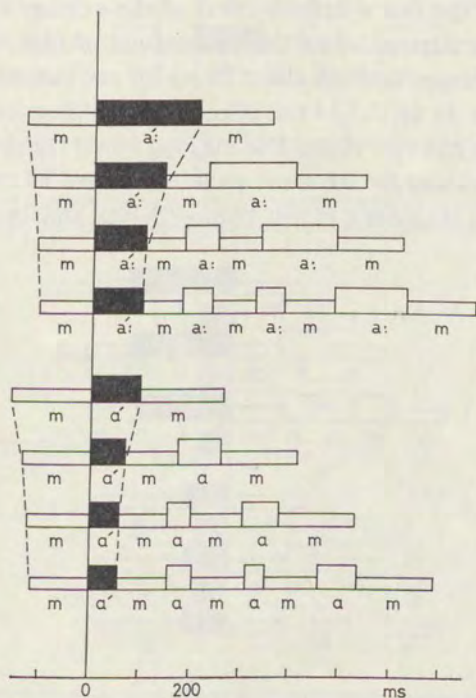


Fig. 2.5.13. Duration of a stressed vowel in an initial syllable as a function of the number of following syllables in the word. The words were spoken in isolation.

taken from tables 13, 14, 15 and 16 in appendix A. The words between brackets are not spoken in the same experimental run as the other words and thus not strictly comparable to the other words. Still there is no doubt about the effect of the number of following syllables. This effect is exemplified once again in fig. 2.5.13, derived from tables 13 and 14. The duration of the long vowel ranges from more than 200 ms in the monosyllable to about 100 ms in the four-syllable word, approximately following a power function. The duration of the short vowel ranges from about 100 to about 60 ms. Thus for the short vowel both the absolute difference and the relative difference between the two positions is less than for the long vowel. It is of interest that the durations of the stressed long and stressed short vowels do not overlap. The duration of the long vowel seems to run to a lower limit set by the longest duration of the short vowel. If we assume that the duration of the vowel in the monosyllable is about the optimal vowel duration, we see that the optimal vowel duration of the long vowel is rather accurately twice as long as the optimal duration of the short vowel. The V/V<sub>1</sub> ratio varies with position.

These data confirm the Lindblom and Rapp findings in that for Dutch also the major part of the effect of the number of syllables in the word seems to stem from the portion of the word that remains to be produced. In the Swedish data there is none the less a definite effect of the number of syllables in the word preceding the stressed vowel on the duration of that vowel. In the final syllable this effect ranges between about 20 ms for one subject and about 50 ms for another subject. In fig. 2.5.14 this effect is exemplified for our Dutch data. It may be seen that it is very slight. For the long vowel the actual durations are 207, 207 and 191 ms and for the short vowel 97, 86 and 86 ms. The differences between the vowel durations in the monosyllables and in the three-syllable

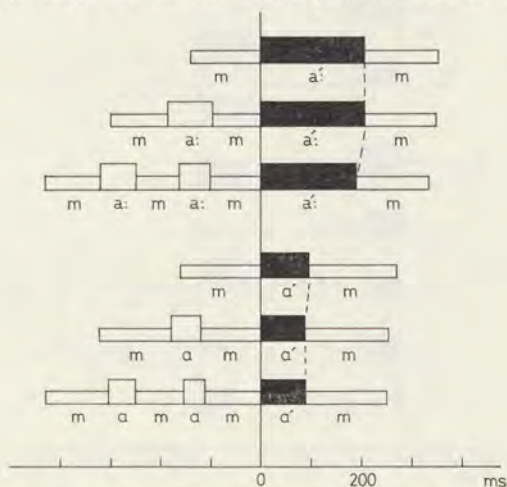


Fig. 2.5.14. Duration of a stressed vowel in a final syllable as a function of the number of preceding syllables in the word. The words were spoken in isolation.

words are significant, however, and this may mean that the same tendency is at work in these data as was more strongly present in the Swedish data.

The shortening effect of the preceding number of syllables is definitely absent in fig. 2.5.15 where the stress is on a non-final syllable and the number of syllables preceding the stressed syllable is varied. Here the duration of the stressed vowel rather seems to have a tendency to increase slightly than to decrease with increasing number of syllables in the word. In fig. 2.5.16, concerned with words with short vowels, there is no significant effect of the number of syllables in the word at all. Thus, the effect of number of syllables in the word on the duration of stressed vowels may be considered to be mainly one

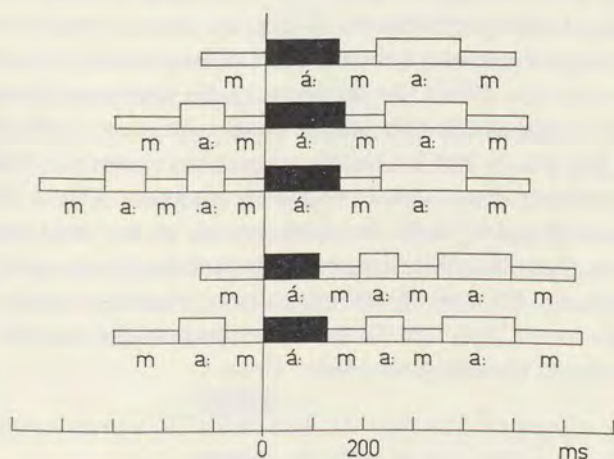


Fig. 2.5.15. Duration of a stressed long vowel in a non-final syllable as a function of the number of preceding syllables in the word. The words were spoken in isolation.

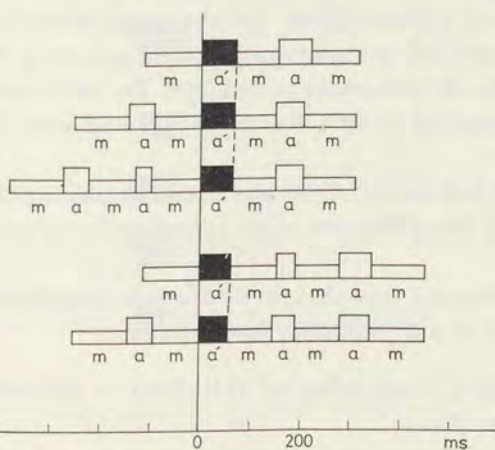


Fig. 2.5.16. Duration of a stressed short vowel in a non-final syllable as a function of the number of preceding syllables in the word. The words were spoken in isolation.

way. The number of syllables following the stressed vowel has a shortening effect on its duration, whereas the number of syllables preceding the stressed vowel has no such effect or a very small effect. The absolute effect in ms and the relative effect in per cent is much less for the short vowels than for the long vowels.

#### 2.5.5.2. The effect of number of syllables in words spoken in isolation on the durations of unstressed vowels

In fig. 2.5.14 it may be seen that the duration of the vowel in the unstressed initial syllables decreases somewhat from the two-syllable word to the three-syllable word. For the long vowel the vowel durations are 90 ms for the two- and 70 ms for the three-syllable word. For the short vowels these durations are 58 and 53 ms respectively. Thus the effect of the number of syllables coming later in the word also affects the durations of the unstressed vowels, but less than those of the stressed vowels. In the words with stress on the penultimate, displayed in figs 2.5.15 and 2.5.16, the respective durations of the vowels in the initial syllable of three- and four-syllable words are 87 and 83 ms for the long vowels and 50 and 53 ms for the short vowels. In the latter case the effect seems to be in the wrong direction, but note that the durations of the preceding consonants are 125 and 120 ms respectively. Apparently the short vowel has run to its lower limit and further shortening of the syllable duration is taking place in the preceding consonant alone.

#### 2.5.5.3. The effect of number of syllables in words spoken in isolation on the consonant durations

In fig. 2.5.13 it may be seen that the number of following syllables not only affects the durations of the stressed vowels but also those of the preceding consonants although to a lesser degree. The shortening is over 100% for the long vowel and only 20% for the consonant preceding it. It is 40% for the short vowel and 25% for the consonant preceding it. For the consonant-vowel combination the shortening is 40% for the long vowel and 30% for the short vowel.

In figs 2.5.14, 2.5.15 and 2.5.16 there is a small shortening effect of the number of syllables on the first consonant in the unstressed initial syllable. This is only a few per cent.

On the whole the data show that the consonants are subject to the same effect as the vowels but to a considerably lower degree.

#### 2.5.5.4. The effect of number of syllables in embedded non-dominant words

The data to be discussed were obtained with nonsense words embedded in a carrier phrase and with no pitch accent on the stressed vowel. In this position

all durations but especially those of the final vowel and consonant are shorter than in the words spoken in isolation, although the lengthening of the final syllable is still definitely present. The data are taken from tables 17, 18, 19 and 20 in appendix A.

In fig. 2.5.17 the effect of the number of following syllables on the duration of the stressed vowel is exemplified. This effect is about 40% for the long vowel and about 20% for the short vowel.

It should be noted that the shortest duration of the stressed long vowel, i.e. the one in the four-syllable word, is not much different from the shortest duration of the stressed long vowel in the words spoken in isolation as displayed in fig. 2.5.13. These durations are 91 and 99 ms respectively. Again the duration of the long vowel seems to run to a lower limit set by the longest duration of the short vowel. Here also there seems to be a strong tendency to keep the durations of the long and short vowels apart.

In fig. 2.5.18 we see a very small tendency of the duration of the vowel in the stressed final syllable to decrease with increasing number of syllables. The greatest difference is 14 ms for the long vowel and only 6 for the short vowel. In figs 2.5.19 and 2.5.20 we see that the small effect of the number of preceding syllables on the duration of stressed vowels is confirmed for stressed vowels in non-final syllables. As one may remember this was not the case for the stressed vowels in non-embedded words.

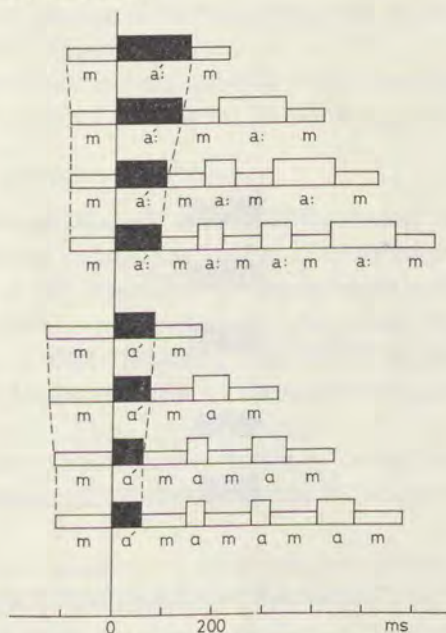


Fig. 2.5.17. Duration of a stressed vowel in an initial syllable as a function of the number of following syllables in the word. The words were spoken in embedded position without pitch accent.

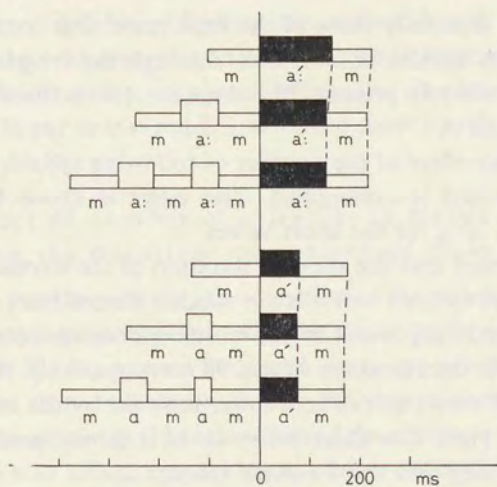


Fig. 2.5.18. Duration of a stressed vowel in a final syllable as a function of the number of preceding syllables in the word. The words were spoken in embedded position without pitch accent.

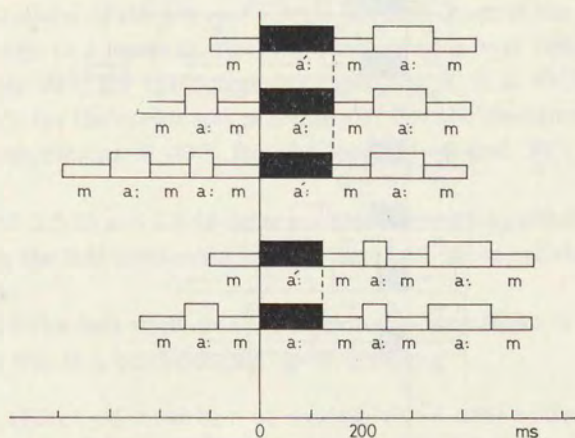


Fig. 2.5.19. Duration of a stressed long vowel in a non-final syllable as a function of the number of preceding syllables in the word. The words were spoken in embedded position without pitch accent.



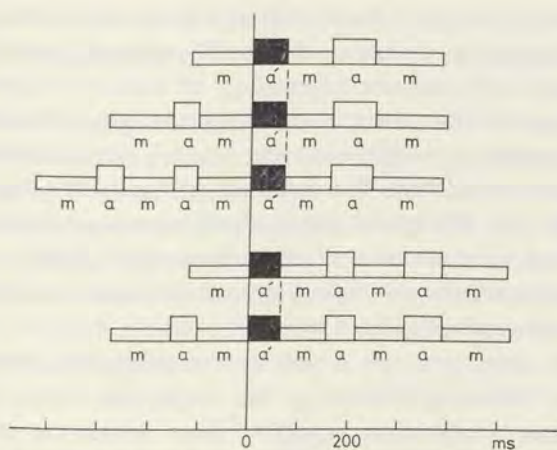


Fig. 2.5.20. Duration of a stressed short vowel in a non-final syllable as a function of the number of preceding syllables in the word. The words were spoken in embedded position without pitch accent.

The data on unstressed vowels in figs 2.5.17, 2.5.18, 2.5.19 and 2.5.20 can be explained by assuming a rather strong effect of final lengthening, a small effect of number of following syllables and a shortening of vowel durations in initial prestress syllables. The latter effect may explain e.g. that in fig. 2.5.19 the first vowel in [ma:ma:ma:m] is shorter than the one in [ma:ma:ma:ma:m] where more syllables follow.

Figure 2.5.17 shows that in the embedded word the consonants also undergo a small shortening effect of the number of following syllables.

### 2.5.6. Towards a quantitative description

The effects as derived from the durational build-up of nonsense words with varying stress placement and varying number of syllables are summarised below.

- (1) The duration of the stressed vowel decreases with increasing number of syllables that remain to be produced in the word. This decrease approximately follows a power function. The effect is stronger for long vowels than for short vowels and stronger for isolated words than for embedded words.
- (2) There is a weak and not-consistent tendency for the duration of the stressed vowel to decrease with increasing number of syllables that precede in the word.
- (3) There is a strong tendency to keep the durations of stressed long and short vowels apart. The duration of the stressed long vowel seems to run to a lower limit set by the longest duration of the stressed short vowel.
- (4) The duration of the unstressed vowel in the final syllable of the word is nearly as long as the duration of the stressed vowel in the same position.

- (5) The unstressed vowel in the final syllable of the word is longer when followed by a pause than when not followed by a pause, although in the latter case there is still a definite lengthening.
- (6) The duration of the vowel in an unstressed syllable is affected by the number of syllables in the word that remains to be produced.
- (7) The duration of the vowel in a word initial unstressed syllable is markedly longer than in a non-initial and non-final unstressed syllable.
- (8) The duration of a vowel in an initial unstressed syllable is shorter when immediately preceding a stressed syllable than when immediately preceding an unstressed syllable.
- (9) Prevocalic consonants are subject to the same durational effect of the number of following syllables as the vowels that follow them but to a lesser degree. Consonants preceding short vowels are somewhat more affected than consonants preceding long vowels.
- (10) Consonants preceding stressed vowels having a pitch accent show an extra increase in duration of about 25 ms.
- (11) Word initial consonants, if not followed by a stressed vowel, are about 15–30 ms longer than non-initial consonants.
- (12) Consonants immediately followed by a pause are about 30–50 ms longer than non-initial consonants in other positions.
- (13) Consonants preceding or following a short vowel are about 10–20 ms longer than consonants preceding or following a long vowel. Immediately before a pause this difference is greatest.
- (14) Consonants preceding the final vowel of a word generally are ca. 10 ms longer than other consonants being not word-initial and not word-final, stress conditions being equal.

In points (9)–(14) “consonants” has to be read as “consonant closures”.

The effect of the number of following syllables in the word has been described by Lindblom and Rapp<sup>74</sup>) with the following formula:

$$V = D/n^{\alpha m/n},$$

in which  $V$  is the vowel duration,  $D$  a constant,  $n$  the number of syllables in the word,  $m$  the number of the syllables counted from the end of the word backwards and  $\alpha$  a factor restricting the effect of  $m/n$ .

They verbalise this rule as follows: “In any given syllable of a word vowel duration is inversely proportional to the effective relative size of the unit that remains to be produced at that point”. For the vowel of the first syllable  $m = n$  and the formula can be written as follows:

$$V = D/m^{\alpha}.$$

This rule fits our results rather nicely.

The elaborated form of the Lindblom and Rapp formula seems to have some disadvantages, however. Firstly, the effect of  $m$  is coupled to the effect of  $n$ , both being controlled by  $\alpha$ . Thus if the effect of  $m$  is great, as in our data, the effect of  $n$  will also be relatively great, which is not found in our data. More seriously, the formula does not predict an increasing effect of  $n$  with increasing value of  $n$ . For instance, if  $D = 200$  and  $\alpha = 1/2$ , the durations of the vowel in a stressed final syllable will be 200, 168, 161, 168 and 171 for a one-, two-, three-, four- and five-syllable word respectively. This is counter-intuitive. It also seems to be in conflict with the Lindblom and Rapp data, although there is much variation between their subjects in this respect.

We propose to change the formula to

$$V = \frac{D}{m^\alpha n^\beta},$$

in which the effect of  $m$  and the effect of  $n$  are separately controlled by  $\alpha$  and  $\beta$  respectively;  $\beta$  is small relative to  $\alpha$ , and has to be smaller for our Dutch data than for the Swedish data of Lindblom and Rapp. In fact, the effect of  $n$  seems to be practically zero in our Dutch data.

To describe the effect of position on the unstressed vowels a few additional rules are needed. The major characteristics of the temporal patterns can be described with the help of the following simple set of rules for vowel durations:

- |     |   |    |                            |
|-----|---|----|----------------------------|
| (1) | $DV = DV_{\text{opt}}/m^\alpha n^\beta$ | if | . . $\bar{V}$ . .          |
| (2) | $DV = DV_{\text{opt}}/1.3^\gamma$       | if | . . $\bar{V}(C)\#$         |
| (3) | $DV = DV_{\text{opt}}/2.2^\gamma$       | if | $\#(\bar{C})\bar{V}CV$ . . |
| (4) | $DV = DV_{\text{opt}}/2.4^\gamma$       | if | $\#(C)\bar{V}CV$ . .       |
| (5) | $DV = DV_{\text{opt}}/3^\gamma$         |    |                            |

In these rules  $DV$  is the vowel duration to be calculated. It is derived from an optimal vowel duration,  $DV_{\text{opt}}$ , plus the rules given.  $DV_{\text{opt}}$  has to be different for long and short vowels. Suitable values seem to be 200 ms for the long and 100 ms for the short vowels.  $m$  is the number of the syllable counted from the end of the word backwards.  $n$  is the total number of syllables in the word.  $\alpha$  is a constant limiting the effect of  $m$ . The value of  $\alpha$  has to be different for long and short vowels in order to describe our data accurately. It may be 0.5 for long vowels and 0.4 for the short ones in describing the vowel durations in words spoken in isolation.  $\beta$  is a constant limiting the effect of  $n$ . Its value has to be very low to describe our data.  $\gamma$  is a constant restricting the extent of the differences between the unstressed vowels. Its value is about 1 for the long vowels and about 0.9 for the short ones.

In the definitions of the conditioning environments  $\bar{V}$  is a stressed vowel,  $V$  an unstressed vowel,  $C$  one or more consonants,  $\#$  a word boundary.

These rules fit the data for words spoken in isolation. The data for embedded words can roughly be accounted for by decreasing the values of  $DV_{opt}$ ,  $\alpha$  and  $\gamma$  with preliminary rules. In figs 2.5.21 and 2.5.22 this is exemplified for the vowels of stressed initial syllables. In fig. 2.5.23 an example is given of the durational build-up of nonsense words as found in the measurements and as generated by our rules. In the latter case fixed values were assumed for the consonant durations. Systematic variations in consonant durations may in principle be described by a similar set of rules.

Evidently the validity of the rules given in this section is very much restricted because they have been found by studying monomorphemic nonsense words with a very simple phonemic make-up. We feel, however, that precisely these limitations may have led us to discover some important characteristics of the underlying patterns which control segment durations and particularly vowel durations in the production of Dutch words. The rules set up for describing these characteristics on the one hand help to generate acceptable sequences of segment durations and, on the other, give a re-definition of the relation between

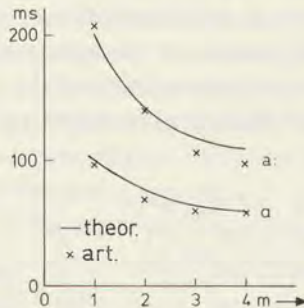


Fig. 2.5.21. Theoretical and measured vowel durations in ms as a function of the number of syllables which remain to be produced in the word. Words were spoken in isolation.  $DV_{opt} = 200$  and  $100$  ms,  $\alpha = 1/2$  and  $2/5$  for long and short vowels respectively.

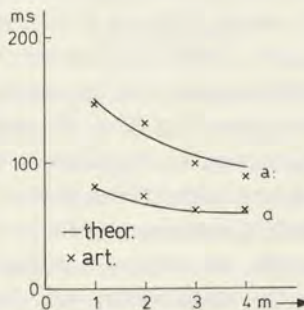


Fig. 2.5.22. Theoretical and measured vowel durations in ms as a function of the number of syllables which remain to be produced in the word. Words were spoken embedded in a carrier phrase.  $DV_{opt} = 150$  and  $80$  ms,  $\alpha = 1/3$  and  $1/4$  for long and short vowels respectively.

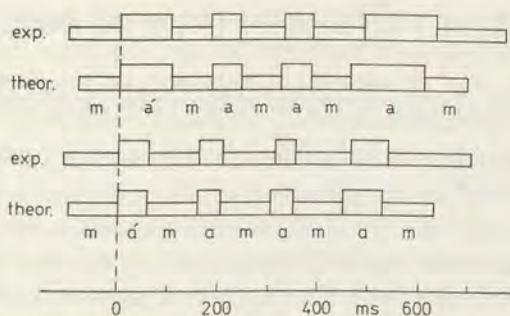


Fig. 2.5.23. Durational build-up of four-syllable nonsense words as measured and as generated by the rules. In the latter versions consonant durations have been kept constant at 80 ms for the word with long vowels and at 100 ms for the word with short vowels.

long and short vowels in Dutch, by describing the effects of stress and position on the realisation of optimal vowel durations.

## 2.6. Summary

The present part of this study was set up as a rather exploratory investigation into the control of vowel durations in speech. It was found that within one experimental session standard deviations from 5 to 10 ms are normal for speech-sound durations and standard deviations may in extreme cases be below 5 ms. This seems to depend on the subject. Although from subject to subject and from session to session for one subject there is some free variation in the durational build-up of identical nonsense words, on the whole subjects seem to adhere to surprisingly fixed temporal patterns. These temporal patterns result from an interplay between phonological quantity, stress, position in the word (and probably other factors). It was shown that not more than two degrees of quantity specification have to be assumed for explaining the durational data, if we assume some additional durational effects of consonant-vowel coarticulation. These coarticulation effects in some cases lead to rather extensive reorganisation of articulatory timing. The domain of such reorganisation can be more than one syllable. The effects of stress and position on vowel durations can be described by some rather simple rules operating on optimal (ideal) vowel durations. For a given speech tempo there are only two such optimal vowel durations, one for the short vowels and one for the long ones. In the stressed positions overlap between durations of long and short vowels is avoided. The duration of the long vowels runs to a lower limit set by the optimal duration of the short vowels. The main effects found were the shortening due to the number of following syllables in the word, shortening of unstressed vowels except in the final syllable of a word, shortening of vowels in prestress syllables and extra shortening of vowels in non-initial and non-final syllables. The same rules essentially apply in dominant and non-dominant words.

### 3. PERCEPTUAL TESTS OF DURATIONAL REGULARITIES

#### 3.1. A method for testing the perceptual reality of durational regularities

Work on speech synthesis-by-rule, in other laboratories as well as in our own, has focussed attention on the perceptual relevance of precise rules for segment duration. In the early work on synthesis-by-rule attention was mainly focussed on producing the correct intrinsic allophones and formant transitions. Even then durational rules for simulating the effect of stress levels were found to be indispensable (Lieberman et al. <sup>75</sup>), Lisker et al. <sup>76</sup>). In later work on synthesis-by-rule it has become increasingly clear that in order to generate reasonably good connected speech rather intricate sets of durational rules, or rather large sets of context-dependent durational allophones are needed (Mattingly <sup>77,78</sup>), Barnwell <sup>79</sup>), Klatt <sup>80</sup>), Slis <sup>81</sup>)). Barnwell specifically notes that a too simple model for segment durations “detracts from the naturalness of the speech but also probably transmits incorrect stress or intonational information to the users of the reading machine” (ref. 79, p. 3). Barnwell worked on a model for segment durations which was meant to be used in a reading machine for the blind, a set-up which takes printed text as an input and delivers connected speech as an output. It is in such a context especially that the need for correct specification of durations is felt. Many such durational rules are never accounted for systematically in linguistic descriptions. Still there seems to be no a priori reason why such durational rules should not be part of the rules of the language. Mattingly <sup>78</sup>) explicitly handles durational rules for vowels depending on “intrinsic vowel length, stress, the phonotactic rules relating to syllable-final voiced clusters, and the rules for prepausal phonemes” as phonological rules. Klatt <sup>64</sup>) wants his durational-rule system to be compatible with other rules of a complete generative phonology of English, although his rules are lower-level rules in the sense that they begin with the output of the Chomsky-Halle rules of phonology.

That for the synthesis of reasonably good speech precise durational rules seem to be indispensable at least suggests that intricate temporal patterns of speech belong to the sound pattern of a language and that a complete account of the implicit knowledge language users possess with regard to the sound pattern of the language, should also contain the rules needed for generating the correct speech-sound durations. It seems appropriate that in an attempt to probe the implicit knowledge of some language users with respect to these temporal patterns use is made of synthetic speech.

Thus we intend to use synthetic speech as a tool in phonological research. The use of synthesis for this purpose was discussed extensively by Lisker, Cooper and Liberman <sup>76</sup>). These authors explicitly suggest that synthesis-by-rule may be used in deriving testable utterances from the phonology of the

language and thus may be used in phonological research in the same way as a linguist in his field work may use his own vocal apparatus "the flexibility and controllability of which is inadequate for the purpose". In the experiments they refer to, however, the task of the subjects is almost exclusively to identify certain linguistic units, which task is somewhat different from the task of informants in linguistic field work. In a more recent article Mattingly<sup>78)</sup> proposes to use synthesis-by-rule in descriptive phonological work with native informants. "Imagine", he writes, "an automatic system, the inputs to which are proposed phonological rules of a language, and a phonemic transcription of an utterance of the language, and the output from which is a synthetic acoustic representation of the utterance. Such a system simulates the phonologist in his generative phase. But it does not make accidental errors, and it applies only rules which have been explicitly stated. A native informant can propose the generation of utterances — or even, if he has learned the transcription system, generate the utterances himself — and report to the phonologist in what respects the synthetic versions are incorrect. In difficult cases, the informant can be invited to compare stimuli produced by alternative versions of the rules differing only with respect to the variable of interest. In the light of the informant's responses, the rules can be revised easily and quickly and the informant can then be confronted with the output of the revised rules".

Thus in Mattingly's proposal the task of the informant is much more complicated than identifying linguistic units. It is also much more complicated than saying whether a particular utterance is acceptable or not. He must state what is wrong with it. Would this proposal work for the kind of durational rules we are concerned with? There are reasons to believe it would not or not well. Mattingly's proposal applies to the case where the informant's language is different from the linguist's language. Where the phonologist describes his own language he can test his rules by generating utterances of his own selection, with his own vocal apparatus or with a synthesiser, and himself judge what is wrong with them. In both cases, however, the assumption is that either the informant or the phonologist is able to specify what is wrong with the utterances, simply by listening or by introspection. The fact that durational rules of the kind we are concerned with have been systematically overlooked in phonological descriptions may indicate that the facts which these rules describe are not easy to specify simply by listening. This may be related to the fact that in most work on speech synthesis, also Mattingly's own<sup>77)</sup>, it is tacitly implied that systematic variations in segment durations as found by analysis of speech produced by humans are also relevant for perception and should be included in the rules for speech synthesis. The same is implied in the work of Barnwell<sup>79)</sup> and that of Klatt<sup>64,80)</sup>. Barnwell tries to find a reasonable model for the relationship between stress, intonation, and the other relevant features and output correlates such as fundamental frequency, pauses, and particularly, segment duration by

“controlled studies of natural speech”. For some reason he did not directly use speech synthesis in the way proposed by Mattingly. The situation is very similar to our own in chapter 2 of this study. It seems to be the case that phenomena in the prosody of speech are not readily available to the introspection of the phonologist. In work on speech synthesis this may lead to the situation that one is very well able to tell *that* something is wrong with the durational properties of the synthesised speech but not to specify *what* is wrong. Therefore one turns to the analysis of natural speech for fresh ideas. This need not imply that a trained phonetician will not be at all able to make reasonable guesses in setting up rules for segment durations. Many rules in current synthesis programmes are found by trial and error. But this does not seem to work all the way towards an optimal set of rules.

Once certain rules for segment duration can be hypothesised from the analysis of natural (= human) speech, how are these rules to be tested? A first test may consist in listening to the output of the rules and judging whether the rules improve the speech or not. But there is a real problem here, in that most synthesised speech sounds unnatural anyway and it is often hard to judge whether one particular rule improves the speech or not.

This situation seems to ask for a perceptual experiment. It is not immediately clear, however, which type of experiment. Perceptual experiments on speech differ in the task the subject has to perform. Perhaps the most common task is that subjects have to identify linguistic units when hearing some acoustic stimulus. This will not do in our case. The recognition of linguistic units does not seem to be immediately dependent on the kind of durational rules we want to test. Another possible task is judging the acceptability of stimuli (e.g. Huggins<sup>65,82</sup>). This could be a possible way of bringing about the subject's internal criteria for segment durations. It was found in preliminary experiments, however, that in using synthetic speech as generated by the available speech synthesiser, some subjects did not find the speech acceptable at all and refused to cooperate in such a task, whereas others, more easy to satisfy, accepted so extensive ranges of segment durations that all relevant differences drowned. Then there is a whole category of types of experiments, in which the subject has to give difference or similarity judgments, which do not seem to be useful for our purpose. To this category belong experiments in which the subject has to judge whether two stimuli are different or not, or which two of three stimuli are the more similar (e.g. Pols, Van der Kamp and Plomp<sup>83</sup>, Terbeek<sup>84</sup>). To this category may also belong scaling methods, if similarity is the criterion for scaling. Scaling for acceptability did not seem a useful method either. A further class of experiments is made up of those in which subjects are asked to match a stimulus, continuously or repeatedly presented, with some reference signal by setting a control or controls (e.g. Cohen and Willems<sup>85</sup>, Cohen and 't Hart<sup>62</sup>). In the case of the durational rules this matching



method at first sight does not seem to be applicable, because it is not at all clear what the reference signal should be. But what we really want to know is whether the durational rules describe some regularities in what the language users know of the perceptual properties of the words of their language. Thus the reference should not be given by some external signal, designed by the experimenter, but rather be provided by the subject's internalised knowledge of the sound pattern of his language. A suitable method then for testing the kind of durational rules we are concerned with may be to ask subjects to match a stimulus to some internal criterion. This formulation perhaps brings to mind a method which has been used for establishing loudness scales, in which the subject hears two tones and is required to change the intensity of one of them until its loudness is a certain ratio of that of the other one (Geiger and Firestone<sup>86</sup>), Stevens and Davis<sup>87</sup>). Here the ratio is an internal criterion. But by using two stimulus tones the situation is different from an experiment in which only one stimulus is presented and the subject is required to change this stimulus according to some internal representation of how this stimulus should sound. Such a method seems to be relatively rare in psycho-acoustic research. It has been used in studies of absolute pitch (Ward<sup>88</sup>). In the study of speech perception the method of matching to internal criterion has been applied by Cohen, Slis and 't Hart<sup>35</sup>) in an investigation of the perceptual characteristics of Dutch vowels. They specifically used it for studying differences in preferred durations between Dutch long and short vowels. They instructed subjects to adjust the control settings of a vowel synthesiser both for the duration of the period of maximum amplitude and for the duration of the decay time "in such a way as to obtain perceptually correct time values without actually looking at the knobs". The subjects made a systematic difference in absolute durations as well as in decay time between long and short vowels. In this way the internalised knowledge Dutch-language users have about the linguistic fact that some vowels are perceptually long and others perceptually short, can be brought to the fore, even where this knowledge is completely subconscious. This study of isolated synthetically generated Dutch vowels has been the inspiration for our own experiments on the perceptual reality of durational characteristics of vowels embedded in words. A similar method was used by Collier<sup>89</sup>) in an investigation of the perceptually optimal timing of a prominence-lending intonational rise. The method of matching to internal criterion has also been used by Blom and Uys, who instructed subjects to adjust the controls for two formant values in order to "match a 'phonemic' utterance of the vowel generating system in terms of their phonological knowledge of Dutch" (ref. 90, pp. 70, 71).

It should be noted that in this type of experiment it is not at all certain that the subjects will behave in the desired way. They have in fact much freedom in handling the controls of the synthesiser and if one wishes to limit this freedom

by the given instruction, one runs the risk that the internal criteria which one hopes to explore in the experiment are replaced by external criteria supplied by the instruction. Perhaps the experimenter's fear of this freedom the subjects have and which they might use in undesired ways can explain why this method is not very much used in speech research, although it has the great advantages of producing results in a relatively fast way and of enabling the experimenter to bypass the conscious knowledge of the subjects with respect to the effects to be investigated. We will try to show that this method may be a useful aid in exploring the implicit knowledge of language users and in testing durational rules by showing that the effects described by them really belong to the sound pattern of the language.

### 3.2. The experimental set-up

The experimental set-up consisted of a computer programme for synthesis-by-rule, a special-purpose electronic memory for storing the information needed for synthesising the stimulus words, a speech synthesiser connected to the electronic memory, an extra knob which made it possible to change the duration of one vowel in the word continuously, a pair of headphones for the subject, a speaker for the experimenter and an electronic counter which visually displayed in ms the chosen vowel duration.

The core of this set-up is the synthesiser, the IPOVOX II, designed by Willems and built by Willems and Loonen, described in Willems<sup>91</sup>). It is a segment-by-segment terminal analog synthesiser. The vowel-like segments are simulated from a buzz source with approximately a sawtooth function and two parallel formant filters. The formant positions in these formants are reached from the formant positions in the preceding segment smoothly. The time constant of the formant transition is variable and the transition starts at the beginning of a segment.

For each segment the following parameters are controllable:

- (1) source: periodic, noise, noise through vowel formants, none;
- (2) intonation: 8 elementary movements;
- (3) onset of intonation: 8 possible moments where the intonation movement can begin within the segment;
- (4) noise filter: 4 possible fixed filters for noise source;
- (5) duration of formant transitions: 10, 20, 30 or 80 ms;
- (6) formant 1 : 15 values, 200 Hz–1200 Hz;
- (7) formant 2 : 15 values, 700 Hz–4200 Hz;
- (8) rise time of amplitude envelope: 15 values;
- (9) rise time plus period of maximum amplitude: 15 values;
- (10) decay time: 8 values;
- (11) decay time plus silent interval: 14 values;
- (12) amplitude: 8 values.

A fixed third formant, not controllable from the memory, can be added.

All parameters needed for the synthesis are stored in a flip-flop memory within the machine. This memory contains information for 10 segments and each segment occupies 44 bits. Information for the segment can be fed into the memory one after another via a read-in desk with push buttons. A block diagram of the synthesiser is given in fig. 3.2.1.

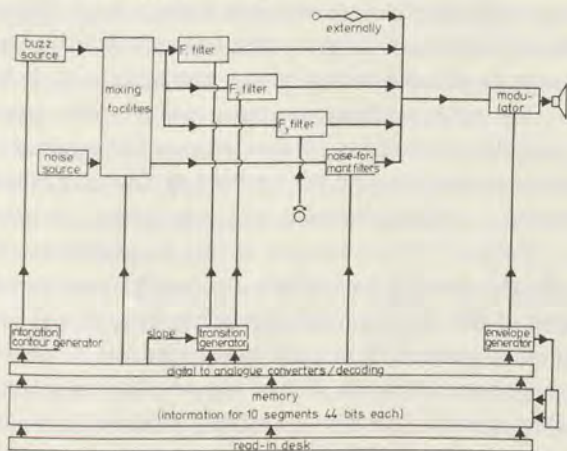


Fig. 3.2.1. Block diagram of IPOVOX II.

The synthesiser has a special feature which is of importance for our experiments. One of the 10 registers in the machine can be marked by a special bit in the memory. This mark denotes that for that particular segment in the sequence of segments the information in the register is overlooked and the information at that moment present in the read-in desk is used instead. This makes it possible to vary all the control parameters of one segment in a word or phrase instantly while listening to it. One may, for instance, ask a subject to adjust the duration of one vowel within a word with the controls on the read-in desk. However, the duration of segments cannot be controlled continuously from the read-in desk, but only in steps which increase with increasing value of the durational parameter. As we wished to give our subjects complete freedom in choosing a preferred duration an extra feature was added to the synthesiser. It was made possible to change the value of the parameter which determines the duration of rise time plus period of maximum amplitude of the segment continuously by means of an external potentiometer. Thus the subject could sit at some distance from the synthesiser, listening to the stimulus word via headphones and varying the duration of one of the vowels with a knob. Turning the knob to the left shortened the duration, turning it to the right increased it. The extreme-left position gave either an absurdly short duration (for stressed vowels) or a zero duration (for unstressed vowels). The extreme-right position of the knob gave an absurdly long duration. The duration as set by the potentiometer

was visualised in ms to the experimenter on an electronic counter. As in all cases the decay time was zero, the value presented on the counter was identical with the total vowel duration.

The IPOVOX II by itself makes it possible to synthesise words of not more than 10 segments, due to the limitations of the built-in flip-flop memory. There is, however, the facility of connecting a special-purpose ferrite-core memory with the synthesiser (Philips C4 magnetic core storage unit). It has 256 memory words of 24 bits. An interface between this memory and the synthesiser and between the memory and a paper-tape reader was built by G.J.J. Moonen and C. A. Lammers. For each synthesis segment two memory words were used, which made it possible to synthesise phrases of up to 127 segments. The memory can be fed from the read-in desk of the synthesiser, but also from a paper-tape reader. The existence of a synthesis-by-rule programme, written by Slis and Muller<sup>92)</sup> for a Philips P 9202 computer makes it possible to obtain punched paper tapes with the control parameters for synthesis of words or phrases rapidly. The input of this programme is formed by a quasi-phonemic transcription plus some extra symbols for syllable division, stress, word division, important versus unimportant words, end of phrase. The output of the programme is a punched paper tape with a code for the control settings of the synthesiser for a sequence of segments (which do not necessarily correspond to the phonemic segments of the input). The punched tape is read into the electronic memory and the information stored.

The information for more than one word or phrase can be stored simultaneously. With the controls of the memory the desired word or phrase can be selected for synthesis. It can then be made audible as many times as one wishes, with a fixed interval between the repetitions.

The words used in the experiments were synthesised by rule and then modified by hand from the read-in desk if the programme gave perceptually unsatisfactory results or if the words to be compared in the same experimental run had to be made as similar as possible.

### 3.3. Accuracy in perceptual timing

As explained earlier, the main aim of our perceptual experiments in this study is to find out whether the rules for vowel durations as formulated in the first part of this study are also valid as a description of part of the knowledge Dutch-language users have about the way words should sound. Before we go into this we will first try to find out the degree of precision with which subjects are able to adjust the duration of a vowel at all in a synthesised word according to some internal criterion. It seemed to us that subjects with much experience in phonetic research would be most likely to keep an internally generated criterion for vowel duration constant. Therefore we used three phonetically experienced subjects in this experiment, viz. the author and two of his colleagues.

Two stimulus words were used, the nonsense words [pəpə:pəp] and [pəpəpəp]. The subjects, one at a time, received the following, spoken, instruction: "You will hear the synthesised word [pəpə:pəp] over your headphones with an absurdly short duration of the stressed vowel [a:]. With the knob in front of you you can change the duration of this vowel; by turning the knob to the right the vowel lengthens, by turning it to the left the vowel shortens. Adjust the duration of the vowel so that it sounds optimal to you in this context. If you are satisfied say so to the experimenter, wait a few seconds and turn the knob again to the extreme-left position. Then try to reproduce your own performance as accurately as possible. In this way the reproducibility of your performance will be tested in 20 successive adjustments of the [a:] duration. After that we will do the same for the vowel [ɑ]. Note that the relation between the duration of the vowel and position of the knob will be changed in an unpredictable way after each individual adjustment".

For simulating the two stimulus words in fact only one word was synthesised with the help of the programme for synthesis-by-rule. This word had the short vowel in the stressed second syllable, because normally long vowels are synthesised with the help of two segments. The vowel of the second syllable was marked in the memory with the special bit which made all its parameters instantly changeable from the read-in desk and its duration variable with the external knob.

The intonation of the word was formed by a slight declination over the whole word and a fall on the stressed vowel, which started at 50% of the vowel duration for the long vowel and 25% for the short vowel and lasted to the end of the word with a maximum of 160 ms duration for the fall. The durational build-up of the word is shown in fig. 3.3.1.

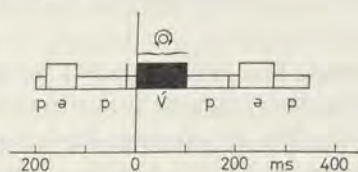


Fig. 3.3.1. Schematic durational build-up of the synthesised nonsense words.

The formant values for the [ə]s were 500 and 1600 Hz, for the [a:] 1200 and 1400, and for the [ɑ] 950 and 1200 Hz. The values of the second formants are much higher than in normal speech because of the absence of third formants.

The results of this preliminary experiment are graphically represented in fig. 3.3.2 *a, b* and *c*, for subjects IS, JtH and SN respectively. It may be seen that for all three subjects the internal criterion for the duration of the short vowel is considerably shorter than that of the long vowel, as was to be expected. The subjects differ in the mean values for both long and short vowels. It may be assumed that in such a nonsense word in isolated position there is much

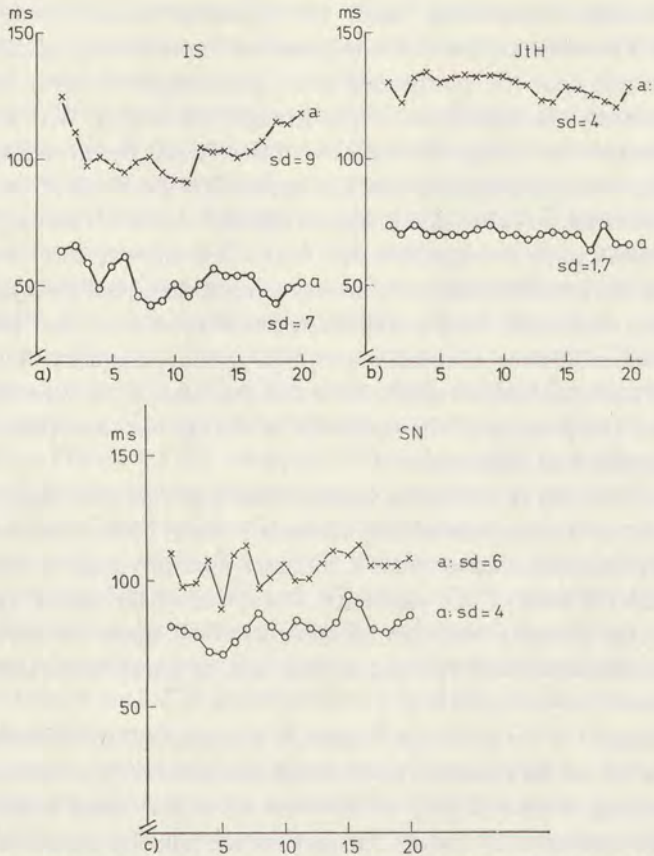


Fig. 3.3.2. 20 successive adjustments of vowel durations of the long vowel [a:] and the short vowel [a]. The data are given for 3 subjects separately in synthesised nonsense words. sd = standard deviation in ms.

freedom in choosing a specific criterion for the vowel duration. Subjects IS and SN declared after the experiment that they had chosen a vowel duration for the long vowel which was the shortest duration acceptable for a long vowel. It is of interest to see that in this case the mean durations do not differ significantly and are close to 100 ms, a value which has been found to be about the lower limit of long vowel durations in stressed position in measurements on speech production. This suggests that, although there is some freedom for individuals in choosing particular vowel durations, there are nevertheless some rather strict limits imposed on this freedom by the sound pattern of the language. Note that in this case the value of 100 ms seems to form a boundary of categories. If the [a:] sound becomes shorter than 100 ms the subjects report that they hear an [a], be it with a somewhat unnatural quality. A question which is still open for investigation is whether the value of this boundary between long and short vowels can be affected by such prosodic parameters as position

in word or phrase and speech tempo. This question will not be pursued here.

The main question in this experiment was the degree of accuracy to which subjects are at all able to adjust the duration of a vowel according to some internal criterion. It may be seen from the standard deviations that this accuracy is in the same order of magnitude as the accuracy found in speech production. This suggests that possibly the accuracy in articulatory timing and the accuracy in perception are related. The accuracy in adjustments of perceptual durations is significantly better for subject JtH than for the other subjects. In sec. 2.3.1 we have seen an exceptional case of accuracy in articulatory timing. This also concerned subject JtH, who furthermore is known in our laboratory as remarkably good in timing tasks. This is a further indication that there is some close relation between accuracy in production and in perception.

It may be seen that, although the task of the subjects in the experiment described does not seem to be particularly suited for studying just-noticeable differences because the subjects had to keep their internally generated criterion constant, the accuracies found are not lower than would be expected on the basis of just-noticeable differences of speech-sound duration as mentioned in the literature (Lehiste, ref. 19, p. 40; Huggins <sup>25,26</sup>). These results have encouraged us in proceeding to using the method of matching to internal criterion for testing durational rules.

#### 3.4. Testing for the perceptual reality of physiologically conditioned effects

In discussing vowel durations, several phoneticians have made a distinction between "articulatorily" or "physiologically conditioned", and "learned" variations in duration (House <sup>93</sup>, Delattre <sup>34</sup>), the first being caused by limitations of the speech organs, the latter resulting from the language system as acquired by the language user. At first sight it seems reasonable to assume that the "learned" variations in duration as defined here belong to the sound pattern of the language which is internalised by the language user (this is probably so per definition), whereas the "physiologically conditioned" variations do not belong to the sound pattern of the language and are not internalised (= learned) by the language user. Such variations automatically result from the way the speech organs are organised and as such must be supposed to be universally present.

An interesting question is whether such small physiologically conditioned effects have perceptual correlates which are known to the language users and perhaps may be used as aids in speech perception.

In order to study this question one should select a durational difference of which it is certain that it is "physiologically conditioned". Thus a model should exist which explains this effect from universal properties of human speech.

The difference in vowel duration due to the voiceless or voiced character of the postvocalic consonant has been shown to be perceptually real by Slis and

Cohen (ref. 16, p. 89; see also Nooteboom<sup>94</sup>). A proposal of Halle and Stevens<sup>95</sup>) that such lengthening could be explained by the time-consuming fine positioning of the vocal cords for voiced consonants as compared to the rapid opening for voiceless consonants does not seem to be satisfactory (Wang<sup>96</sup>)).

Thus a model which can explain this durational difference in terms of the universal properties of the organs of speech is still lacking.

Such a model is available for explaining the fact that open vowels have a longer duration than close ones, at least for the bilabial environment, as provided by the earlier-mentioned study of Lindblom<sup>41</sup>) on lip mandible coordination. The plausibility of the effect of vowel height on vowel duration really being a physiologically conditioned and universal effect seems to be much better established than that for other effects of this kind. It would be of interest, then, to find out whether language users do indeed know in some way the perceptual results of the durational effect or not. We have tried to do this in the following experiment.

In the articulatory measurements described in sec. 2.4.2.3 of this study it was found that in the stressed second syllable of words with the form [pVpVpVp] the duration of the open vowel [a:] is about 10 ms longer than that of the more close vowel [e:]. We have tried to find out whether this difference shows up in preferred vowel durations as established with the method of matching to internal criterion. Three stimulus words were used, viz. [pəpə:pəp], [pəpe:pəp] and [pəpə:pəp]. The third one was added to see whether the lengthening due to lip rounding as found in the articulatory measurements would show up. The subject received the following spoken instruction: "You will hear over your headphones one of the following three words: [pəpə:pəp], [pəpe:pəp] and [pəpə:pəp]. The vowel of the second syllable is supposed to be stressed. When you hear each word for the first time the duration of the stressed vowel is absurdly short and the perceptual effect of the stress may be absent. By turning the knob in front of you to the right you can increase the duration of this vowel continuously. When you turn the knob to the left again you can decrease the vowel duration. You are asked to find a duration of this vowel such that the overall rhythmical pattern of the word is optimised. When you are satisfied with your adjustment, warn the experimenter. We will then proceed to the following word. We will continue until you have made 10 adjustments for each word. Note that the relation between the position of the knob and the duration of the vowel will be changed after each individual adjustment randomly. Are there any questions?"

Three naive subjects, a psychologist, a computer programmer and a research assistant participated in the experiment. These subjects had no knowledge of the outcome of our earlier articulatory measurements. They had on the whole no conscious knowledge of the durational regularities to be investigated.

For simulating the three words in fact only one word was synthesised with the



help of the programme for synthesis-by-rule. The values of the synthesis parameters of this word have been described in sec. 3.3. The formant values for the stressed vowels were 1200 and 1400 Hz for [a:], 400 and 2300 for [e:] and 400 and 950 Hz for [o:]. These vowels had the same intensity. The results for [a:] and [e:] are graphically represented in fig. 3.4.1. The results for [o:] were very in-

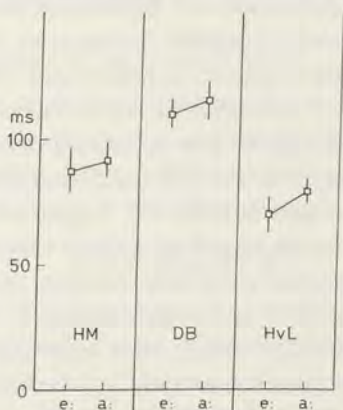


Fig. 3.4.1. Adjusted vowel durations of the vowels [e:] and [a:] in synthesised nonsense words. The means over 10 adjustments are indicated with the standard deviations on both sides for three subjects separately. Amplitude of the vowel sound was the same for [e:] and [a:].

consistent and are omitted here. All subjects bitterly complained about the unnatural character of the [o:], which sounded more like [u] than like [o:]. It seems to be the case that the diphthongal character of [o:] cannot be omitted in synthesis. For [e:], which is also somewhat diphthongal in natural speech, this does not seem to be a necessary attribute in speech synthesis.

On the whole the subjects complained during and after the experiment that the speech sounded unnatural and that the task was very difficult, the more so because they tended to hear many different real or nonsense words in the stimulus word. They found it often impossible to concentrate on the given interpretation of the stimulus word. They said that the rhythmical pattern of the word changed with every interpretation. They found the task definitely unpleasant.

Given the complaints of the subjects it is surprising to see how consistently they behaved in their durational adjustments. The standard deviations range between 4.5 and 9 ms. At first sight it also seems to be the case that the durational effect of vowel height shows up in the preferred vowel durations: for all three subjects the mean of the adjustments for [a:] has a higher value than that for [e:]. The difference is not significant, however, for subjects HM and DB. For HvL the difference is significant at the 5% level (student *t*-test). The significant difference for one subject plus the non-significant tendency in the same direction for the two other subjects might be taken as a slight indication that

the difference in duration due to vowel height played a role in the duration of the vowel adjustments.

Note, however, that the amplitude of the two vowels was the same. This is not so in natural speech, where the amplitude of the [a:] sound is greater than that of [e:], other things being equal. It is not a priori impossible that there is some trade-off between amplitude and duration in the perception of vowels, which could lead to a shorter preferred duration for [e:] when the amplitude is greater. Thus we decided to repeat the experiment with a difference in amplitude between [a:] and [e:]. This difference should ideally be as close as possible to the situation in normal speech. For the Dutch vowels [a:] and [e:] no data are available. For English differences in vowel amplitudes have been measured and described by Lehiste and Peterson<sup>97</sup>). The vowel pair [a:]–[e:] may perhaps best be compared to the American English vowel pair [ɔ]–[ɪ].

The difference in sound-pressure level between these vowels as spoken in monosyllables was found to be somewhat less than 3 dB. This value was taken to be a rough estimate of the difference in amplitude between [a:] and [e:]. Thus we repeated our experiment with [e:] being 3 dB less in amplitude than [a:].

The results are presented in fig. 3.4.2. They look rather chaotic. For HM [e:] is significantly longer than [a:], for DB and HvL there is no significant difference between the vowel durations. It must be stated, however, that the motivation of the subjects in this experiment was even less than in the former one. It may be the case that the motivation of the subjects was much more important than the difference in amplitude. In any case, from these results we cannot conclude that the longer duration of open vowels is reflected consistently in the preferred durations in this experiment. The surprisingly consistent behaviour of the

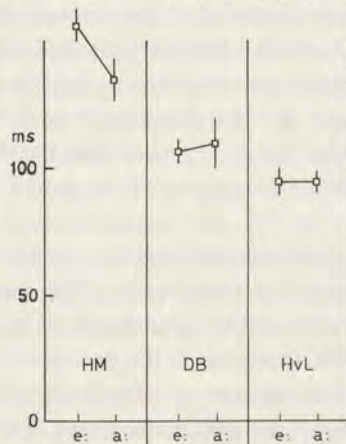


Fig. 3.4.2. Adjusted vowel durations of the vowels [e:] and [a:] in synthesised nonsense words. The means over 10 adjustments are indicated with the standard deviations on both sides for three subjects separately. Amplitude of the [e:] sound was three dB lower than of the [a:] sound.

subjects as shown by the low standard deviations shows that this type of experiment may be useful for the investigation of this type of question. In future experiments in this direction care should be taken that the speech is of a natural quality. Probably the use of nonsense words should be avoided.

### 3.5. Testing rules for vowel durations

#### 3.5.1. *Testing a durational rule for stressed vowels*

In sec. 2.5 of this study we found among other things that the duration of a stressed vowel decreases as the number of syllables which remain to be produced in the word increases. Also a slight effect of the total number of syllables in the word was found. A formal rule to this effect was formulated as follows:

$$(1) \quad DV = DV_{opt}/m^{\alpha} n^{\beta}.$$

We assume that this rule describes some important aspect of the rhythmical organisation of the polysyllabic word in Dutch. This assumption implies that the effect of word length as it is found in production and is described in this rule, results from some underlying mental structure which does not only affect the timing of opening and closing movements of the mouth in the production of a word, but also serves some function in the perception of speech. It should be noted that the existence of such a mental structure is not easily, if at all, revealed by introspection. Thus to find out whether such a mental structure exists or not, a method is required which bypasses the introspection. It is precisely in cases like this that the method of matching to internal criterion seems to be most appropriate. In this section we have tested the hypothesis that the effect of word length as formulated in rule (1) will show up in preferred durations as found by asking subjects to adjust the durations of stressed vowels in words which differ in the amount of syllables following the stressed vowel.

From the experiments described in the preceding sections we learned that subjects strongly object to nonsense words as produced by a synthesiser. Furthermore it seems important to test the rule found for nonsense words, for real words also. Thus we decided to use existing Dutch words for testing the rule.

We ran several tests, one in which the number of syllables following a syllable with stressed long vowel was varied, a similar one with a stressed short vowel, one in which the position of a stressed long vowel in the word was varied, and one in which the number of syllables preceding a syllable containing a stressed long vowel was varied. All words were presented in isolation.

We attempted to make the environment for the vowels under investigation as similar as possible in the synthesised words. The durational build-up of the synthesised words will be shown below. The intonation pattern of all words was constituted by a slight declination over the whole word and a rather rapid fall on the stressed vowel. This fall started at 50% of the vowel duration for the long

vowel and at 25% for the short one. Although we do not know in which ways the intonation pattern may affect the perception of vowel duration we nevertheless preferred a more or less natural intonation pattern to a monotone or declination only, in order to let the word sound as natural as possible. In any case it seems probable that the effect of the intonation was the same in all cases under comparison. The same three naive subjects who participated in the experiment described in sec. 3.4 took part in this experiment. The written instruction was as follows: "You are going to participate in a perceptual experiment with synthetic speech. You will hear via your headphones one of the following words:

..... (see below)

The word you hear will be repeated as many times as you wish. The underlined vowel in the word will not have the correct duration. You can change this duration yourself with the knob in front of you. By turning the knob to the right the duration gets longer, by turning the knob to the left it gets shorter. You are asked to adjust the duration so that the word sounds as natural as possible as said on a neutral tone. Such an adjustment does not need to take long. When you are satisfied with the result, say so to the experimenter and do not touch the knob for a moment. The first adjustment will be made starting from a duration which is too short. You will do this for all different words. The second adjustments for all words will be made starting from a duration which is too long. We will continue switching in this way until you have made 10 adjustments for each word. Note that the relation between the position of the knob and the vowel duration will be changed in an unpredictable way after each individual adjustment".

The words filled in in the instruction were for the four tests respectively:

- |     |                  |                  |                    |                      |
|-----|------------------|------------------|--------------------|----------------------|
| (1) | <u>ma</u> t      | ma <u>t</u> e    | ma <u>t</u> eloos  | ma <u>t</u> eloze    |
|     | ([ma:t])         | [ma:tə]          | [ma:təlo:s]        | [ma:təlozə]          |
| (2) | <u>pa</u> n      | pa <u>n</u> ne   | pa <u>n</u> nekoek | pa <u>n</u> nekoeken |
|     | ([pan])          | [panə]           | [panəkuk]          | [panəkukə]           |
| (3) | auto <u>ma</u> t | to <u>ma</u> ten | ma <u>t</u> eloos  |                      |
|     | ([o:to:mat])     | [to:ma:tə]       | [ma:təlo:s]        |                      |
| (4) | <u>ma</u> t      | to <u>ma</u> t   | auto <u>ma</u> t   |                      |
|     | ([ma:t])         | [to:ma:t]        | [o:to:ma:t]        |                      |

In the instruction these words were given in orthography only.

It was hoped that this instruction was sufficiently precise to let the subjects know what they had to do and yet sufficiently vague to give the subjects much freedom to let their own internal representation of how these words should sound govern their behaviour. The instruction was found to work satisfactorily in some preliminary sessions with other subjects.

We also made tape recordings of all these words as spoken by the same subjects who took part in the adjustment tests. Each subject spoke each word twice. From these tape recordings spectrograms were made with a Kay Sonagraph and the durations of the stressed vowels were measured. In this way we were able to compare the vowel-duration adjustments with the durations of the vowels in spoken realisations of the words. It should be noted, however, that the rest of the durational build-up of the synthesised words was not determined by the subjects, so that their behaviour in the adjustments may have been influenced by an overall durational build-up which was foreign to their own way of speaking. The results of the adjustments and spectrographic measurements are presented fully in tables 21–24 in appendix B.

### 3.5.1.1. The effect of the number of syllables following in the word on the preferred duration of a stressed vowel

In figs 3.5.1 and 3.5.2 the results for one subject are graphically represented. It may be seen that the standard deviations, although somewhat inconsistent, are on the whole rather low, if calculated for a particular word and for the adjustments with the same initial duration. There is a systematic difference, however, between the adjustments starting from initially short and those starting from initially long. This systematic difference was found for all subjects. It may also be seen that the effect of the number of syllables following in the word

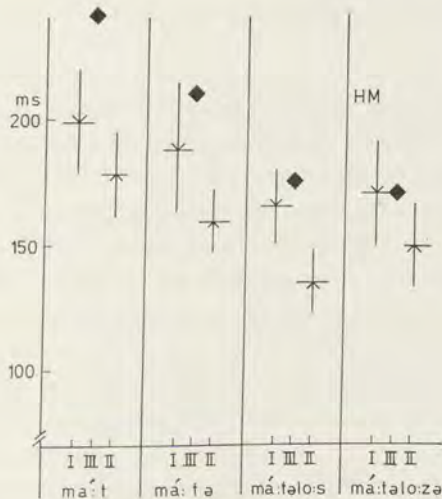


Fig. 3.5.1. Adjusted durations of stressed [a:] sounds in synthesised real words and in actually spoken words for one subject.

I: Adjustments from initially long durations.

II: Adjustments from initially short durations. The means over 5 adjustments and the standard

deviations on both sides are indicated.

III: Durations as measured in spectrograms averaged over two realisations.

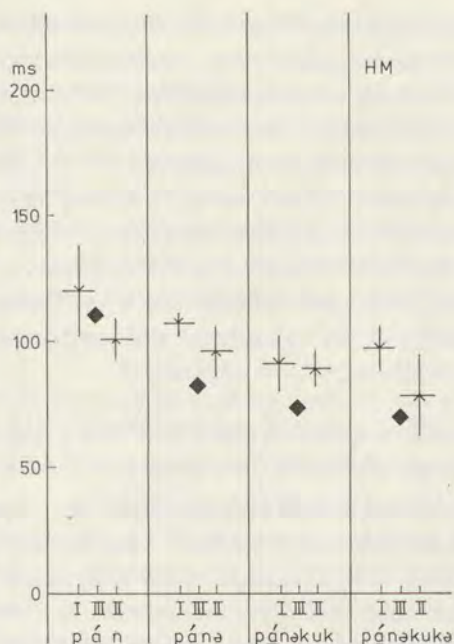


Fig. 3.5.2. Adjusted durations of stressed [a] sounds in synthesised real words and in actually spoken words for one subject.

I: Adjustments from initially long durations.

II: Adjustments from initially short durations. The means over 5 durations and the standard deviations on both sides are indicated.

III: Durations as measured in spectrograms averaged over two realisations.

is definitely present, both in the adjustments and in the spoken vowel durations. This is also true for the other subjects. In figs 3.5.3 and 3.5.4 the results averaged over the two initial situations and over the three subjects are graphically represented together with the durational build-up of the synthesised words. The stressed vowels, black in the diagrams, were adjusted by the subjects. The effect of number of following syllables is clearly present, although to a lesser extent than in the articulatory measurements on the nonsense words.

In fig. 3.5.5 the theoretical, spoken and adjusted vowel durations may be compared. It is rather satisfactory that the values of  $DV_{opt}$ , 200 and 100 ms, which were found in the measurements on nonsense words, fit these new data so well. The value of  $\alpha$ , however, had to be adapted to fit these data. The same value holds good for long [a:] and short [a]. The values of spoken and adjusted vowel durations are remarkably similar. In speaking and in adjusting the vowel durations the subjects seem to obey the same internal patterns. This is the more significant as the spoken versions were recorded about two months after the adjustments were made. From these results we may conclude that the effect concerned is not a sole attribute of the speech-production system, but clearly belongs to the subjects' knowledge of how the words should sound.

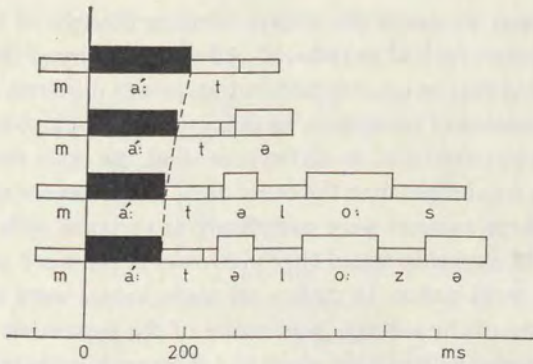


Fig. 3.5.3. Schematic durational build-up of synthesised words. The durations of the stressed vowels were found in adjustments test. The values give the mean over 10 trials for each of three subjects.

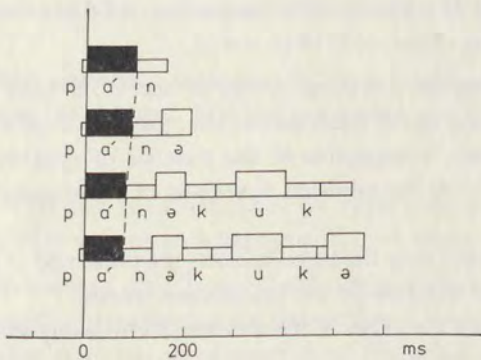


Fig. 3.5.4. Schematic durational build-up of synthesised words. The durations of the stressed vowels were found in adjustment tests. The values give the mean over 10 trials for each of three subjects.

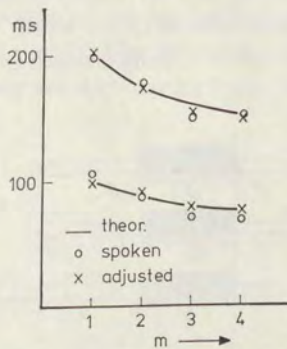


Fig. 3.5.5. Theoretical, spoken and adjusted durations of stressed [a:] and [a] as a function of the number of syllables which remain to be produced in the word.  $DV_{opt} = 200$  and  $100$  ms for long and short vowels respectively,  $\alpha = 1/5$  for both. Circles refer to mean spoken durations over 2 trials, crosses refer to mean adjusted durations over 10 trials for each of three subjects.

After each session we asked the subject what he thought of the task and of the kind of differences he had introduced. All subjects agreed that the task was not too difficult but that an extreme concentration was required, mainly because of the unnatural quality of the speech. In the case of the long vowels subject HM thought that he had introduced no differences at all, the other two assumed that the monosyllable was longer than the other ones. In the case of the short vowels the ideas of all three subjects were completely at variance with what they had done. Subject HM explicitly stated that the vowel in the word *pan* was shorter than that in the word *panne*. In reality his mean values were 110 and 101 ms respectively. None of the subjects was aware of the systematic differences between the adjustments with initially short and those with initially long durations. The results of these interviews show that the mental patterns underlying the behaviour of our subjects in this task mainly operate on a subconscious level.

### 3.5.1.2. The effect of position on the preferred duration of a stressed vowel in a three-syllable word

The measurements on nonsense words in sec. 2.5.5 have shown that the duration of a stressed vowel decreases as the number of syllables coming later in the word increases, irrespective of the number of syllables which precede. We have tested this for the adjusted durations of the stressed vowels in three-syllable words.

The results averaged over the three subjects are presented in fig. 3.5.6 together with the durational build-up of the synthesised words.

It may be seen that the effect of the number of following syllables in the word is again present.

In fig. 3.5.7 the theoretical, spoken and adjusted durations may be compared. There seems to be a fairly good agreement. That both the spoken and the adjusted durations are somewhat shorter than the theoretical duration in the case that  $m = 1$ , i.e. in the final syllable of the word, may indicate that there is some slight effect of the number of syllables which precede the stressed vowel.

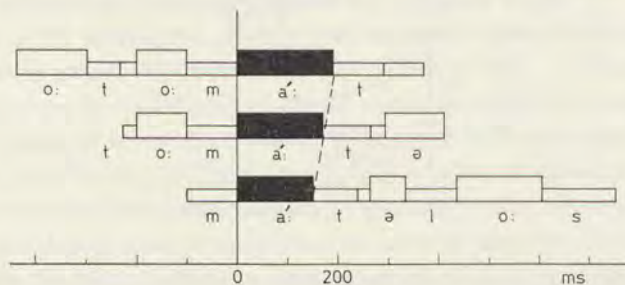


Fig. 3.5.6. Schematic durational build-up of synthesised words. The durations of the stressed vowels were found in adjustment tests. The values give the mean over 10 trials for each of three subjects.



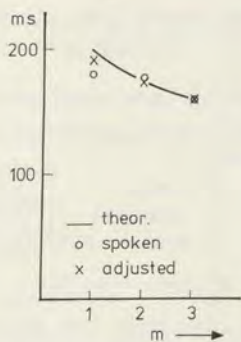


Fig. 3.5.7. Theoretical, spoken and adjusted durations of stressed [a:] as a function of position in a three-syllable word.  $DV_{opt} = 200$  ms,  $\alpha = 1/5$ . Circles refer to mean spoken durations over 2 trials, crosses refer to mean adjusted durations over 10 trials for each of three subjects.

The order of magnitude of 10 ms agrees with that found in the articulatory measurements.

In the interview HvL explicitly stated that the [a:] of *mateloos* was the longest, whereas in fact it was the shortest. HM did not know, and DB did know what he had done.

### 3.5.1.3. The effect of the total number of syllables on the preferred duration of a vowel in a stressed final syllable

Both the measurements of articulatory durations and the adjusted durations discussed in the preceding paragraph have shown that if there is an effect of the total number of syllables in the word apart from the effect of the number of syllables coming later in the word, it is extremely slight. We have again studied this effect separately by comparing the adjusted durations of the vowels in the final syllables of the words *maat*, *tomaat*, *automaat*, all with stress on the final syllable. The results are presented in fig. 3.5.8, together with the durational build-up of the synthesised words. The differences in the adjusted vowel durations are not significant. They are shown once more, together with the theoretical

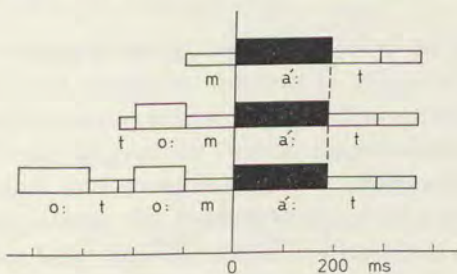


Fig. 3.5.8. Schematic durational build-up of synthesised words. The durations of the stressed vowel were found in adjustment tests. The values give the mean over 10 trials for each of three subjects.

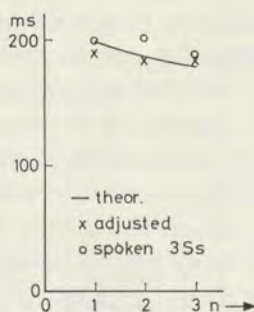


Fig. 3.5.9. Theoretical, spoken and adjusted durations of stressed [a:] as a function of the number of syllables which precede in the word.  $DV_{opt} = 200$  ms,  $\beta = 0.05$ . Circles refer to mean spoken durations over 2 trials, crosses to mean adjusted durations over 10 trials for each of three subjects.

and spoken durations, in fig. 3.5.9. There it may be seen that in both the spoken and the adjusted durations there is a weak (and not significant) tendency towards a decrease in vowel duration with increasing  $n$ . For the spoken vowel durations this is mainly due to one subject, HM. His data are plotted separately in fig. 3.5.10. His spoken vowels show a clear tendency towards decreasing durations with increasing  $n$ . Thus it seems to be the case that the effect of the total number of syllables is subject-specific. It may also be the case, however, that the effect increases with increasing vowel durations. In the Lindblom and Rapp data for at least five of their six subjects the vowel durations for the final syllable were above 250 ms, and there a definite effect of  $n$  was present. Such long vowel durations are extremely rare for Dutch speakers, and this may be the explanation why the effect of  $n$  is so weak in our data.

In the interview two subjects, HM and HvL stated that the vowel in *tomaat* was longer than the one in *maat*. HvL thought that the vowel of *automaat* was the longest. All three subjects objected to the unnatural quality of the final [t].

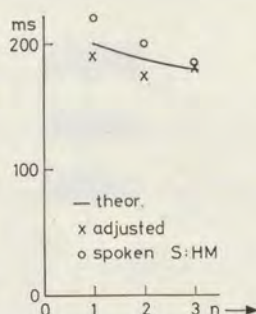


Fig. 3.5.10. As fig. 3.5.9. for subject HM alone.

### 3.5.2. Testing for the perceptual reality of some durational rules for unstressed vowels

In sec. 2.5.6 of this study we formulated the following rules for generating correct durations for long vowels in unstressed syllables:

- (2)  $DV = DV_{opt}/1.3$  if . . Y(C) #
- (3)  $DV = DV_{opt}/2.2$  if # (C)VCV . .
- (4)  $DV = DV_{opt}/2.4$  if # (C)YCV . .
- (5)  $DV = DV_{opt}/3$

We have tested the hypothesis that these rules also apply to describing the preferred durations of unstressed vowels in adjustment tests and have thus tested the perceptual reality of these rules.

We first tested rules (2), (4) and (5) with the vowel in the syllable [to:m] forming part of words of different structure, such as [o:vərto:m] [to:ma:t] [o:to:ma:t]. The duration of [o:] in the word [a:na:to:m] served as a reference point as it was thought that this vowel duration would be close to the optimal vowel duration. The same three naive subjects who took part in the previously described adjustment tests participated in this experiment. Furthermore three phonetically trained people did the same task, the reason for which will be explained below.

We then tested the hypothesis that the vowel in the context # CVCV . . will get a longer duration in an adjustment test than the vowel in the context # CVCV . . as predicted by rules (3) and (4). This was done again with the same three naive subjects. The words used were [me:todik] and [me:tə:də].

The durational build-up of the words used will be shown below. The intonation pattern was the same as that of the words in sec. 3.5.1.

We again made tape recordings of all the words used in these tests as spoken twice by the three naive subjects. Spectrograms were made and the durations of the vowels under investigation measured. The results of adjustments and spectrographic measurements are presented fully in tables 25–30 in appendix B.

#### 3.5.2.1. Testing rules (2), (4) and (5)

In a first test we compared the adjusted durations of the underlined vowels in Overtoom, tomaten and automaat with that in anatoom. In fact, the word anatoom was adjusted at separate later sessions. This made the sessions shorter and less boring for the subjects. In view of the consistent behaviour of the subjects in the previous tests it was thought that adjustments made in different sessions are still comparable. The first test with the three unstressed [o:]s was done by two subjects only. The instruction was the same as that described in sec. 3.3.5. The results are graphically represented in fig. 3.5.11 together with the result for anatoom, within the frames of the durational build-up of the synthesised words. In fig. 3.5.12 the results may be compared with the theoretical

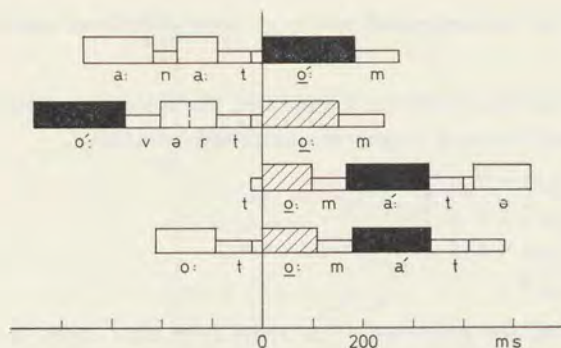


Fig. 3.5.11. Schematic durational build-up of synthesised words. The durations of the underlined vowels are found in adjustment tests. The values give the mean over 10 trials for each of 2 subjects HM and HvL.

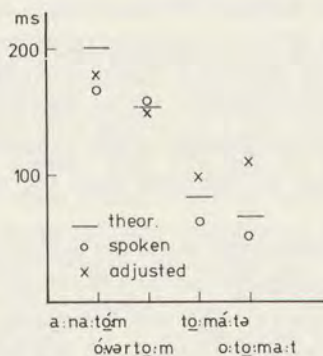


Fig. 3.5.12. Theoretical, spoken and adjusted durations of [o:] as a function of stress and position. Circles refer to mean spoken durations over 2 trials, crosses refer to mean adjusted durations over 10 trials for each of 2 subjects.

and spoken durations. It is evident that the adjusted durations differ rather much from both the spoken durations and the theoretical durations. The spoken durations and theoretical durations seem to differ much less. One may note that in the adjusted durations the subjects do not go much lower than 100 ms, a value of which we know that it may correspond to an important perceptual boundary between long and short vowels. And indeed, when asked to describe what they thought during the session both subjects stated that they made the [o:] as short as possible in the words *tomaten* and *automaten*, without making it so short that it lost its typical [o:]-like sound. "If you make it shorter", they said, "it is not an [o:] anymore". It seems to be the case that the experimental situation focussed the attention of the subject too much on the unstressed vowel, which normally gets no attention, and the subjects tried to give the unstressed vowel those perceptual characteristics which it normally loses through lack of stress. We thought that perhaps phonetically more-sophisticated people would not succumb to the same temptation. Therefore we repeated the experiment with

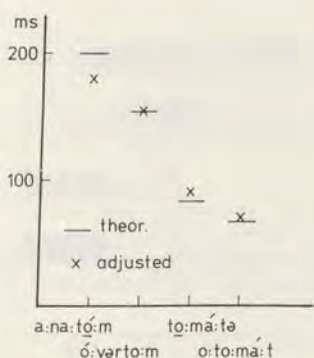


Fig. 3.5.13. As fig. 3.5.12, without the spoken durations, for three phonetically non-naive subjects.

three phonetically experienced people from our institute. Their results are plotted in fig. 3.5.13 together with the theoretical values. Here the agreement is much better, so close even that we feel no need for revising the rules in the absence of further evidence. After this we decided to do the experiment once more with the three naive subjects. This time, however, we explicitly added to the instruction that the way the word sounded as a whole was to be considered more important than the realisation of the [o:]. Furthermore we replaced the word *automaat* by the word *automaten*, in order to make the position of the [o:] still more comparable to the one in *tomaten*. The results are represented first in fig. 3.5.14 with the schematic durational build-up of the words, and then in fig. 3.5.15 with the theoretical and spoken values. The agreement is much better than the first time, but still the duration of the [o:] in *automaten* is not made significantly shorter than that in *tomaten* as predicted by the rules. This is due to subjects HvL and DB. Subject HM this time did make the difference according to the rules, as may be seen from table 29 in appendix B.

It may be noted from fig. 3.5.14 that the other subjects at least made the second [o:] in *automaten* markedly shorter than the first [o:] in the same word. Thus the pattern of vowel durations as predicted by our rules is borne out within the frame of the same word, be it in relative terms and not in absolute terms. It does not seem unlikely that the rather artificial durational build-up of the synthesised words has interfered with the subjects' internal representation of how the word should sound. In this case it would be particularly interesting to see what would happen when the adjustable vowel was part of a naturally spoken word. Perhaps we should be surprised that the results of the adjustment tests agreed so remarkably well with the predictions of the rules, and with the spoken durations, in spite of the poor quality and artificial durational build-up of the synthesised words.

The adjusted and the spoken durations of stressed [o:] were ca 20 ms shorter than the predicted value of 200 ms, for both the naive and the non-naive sub-

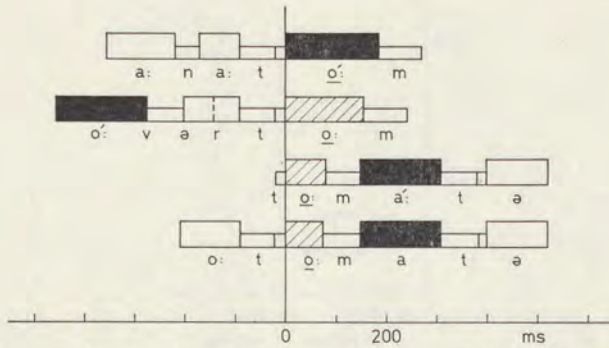


Fig. 3.5.14. Schematic durational build-up for synthesised words. The durations of the underlined vowels were found in adjustment tests. The values give the mean over 10 trials for each of three subjects.

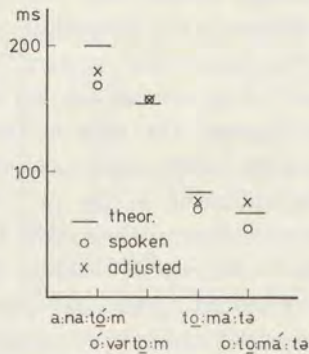


Fig. 3.5.15. As fig. 3.5.12, for three subjects. The instruction was slightly changed (see text).

jects. This may be due to a difference between [a:], from which the predicted durations were derived, and [o:], and/or a difference due the following consonant.

Further study may perhaps lead to changes in the rules as stated, but on the whole they seem to describe fairly well what the subjects do with their vowel durations in speaking and what they know about them in perception.

### 3.5.2.2. Testing rules (3) and (4)

In a final test we investigated the perceptual relevance of the difference between the vowel duration in an unstressed initial syllable before a stressed and before an unstressed syllable. It is predicted by rules (3) and (4) that for instance in the words *methode* [me:tə:də] and *methodiek* [me:tə:dik] the vowel in the first syllable has a shorter duration in the first than in the second. We used these two words to test whether this difference would be introduced in the adjustment task. The same three naive subjects took part in the test. The instruction was the same as in the last test described, included the addition that the way the word

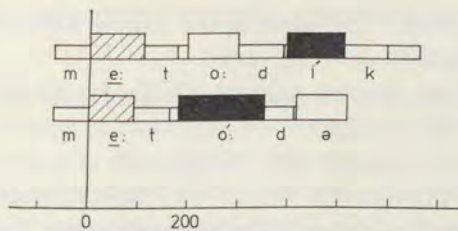


Fig. 3.5.16. Schematic durational build-up of synthesised words. The durations of the underlined vowels were found in adjustment tests. The values give the mean over 10 trials for each of three subjects.

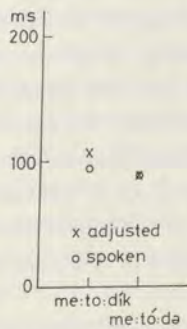


Fig. 3.5.17. Spoken and adjusted durations of unstressed [e:] as a function of stress in the following syllable. Circles refer to mean spoken durations over 2 trials, crosses refer to mean adjusted durations over 10 trials for each of three subjects.

sounded as a whole was to be considered more important than the realisation of the vowel to be adjusted.

The results are represented in fig. 3.5.16, in the frame of the durational build-up of the synthesised words. In fig. 3.5.17 the results are presented together with the spoken durations. It may indeed be seen that the difference between the two situations shows up in the adjusted vowel durations. The absolute vowel durations are somewhat longer than predicted by the rules if we assume an optimal vowel duration of about 200 ms. We can only speculate why this is the case.

### 3.6. Summary

This chapter was devoted to testing the durational rules formulated in chapter 2 for perceptual reality using a method which may be called the method of matching to internal criterion. Essential for the method is that subjects are asked to adjust some acoustic parameter until the perceptual result is satisfactory according to some internal criterion. A great advantage of the method is that it may bring to the fore aspects of the internalised knowledge concerning the way speech should sound even where this knowledge is completely subconscious. In the tests for vowel durations use was made of synthetic speech. The subjects were asked to adjust the duration of vowels embedded in words by

means of a potentiometer. By selecting the words, stress and position of the vowel under investigation could be varied.

In a preliminary test directed towards establishing the accuracy with which subjects are at all able to adjust vowel durations in such a task, it was found that this accuracy is of the same order of magnitude as that found in articulatory measurements. Furthermore the subject with highest accuracy in articulation also showed the highest accuracy in this perceptual task. This indicates that the accuracy in both production and perception may be closely related.

Several phoneticians have made a distinction between "articulatorily" or "physiologically conditioned" and "learned" variations in duration. There are some indications that physiologically conditioned variations in vowel duration belong to the implicit knowledge language users have about the way speech sounds in their language. Some data are discussed which seem to point in this direction. The effects concerned, however, are not clearly shown to be physiologically conditioned. This has been shown rather convincingly for the effect of vowel height on vowel duration in a bilabial environment. Therefore the attempt was made to assess the perceptual reality of the systematic variations in vowel duration due to vowel height in a bilabial environment. The results of the test were inconclusive.

The durational rule for the duration of stressed vowels was tested and found to have perceptual reality for three naive subjects. The results of the adjustments done with synthesised real words were in good agreement with the values predicted by the rules. In some cases, namely when the following consonant was not the same as in the articulatory measurements, the value of the exponent of the power function in the rules had to be different from the one found in the articulatory measurements. The results were in good agreement with those obtained in measuring durations in spectrograms of spoken versions of the words used in the adjustment tests.

The durational rules (2), (4) and (5) for unstressed vowels were tested for three naive subjects. Both the results and the interviews with the subjects showed that they avoided durations of a long vowel shorter than 100 ms. The same test repeated with three phonetically non-naive subjects showed good agreement with the predictions. When, then, the test with the naive subjects was repeated with a somewhat different instruction the results showed better agreement with the predictions than the first time. The results obtained with durational measurements of spectrograms of spoken versions of the words showed good agreement with the predictions.

Durational rules (3) and (4), generating different vowel durations in an unstressed initial syllable due to the stress on the second syllable, were tested against each other with three naive subjects. The difference was found to have perceptual reality.



#### 4. SOME INTERPRETATIONS AND SPECULATIONS

##### 4.1. The place of vowel-duration rules in linguistic theory: the phonetic component

In phonology the sounds of speech or phonemes are looked upon as abstract entities showing serial order but having essentially no extension in time. For instance, if Chomsky says that "a language associates sound and meaning in a particular way; to have command of a language is to be able, in principle, to understand what is said and to produce a signal with an intended semantic interpretation" (ref. 98, p. 397) with the terms "sound" and "signal" something is meant that can be described in terms of a sequence of abstract symbols. This may become evident from the following quotation: "The theory of universal phonetics attempts to establish a universal phonetic alphabet and a system of laws. The alphabet defines the set of possible signals from which the signals of a particular language are drawn. If the theory is correct, each signal of a language can be represented as a sequence of symbols of the phonetic alphabet" and "Representation in terms of the universal alphabet should provide whatever information is necessary to determine how the signal may be produced, and it should, at the same time, correspond to a refined level of perceptual representation" (ref. 98, p. 403). According to this view, the universal alphabet, consisting of symbols which themselves are made up of phonetic features "properties such as voicing, frontness, backness, stress, etc." (Chomsky, ref. 98, p. 403), should provide the possibility of giving the necessary information for shaping the correct durational build-up. At present this possibility seems to be provided for in the Chomsky and Halle list of features (ref. 29, pp. 299-300) by the prosodic feature of length. On prosodic features these authors state: "Our investigations of these features have not progressed to a point where a discussion in print would be useful" (ref. 28, p. 329). Thus we are left out in the cold, so to speak.

One may imagine that each vowel, and perhaps each segment, is assigned a specification for the length feature on the level of phonetic representation. The rules for such specifications would then have to take account of vowel quantity, stress, position in the word and perhaps a number of other factors such as tempo, semantic import of the word, etc. In this way the full information necessary for shaping the correct durational build-up could be given on the level of phonetic representation, in terms of a feature specification assigned to each of the symbols in each "signal".

There are reasons to believe, however, that this would not do. What we normally call a "vowel duration", may be something else than the duration of the vowel phoneme as specified on the level of phonetic representation. One possible way of speaking of vowel duration would be to define the vowel duration as the

interval of time during which attributes of this vowel affect articulation. In pronouncing the English word *noon* the rounding of the lips, being an attribute of the vowel, is already present before the initial [n] has started, and probably often lasts until after the second [n] has stopped, thus giving a vowel duration which is at least as long as the total duration of the word. This evidently is not the way we normally talk about vowel durations. We may try it another way. We may define the vowel duration as the interval of time between the release of consonantal closure, or a point of maximum increase in intensity (Fant, ref. 99, p. 222) and some similar break at the end of the vowel. But note that on the level of phonetic representation the vowel in the English word *noon* is made up of two symbols [u] and [ɔ]. If we want to define the duration of each of these sounds we run into difficulties. There is no useful discontinuity in the articulation to define the beginning or ending of a durational interval. If we talk about vowel durations we normally mean the durations of syllabic nuclei as a whole, and a syllabic nucleus may consist of more than one phonetic segment. In present-day generative phonology there seems to be no level where one can assign a feature specification to a syllabic nucleus as a whole. Now one may perhaps argue that diphthongs should not be described as consisting of two segments, but rather as being one segment on the level of phonetic representation. This has been done by Cohen<sup>53</sup>; see also this study, sec. 2.4.2.8. But the syllabic nucleus may not only consist of a diphthong, it can also be made up of a vowel plus preceding or following [r] or [l]. This has been argued recently on the basis of durational measurements by Ilse Lehiste<sup>100</sup>). Evidence may also be found in the behaviour of the short vowel plus [r] as described in sec. 2.4.2.7 of this study. We think it goes rather far to propose describing such combinations as only one segment on the level of phonetic representation. What seems to be the case is that the "vowel duration" or "syllabic-nucleus duration" cannot be assigned to any particular segment on the level of phonetic representation.

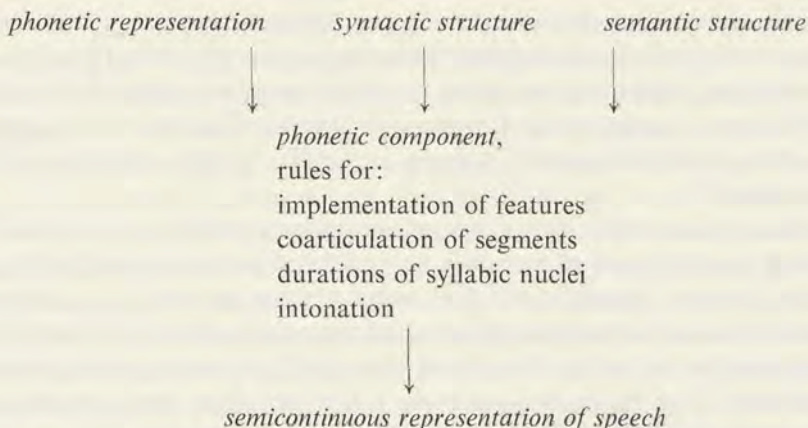
What is lacking in the Chomsky and Halle approach is the following idea, expressed by Gunnar Fant: "Before we can accomplish the happy marriage between phonology and phonetics we have to work out the rules for predicting the speech event given the output of the phonological component of the grammar", and "the derivation of the rules of the 'phonetic component' of language aims at describing the speech production, speech wave, or perception correlates of each feature given the 'context' in a very general sense of co-occurring features within the phonological segment as well as those of following and preceding segments" (ref. 99, p. 221).

Fant specifically mentions "rules for modifications dependent on stress patterns, intonation, tempo, speaker, sex, type and dialect, attitude, etc. Rules for speech segment durations and sound shapes have to be expressed in terms of larger phonological segments, generally several syllables defining a natural rhythmical unit in terms of stress and intonation" (ref. 99, p. 222).

According to Fant such rules belong in a "phonetic component" which has no place in the theory of universal phonetics as seen by Chomsky and Halle. The idea that a phonetic component should be described linking the output of phonology to a specification of the acoustic waveform has also been expressed by others, cf. Ladefoged<sup>101</sup>), Öhman et al.<sup>102</sup>), Tatham and Morton<sup>103</sup>), Lieberman<sup>104</sup>).

One may notice that such a phonetic component readily seems to take the form of a speech-production model. So Ladefoged proposed to include in linguistic theory a specification of a speech synthesiser (ref. 101, p. 58) and Öhman et al. wrote that "the phonological output of generative grammar may be regarded as an abstract description of a set of input commands to a speech synthesiser. This synthesiser should be a true model of human speech production" (ref. 102, p. 15). Tatham and Morton, in the same line, are concerned with "establishing a model of speech production as the output of a generative grammar" (ref. 103, p. 39).

What seems to be needed, however, is not a phonetic component which relates a phonetic representation to an acoustic signal. This leaves us with the phonetic representation as the lowest level of representation where language structure can be specified, regardless whether a speaker is actually speaking or not. We must rather look for a phonetic component which has essentially a status similar to the phonological component, and which specifies the relation between the phonetic representation and another mental representation of speech, which is not in terms of phonetic segments, but rather semicontinuous in nature, and closer to the acoustic signal. In the rules generating this mental representation of the more or less continuous properties of speech, units such as morphemes, words and small word groups, may play a part, but also units corresponding to portions of speech which are louder as opposed to portions of speech which are less loud, or in articulatory terms, units corresponding to the intervals of time the mouth is open as opposed to the intervals of time the mouth is closed. The semicontinuous mental representation of speech may be derived from the phonetic representation plus the rules for intonation, for prosodic duration, for overall loudness etc., plus information on the syntactic and semantic structure of the phrase concerned, as is shown schematically in the following diagram:



That the phonetic representation alone does not suffice as an input for the phonetic component will not be argued extensively here. Phonetic arguments concerning intonation can be found in Collier<sup>105</sup>). Bresnan<sup>106</sup>) has shown in a purely linguistic way that the surface structure is insufficient for generating many neutral stress contours in American English.

We feel that rules for modifications dependent on speaker, sex, type, attitude, etc. need not to be included in the phonetic component, as suggested by Fant (ref. 99, p. 222). The same idealisation which holds for all components of language can be applied to the phonetic component. The rules of the phonetic component may be closely linked to the mechanisms of both speech production and speech perception, but they cannot be identified with these mechanisms. The phonetic component is a model of an important part of the knowledge a language user has about the way speech of his language sounds, and this knowledge is present also when no production or perception is going on. The confusion between a model for speech production and a phonetic component of language is very understandable. We may imagine that in actual speech production a perceptual result is created from some linguistic specification of a phrase plus the mechanisms of speech production. It seems at least plausible that the phonetic component derives its properties from what happens in actual speech production, where the link between articulatory properties of speech and perceptual properties of speech is laid. Whether we know how a word or phrase sounds without actually speaking or hearing it has to do with the fact that we are able to speak and hear this phrase. But this does not mean that the rules which tell us how a phrase or word sounds in the absence of a performance act are to be identified with the actual mechanisms of speaking and hearing. The rules of the phonetic component are mental rules. It may be that they closely parallel the actual performance mechanisms, but it may also be that they are more abstract and skip a lot of operational details which have no perceptual results.

The above does not pretend to give much that is new or surprising. We think

that it is essentially in line with the ideas expressed by Ladefoged<sup>101)</sup> and by Fant<sup>99)</sup>. It has also something in common with the ideas of Lieberman, who recently proposed a unified phonetic theory "in which the ensemble of phonologic features reflects the constraints imposed by the human vocal tract and the human perceptual system. The physical basis of particular features are 'matches' to auditory detectors and to articulatory manoeuvres that are 'easy' to effect" (ref. 104).

Furthermore it is unfortunately the case that the formal properties of the phonetic component are unknown and it seems certain that they will be very different from the formal properties of other parts of grammar. This must be so because of the semicontinuous nature of the output. Perhaps the closest we can come at present to generating an output with such properties is using a speech synthesiser. But it must be kept in mind that the mental representation of speech is not the same as an acoustic waveform and there certainly are acoustic properties of speech which are not part of the mental representation.

As long as no formal phonetic component has been designed, the ideas concerning such a phonetic component are not much more than a long-term goal which forms a convenient frame of reference for discussing a number of phonetic properties of speech which do not belong in the phonological component of generative grammar. It seems possible, however, to state some properties which the phonetic component should have. The phonetic component should at some stage reorganise the sequence of speech sounds, i.e. the columns of features in the phonetic representation, into a semicontinuous representation of speech, presumably with all the coarticulation phenomena accounted for. We think that it is this mental representation we have been investigating in the adjustment tests. The result of this reorganisation is, among other things, an alternation of syllabic nuclei and non-nuclear portions, more or less corresponding to intervals the mouth is open and intervals the mouth is closed. This sequence of nuclear and non-nuclear portions is subjected to rules which assign a duration to each of them on the basis of a number of factors such as the quantity of the segments going into these portions, stress and position in word and phrase, the articulatory properties of the portion itself and the adjacent portions. The synchronisation of intonation movements with the other aspects of speech may also take place at this stage. In as far as the durational properties of such portions of speech derive from the organisation of articulation they can perhaps most easily be modelled with the help of a speech-production model, but it should again be recognised that then it must be assumed that knowledge concerning the perceptual results of the working of speech production is present in the language user's brain. We may also think of abstract rules associating a particular durational change with a particular phonetic environment.

The above ideas may have some consequences for the organisation of a synthesis-by-rule system. If it is correct that there are portions of speech made up

of more than one phonetic segment, but none the less behaving as a unit with respect to rules of duration, it should be useful to include this knowledge in a synthesis-by-rule system and create an intermediate level where the sequence of input segments, in most cases corresponding to the speech sounds of the language, are reorganised into such units. Not using this intermediate level would possibly make the rules for generating the correct durational build-up needlessly complex. Henceforth, when speaking of vowel durations, we do not mean durations of vowel phonemes but durations of the syllabic nuclei associated with the vowel phonemes.

#### **4.2. Accuracy in production and perception of vowel durations and the coding of timing information in the brain**

In the articulatory measurements it was found that subjects are able to repeat the same articulatory programme with a fairly high degree of accuracy. Standard deviations between 5 and 10 ms for speech-segment durations are not abnormal and sometimes standard deviations below 5 ms are found. This shows that speakers have a good command over the control of timing in speech.

In the adjustment test described in sec. 3.3 it was found that the reproducibility of a vowel adjustment shows an accuracy of the same order of magnitude as that found in production. It seems not unreasonable to assume that these two accuracies, the one in production and the one in perception are in some way connected. The more so because it was found that the subject who showed the highest accuracy in production also showed the highest accuracy in perception. The relation between accuracy in the time domain in production and perception is open to further investigation. Further research on this question may be directed towards the correlation between timing accuracy in spoken repetitions of one word or phrase and timing accuracy in durational adjustments in the same word or phrase. In order to make the two tasks comparable use should be made of recordings of spoken versions in the adjustment task. At present the set-up for controlling the duration of a speech segment in a recording of real speech with an external knob is not available to us.

It may be of interest to speculate on the relation between accuracy in production and accuracy in perception. There are principally three possibilities. The accuracy in production may be derived from the accuracy in perception. One may imagine that there is no need for the articulatory system to be any more precise than the perceptual system. Particularly, it may be the case that the accuracy in production is derived from the accuracy in perceptual feedback. If this were the case one would never expect accuracy in production to exceed the limits of accuracy in perception. A second possibility would be that accuracy in perception derives its limits from accuracy in production. This would be the case specifically if the perceptual mechanism operated to a high extent with reference to the production mechanisms as suggested by the supporters of the

motor theory of speech perception (Lieberman et al. <sup>107,108</sup>). In that case one would never expect accuracy in perception to exceed the limits of accuracy in production.

Our data weakly support the first of these two possibilities as the highest accuracy in all our experiments was found for perception, viz. in the adjustment accuracy of subject JtH. But further study of the limits of accuracy of production and perception could possibly give other results, particularly because our production experiments were not designed for studying accuracy and were always done with more than one word type in an experimental session, so that there may have been interference between different word types leading to a loss of accuracy.

There is still a third possibility. It may be the case that accuracy in both production and perception derives from the same central mechanism for handling and storing timing information which is not specific to the task involved. It seems reasonable that such a mechanism is present anyway as timing seems to be involved in all human motor and perceptual skills. It might be the case, however, that such a central mechanism has an accuracy which is much higher than the one found in a particular task, because this task does not require the highest possible accuracy. Thus the presence of such a central mechanism does not principally exclude our first two proposals. Assuming for a moment that it is the central mechanism for timing which determines the accuracy in production and perception, one would expect the measurable accuracy in production to be somewhat lower than that in perception, because in production variability would be added on the way from the central brain mechanisms towards the articulatory movements due the grossness of the articulatory structures (and measuring devices). On the perception side there seems to be no need to assume an extra variability of the same order of magnitude. Likewise our results do not conflict with the assumption that the accuracy both in production and perception derives from the accuracy of a central mechanism for the handling of timing.

We will now briefly discuss the possible forms the coding of timing information in the brain may take. Throughout this study we have adduced evidence that we cannot view the process of speech production as a sequence of articulatory targets which is run off at a certain rate. We must assume that intervals between subsequent articulatory events are calculated in the brain from a number of different factors, such as vowel quantity, stress and syllable position, etc., as expressed in the rules found. In a recent paper George Allen <sup>109</sup>) has proposed two possible forms this timing information could take in the brain. Using the duration of a vowel as an example Allen states as a first possibility that the command to produce a vowel articulation could include the information: "simultaneous with the start of the commands for this vowel, send out a neutral impulse along a nerve-net type loop, known to the motor control program, specific to vowels with this particular duration; continue to issue com-

mands for this vowel production until impulses arrive back on the return branch of this loop". According to Allen "the neural loop would act like a delay line, with each class of articulatory durations having its own fixed delay between initiation and cessation of neuromotor command".

The second possible model for timing information in the brain involves a clock and a count-down number. The vowel command would include the instruction "simultaneous with the start of the commands for this vowel, begin counting down  $n$  cycles of the speech time clock; when  $n$  cycles are complete, the vowel is complete". The number  $n$  would be computed by the brain for each articulatory duration.

Because the segment durations in speech depend on a number of different factors apart from tempo, the number of possible vowel durations controlled as such by the brain form an essentially continuous scale of possibilities. For the delay-line model this necessitates an "unreasonably large number of delay lines, organized in a complicated switching network controlled by stress, word structure and tempo".

Such difficulties do not exist in the clock count-down model, which can easily accommodate the known data on segment-duration variability. As Allen says "the speech motor control program would accept as inputs the intrinsic segmental duration along with information about stress, number of syllables in the word, and overall rate, and from this information compute the single count down number,  $n$ . The same clock would be referred to in all cases, since this model assumes for each articulation a memory buffer keeping track of the count. Timing information could be simultaneously available to all such buffers without causing interaction and confusion among them". Thus Allen opts for the clock count-down model of timing control. In support of his idea of digitised time in motor-perceptual tasks Allen cites evidence from a number of experiments on time in motor and perceptual behaviour. Models involving time quanta have been developed by Creelman<sup>110</sup>), Treisman<sup>111</sup>) and Michon<sup>112</sup>). Treisman suggests a "pace-maker", or clock-like generator, whose rate may be subject to small errors; Creelman suggests a Poisson source and says that "no constantly running 'internal clock' will account for the data" of his experiments. Michon assumes a pace-maker source whose rate of pulsation is highly task-dependent, to account for a variety of periodicities in his data. These data stemmed from an experiment in which subjects had to tap a key in synchrony with an external click train, the frequency of which was either constant or modulated.

The clock count-down model, whatever its physiological plausibility may be, offers an interesting possibility of modelling the storing of durational information in the brain. A third possible model, not mentioned by Allen, is what one could call the condenser model. Its essential idea is that a duration may be stored in terms of the interval of time a given condenser with a given charge



needs to discharge to a certain threshold level. Thus each unit to be assigned a duration is assigned a charge for a condenser, and durational rules in effect would be operations on the charge of condensers. This idea was mentioned to us by Huggins (personal communication). In fact, our hardware synthesiser controls segment durations in just this way.

At present there seems to be no valid reason to accept or reject any of these models as a possible mechanism for the storage of timing information in speech behaviour. Even the rejection of the delay-line model by Allen seems unjust. It is not necessary to assume that each possible duration from an infinite number of possible durations has its own delay line. In the extreme it is even thinkable that one delay line forming a closed loop serves as the clock in the clock count-down model.

At present we can state at least that a model for the storage of timing information must give the possibility of storing a duration with an accuracy of a few milliseconds and must have a higher absolute accuracy for shorter intervals than for longer intervals.

It may be expected that further quantitative research on accuracy in production and perception of speech durations may lead to the formulation of more-specific properties of the mechanism for the storage of timing in speech. Specifically it may be the case that it will be possible to derive conflicting and testable hypotheses from the three models proposed.

#### **4.3. On the perceptual knowledge of physiologically conditioned effects**

In sec. 3.4 we have suggested that it may be the case that small and presumably universal, physiologically conditioned effects on vowel durations are in some sense, as far as their perceptual results are concerned, known to the language users. Evidence of this would be that subjects bring such effects about in adjusting vowel durations. Unfortunately, in the case of the durational effect of vowel height, which has been explained from a universal property of speech, our attempt to show that this effect belongs to the implicit knowledge language users have about speech, failed. We hope that future attempts in this direction will be more successful.

Still it may be of interest to speculate on the theoretical importance of the idea that universal physiologically conditioned effects on vowel duration are known to individual language users as part of their implicit knowledge of the sound pattern of their language. The existence of such knowledge might provide a powerful explanation for the origin of some language-particular phenomena. Once such small universal effects are known by the individual speakers of a language, in principle it is possible that they will be exaggerated and acquire a language-particular function in linguistic communication. Perhaps an example may be vowel lengthening before voiced stops in English. This is specific to English and may be considered part of the phonology of English (House<sup>93</sup>)).

It could be related, however, to a universal effect of voicing of the following consonant on the vowel duration which is much smaller. That it is not unlikely that such an effect exists may be concluded from the fact that a much smaller difference in vowel duration before voiced and voiceless consonants was found for a number of languages, viz. for Spanish (Zimmerman and Sapon<sup>113</sup>), for French (Delattre<sup>34</sup>), for Swedish (Elert<sup>15</sup>), for Dutch (Slis and Cohen<sup>16</sup>). We may speculate that in these languages the lengthening of a vowel before a voiced consonant is caused by some universal properties of human speech. Thus we may assume that in the historical development of a language there has been an interaction between knowledge concerning universal aspects of speech and knowledge concerning language-particular aspects of speech. If such interaction could be shown to take place it could provide an explanation for the naturalness of certain language-particular phenomena.

#### 4.4. On the comparability of vowel quantities in different languages

##### 4.4.1. *The specification of Dutch vowel quantity on the level of phonetic representation*

As vowel quantity seems to be one of the main determinants of measurable durations of syllabic nuclei in Dutch we have taken pains to find out how vowel quantity may be organised on the linguistic level of phonetic representation. In order to do this we have assumed, notably in sec. 2.4 of this study, that in the actual production of speech there is some brain level corresponding to the linguistic level of phonetic representation, and furthermore that, if we minimised all disturbing effects of lower levels, the phonetic organisation of quantity categories would become apparent in the measurable durations. The data obtained in the articulatory measurements can be explained satisfactorily by assuming that on the level of phonetic representation there are only two possible specifications of quantity in Dutch. It was specifically found that the vowels [u, y, i] which traditionally have caused Dutch phonologists trouble in describing their quantity relation to the other vowels, behave as phonetically short vowels before [p] and [t] and as phonetically long vowels before [r].

It seems reasonable to assume that they generally behave as phonetically short, except before [r] and perhaps [l].

Our results concerning vowel quantity have an implication for the linguistic description of Dutch vowels, as far as such a description were to provide a basis for comparing the phonetic properties of languages, and in this case specifically for comparing vowel quantity in different languages.

Chomsky and Halle state that "whereas in the representations that constitute the surface structure (the output of the syntactic rules), specified features will be marked as plus or minus, the phonological rules, as they apply to these representations, gradually convert these specifications to positive integers" (ref.

29, p. 65). In most of their examples they do not actually use this facility, but it should be there if at least "theoretically" the theory provides us with the possibility to compare the phonetic representations taken from different languages. Thus in the case of quantity differences in different languages, in the phonological representation it suffices to distinguish between long and short vowels, or more precisely, between vowels which are + long and vowels which are -long (the feature *long* here is not to be identified with the Chomsky and Halle prosodic feature of *length*, it rather is a segmental feature resembling very much the Chomsky and Halle feature of *tenseness*). In the phonetic representation, however, this will not sufficiently define the way vowel quantity is used in different languages.

Both the absolute and the relative durations of long and short vowels can be rather different for languages having vowel quantity. This could be described by assigning numerical values (positive integers) to the quantity feature. These numerical values can differ from language to language. Our Dutch data could be explained by assuming only two of such values, but in principle it is possible that more than two of such numerical values may be necessary to account for the way the vowels of the language are realised with respect to articulatory timing. Thus it might be the case that in Swedish the quantity specifications for the [u, u:] vowel pair are different from those for other vowel pairs (see sec. 2.5.1).

In order to provide the possibility of actually comparing vowel quantity in different languages there should be an agreed way of defining quantity relationship between vowels for different languages. We will explain in the next section why we think that the standard way of doing this in terms of the V/V: ratio is not satisfactory.

#### 4.4.2. *The quantitative relation between long and short vowels*

It seems desirable to express the relation between long and short vowels for a given language quantitatively. Because actual vowel durations are very dependent on such factors as speaking tempo, prominence, stress, position, etc., it seems unpractical to state the quantitative relationship between long and short vowels in terms of absolute durations. A common way of stating this relationship is in terms of the V/V: ratio, defined as  $(V/V:) \times 100$ . A survey of the V/V: ratios in a number of languages is given by Elert (ref. 15, pp. 109–113). From Elert's discussion of the literature it becomes apparent that the V/V: ratio is not a fixed measure for each language with a short-long opposition. For example Abramson (ref. 114, pp. 81 ff.) found for Thai that the V/V: ratio of vowels spoken in isolation is different from the V/V: ratio of vowels in running discourse. Data taken from Meyer and Gombocz<sup>115</sup>) and from Lazicsius<sup>116</sup>) show that in Hungarian the V/V: ratio in monosyllables is different from this ratio in disyllabic words. Stefán Einarsson<sup>117</sup>) found in Icelandic single words

a V/V: ratio of about 53%. Bergsveinsson<sup>118</sup>) found that in Icelandic in connected speech the V/V: ratio of stressed vowels was 83%, but in unstressed vowels there was practically no difference. Elert's own data show that the V/V: ratio in Swedish is different for the various long-short vowel pairs. In single words it varies from 59% for the long and short allophone of /ö/, /y/ and /a/ to 83% for the long and short allophone of /u/. It may be noticed here that in Swedish long and short allophones differ considerably in quality. The short allophone of /u/ is much opener than the long one, which may explain its relatively long duration.

Lehtonen (ref. 119, p. 90) found for Finnish in a series of two-syllable and three-syllable words of varying phonemic structure V/V: ratios ranging from 1 : 1.38 (= 72%) to 1 : 3.19 (= 32%), dependent on both syllable position and phonemic make-up of the rest of the word.

Delattre and Hohenberg (ref. 30, pp. 380-381) report that the V/V: ratio for unstressed syllables in German, although different from the V/V: ratio in stressed syllables, is constant for the various vowel pairs, and for the initial, medial and final syllable position in three-syllable words. Their results are difficult to compare with the results of other studies, however, because they defined vowel durations in terms of steady-state portions in spectrograms.

The overall conclusion of the observations presented above may be that the V/V: ratio as commonly used in the literature is not a fixed property of a given language but is dependent on such factors as utterance length, word length, stress, vowel quality.

For comparing V/V: ratios of different languages these ratios should be measured in sharply defined conditions which are identical for the languages under comparison. Very often the V/V: ratios to be found in the literature are not comparable.

But even if one strictly defines the conditions under which the V/V: ratios of the languages one wishes to compare will be measured, this will not give complete information on the relation between long and short vowels. If, for example, one measures the durations of long and short vowels in monosyllables, keeping phonemic environment and speaking tempo constant as closely as possible, one gets a certain value for the V/V: ratio. This ratio might accidentally be exactly the same for the two different languages. Still in that fictitious case, the relation between long and short vowels taken in a broader sense can be very different in these two languages, e.g. because the effect of stress and position on vowel duration is very different for the two languages. A more complete description of the relation between long and short vowels in a given language would consist of some abstract quantity specifications plus a set of rules which govern the actualisation of vowel durations under different conditions of stress, position within the word, position within the sentence, speaking tempo, etc. It may be noticed that it is not necessary for the abstract specifications to be identical with

some really measured durations. It would be sufficient if these abstract specifications plus the set of rules would explain the actual durations measured.

However, as the rules have to be tested in real speech going on in real time, and as the differences between languages in the way they use vowel quantity, may show up in measurable durations, we feel that the abstract pair of values should be given in terms of measurable durations, e.g. in terms of milliseconds. This would provide a healthy basis for comparing languages as to the way they shape the correct durational build-up of words and phrases. We feel that there are real differences between languages in this respect which are not only to be explained as individual differences. For example durations of long vowels which seem to be normal in Swedish, seem to be extremely rare in Dutch.

In the present study we have shown that the relation between long and short vowels in Dutch may be described by two optimal values in the neighbourhood of 200 and 100 ms respectively, plus a set of quantitative rules for describing the effect of stress and position. We propose that such a description will give a more solid basis for comparing quantity phenomena in different languages than the traditional V/V: ratio. In this way we express the relation between long and short vowels in terms of absolute durations instead of relative durations. Probably our claim is somewhat too strong, because the optimal values will be affected by variations in speech tempo, emotional state, etc. Nevertheless, a description in terms of absolute durations seems more satisfying because the range within which the optimal vowel durations may vary seems to be rather limited, especially for the short vowel.

One may note that the quantitative relation between long and short vowels as defined here does not concern the level of phonetic representation, but rather the level of semicontinuous representation which is thought to be the output of the phonetic component. The specifications of quantity on the level of phonetic representation are underlying determinants of the quantitative relation. The quantity specifications on the level of phonetic representation may be identified with the optimal durations, which are then passed on to the phonetic component where they, together with other determinants and a set of rules, will determine the specified durations of syllabic nuclei.

#### 4.5. On the durational rules for vowels

##### 4.5.1. *The generality of the rules*

In sec. 2.5 we found a number of regularities in the behaviour of vowel durations in nonsense words. These regularities could be described by a simple set of rules. All these rules have the form, inspired by a paper by Lindblom and Rapp<sup>74</sup>), that an optimal vowel duration is divided by a number to a certain power. If we took as the optimal vowel duration the duration of the vowel in

the monosyllable spoken in isolation, the rules would reasonably have fitted the data. But clearly the data were very limited, consisting of a number of nonsense words with bilabial consonants only and with a limited amount of stress patterns. Part of the words was spoken by only one subject, i.e. the author. Thus one could fear that the rules were only of very restricted validity.

In testing the rules in sec. 3.5 we took three other, naive subjects and words of a very different phonemic make-up. It would have been surprising if the rules were found valid without any adaptation. What we did find in the first place was that the optimal durations of 200 and 100 ms fitted the data surprisingly well. The subjects differed somewhat but not very much. It seems to be the case that in neutral realisations of words spoken in isolation our subjects use about the same optimal vowel durations, having values of roughly 200 and 100 ms for long and short vowels respectively. This is confirmed in the spoken versions of the same words. Whether this holds good for Dutch speakers in general is not yet certain.

One may note that we have made an oversimplification in operationally equating the optimal duration with the duration of a vowel in a monosyllable spoken in isolation. The duration of such a vowel depends on more factors than an abstract optimal duration alone. It is well known that the duration of a vowel in a monosyllable depends on the interaction between the vowel and the consonantal make-up of that syllable and especially on the following consonant or cluster. We have deliberately neglected this, but naturally a complete set of rules for vowel durations will have to take this into account.

For testing rule (1) for long vowels we used the syllable [ma:t]. What we found was that this rule ( $DV = DV_{opt}/m^2$ ) gave a useful description of the results obtained in both the adjustment tests and in spectrographic recordings, provided we lowered the value of  $\alpha$ . One possible interpretation of this is that the effect of the interaction between vowel and consonant duration may at least be roughly described by lower-level rules operating on the values of the exponents in the rules for position and stress. This would result in an increased relative effect of consonantal environment as vowel duration decreases. Another possible cause of the discrepancy between the data obtained with nonsense words and those obtained with real words may be the difference in rhythmical pattern. The three- and four-syllable real words had a secondary stress on the third syllable.

The assumption that the consonantal environment affects the exponent of the power functions in the durational rules is weakly supported by the tests of the perceptual reality of the rules for unstressed vowels. In testing rules (2), (4) and (5) we used a bilabial consonant following the vowel and there the prediction of the rules was rather good without changing the exponent. In testing rules (3) and (4) against each other we used a [t] following the vowel and found absolute durations which were rather long. This suggests that the exponent  $\gamma$  in the rules (3) ( $DV = DV_{opt}/2 \cdot 2^\gamma$ ) and (4) ( $DV = DV_{opt}/2 \cdot 4^\gamma$ ) was somewhat lower

than 1, which may have been induced by the [t] environment. This evidence, however, is at this stage insufficient to allow of any definite conclusions.

#### 4.5.2. *The boundary between Dutch long and short vowels*

A particular point of interest seems to be the effect of rule (1) on the durations of long and short vowels. As  $m$ , the number of the syllable counted from the end of the word backward, increases, the duration of the stressed vowel decreases. But when  $m = 4$  the duration of the stressed vowel may already be very near its lower limit. For the short vowel this lower limit may be the shortest vowel duration in that particular consonantal context that is possible for a stressed vowel. For the long vowel, however, this lower limit seems to be about the longest duration of the short vowel. Thus overlap between the durations of stressed long and short vowels seems to be avoided. The boundary between Dutch long and short vowels set by the optimal duration of the short vowel may be an interesting topic for further research. One would like to know whether this boundary in terms of the duration in ms is fixed for a particular speaker or can shift due to changes in tempo. It may also be of interest to see whether the perceptual boundary between long and short vowels can shift under the influence of position in the word. Thus it would be possible to ask listeners to identify the length category of a vowel sound with some quality between that of [a:] and [a], the duration of which is systematically varied. This could be done for a vowel in a monosyllabic and for the vowel in the stressed initial syllable of polysyllabic words. One could expect that the duration where short vowel judgments are replaced by long vowel judgments is somewhat longer for the monosyllabic than for the polysyllabic words. This would mean that the organisation of the whole word affects the perceptual boundary between long and short vowels. Although a recent study by Ainsworth<sup>120</sup>) suggests that speech rhythm may affect the perceptual boundary between long and short vowels, the rather strict separation between long and short vowel durations in our experiments (weakly) suggests that this boundary may be independent of the number of syllables in the word.

#### 4.5.3. *Other determinants of vowel duration*

There must also be other effects on the duration of a vowel than those we have been discussing, for instance it evidently is not the case that one person always speaks the same monosyllable in isolation with the same duration. But we may assume that when pronunciation is neutral with respect to emotional state, required loudness, rhetorical effects, intonation, etc., the duration will always be about the same. When the pronunciation is not neutral with respect to such factors the resulting deviation in principle can be described by a rule operating on the optimal vowel duration. Thus this presupposes that each speaker has only one neutral vowel duration for each monosyllable. Whether

this is a very realistic assumption or not, it certainly simplifies a quantitative formulation of durational rules.

For some of such factors it will be difficult to describe their effects in quantitative rules, because they are difficult to investigate under laboratory conditions, and besides, especially as regards the emotional factors, their effect may be rather unsystematic. Some of them seem to lend themselves rather well to systematic description. We think specifically of the interaction between duration and intonation. The interaction between intonation and vowel duration has not been studied systematically for Dutch as far as we know. The method of matching to internal criterion seems to be particularly suited for this, by asking subjects to adjust vowel durations and systematically varying the intonation of the stimuli. This seems a promising and important line of future research.

Another factor which may affect vowel duration is speech tempo. Although it is known that the main difference between slow, fast and normal speech is in the number and durations of the speech pauses, none the less tempo may also affect articulatory durations. The rules for speech tempo may be different for different languages. According to Kozhevnikov and Chistovich in Russian the relative durations of syllables are preserved when absolute durations are changed due to tempo (ref. 27, p. 89). Within the syllable, segmental durations do not keep their relative durations constant. For English it has been found that stressed syllables are less affected by tempo changes than unstressed syllables (Lehiste, ref. 19, pp. 39–40; but note that the conditions were rather artificial in that subjects had to speak in synchronism with a periodic pulse which had to coincide with stressed syllables).

All these factors which may affect vowel durations can be studied in relation to the rules already formulated. For a given factor, let us say speech tempo, it may be found that it affects the optimal durations and leaves the rules intact. It may also be found that it does not affect the optimal durations — or if so, only slightly — but does affect the exponents of the power functions in the rules drastically. A third possibility is that the rules as formulated are not valid at all under the influence of such a factor. But in all cases the rules for vowel durations may provide a convenient starting point for further research.

#### 4.5.4. *On the universality of the rules proposed*

The question arises whether the phenomena described by rules (1)–(5) have to be accounted for in a complete linguistic description of the Dutch language. An alternative is that these phenomena are in some sense universal and thus to be accounted for in a general phonetic theory. Let us consider the effect of the number of syllables which remain to be produced in the word. This effect cannot be universal in the sense that it is present in all languages of the world. There is evidence that in Finnish, for instance, the number of syllables following in the word has no effect on vowel duration (Lehtonen, ref. 119, pp. 138–139).



On the other hand there is evidence of the manifestation of this effect in so many widely diverging languages that it seems at least plausible that it is closely linked to universal tendencies of human speech. It has been found for German (Meyer<sup>67</sup>), Lappish (Äimä<sup>121</sup>), French (Roudet<sup>66</sup>), Hungarian (Tarnóczy<sup>73</sup>), English (Jones<sup>70</sup>), Swedish (Lindblom<sup>63</sup>), Lindblom and Rapp<sup>74</sup>).

We propose as a possible view that this durational effect follows automatically from the universal tendencies of human speech, given the requirements on the durational build-up of the language in question. These requirements are different for different languages and, in particular, may be different in a language with phonemic quantity from those in a language without phonemic quantity. They may modify or "filter" the effect of the universal durational tendencies. In a language with a high functional load on quantity patterns such as Finnish these tendencies may even be completely or almost completely counteracted. This view makes it possible to assume that the effect of word length on the duration of a stressed vowel is different for different languages and sometimes even absent and yet has not to be accounted for in the descriptions of the particular languages. It presupposes, however, that the effect of word length can be predicted once the phonology of the language is described. This has not been attempted.

The hypothesis of the universality of the effect implies the hypothesis that the effect is in some way determined by innate properties of the human-speech-producing system. If independent evidence for the existence of such properties could be adduced this would lend strong support to the hypothesis of universality. Such evidence is not available. The same applies to the durational effects in unstressed vowels. The universality of these effects seems to be even less well established than that of the position effect for stressed vowels, but still, it might be the case that given the presence of stress in a language, the phenomena described by rules (2)–(5) automatically result from the universal organisation of human speech. We do not know and leave the question open for further study.

Independent of the questions as to universality is the need for the inclusion of a quantitative specification of the effects concerned in a synthesis-by-rule system for a particular language, as long as we do not have at our disposal a synthesis system with all the inbuilt properties of human speech.

#### 4.6. On pitch accent, stress and duration

A number of investigators have shown that pitch and particularly pitch variation is the main cue for what normally is called "stress" (for English this has been shown by Fry<sup>122</sup>), Bolinger<sup>123</sup>), Lieberman<sup>124</sup>), Jassem, Morton and Steffen-Batóg<sup>125</sup>); for Dutch this has been shown by Van Katwijk and Go-vaerts<sup>126</sup>). What is meant by "stress" here may be interpreted as perceptual

prominence. Throughout this study we have, in line with Bolinger's terminology, spoken of pitch accent in the case of stressed syllables made perceptually prominent by pitch variations and of stress in the case of word stress or lexical stress. Thus the class of lexical stresses in a phrase contains a subclass of lexical stresses with a pitch accent, and it is members of this subclass which are normally called "stressed" in articles dealing with stress perception.

It has been shown that in Dutch the pitch variations due to pitch accents are important parts of the intonation patterns (Cohen and 't Hart <sup>62</sup>) and members of the set of possible pitch variations seem to obey their own rather strict concatenation rules (Collier and 't Hart <sup>127</sup>), Collier <sup>105</sup>).

In this study we found that in order to assign the correct duration to a vowel we must know whether the syllable is lexically stressed or not. It does not seem to make much difference whether the syllable has a pitch accent or not (see sec. 2.4). Thus the pitch accent seems to be superimposed on the durational realisation of lexical stress. In order to assign both the correct vowel durations and the correct pitch variations to the stressed syllables in a phrase we must know of all words in the phrase which syllable bears lexical stress and furthermore we must know which words are "dominant", i.e. bound to be intonationally marked. The dominance of a word may in principle follow from the syntactic and semantic structure of the phrase. As yet the rules for predicting the dominance of a word have not been formulated. The lexical stress in a word may either be lexically given, or follow from stress-assignment rules. The latter have not been formulated for Dutch.

We have not dealt with compound words. In compound words more than one syllable may bear lexical stress. For example, in the word *spraaksynthese* [sprɑ:kstɪntɛ:zə] (speech synthesis) both the [ɑ:] and the [ɛ:] bear lexical stress, and both have to be assigned the duration of a stressed vowel (given the position in the word). If in speech synthesis only the [ɑ:] is assigned the duration of a stressed vowel the word will sound like [sprɑ:kstɪntɪzə]. The difference between the first and the third syllable of the compound word *spraaksynthese* is that the former is a potential bearer of pitch accent and the latter is not (we do not consider here the special case of contrast accent). Thus when the word is spoken in isolation the first syllable normally is intonationally marked with both a rise and a fall and the third syllable is not intonationally marked. When the word is spoken embedded in a phrase and is not dominant, for instance because the word has been recently mentioned in the conversation, neither of the stressed syllables is intonationally marked. The question arises whether in such cases the stresses differ in degree or not. Thus it might be the case that the vowel in the syllable with primary lexical stress has a longer duration than the vowel in the syllable with secondary lexical stress, other things being equal. This we do not know, but as a first approximation assigning both syllables the same duration will give reasonable results in speech synthesis.

In our investigations we did not separately study words from different word classes. In practice this meant that we only dealt with words behaving like semantically important words such as substantives and adjectives. It seems to be the case that in order to assign the correct segment durations we should minimally distinguish between "important" and "unimportant" words. In "important" words (at least) one syllable is assigned lexical stress (Slis<sup>81</sup>). Unimportant words are mainly function words. The term "important" should not be confused with "dominant". "Important" words may be either "dominant" or "non-dominant". As indicated by work on American English (Flanagan et al.<sup>128</sup>), Klatt<sup>80</sup>) a further differentiation into more than two categories may be needed, but as yet systematic studies of the regularities involved are lacking for Dutch.

#### 4.7. Other applications of the method of matching to internal criterion

Looking at the results of our adjustment tests one might get the idea that they are in some way trivial. Why after all should a subject not know how a word with its complete durational build-up should sound? Still we had not expected to get such clear-cut results, precisely because on a conscious level a subject does not seem to know how a word should sound with all its durational details. It is only in the adjustment task where he is thrown back on turning a knob to the right or to the left and finding a setting which leads to the most natural realisation of the word, that he gives away an aspect of how a word should sound, and even then, in most cases, he is aware only of the naturalness of the whole word, not of the duration of the vowel. The results of the adjustment tests, then, show that the method of matching to internal criterion may successfully be used to study the implicit knowledge language users have with respect to those aspects of speech which are not easily studied by introspection.

As the results of the tests on the whole show that the subjects fairly precisely duplicate all durational effects which are found in production, one may conclude that some small durational effects, which perhaps may be attributed to the organisation of the production of speech, are part of the perceptual knowledge, and probably may also play a role in the decoding of speech. For many other small effects such as the effect of vowel height on vowel duration this has still to be shown. In principle the method of matching to internal criterion seems to be very well suited for the investigation of such questions. In the same way it may be of interest to see whether the adjustments of consonant durations would give essentially the same kind of results as the adjustments of vowel durations. It has been found by Huggins<sup>65</sup>) that subjects are less sensitive to changes in consonant duration than to changes in vowel duration. It may be of interest to see how this would affect their behaviour in an adjustment task.

The adjustment method is in principle not limited to durational aspects of speech. It has already been used for studying spectral characteristics of isolated

vowels (Blom and Uys <sup>90</sup>) and for the timing of pitch movements (Collier <sup>89</sup>). It may also be used, for example, for studying such perceptual cues as the direction and speed of formant transitions. These are often particularly difficult to study in real speech but may be rather easily studied if they are made controllable continuously with one or more external knobs. We anticipate that this would offer a means of exploring the perceptual relevance of such cues much faster than is possible with the traditional method of preparing stimulus tapes and eliciting responses from the subjects.

On the whole we think that the present results of our perceptual experiments, though themselves rather modest, give good reason to believe that the method of matching to internal criterion may be useful in several interesting ways in future speech research. If it were possible to apply the method to high-quality speech instead of the synthetic speech of somewhat poor quality we used, probably subjects will be even more sensitive to minor changes and the method will gain in power. The results obtained in this way may in turn prove to be useful for improving synthetic speech.

#### 4.8. Summary

This chapter contained some interpretations and speculations on the durational phenomena we have been exploring in chapters 2 and 3. What we normally call a vowel duration is, in fact, not the duration of a vowel phoneme, but rather the duration of a syllabic nucleus which may consist of more than one phonetic segment. This syllabic nucleus has no place in the level of phonetic representation as defined within generative phonology. We suggest that the syllabic nucleus corresponds to a relatively loud portion in a semicontinuous representation of speech, which is the output of the phonetic component of language. This semicontinuous representation is a mental representation and not to be identified with the acoustic waveform, although it is much closer to the acoustic waveform than the phonetic representation is. The phonetic component takes as its input the segmental phonetic representation, the syntactic structure and the semantic structure of a phrase. It contains rules generating the effects of implementation of features, coarticulation, duration and intonation. The present study has been concerned with the discovery and formulation of durational regularities which have to be accounted for in the phonetic component of language.

The results of our measurements concerning accuracy in the production and perception of durations in speech suggest that these two accuracies are in some way related. This may be so in one of three ways. Accuracy in production may derive its limits from accuracy in perception, accuracy in perception may derive its limits from accuracy in production and accuracy in both production and perception may derive its limits from a central mechanism for the storage and

handling of timing information. Possible forms of such a mechanism are discussed.

Some of our results suggest that language users have an implicit knowledge of small, universal, physiologically conditioned effects on vowel duration. Interaction between such knowledge and knowledge concerning language-particular aspects of duration may provide an interesting explanation for the naturalness of some language-particular phenomena, such as the lengthening of vowels before voiced stops in English.

Vowel quantity should be specified on the level of phonetic representation. For Dutch two possible specifications of the quantity feature would suffice. In order to compare languages with respect to vowel quantity these specifications should be positive integers which may differ from language to language.

The traditional way of comparing languages with respect to vowel quantity by means of the V/V: ratio is not satisfactory. A more satisfactory method of comparison would be obtained by a pair of abstract optimal vowel durations plus a set of rules for generating vowel durations under different conditions of stress, position, tempo, etc. The vowel-quantity specifications on the level of phonetic representation could be identified with these optimal vowel durations.

In the experiments described in this study the effect of consonantal environment on vowel duration has been neglected. The assumption that the consonantal environment affects the exponent of the power function in the durational rules for stress and position may provide an interesting basis for future experiments.

Our results have shown that there seems to be a fairly fixed boundary between the duration of long and short vowels. Future experiments may show whether the durational value of this boundary may be affected by such factors as the number of syllables in the word.

Determinants of vowel duration not discussed in this study include a variety of factors such as emotional state, required loudness, rhetorical effect, intonation. The effect of some of these factors will be difficult to describe quantitatively, because it may be rather unsystematic. The interaction between duration and intonation seems to lend itself rather well to systematic study, particularly with the method of matching to internal criterion. The effect of speech tempo may also be described by quantitative rules. The rules for vowel durations as proposed in this study may provide a convenient starting point for further research on systematicity in vowel durations.

The rules for vowel duration are to be included in a description of the Dutch language in as far as they describe language-particular phenomena. It may also be the case, however, that these rules describe regularities in vowel durations which automatically follow from the universal properties of human speech, given the requirements of the language on the durational build-up of speech. Such requirements may, as it were, modify or "filter" the effect of the universal

tendencies and perhaps even completely counteract them. At present we are not able to explain the durational regularities found from universal properties of human speech.

The method of matching to internal criterion, used in this study for assessing the perceptual reality of the durational effects found in the articulatory measurements, has proved to be useful in this respect. The same method may be applied to a number of different problems and provide much information on the perceptual reality of acoustic cues rather quickly. This may be particularly interesting for research directed towards improving the rules for speech synthesis. An improved speech synthesis may, in turn, provide us with a better tool for studying the phonetic component of language.

## APPENDIX A

**Tables 1-20 with data from the articulatory measurements****Tables 1.A-5.A**

Durations in ms as defined by the moments of lip opening and closure in the speaking of nonsense words. The data are presented for three or two subjects separately. In the first columns the actual forms of the nonsense words are given. Columns  $n$  give the number of times the word has been spoken by the subject. Columns  $p_1$ - $p_4$  and  $V_1$ - $V_3$  give the mean durations over  $n$  occurrences of the periods, the lips are closed and open, respectively. Columns tot. give the mean total duration of the period between the first closure and the last opening. All mean durations are followed by their standard deviations in the columns sd.

**Tables 1.B-5.B**

Durations in ms of the intervals between lip opening and voice onset and between lip closure and voice end, in the speaking of nonsense words. The moments of lip opening and closure have been measured by means of a lip contact, the moments of voice onset and end by means of a throat microphone. The data are presented for three or two subjects separately. In the first columns the actual forms of the nonsense words are given. Columns  $n$  give the number of times the word has been spoken by the subject. Columns  $p_1V_1$ - $p_3V_3$  give the mean durations of the voice onset time in the three syllables. Columns  $V_1p_2$  and  $V_2p_3$  give the mean durations of the voice tail after the first two vowels. All mean durations are followed by their standard deviations in the columns sd.

**Tables 6-8**

Durations in ms as defined by the moments of tongue opening and closure in the speaking of nonsense words. The data are presented for two subjects separately. In the first columns the actual forms of the nonsense words are given. The columns  $n$  give the number of times the word has been spoken by the subject. Columns  $t_1$ - $t_4$  and  $V_1$ - $V_3$  give the mean durations over  $n$  occurrences of the periods the tongue makes contact with the palate and the periods it does not respectively. Columns tot. give the mean total duration of the period between the first closure and the last opening. All mean durations are followed by their standard deviations in the columns sd.

**Tables 9-10**

Durations in ms as defined by the moments of tongue and lip opening and closure in the speaking of nonsense words. The data are presented for two

subjects separately. In the first columns the actual forms of the nonsense words are given. Columns  $n$  give the number of times the word has been spoken by the subject. Columns  $p$  give the mean durations of the periods the lips are closed, columns  $Vr$  give the mean durations of the periods between lip opening and tongue closure, columns  $t$  give the mean durations of the periods the tongue makes contact with the palate. Columns  $tot.$  give the mean durations of the periods between first lip closure and last tongue opening. All mean durations are followed by their standard deviations in the columns  $sd.$

#### Table 11

Durations in ms as defined by the moments of lip opening and closure in the speaking of three-syllable nonsense words with varying stress placement. The words were spoken in embedded position with pitch accent on the stressed syllable. In the first column the actual forms of the words are given. Column  $n$  gives the number of times the words have been spoken by the subject. Columns  $p_1-p_4$  and  $V_1-V_3$  give the mean durations over  $n$  occurrences of the periods the lips are closed and open respectively. The columns  $tot.$  give the mean total durations of the periods between the first closure and the last opening. All mean durations are followed by their standard deviations in columns  $sd.$

#### Table 12

As table 11, the only difference being that the nonsense words were spoken without a pitch accent on the stressed syllable.

#### Tables 13-16

Durations in ms as defined by the moments of lip opening and closure in the speaking of nonsense words with varying number of syllables and varying stress placement. The data were obtained for one subject only. The words were spoken in isolation with pitch accent on the stressed syllable. In the first columns the actual forms of the words are given. Columns  $n$  give the number of times the word has been spoken by the subject. The columns  $m_1-m_5$  and  $V_1-V_4$  give the mean durations over  $n$  occurrences of the periods the lips are closed and open respectively. Columns  $tot.$  give the mean total durations of the periods between the first closure and the last opening. All mean durations are followed by their standard deviations in the columns  $sd.$

#### Tables 17-20

As tables 13-16, the only difference being that the nonsense words were spoken without a pitch accent on the stressed syllable.



TABLE I.A

IS	<i>n</i>	<i>p</i> <sub>1</sub>	<i>sd</i>	<i>V</i> <sub>1</sub>	<i>sd</i>	<i>p</i> <sub>2</sub>	<i>sd</i>	<i>V</i> <sub>2</sub>	<i>sd</i>	<i>p</i> <sub>3</sub>	<i>sd</i>	<i>V</i> <sub>3</sub>	<i>sd</i>	<i>p</i> <sub>4</sub>	<i>sd</i>	tot.	<i>sd</i>
pa:pa:pa:p	11	126	12.8	80	9.7	90	7.6	125	11.0	92	7.1	125	11.6	98	5.0	736	30.7
pe:pe:pe:p	17	128	12.3	82	9.6	104	7.8	107	6.2	93	7.0	122	11.4	103	7.4	738	20.9
peipeipeip	15	135	14.5	86	8.2	95	7.5	113	8.6	94	6.9	127	9.7	103	10.7	752	25.8
pepepep	16	126	10.7	75	6.9	103	7.4	80	5.1	101	7.3	93	6.1	106	11.0	683	21.1
pIpIpIp	19	128	11.8	71	7.2	108	7.0	80	3.7	104	11.0	87	6.7	109	9.0	686	25.1
pipipip	17	127	11.4	73	6.1	107	6.9	83	5.8	101	6.5	95	6.1	111	9.2	696	20.9
JtH																	
pa:pa:pa:p	12	132	19.4	79	8.6	97	4.7	135	5.8	95	6.3	136	7.9	107	7.9	781	30.5
pe:pe:pe:p	20	130	18.6	91	6.5	98	6.5	123	5.3	108	5.9	128	7.2	113	6.2	791	33.8
peipeipeip	21	129	12.3	91	8.8	98	5.9	129	7.7	103	7.6	134	6.5	108	8.6	791	29.3
pepepep	18	133	16.6	73	5.1	109	5.7	79	5.7	107	6.6	92	5.7	114	5.9	707	22.0
pIpIpIp	17	129	12.3	68	3.9	111	7.0	72	3.9	112	6.5	83	5.8	117	4.2	693	22.6
pipipip	18	129	16.7	76	5.5	109	6.7	80	3.8	111	6.0	90	5.3	120	6.6	714	28.4
SN																	
pa:pa:pa:p	22	126	12.4	102	11.9	120	8.4	122	11.2	113	4.9	132	7.7	106	10.7	821	46.9
pe:pe:pe:p	16	141	16.5	101	8.0	126	5.2	114	5.1	120	4.7	128	12.0	104	7.0	834	34.7
peipeipeip	20	130	15.0	112	8.1	124	7.2	122	7.7	113	7.1	132	6.3	101	8.8	832	35.0
pepepep	17	137	14.6	79	6.9	142	10.2	82	5.4	136	7.7	95	6.6	123	7.8	794	32.8
pIpIpIp	18	143	12.3	70	9.8	148	12.2	76	6.9	145	10.5	88	11.7	126	11.3	797	38.6
pipipip	16	141	14.0	74	7.3	147	14.0	83	8.5	142	12.5	89	11.7	125	10.3	801	44.3

TABLE 2.A

IS	<i>n</i>	<i>p</i> <sub>1</sub>	sd	<i>V</i> <sub>1</sub>	sd	<i>p</i> <sub>2</sub>	sd	<i>V</i> <sub>2</sub>	sd	<i>p</i> <sub>3</sub>	sd	<i>V</i> <sub>3</sub>	sd	<i>p</i> <sub>4</sub>	sd	tot.	sd
pa:pa:pa:p	20	97	11.6	54	7.6	66	8.5	145	7.5	87	14.4	143	12.3	100	9.7	692	40.6
pø:pø:pø:p	14	94	14.1	72	8.4	70	17.1	145	12.6	77	12.6	153	13.8	97	8.6	708	38.0
p^yp^yp^yp	14	104	12.7	67	9.6	71	8.7	145	13.2	84	8.3	147	14.3	106	14.0	723	24.6
pæpæpæp	12	100	19.7	57	12.2	78	16.8	86	10.5	77	18.5	100	8.2	101	19.1	598	59.4
pyypyyp	18	99	14.4	66	9.4	76	14.6	110	25.5	64	22.6	111	16.6	99	14.1	623	39.9
JtH																	
pa:pa:pa:p	17	99	5.6	75	6.7	84	6.4	137	10.3	85	7.0	133	7.5	99	9.4	711	20.8
pø:pø:pø:p	18	115	11.9	84	7.3	85	6.0	124	7.2	85	5.5	128	7.3	89	7.3	710	21.6
p^yp^yp^yp	22	104	7.7	88	7.9	81	4.8	132	7.0	83	7.7	131	7.4	93	6.4	713	20.4
pæpæpæp	18	113	9.6	66	5.3	94	6.8	80	6.4	88	4.7	80	7.0	95	5.8	616	16.9
pyypyyp	21	123	12.5	76	8.0	96	8.6	81	7.4	89	5.4	86	7.7	96	7.3	647	19.0
SN																	
pa:pa:pa:p	17	123	15.5	88	8.4	110	8.4	131	6.6	107	4.0	140	8.8	106	11.5	803	34.9
pø:pø:pø:p	17	121	14.2	100	18.2	100	10.1	138	9.9	95	8.1	142	11.9	94	12.5	790	28.7
p^yp^yp^yp	20	114	10.1	102	7.9	103	9.5	139	7.7	92	6.1	147	6.4	95	11.4	790	24.5
pæpæpæp	19	112	16.1	89	18.8	108	12.8	89	15.0	98	12.8	119	14.1	101	16.2	725	26.7
pyypyyp	19	116	21.8	94	22.2	106	15.6	111	23.2	96	15.8	124	18.0	100	17.4	747	31.0

TABLE 3.A

IS	<i>n</i>	<i>p</i> <sub>1</sub>	<i>sd</i>	<i>V</i> <sub>1</sub>	<i>sd</i>	<i>p</i> <sub>2</sub>	<i>sd</i>	<i>V</i> <sub>2</sub>	<i>sd</i>	<i>p</i> <sub>3</sub>	<i>sd</i>	<i>V</i> <sub>3</sub>	<i>sd</i>	<i>p</i> <sub>4</sub>	<i>sd</i>	tot.	<i>sd</i>
pa:pa:pa:p	19	95	9.1	72	13.7	91	6.8	113	5.4	82	6.2	137	9.1	99	7.7	690	25.2
po:po:po:p	16	98	7.8	86	18.9	84	7.6	108	8.1	86	5.0	134	7.3	98	8.0	694	27.8
paupaupaup	23	103	8.6	77	12.4	90	7.5	111	6.4	86	5.0	137	7.7	94	6.3	697	23.9
paapaapa	14	96	8.4	65	7.3	102	10.6	71	3.7	102	5.6	90	5.2	107	8.9	631	23.9
paapaapa	19	102	11.3	72	8.8	93	10.8	81	4.7	97	7.0	95	5.6	104	8.4	643	22.3
paupaupa	14	100	11.6	71	6.2	93	7.7	87	6.0	82	5.1	102	8.0	95	9.6	629	18.7
JtH	.																
pa:pa:pa:p	18	109	14.5	92	7.2	94	5.9	134	8.9	84	3.9	136	7.9	97	6.5	747	23.4
po:po:po:p	17	128	18.6	87	6.9	105	4.8	106	12.0	101	5.7	124	15.2	101	7.7	751	39.2
paupaupaup	22	116	14.4	99	8.0	94	4.7	122	7.9	92	5.2	138	6.5	96	7.7	757	26.8
paapaapa	17	111	8.7	78	3.3	107	5.9	81	3.9	97	4.3	96	4.0	102	6.4	670	16.5
paapaapa	17	120	16.0	80	5.3	109	4.4	77	3.1	102	3.5	89	5.4	103	5.4	680	19.1
paupaupa	20	127	12.7	77	7.3	112	7.3	75	4.9	108	3.6	86	5.8	111	5.6	697	20.1
SN																	
pa:pa:pa:p	16	110	12.1	88	5.6	106	3.6	107	9.2	100	4.7	123	7.6	94	5.1	728	21.6
po:po:po:p	12	112	10.0	90	7.4	107	5.4	107	8.9	95	6.1	118	9.0	91	8.7	720	21.2
paupaupaup	18	109	10.7	91	6.8	107	7.8	112	9.3	96	5.5	120	10.8	91	15.2	726	23.2
paapaapa	19	117	17.1	62	5.1	118	5.1	68	5.3	115	5.8	80	7.3	116	9.7	676	21.7
paapaapa	17	119	12.6	67	8.6	119	10.0	68	6.6	112	8.0	81	8.1	118	21.6	683	34.6
paupaupa	11	111	11.4	78	5.7	112	8.3	77	6.7	110	10.7	88	9.3	108	13.1	683	23.5

TABLE 4.A

IS	<i>n</i>	<i>p</i> <sub>1</sub>	<i>sd</i>	<i>V</i> <sub>1</sub>	<i>sd</i>	<i>p</i> <sub>2</sub>	<i>sd</i>	<i>V</i> <sub>2</sub>	<i>sd</i>	<i>p</i> <sub>3</sub>	<i>sd</i>	<i>V</i> <sub>3</sub>	<i>sd</i>	<i>p</i> <sub>4</sub>	<i>sd</i>	tot.	<i>sd</i>
pe:pe:pe:p	22	126	12.7	79	9.4	95	7.5	107	7.9	92	7.1	127	6.2	100	7.5	722	21.8
pø:pø:pø:p	20	112	13.6	91	6.4	91	5.5	116	8.3	88	8.1	136	12.2	98	8.2	731	21.7
po:po:po:p	17	120	15.7	75	8.5	92	6.7	110	8.6	89	7.1	133	8.9	104	8.5	722	23.9
pIpIpIp	19	137	12.6	71	9.0	105	7.4	77	5.7	105	11.1	87	7.7	109	8.7	691	41.9
pæpæpæp	18	120	12.2	74	8.3	100	7.4	80	6.4	99	7.7	97	5.3	105	9.4	675	26.8
pəpəpəp	17	121	18.3	72	8.0	97	8.7	83	6.6	102	8.5	93	9.0	107	10.8	674	29.0
JtH																	
pe:pe:pe:p	19	117	8.9	76	5.4	98	5.7	124	7.5	103	6.2	127	6.3	109	5.5	753	21.2
pø:pø:pø:p	17	120	7.4	85	6.4	93	7.0	131	7.7	97	8.3	134	7.8	99	8.0	758	24.1
po:po:po:p	18	117	6.6	80	5.4	98	5.5	127	7.7	99	7.0	132	6.3	101	8.2	754	22.4
pIpIpIp	18	121	10.0	61	3.6	104	5.3	67	5.6	100	5.1	76	5.8	115	6.2	643	13.5
pæpæpæp	21	125	12.5	62	5.6	103	3.3	72	6.6	96	5.4	81	6.8	107	5.4	647	15.8
pəpəpəp	16	123	6.6	65	6.5	107	4.7	69	4.3	101	4.7	78	4.7	111	5.9	655	15.1
SN																	
pe:pe:pe:p	19	140	19.7	87	9.1	119	6.2	104	11.6	118	7.0	111	9.7	125	14.3	804	33.1
pø:pø:pø:p	20	130	13.8	104	9.0	99	9.2	124	11.5	99	9.6	130	10.3	102	11.7	789	31.0
po:po:po:p	14	133	15.5	92	7.0	108	7.0	114	8.2	101	5.8	125	7.9	104	10.9	777	41.4
pIpIpIp	16	137	17.6	69	8.5	129	8.3	73	8.4	128	10.6	75	11.7	131	14.7	740	41.4
pæpæpæp	16	130	12.2	84	11.0	112	8.4	91	11.3	108	9.3	99	12.5	112	11.5	736	33.8
pəpəpəp	11	130	10.8	80	10.5	113	7.3	86	7.0	109	4.5	98	10.8	119	14.0	734	33.4

TABLE 5.A

JtH	n	p <sub>1</sub>	sd	V <sub>1</sub>	sd	p <sub>2</sub>	sd	V <sub>2</sub>	sd	p <sub>3</sub>	sd	V <sub>3</sub>	sd	p <sub>4</sub>	sd	tot.	sd
pa:pa:pa:p	15	113	12.8	71	6.3	86	7.9	145	5.9	85	7.1	138	9.5	98	5.2	736	29.0
pəpa:pəp	13	119	12.8	54	5.0	88	5.4	158	7.7	90	6.8	56	5.1	112	4.3	676	22.3
paɾaɾaɾaɾ	19	120	24.6	69	5.1	90	7.4	81	4.4	87	8.3	87	4.8	109	8.0	642	35.9
pəpaɾəp	19	129	14.8	61	5.5	92	6.2	86	5.2	93	7.5	61	7.3	109	5.9	630	24.0
pIpIpIp	17	126	10.2	67	4.4	91	7.8	74	6.6	91	7.2	76	3.8	109	5.3	634	20.4
pəpIpəp	20	130	11.6	66	5.9	89	8.9	79	5.6	98	6.0	59	5.5	111	6.0	630	23.8
SN																	
pa:pa:pa:p	22	110	14.3	58	6.5	98	4.2	120	5.8	96	4.9	116	7.5	93	6.7	692	24.3
pəpa:pəp	19	124	13.9	40	6.4	109	6.2	126	9.0	97	4.5	59	6.9	105	12.8	661	20.5
paɾaɾaɾaɾ	19	122	15.8	49	6.5	108	4.0	65	5.0	102	4.4	80	5.9	116	11.5	643	20.4
pəpaɾəp	11	120	15.4	40	6.7	116	6.7	77	6.9	105	13.7	60	7.7	118	14.9	636	24.4
pIpIpIp	21	132	19.0	43	7.9	119	7.6	59	6.6	118	7.7	64	8.1	125	8.5	661	25.0
pəpIpəp	22	124	16.7	47	8.2	112	10.9	65	7.4	113	10.2	69	12.8	116	16.5	646	28.7

TABLE 1.B

IS	<i>n</i>	$p_1V_1$ sd		$V_1p_2$ sd		$p_2V_2$ sd		$V_2p_3$ sd		$p_3V_3$ sd	
pa:pa:pa:p	11	28	8.9	18	3.7	15	2.8	17	3.5	16	2.5
pe:pe:pe:p	17	22	14.1	22	5.4	12	3.5	23	5.4	15	2.5
peipeiipeip	15	24	8.7	19	3.9	15	1.4	21	3.4	16	1.6
pepepep	16	20	9.3	16	4.4	16	2.3	19	2.3	16	2.1
pIpIpIp	19	19	8.3	22	6.3	12	3.9	25	4.7	14	3.8
pipipip	17	20	7.9	26	8.0	14	5.1	30	7.0	13	4.8
JtH											
pa:pa:pa:p	12	15	1.7	33	3.0	9	1.3	27	4.3	13	1.1
pe:pe:pe:p	20	15	2.4	39	4.4	11	0.9	38	3.2	14	1.6
peipeiipeip	21	19	3.5	37	3.1	12	1.4	33	3.3	15	1.8
pepepep	18	18	2.3	37	3.6	12	3.3	34	4.9	15	2.9
pIpIpIp	17	14	2.1	41	3.3	10	1.4	38	3.2	14	1.7
pipipip	18	16	3.7	47	6.9	11	3.9	45	4.4	15	3.5
SN											
pa:pa:pa:p	22	11	3.1	21	2.4	7	2.6	19	2.8	8	3.3
pe:pe:pe:p	16	9	1.9	28	2.4	9	2.4	28	4.5	10	2.0
peipeiipeip	20	10	1.6	24	4.4	8	2.7	20	2.7	9	1.5
pepepep	17	9	1.3	27	2.6	8	1.9	22	3.3	9	1.7
pIpIpIp	18	8	2.2	33	4.2	9	2.7	29	3.8	10	2.0
pipipip	16	9	2.4	36	4.3	9	2.9	33	4.3	10	1.5

TABLE 2.B

IS	<i>n</i>	$p_1V_1$ sd		$V_1p_2$ sd		$p_2V_2$ sd		$V_2p_3$ sd		$p_3V_3$ sd	
pa:pa:pa:p	20	23	5.4	28	2.6	14	2.7	29	6.2	13	1.7
pø:pø:pø:p	14	29	6.1	24	6.1	16	8.9	30	5.3	17	3.4
p^yp^yp^yp	14	20	5.5	28	3.9	13	3.9	31	4.2	15	2.8
pærpærpæp	12	25	6.6	26	5.3	15	5.5	34	4.0	14	2.8
pyypyyp	18	32	5.9	30	6.9	22	8.8	29	6.5	28	8.7
JtH											
pa:pa:pa:p	17	18	5.4	32	2.9	10	2.1	27	3.6	15	4.6
pø:pø:pø:p	18	16	4.3	35	4.8	12	4.3	35	3.2	16	2.3
p^yp^yp^yp	22	14	2.1	35	4.8	13	3.5	33	3.7	15	2.2
pærpærpæp	18	16	4.3	34	4.0	12	2.5	36	2.6	17	3.3
pyypyyp	21	19	5.0	42	6.2	14	4.1	42	5.0	16	2.8
SN											
pa:pa:pa:p	17	14	1.9	22	3.7	9	1.9	19	2.6	9	2.3
pø:pø:pø:p	17	18	6.6	19	2.5	18	3.6	23	4.5	20	3.7
p^yp^yp^yp	20	15	4.9	23	2.9	14	1.9	20	5.4	15	2.8
pærpærpæp	19	24	4.1	19	6.4	20	5.6	19	10.1	25	8.3
pyypyyp	19	22	5.9	19	11.2	24	6.2	17	14.0	33	13.9

TABLE 3.B

IS	<i>n</i>	$p_1V_1$ sd		$V_1p_2$ sd		$p_2V_2$ sd		$V_2p_3$ sd		$p_3V_3$ sd	
pa:pa:pa:p	19	12	9.7	27	4.9	6	4.7	32	9.5	7	4.9
po:po:po:p	16	17	11.8	30	8.7	10	6.1	36	3.1	8	4.3
paupaupaup	23	12	9.1	33	4.0	8	4.8	40	3.4	8	4.4
paapaapa	14	11	9.4	29	4.7	6	4.6	34	2.3	7	3.7
pəpəpəp	19	15	8.4	27	5.9	9	4.5	34	4.4	8	4.5
puupuup	14	10	11.2	38	10.2	5	9.7	37	5.0	5	6.3
JtH											
pa:pa:pa:p	18	18	4.9	31	2.6	12	3.6	27	2.6	13	2.5
po:po:po:p	17	19	3.6	39	3.8	10	3.1	40	3.7	12	2.1
paupaupaup	22	17	4.7	37	2.9	12	3.8	36	3.3	13	1.3
paapaapa	17	17	3.8	32	3.1	10	2.8	31	2.1	12	1.7
pəpəpəp	17	21	4.4	34	3.9	11	5.0	36	3.0	11	2.3
puupuup	20	20	5.7	43	2.7	14	5.7	45	3.2	16	3.4
SN											
pa:pa:pa:p	16	10	2.0	24	3.0	6	1.6	24	3.0	6	1.6
po:po:po:p	12	12	2.8	35	3.7	12	1.8	28	2.9	13	1.9
paupaupaup	18	8	3.1	34	4.7	8	2.2	36	3.0	9	4.2
paapaapa	19	9	2.8	29	4.3	8	2.9	31	3.7	7	2.6
pəpəpəp	17	11	3.9	31	3.4	13	2.8	30	4.9	12	2.7
puupuup	11	18	4.3	38	5.6	19	3.6	30	3.2	20	3.7



TABLE 4.B

IS	<i>n</i>	$p_1V_1$ sd		$V_1p_2$ sd		$p_2V_2$ sd		$V_2p_3$ sd		$p_3V_3$ sd	
pe:pe:pe:p	22	30	9.2	19	2.1	16	2.4	21	3.2	16	1.9
pø:pø:pø:p	20	39	13.4	20	4.3	21	3.3	25	3.3	19	4.1
po:po:po:p	17	29	4.6	18	3.6	19	2.5	23	2.7	17	2.5
pIpIpIp	19	28	2.8	17	3.3	18	1.9	19	2.9	17	1.9
pæpæpæp	18	37	10.9	16	3.9	20	2.8	23	2.8	18	1.8
pöpöpöp	17	28	3.6	15	3.9	18	2.2	20	3.0	16	2.7
JtH											
pe:pe:pe:p	19	10	3.4	43	2.9	6	2.5	41	4.2	5	3.4
pø:pø:pø:p	17	14	3.3	42	2.4	8	2.3	40	2.8	9	2.6
po:po:po:p	18	12	4.7	39	3.4	6	2.1	36	3.5	9	2.4
pIpIpIp	18	12	4.1	45	5.0	6	2.5	44	3.7	8	3.0
pæpæpæp	21	13	4.6	42	3.4	9	3.1	44	3.6	10	2.6
pöpöpöp	16	12	4.7	36	3.6	5	2.9	38	1.7	5	2.0
SN											
pe:pe:pe:p	19	5	4.2	38	3.8	4	4.7	41	4.4	3	4.4
pø:pø:pø:p	20	12	3.5	26	4.7	11	3.6	30	5.4	11	3.6
po:po:po:p	14	7	9.1	28	2.6	10	2.4	30	5.4	14	4.2
pIpIpIp	16	6	2.7	38	6.8	5	3.4	41	3.5	3	5.3
pæpæpæp	16	12	4.3	27	5.2	12	3.2	27	6.2	13	3.3
pöpöpöp	11	18	3.8	28	3.3	13	2.0	28	3.4	18	3.6

TABLE 5.B

JtH	<i>n</i>	$p_1V_1$	sd	$V_1p_2$	sd	$p_2V_2$	sd	$V_2p_3$	sd	$p_3V_3$	sd
pa:pa:pa:p	15	18	2.9	38	2.5	11	2.8	33	4.1	15	3.1
pəpa:pəp	13	18	4.4	40	4.6	13	3.0	32	3.0	15	2.3
papapap	19	19	3.7	36	3.3	10	3.8	37	3.5	14	3.4
pəpəpəp	19	18	5.3	41	3.6	11	2.7	34	3.1	19	13.6
pIpIpIp	17	15	3.0	44	4.3	10	2.3	46	3.1	12	3.1
pəpIpəp	20	18	5.0	43	4.9	13	4.9	43	3.9	13	1.9
SN											
pa:pa:pa:p	22	13	3.2	26	2.2	8	2.0	20	1.8	7	2.2
pəpa:pəp	19	13	3.3	31	5.0	12	2.3	19	2.1	12	3.3
papapap	19	12	2.9	27	5.8	10	1.9	25	2.6	12	3.1
pəpəpəp	11	12	4.5	30	3.3	11	2.7	21	3.3	14	4.1
pIpIpIp	21	9	4.2	35	5.9	9	2.2	35	6.6	8	1.8
pəpIpəp	22	14	3.0	32	5.5	12	2.6	32	4.3	17	8.5

TABLE 6

SN	<i>n</i>	<i>t</i> <sub>1</sub>	sd	<i>V</i> <sub>1</sub>	sd	<i>t</i> <sub>2</sub>	sd	<i>V</i> <sub>2</sub>	sd	<i>t</i> <sub>3</sub>	sd	<i>V</i> <sub>3</sub>	sd	<i>t</i> <sub>4</sub>	sd	tot.	sd
ta:ta:ta:t	18	147	19.4	92	7.3	104	6.4	115	7.0	98	4.2	129	4.6	126	18.4	812	29.3
te:te:te:t	20	153	24.0	92	5.8	105	6.4	102	7.0	118	7.2	100	13.3	115	8.3	785	33.8
teiteiteit	16	153	17.3	101	4.4	95	9.3	113	5.6	100	4.6	127	5.7	116	15.9	806	36.0
tetetet	13	163	23.4	76	4.6	117	12.1	81	7.0	120	4.1	93	4.8	131	25.6	781	40.1
tItItIt	17	148	16.1	65	7.1	127	9.1	66	7.0	131	9.4	75	10.2	122	10.2	735	29.8
tititit	15	164	22.0	64	6.8	130	10.5	58	12.4	134	8.0	75	14.6	118	10.1	743	29.0
IS																	
ta:ta:ta:t	21	119	20.1	80	11.2	49	13.7	152	15.4	65	11.5	165	11.2	75	12.7	705	23.0
te:te:te:t	21	125	20.4	88	12.3	57	12.9	141	16.8	69	11.4	162	9.1	76	9.6	716	23.1
teiteiteit	21	126	21.6	95	10.0	51	13.4	147	14.4	73	8.7	167	7.4	80	8.2	737	31.4
tetetet	17	114	17.8	78	10.4	54	8.4	110	11.8	70	15.7	121	9.5	81	16.0	626	32.2
tItItIt	14	125	24.2	84	8.7	50	17.6	113	16.9	58	17.8	125	12.6	75	12.6	630	33.8
tititit	19	122	19.4	81	11.6	63	13.8	111	17.2	67	21.2	118	20.4	71	11.7	632	39.4

TABLE 7

SN	<i>n</i>	<i>t</i> <sub>1</sub>	sd	V <sub>1</sub>	sd	<i>t</i> <sub>2</sub>	sd	V <sub>2</sub>	sd	<i>t</i> <sub>3</sub>	sd	V <sub>3</sub>	sd	<i>t</i> <sub>4</sub>	sd	tot.	sd
ta:ta:ta:t	10	169	16.9	100	11.9	113	7.5	122	9.9	105	5.9	130	8.3	126	14.9	866	44.4
tø:tø:tø:t	22	151	19.8	103	9.6	119	10.2	115	8.0	113	9.4	130	9.0	121	25.5	852	53.9
t^yt^yt^yt	14	161	23.2	103	8.7	109	12.7	119	10.4	101	6.2	133	11.8	124	20.7	849	45.1
tætætæt	20	151	19.0	74	7.8	137	12.8	71	7.7	128	10.3	83	10.6	146	28.0	790	59.5
tytytyt	14	155	16.8	72	5.4	143	10.4	66	10.3	135	13.9	74	9.1	137	19.1	783	36.3
IS																	
ta:ta:ta:t	20	131	18.5	67	12.2	53	13.7	147	13.5	68	9.9	161	12.5	64	13.3	690	24.2
tø:tø:tø:t	16	123	12.8	76	6.2	59	8.6	131	11.3	68	9.4	157	10.7	73	6.0	685	33.0
t^yt^yt^yt	24	128	21.7	84	12.1	54	13.5	138	20.5	63	11.1	154	12.9	80	12.1	700	39.0
tætætæt	19	128	23.2	65	9.7	68	9.7	91	5.4	59	11.1	109	10.7	78	18.1	599	33.3
tytytyt	18	121	17.0	64	8.2	69	8.6	98	7.5	49	9.5	122	11.3	80	28.3	601	36.0

TABLE 8

SN	<i>n</i>	<i>t</i> <sub>1</sub>	sd	<i>V</i> <sub>1</sub>	sd	<i>t</i> <sub>2</sub>	sd	<i>V</i> <sub>2</sub>	sd	<i>t</i> <sub>3</sub>	sd	<i>V</i> <sub>3</sub>	sd	<i>t</i> <sub>4</sub>	sd	tot.	sd
ta:ta:ta:t	13	137	17.1	118	7.9	108	7.9	136	5.9	102	8.7	148	6.7	116	15.2	863	35.2
to:to:to:t	21	126	15.9	117	7.5	108	6.3	128	6.5	103	7.1	143	8.7	92	11.0	817	26.0
tautautaut	23	141	15.5	116	6.0	111	7.4	131	6.8	102	5.5	145	7.0	92	10.8	839	31.1
tatatat	16	146	21.8	90	4.0	126	8.8	90	4.6	127	7.5	104	6.1	139	22.7	823	32.1
tōtōtōt	17	138	13.5	94	4.3	127	6.9	90	7.3	130	10.0	109	6.8	107	9.3	796	28.9
tututut	19	121	16.0	96	4.3	136	11.0	90	5.5	125	13.3	106	6.5	110	12.5	784	35.4
IS																	
ta:ta:ta:t	24	121	20.7	84	7.3	69	7.2	132	6.5	73	8.2	160	9.0	80	8.9	718	23.7
to:to:to:t	12	110	11.3	90	12.0	77	9.3	118	4.8	76	6.8	155	10.2	73	9.3	699	20.8
tautautaut	18	115	17.3	102	19.1	67	21.2	134	12.5	71	9.4	159	13.3	77	11.3	726	30.4
tatatat	19	132	21.2	74	8.4	87	9.7	94	6.5	79	8.3	111	8.7	83	9.2	660	23.7
tōtōtōt	19	124	35.7	81	7.4	77	10.0	92	7.3	80	8.5	120	14.9	71	19.6	643	34.0
tututut	18	120	7.5	75	5.6	88	7.6	91	4.0	82	8.5	117	9.6	70	9.6	643	17.8

TABLE 9

SN	<i>n</i>	<i>p</i>	<i>sd</i>	Vr	<i>sd</i>	<i>t</i>	<i>sd</i>	tot.	<i>sd</i>
pa:rt	24	125	15	217	15	144	20	485	31
po:rt	20	121	10	235	15	142	28	497	37
pu:rt	20	121	14	217	17	144	20	482	29
pa:rt	22	132	18	131	12	182	28	445	41
pe:rt	21	133	20	127	10	193	34	453	46
po:rt	19	128	14	131	15	189	30	447	36
IS									
pa:rt	21	134	13	192	9	139	16	465	25
po:rt	20	143	18	194	10	117	19	454	30
pu:rt	20	151	30	199	20	112	20	463	36
pa:rt	25	147	18	132	8	161	31	440	42
pe:rt	18	138	10	131	11	165	30	435	37
po:rt	19	138	16	122	13	148	21	408	30

TABLE 10

SN	<i>n</i>	<i>p</i>	<i>sd</i>	Vr	<i>sd</i>	<i>t</i>	<i>sd</i>	tot.	<i>sd</i>
pø:rt	22	126	16	239	20	149	23	514	37
pe:rt	23	127	12	236	16	143	16	506	26
py:rt	16	121	16	232	12	143	22	496	37
pi:rt	22	132	14	229	17	150	19	511	34
pø:rt	19	129	10	130	11	210	31	469	35
pi:rt	20	139	20	134	13	211	32	484	37
IS									
pø:rt	22	143	28	203	14	131	14	476	35
pe:rt	21	139	12	203	8	142	17	484	25
py:rt	19	140	16	209	17	138	20	486	33
pi:rt	19	138	15	215	14	117	23	470	19
pø:rt	15	141	15	133	13	167	20	440	29
pi:rt	23	150	22	143	11	163	18	456	36

TABLE 11

SN	<i>n</i>	<i>p</i> <sub>1</sub>	sd	<i>V</i> <sub>1</sub>	sd	<i>p</i> <sub>2</sub>	sd	<i>V</i> <sub>2</sub>	sd	<i>p</i> <sub>3</sub>	sd	<i>V</i> <sub>3</sub>	sd	<i>p</i> <sub>4</sub>	sd	tot.	sd
pa:pa:pa:p	20	100	8·8	118	7·2	73	5·4	56	7·1	85	4·9	124	7·2	75	6·3	630	28·9
pa:pa:pa:p	20	90	7·6	67	9·8	93	9·8	127	15·5	77	6·6	120	11·0	79	3·7	653	23·9
pa:pa:pa:p	18	87	10·6	85	8·9	73	4·9	65	9·0	101	7·9	136	9·5	80	6·5	628	23·2
pa:pa:pa:p	19	108	13·5	72	8·5	83	10·5	59	8·9	90	7·1	79	4·9	88	6·4	578	21·2
pa:pa:pa:p	22	94	9·5	59	4·3	104	7·5	77	4·7	86	6·1	81	6·7	83	8·4	583	20·7
pa:pa:pa:p	21	93	6·9	66	6·1	83	5·2	57	9·3	110	9·2	80	6·2	98	6·2	586	21·3
IS																	
pa:pa:pa:p	24	101	9·9	116	8·1	58	7·9	55	9·8	83	6·1	127	6·0	74	7·2	615	25·3
pa:pa:pa:p	19	99	13·3	72	4·8	83	9·2	125	6·1	74	6·3	133	5·4	76	7·9	662	24·1
pa:pa:pa:p	12	93	12·3	89	9·2	58	11·6	71	8·2	90	6·8	127	5·9	89	10·9	617	25·6
pa:pa:pa:p	21	111	11·9	71	4·5	70	4·8	59	6·9	87	10·4	86	4·1	82	9·9	567	27·3
pa:pa:pa:p	22	103	12·9	71	4·2	93	7·5	81	4·5	81	9·6	89	5·1	89	8·2	606	32·7
pa:pa:pa:p	24	97	8·3	74	5·6	70	11·0	67	6·6	96	9·6	77	3·8	93	12·7	574	25·5

TABLE 12

IS	<i>n</i>	<i>p</i> <sub>1</sub>	sd	<i>V</i> <sub>1</sub>	sd	<i>p</i> <sub>2</sub>	sd	<i>V</i> <sub>2</sub>	sd	<i>p</i> <sub>3</sub>	sd	<i>V</i> <sub>3</sub>	sd	<i>p</i> <sub>4</sub>	sd	tot.	sd
pa:pa:pa:p	20	85	5.0	104	7.0	52	10.0	48	8.0	66	7.3	109	8.2	61	6.8	525	16.6
pa:pa:pa:p	21	77	9.7	60	9.4	60	7.0	114	12.4	56	10.5	104	8.3	59	11.2	530	26.5
pa:pa:pa:p	17	80	5.0	83	7.6	47	8.7	57	6.9	68	7.6	119	10.9	72	7.1	525	17.0
pa:pa:pa:p	23	92	10.0	67	4.6	63	6.9	49	7.7	68	6.4	74	7.1	68	5.9	482	15.6
pa:pa:pa:p	20	84	6.4	60	4.9	60	11.5	79	9.1	65	9.4	69	7.9	78	6.0	494	20.2
pa:pa:pa:p	16	85	8.0	68	6.7	55	8.7	60	11.1	66	8.8	75	8.6	75	11.3	484	20.8
SN																	
pa:pa:pa:p	17	96	8.8	115	6.8	78	4.0	36	10.9	85	4.3	120	5.9	85	6.2	615	19.6
pa:pa:pa:p	18	98	9.0	50	9.1	90	5.8	118	6.9	79	4.8	103	7.0	88	8.3	626	24.4
pa:pa:pa:p	18	93	10.6	77	10.8	81	4.5	37	8.0	95	5.1	126	9.4	105	38.5	614	45.5
pa:pa:pa:p	23	110	12.4	67	3.6	87	5.3	45	7.5	90	5.4	74	5.4	96	7.5	568	22.3
pa:pa:pa:p	26	106	8.0	47	5.7	100	5.8	70	4.7	83	4.5	71	7.9	98	9.1	575	25.7
pa:pa:pa:p	19	103	10.3	55	4.3	88	4.8	37	6.2	101	6.4	74	7.1	101	12.5	559	31.2



TABLE 13

SN	<i>n</i>	<i>m</i> <sub>1</sub>	sd	<i>a</i> : <sub>1</sub>	sd	<i>m</i> <sub>2</sub>	sd	<i>a</i> : <sub>2</sub>	sd	<i>m</i> <sub>3</sub>	sd	<i>a</i> : <sub>3</sub>	sd	<i>m</i> <sub>4</sub>	sd	<i>a</i> : <sub>4</sub>	sd	<i>m</i> <sub>5</sub>	sd	tot.	sd
ma:m	19	140	12.2	207	12.5	149	12.4													496	17.9
ma:ma:m	20	126	13.9	140	18.1	87	4.7	174	12.8	141	23.7									664	35.1
ma:ma:ma:m	23	116	14.9	106	10.3	82	4.4	64	7.6	90	6.2	151	11.4	133	14.7					741	37.3
ma:ma:ma:ma:m	23	110	9.8	99	12.4	82	3.5	55	8.8	90	5.7	57	10.8	100	4.1	151	10.1	138	13.0	882	40.4
ma:ma:m	21	117	13.4	90	15.1	95	7.9	207	15.9	144	19.3									653	33.2
ma:ma:ma:m	18	112	9.0	70	10.3	87	4.3	61	10.8	102	6.3	191	20	145	9.3					768	31.8

TABLE 14

SN	<i>n</i>	<i>m</i> <sub>1</sub>	sd	<i>a</i> <sub>1</sub>	sd	<i>m</i> <sub>2</sub>	sd	<i>a</i> <sub>2</sub>	sd	<i>m</i> <sub>3</sub>	sd	<i>a</i> <sub>3</sub>	sd	<i>m</i> <sub>4</sub>	sd	<i>a</i> <sub>4</sub>	sd	<i>m</i> <sub>5</sub>	sd	tot.	sd
mam	17	159	18.8	97	9.0	174	14.3													430	22.0
mamam	21	139	18.1	69	7.2	104	6.4	88	12.5	156	16.4									557	28.2
mamamam	15	130	14.0	60	7.8	94	6.0	44	8.2	103	4.9	72	12.1	154	18.5					657	31.6
mamamamam	17	120	10.2	59	6.0	93	4.6	40	6.5	106	6.6	41	8.3	111	5.7	79	7.2	148	11.7	796	18.7
mamam	20	140	17.7	58	9.3	119	5.7	86	8.9	167	11.8									570	26.1
mamamam	18	125	18.4	53	5.8	98	4.9	43	9.5	113	5.2	86	5.1	161	10.9					687	23.2

TABEL 15

SN	<i>n</i>	<i>m</i> <sub>1</sub>	sd	<i>a</i> : <sub>1</sub>	sd	<i>m</i> <sub>2</sub>	sd	<i>a</i> <sub>2</sub> :	sd	<i>m</i> <sub>3</sub>	sd	<i>a</i> : <sub>3</sub>	sd	<i>m</i> <sub>4</sub>	sd	<i>a</i> : <sub>4</sub>	sd	<i>m</i> <sub>5</sub>	sd	tot.	sd
ma:ma:m	16	127	14.2	143	7.4	78	5.9	168	10.7	110	14.3									625	30.6
ma:ma:ma:m	15	132	16.2	87	16.0	83	7.6	159	7.6	80	5.2	164	8.0	118	16.6					822	33.2
ma:ma:ma:ma:m	20	129	23.0	83	10.4	78	5.8	75	13.1	81	5.0	149	7.0	79	5.0	168	11.3	125	22.0	965	34.5
ma:ma:ma:m	22	129	14.2	110	11.1	79	3.4	79	13.7	77	5.4	146	21.3	130	33.6					750	49.1
ma:ma:ma:ma:m	18	130	21.4	91	15.7	80	7.5	127	11.4	79	4.4	74	15.8	79	4.5	145	12.9	133	22.4	937	66.5

TABLE 16

SN	<i>n</i>	<i>m</i> <sub>1</sub>	sd	<i>a</i> <sub>1</sub>	sd	<i>m</i> <sub>2</sub>	sd	<i>a</i> <sub>2</sub>	sd	<i>m</i> <sub>3</sub>	sd	<i>a</i> <sub>3</sub>	sd	<i>m</i> <sub>4</sub>	sd	<i>a</i> <sub>4</sub>	sd	<i>m</i> <sub>5</sub>	sd	tot.	sd
mamam	31	123	9.4	68	8.6	93	5.6	85	5.7	134	13.3									503	21.3
mamamam	29	125	9.4	50	7.0	106	7.4	68	4.7	95	8.5	84	6.5	145	13.5					673	25.6
mamamamam	28	120	7.4	53	6.7	100	6.1	44	6.8	110	7.6	67	8.3	94	5.9	81	6.2	141	16.8	810	17.1
mamamam	28	119	7.4	62	6.0	94	4.7	55	8.4	100	7.8	73	5.6	132	15.3					635	17.8
mamamamam	33	121	10.5	52	6.5	103	6.1	63	5.2	95	6.2	58	7.5	99	7.8	72	6.5	139	14.0	803	25.8

TABLE 17

SN	<i>n</i>	<i>m</i> <sub>1</sub>	sd	<i>a</i> : <sub>1</sub>	sd	<i>m</i> <sub>2</sub>	sd	<i>a</i> : <sub>2</sub>	sd	<i>m</i> <sub>3</sub>	sd	<i>a</i> : <sub>3</sub>	sd	<i>m</i> <sub>4</sub>	sd	<i>a</i> : <sub>4</sub>	sd	<i>m</i> <sub>5</sub>	sd	tot.	sd
ma:m	22	100	11.0	147	10.4	78	5.3													325	20.1
ma:ma:m	24	91	8.2	131	11.5	73	4.8	124	13.1	76	7.2									495	37.0
ma:ma:ma:m	20	89	7.2	102	9.5	75	5.1	63	12.3	76	6.1	126	10.8	85	24.9					616	47.5
ma:ma:ma:ma:m	18	89	6.1	91	7.4	75	4.1	51	4.7	77	4.3	62	9.8	80	3.8	128	5.7	80	10.4	732	35.9
ma:ma:m	19	92	8.6	74	17.5	80	4.9	135	18.5	80	9.9									461	37.1
ma:ma:ma:m	16	93	9.0	70	11.1	77	4.5	54	7.0	80	8.6	133	11.6	83	6.1					590	38.0

TABLE 18

SN	<i>n</i>	<i>m</i> <sub>1</sub>	sd	<i>a</i> <sub>1</sub>	sd	<i>m</i> <sub>2</sub>	sd	<i>a</i> <sub>2</sub>	sd	<i>m</i> <sub>3</sub>	sd	<i>a</i> <sub>3</sub>	sd	<i>m</i> <sub>4</sub>	sd	<i>a</i> <sub>4</sub>	sd	<i>m</i> <sub>5</sub>	sd	tot.	sd
mam	25	133	18.0	81	9.4	93	5.6													307	24.2
mamam	17	126	23.7	74	7.2	84	4.7	70	12.2	100	7.2									454	25.3
mamamam	16	113	12.3	62	5.3	87	5.5	41	8.7	90	7.2	68	8.2	95	8.4					558	21.4
mamamamam	18	109	10.8	63	5.4	87	4.3	36	6.2	97	8.9	38	9.9	93	3.9	75	4.7	94	4.1	692	18.3
mamam	25	121	18.0	48	6.2	94	6.3	75	7.2	93	7.1									431	19.6
mamamam	15	114	14.4	53	6.0	92	7.0	33	8.5	96	5.3	77	5.9	92	5.4					555	17.7

TABLE 19

SN	<i>n</i>	<i>m</i> <sub>1</sub>	sd	<i>a</i> : <sub>1</sub>	sd	<i>m</i> <sub>2</sub>	sd	<i>a</i> : <sub>2</sub>	sd	<i>m</i> <sub>3</sub>	sd	<i>a</i> : <sub>3</sub>	sd	<i>m</i> <sub>4</sub>	sd	<i>a</i> : <sub>4</sub>	sd	<i>m</i> <sub>5</sub>	sd	tot.	sd
ma:ma:m	20	98	6.8	145	8.6	81	3.4	121	8.8	85	5.6									529	18.1
ma:ma:ma:m	22	96	8.7	63	9.2	87	5.2	146	11.0	80	4.8	106	12.8	89	10.1					667	26.9
ma:ma:ma:ma:m	22	92	9.0	77	8.2	82	4.9	46	7.0	94	8.1	140	7.3	80	5.0	107	10.0	87	6.1	805	29.5
ma:ma:ma:m	17	100	8.1	123	8.6	80	3.3	47	7.5	84	5.4	123	6.5	83	4.1					640	16.7
ma:ma:ma:ma:m	20	91	7.3	67	7.3	83	5.9	126	9.0	79	3.6	45	5.1	84	4.8	123	8.2	80	5.6	780	19.0

TABLE 20

SN	<i>n</i>	<i>m</i> <sub>1</sub>	sd	<i>a</i> <sub>1</sub>	sd	<i>m</i> <sub>2</sub>	sd	<i>a</i> <sub>2</sub>	sd	<i>m</i> <sub>3</sub>	sd	<i>a</i> <sub>3</sub>	sd	<i>m</i> <sub>4</sub>	sd	<i>a</i> <sub>4</sub>	sd	<i>m</i> <sub>5</sub>	sd	tot.	sd
mamam	21	114	9.4	70	6.2	83	4.4	67	6.4	95	17.0									429	24.9
mamamam	14	109	9.4	48	6.9	96	4.5	64	4.1	84	3.7	59	6.3	105	6.1					563	14.5
mamamamam	17	107	6.1	51	6.6	96	4.8	31	5.1	99	5.8	65	5.2	85	4.1	57	6.5	105	6.8	696	12.0
mamamam	17	114	8.3	62	5.3	91	3.9	38	5.7	92	5.7	64	5.8	103	8.2					564	13.3
mamamamam	22	110	5.8	51	5.5	95	3.7	55	7.5	90	5.4	46	9.2	94	5.4	59	7.5	106	9.4	704	17.7

APPENDIX B

Tables 21–30 with data from adjustment tests and spectrographic measurements

Vowel durations as found in adjustment tests with synthetic speech and in spectrographic measurements of spoken words. The data from the adjustment tests are given for each subject separately and for each initial duration separately. In the head of the columns the words concerned are given in phonetic symbols. The adjusted vowels are underlined. In the columns the mean durations over 5 adjustments are given together with the standard deviations. In the lower half of the tables the mean durations of two spoken versions as measured from spectrograms are given for each subject separately.

TABLE 21

	sub- ject	initial duration	<u>ma</u> :t		<u>ma</u> :tə		<u>ma</u> :təlo:s		<u>ma</u> :təlo:zə	
				sd		sd		sd		sd
adjusted vowel durations	HvL	long	217	12	189	7	195	19	162	13
		short	215	16	167	15	132	23	131	15
	DB	long	202	34	180	10	160	32	142	10
		short	202	25	155	12	137	10	138	6
HM	long	199	20	188	25	165	14	170	21	
	short	178	17	159	12	135	12	149	16	
	m		202		173		154		149	
spoken vowel durations	HvL		175		170		145		160	
	DB		175		152		130		122	
	HM		243		210		175		170	
	m		198		178		150		150	

TABLE 22

	subject	initial duration	pa <u>n</u>		pa <u>nə</u>		pa <u>nəkuk</u>		pa <u>nəkukə</u>	
				sd		sd		sd		sd
adjusted vowel durations	HvL	long	105	5	100	5	89	12	91	7
		short	101	5	98	3	73	5	75	6.5
	DB	long	86	10	78	9	71	14	61	19
		short	76	4	73	14	65	14	60	15
	HM	long	120	18	107	4.5	91	11	97	8.5
short		101	8	96	10	89	7	78	16	
m		98		92		80		77		
spoken vowel durations	HvL		105		90		75		65	
	DB		110		92		65		75	
	HM		103		82		72		67	
	m		106		88		71		69	

TABLE 23

	subject	initial duration	o:to:ma:t		to:ma:tə		ma:təlo:s	
				sd		sd		sd
adjusted vowel durations	HvL	long	183	6	154	12	145	4.3
		short	156	5.5	139	12	135	9
	DB	long	205	14	189	15	165	11
		short	196	12	172	15	142	6
	HM	long	218	12	208	5	178	5
short		190	15	181	19	137	7	
m		191		174		150		
spoken vowel durations	HvL		160		175		145	
	DB		200		205		175	
	HM		180		150		130	
	m		180		177		150	

TABLE 24

	subject	initial duration	ma:t		to:ma:t		o:to:ma:t	
				sd		sd		sd
adjusted vowel durations	HvL	long	185	19	171	10	178	14
		short	172	17	161	13	153	6
	DB	long	200	11	210	12	213	12
		short	198	10	206	5.5	198	7.3
	HM	long	201	16	195	17	202	16
short		179	12	154	11	159	11	
m		189		183		183		
spoken vowel durations	HvL		180		185		180	
			200		220		200	
	HM		220		200		185	
		m	200		201		188	

TABLE 25

	subject	initial duration	a:na:to:m	
				sd
adjusted vowel durations	HvL	long	192	11
		short	188	8
	DB	long	194	12
		short	187	10
	HM	long	180	15
short		172	10	
m		186		
spoken vowel durations	HvL		185	
	DB		168	
	HM		175	
	m		176	

TABLE 26

	subject	initial duration	o:vərt <u>o</u> :m		t <u>o</u> :mɑ:tə		o:t <u>o</u> :mɑ:t	
				sd		sd		sd
adjusted vowel durations	HvL	long	169	21	107	7.5	127	13
		short	141	14	92	5	101	9
	HM	long	155	8	102	10	105	17
		short	129	16	92	8	107	7
m		149		98		110		
spoken vowel durations	HvL		155		57		35	
	HM		160		65		62	
	m		158		61		49	

TABLE 27

	subject	initial duration	a:nɑ:t <u>o</u> :m	
				sd
adjusted vowel durations	IHS	long	189	16
		short	174	8
	JtH	long	197	10
		short	201	9
	AvK	long	167	4.5
short		149	8	
m		180		



TABLE 28

	subject	initial duration	o:vərt <u>o</u> :m		t <u>o</u> :ma:tə		o:t <u>o</u> :ma:t	
				sd		sd		sd
adjusted vowel durations	IHS	long	135	4.5	104	16	80	19
		short	131	16	76	11	75	11
	JtH	long	175	6.5	106	5	92	6.5
		short	163	9	99	1.5	70	8
	AvK	long	165	14	93	6.5	53	14
short		169	11	76	14	50	8	
m		156		92		70		

TABLE 29

	subject	initial duration	o:vərt <u>o</u> :m		t <u>o</u> :ma:tə		o:t <u>o</u> :matə	
				sd		sd		sd
adjusted vowel durations	HvL	long	159	13	61	9	79	15
		short	153	22	63	6	72	19
	DB	long	160	7	94	14	95	9
		short	159	2	71	11	77	16
	HM	long	159	9	96	7	76	14
short		153	17	86	18	64	18	
m		157		79		77		
spoken vowel durations	HvL		155		53		33	
	DB		155		90		52	
	HM		160		65		78	
	m		157		71		54	

TABLE 30

	subject	initial duration	me:to:dik		me:to:də	
				sd		sd
adjusted vowel durations	HvL	long	107	23	94	18
		short	93	9	93	15
	DB	long	111	11	93	15
		short	120	15	99	13
	HM	long	106	11	77	21
short		103	19	75	19	
	m		107		89	
spoken vowels durations	HvL		88		85	
	DB		95		93	
	HM		103		85	
	m		95		88	

## REFERENCES

- 1) K. N. Stevens and A. S. House, Perturbations of vowel articulations by consonantal context: an acoustical study. *J. Speech and Hearing Res.* **6**, 111-128, 1963.
- 2) K. N. Stevens, A. S. House and A. P. Paul, Acoustical description of syllabic nuclei: an interpretation in terms of a dynamical model of articulation. *J. acoust. Soc. Amer.* **40**, 123-132, 1966.
- 3) B. E. F. Lindblom, Spectrographic study of vowel reduction. *J. acoust. Soc. Amer.* **35**, 1773-1781, 1963.
- 4) G. E. Peterson and I. Lehiste, Duration of syllable nuclei in English. *J. acoust. Soc. Amer.* **32**, 693-703, 1960.
- 5) D. Abercrombie, *Elements of general phonetics*. Edinburgh University Press, Edinburgh, 1967.
- 6) L. Lisker, On time and timing in speech. Haskins Status Report on Speech Research, **SR 23**, 151-178, 1970. Also in: T. A. Sebeok (ed.), *Current trends in linguistics*, Mouton & Co., The Hague, 1971, vol. 12.
- 7) A. Cohen, Errors of speech and their implication for understanding the strategy of language users. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* **21**, 177-181, 1968.
- 8) S. G. Nooteboom, The tongue slips into patterns. In: *Nomen, Leyden studies in linguistics and phonetics*, Mouton & Co., The Hague, 1969, pp. 114-132.
- 9) D. G. Mackay, Spoonerisms; the structure of errors in the serial order of speech. *Neuropsychologia* **8**, 323-350, 1970.
- 10) V. A. Fromkin, The non-anomalous nature of anomalous utterances. *Language* **47**, 27-52, 1971.
- 11) D. S. Boomer and J. D. M. Laver, Slips of the tongue. *British Journal of Disorders in Communication* **3**, 2-12, 1968.
- 12) L. F. Willems, Measuring circuit for tongue and lip closure duration. Institute for Perception Research, Eindhoven, Annual Progress Report **5**, 185-187, 1970.
- 13) I. H. Slis, Articulatory measurements on voiced, voiceless and nasal consonants. *Phonetica* **21**, 193-210, 1970.
- 14) J. F. Schouten and J. Domburg, De DONDEERS, een elektronisch meetsysteem voor menselijke reacties. *Philips techn. Tijdschrift* **24**, 372-382, 1962.
- 15) C.-C. Elert, Phonologic studies of quantity in Swedish. Monografier utgivna av Stockholms kommunalförvaltning **27**, Almqvist & Wiksell, Stockholm, Göteborg, Uppsala, 1964.
- 16) I. H. Slis and A. Cohen, On the complex regulating the voiced-voiceless distinction I. *Language and Speech* **12**, 80-102, 1969.
- 17) I. H. Slis, Experiments on consonant durations related to the time structure of isolated words. Institute for Perception Research, Eindhoven, Annual Progress Report **3**, 71-80, 1968.
- 18) W. G. Moulton, The vowels of Dutch; phonetic and distributional classes. *Lingua* **11**, 294-312, 1962.
- 19) I. Lehiste, *Suprasegmentals*. M.I.T. Press, Cambridge Mass. and London, 1970.
- 20) L. Lisker, Closure duration and the intervocalic voiced-voiceless distinction in English. *Language* **33**, 42-49, 1957.
- 21) L. Lisker and A. S. Abramson, Voice onset time in the production and perception of English stops. Haskins Status Report on Speech Research **SR1**, 3.1-3.15, 1965.
- 22) A. S. Abramson and L. Lisker, Voice onset time in stop consonants; acoustic analysis and synthesis. Haskins Status Report on Speech Research **SR3**, 1.1-1.8, 1965.
- 23) I. H. Slis and A. Cohen, On the complex regulating the voiced-voiceless distinction II. *Language and Speech* **12**, 137-155, 1969.
- 24) A. Cohen, J. F. Schouten and J. 't Hart, Contribution of the time parameter to the perception of speech. *Proceedings of the Fourth International Congress of Phonetic Sciences (Helsinki Sept. 4-9, 1961)*, A. Sovijärvi and P. Aalto (eds), *Janua Linguarum*, series major **10**, Mouton & Co., The Hague, 1962, pp. 555-560.
- 25) A. W. F. Huggins, How accurately must a speaker time his articulations? *IEEE Trans. Audio AU-16*, 112-117, 1968.
- 26) A. W. F. Huggins, Just noticeable differences in segment duration in speech. *J. acoust. Soc. Amer.* **51**, 1270-1278, 1972.
- 27) V. A. Kozhevnikov and L. A. Chistovich, *Speech: articulation and perception*. 2nd print, Clearinghouse, Washington D.C., 1966 (orig. in Russian).

- 28) I. H. Slis, The influence of articulatory effort on the timing of speech. Paper read at the Symposium on Experimental and Theoretical Approaches to the Role of Time in Speech, Language Centre, University of Essex, January 1972.
- 29) N. Chomsky and M. Halle, The sound pattern of English. Harper and Row, New York, Evanston and London, 1968.
- 30) P. Delattre and M. Hohenberg, Duration as a cue to the tense/lax distinction in German unstressed vowels. *International Rev. appl. Linguistics* 6, 367-390, 1968.
- 31) Ch. F. Hockett, A manual of phonology. *International J. Amer. Linguistics* 21, no. 4, part 1, 1955.
- 32) J. S. Perkell, Physiology of speech production: results and implications of a quantitative cineradiographic study. Research Monographs no. 53, M.I.T. Press, Cambridge Mass. and London, 1969.
- 33) A. Cohen, C. L. Ebeling, K. Fokkema and A. G. F. van Holk, *Fonologie van het Nederlands en het Fries*. Martinus Nijhoff, The Hague, 1961.
- 34) P. Delattre, Some factors of vowel duration and their cross-linguistic validity. *J. acoust. Soc. Amer.* 34, 1141-1143, 1962.
- 35) A. Cohen, I. H. Slis and J. 't Hart, Perceptual tolerances of isolated Dutch vowels. *Phonetica* 9, 65-78, 1963.
- 36) A. W. de Groot, De wetten der fonologie en hun betekenis voor de studie van het Nederlands. *De nieuwe Taalgids* 25, 225-243, 1931.
- 37) K. Hadding-Koch and A. S. Abramson, Duration versus spectrum in Swedish vowels: some perceptual experiments. *Studia linguistica* 18, 94-107, 1964.
- 38) H. Zwaardemaker en L. P. H. Eykman, *Leerboek der fonetiek*. De Erven F. Bohn, Haarlem, 1928.
- 39) E. Fischer-Jørgensen, Sound duration and place of articulation. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 17, 175-207, 1964.
- 40) Q. McNemar, *Psychological statistics*. John Wiley and Sons, Inc., New York and London, 1962, 3rd edition.
- 41) B. E. F. Lindblom, Vowel duration and a model of lip mandible coordination. Royal Institute of Technology, Stockholm, Speech Transmission Laboratory, Quarterly Progress and Status Report 4, 1-29, 1967.
- 42) A. W. de Groot, *Inleiding tot de algemene taalwetenschap*. J. B. Wolters, Groningen, 1962.
- 43) O. Jespersen, *Lehrbuch der Phonetik*. B. G. Teubner, Leipzig and Berlin, 1913, 2 Aufl.
- 44) N. van Wijk, *Phonologie, een hoofdstuk uit de structurele taalwetenschap*. Nijhoff, The Hague, 1939.
- 45) N. S. Troubetzkoy, *Principes de phonologie*. Librairie C. Klincksieck, Paris, 1957. Orig. N. S. Trubetzkoy, *Grundzüge der Phonologie*. Travaux du Cercle linguistique de Prague 7, Prague, 1939.
- 46) R. Jakobson, Die Betonung und ihre Rolle in der Wort- und Syntagmaphonologie. *Selected Writings* 1, 117-136, Mouton & Co., The Hague, 1962. Also in *Travaux du Cercle linguistique de Prague* 4, Prague, 1931, pp. 164-182.
- 47) N. van Wijk, Quantiteit en intonatie. *Mededelingen der Koninklijke Nederlandse Academie van Wetenschappen, Afdeling Letterkunde, Nieuwe Reeks*, deel 3, no. 1, 1940.
- 48) A. L. Fliflet, Einige Beobachtungen über Anschluss und Silbe. *Proceedings of the Fourth International Congress of Phonetic Sciences (Helsinki Sept. 4-9, 1961)*, A. Sovijärvi and P. Aalto (eds), *Janua Linguarum, series major* 10, Mouton & Co., The Hague, 1962, pp. 610-615.
- 49) E. Fischer-Jørgensen, Untersuchungen zum sogenannten festen und losen Anschluss. *Kopenhager germanistische Studien* 1, 138-164, 1969.
- 50) I. H. Slis, Articulatory effort and its durational and electromyographical correlates. *Phonetica* 23, 171-188, 1971.
- 51) C. B. van Haeringen, De plaats van ie, oe en uu in het Nederlandse klinkerstelsel. *Album Edgard Blancquaert*, 159-164. Michiels, Tongeren, 1958.
- 52) B. van den Berg, *Fonetiek van het Nederlands*. Van Goor Zonen, The Hague, 1960.
- 53) A. Cohen, Diphthongs, mainly Dutch. In: *Form and substance*, presented to Eli Fischer-Jørgensen February 11th, 1971; L. L. Hammerich, R. Jakobson and E. Zwirner (eds), *Akademisk Forlag, Copenhagen*, 1971, pp. 277-289.
- 54) H. Mol, *Fonetische zekerheden*. *De nieuwe Taalgids* 62, 161-167, 1969.
- 55) F. J. Koopmans-van Beinum, *Nog meer fonetische zekerheden*. *De nieuwe Taalgids* 62, 245-250, 1969.
- 56) I. H. Slis and A. F. V. van Katwijk, *Onderzoek naar Nederlandse tweeklanken*.

- Institute for Perception Research, Eindhoven, Rapport No. 31, 1963.
- 57) T. Gay, A perceptual study of English diphthongs. Haskins Status Report on Speech Research, Supplement, 1967.
  - 58) T. Gay, Effect of speaking rate on diphthong formant movement. *J. acoust. Soc. Amer.* **44**, 1570-1573, 1968.
  - 59) T. Gay, A perceptual study of American English diphthongs. *Language and Speech* **13**, 65-88, 1970.
  - 60) R. H. Stetson, *Motor phonetics; a study of speech movement in action*. North-Holland Publ. Comp., Amsterdam, 1951, 2nd ed.
  - 61) G. Heike, *Sprachliche Kommunikation und linguistische Analyse*. Sprachwissenschaftliche Studienbücher 2, Carl Winter Universitätsverlag, Heidelberg, 1969.
  - 62) A. Cohen and J. 't Hart, On the anatomy of intonation. *Lingua* **19**, 177-192, 1967.
  - 63) B. E. F. Lindblom, Temporal organization of syllable production. Royal Institute of Technology, Stockholm, Speech Transmission Laboratory, Quarterly Progress and Status Report 2-3, 1-5, 1968.
  - 64) D. H. Klatt, On predicting the duration of the phonetic segment [s] in English. Massachusetts Institute of Technology, Cambridge, Mass., Research Laboratory of Electronics, Quarterly Progress Report **103**, 111-126, 1971.
  - 65) A. W. F. Huggins, On the perception of temporal phenomena in speech. *J. acoust. Soc. Amer.* **51**, 1279-1290, 1972.
  - 66) L. Roudet, *Éléments de phonétique générale*. Librairie Universitaire, H. Welter, Paris, 1910.
  - 67) E. A. Meyer, Zur Vokaldauer im Deutschen. In: *Nordiska studier tillegnade A. Noreen*, 347-356, Uppsala, 1904.
  - 68) B. Collinder, Das Wort als phonetische Einheit, Sprachwissenschaft und Wahrscheinlichkeit. *Acta Universitatis Upsaliensis, Studia uralica et altaica Upsaliensia* 1, Uppsala, 1939.
  - 69) B. Malmberg, Die Quantität als phonetisch-phonologischer Begriff. *Lunds Universitets Årsskrift, N. F., Avd. 1, Bd. 41, Nr. 2*, 1944/1945.
  - 70) D. Jones, Chronemes and tonemes. *Acta linguistica* **4**, 1-10, 1944 (1948).
  - 71) I. Fónagy and K. Magdics, Speed of utterance in phrases of different lengths. *Language and Speech* **3**, 179-192, 1960.
  - 72) J. Gaitenby, The elastic word. Haskins Status Report on Speech Research **SR2**, 3.1-3.12, 1965.
  - 73) T. Tarnóczy, Can the problem of automatic speech recognition be solved by analysis alone? *Rapports du 5e Congrès internationale d'Acoustique (Liège 7-14 sept., 1965)*, Liège, 1965, vol. 2, pp. 371-387.
  - 74) B. Lindblom and K. Rapp, Reexamination of the compensatory adjustment of vowel duration in Swedish words. Paper submitted to the Symposium on Experimental and Theoretical Approaches to the Role of Time in Speech, Language Centre, University of Essex (January 1972).
  - 75) A. M. Liberman, F. Ingemann, L. Lisker, P. Delattre and F. S. Cooper, Minimal rules for synthesizing speech. *J. acoust. Soc. Amer.* **31**, 1490-1499, 1959.
  - 76) L. Lisker, F. S. Cooper and A. M. Liberman, The uses of experiment in language description. *Word* **18**, 82-106, 1962.
  - 77) I. G. Mattingly, Synthesis by rule of prosodic features. *Language and Speech* **9**, 1-13, 1966.
  - 78) I. G. Mattingly, Synthesis by rule as a tool for phonological research. *Language and Speech* **14**, 47-56, 1971.
  - 79) T. P. Barnwell, An algorithm for segment durations in a reading machine context. Massachusetts Institute of Technology, Cambridge, Mass., Research Laboratory of Electronics, Technical Report No. 479, 1971.
  - 80) D. H. Klatt, A theory of segmental duration in English. Paper presented at the 82nd meeting of the Acoustical Society of America, Denver, Colorado, October 19-22, 1971.
  - 81) I. H. Sliis, Rules for the synthesis of speech. Institute for Perception Research, Eindhoven, Annual Progress Report **6**, 28-31, 1971.
  - 82) A. W. F. Huggins, The perception of timing in natural speech I: compensation within the syllable. *Language and Speech* **11**, 1-11, 1968.
  - 83) L. C. W. Pols, L. J. Th. van der Kamp and R. Plomp, Perceptual and physical space of vowel sounds. *J. acoust. Soc. Amer.* **46**, 458-467, 1969.
  - 84) D. Terbeek and R. Harshman, Cross-language differences in the perception of natural vowel sounds. University of California, Los Angeles. Working Papers in Phonetics **19**, 26-38, 1971.

- 85) A. Cohen and L. F. Willems, Some aspects of the distinctive feature theory in the light of psychoacoustic evidence: the perception of formants of front and back vowels. Institute for Perception Research, Eindhoven, Annual Progress Report 2, 93-99, 1967.
- 86) P. H. Geiger and F. A. Firestone, The estimation of fractional loudness. *J. acoust. Soc. Amer.* 5, 25-30, 1933.
- 87) S. S. Stevens and H. Davis, *Hearing, its psychology and physiology*. John Wiley and Sons, New York, 1939.
- 88) W. Dixon Ward, Absolute pitch, part II. *Sound; its Uses and Control* 2, 33-41, 1963.
- 89) R. Collier, De plaats van de top van een toonhoogtestijging bij optimale klemtoonwaarneming. Institute for Perception Research, Eindhoven, Rapport no. 184, 1970.
- 90) J. G. Blom and J. Z. Uys, Some notes on the existence of a "universal concept" of vowels. *Phonetica* 15, 65-85, 1966.
- 91) L. F. Willems, IPOVOX II: A speech synthesizer. Institute for Perception Research, Eindhoven, Annual Progress Report 1, 120-123, 1966.
- 92) I. H. Slis and H. F. Muller, A computer programme for synthesis by rule. Institute for Perception Research, Eindhoven, Annual Progress Report 6, 24-28, 1971.
- 93) A. S. House, On vowel duration in English. *J. acoust. Soc. Amer.* 33, 1174-1178, 1961.
- 94) S. G. Nootenbooth, Some timing factors in the production and perception of vowels. Paper read at the Symposium on Experimental and Theoretical Approaches to the Role of Time in Speech, Language Centre, University of Essex, January 1972.
- 95) M. Halle and K. N. Stevens, On the mechanism of glottal vibrations for vowels and consonants. Massachusetts Institute of Technology, Cambridge, Mass., Research Laboratory of Electronics, Quarterly Progress Report 85, 267-271, 1967.
- 96) W. S. Y. Wang, The many uses of  $F_0$ . Paper presented at the Kyoto Speech Communication Symposium 1968. Project on Linguistic Analysis Report 8, 2. Series, W1-W35. University of California, Berkeley, 1968.
- 97) I. Lehiste and G. E. Peterson, Vowel amplitude and phonemic stress in American English. *J. acoust. Soc. Amer.* 31, 428-435, 1959.
- 98) N. Chomsky, The formal nature of language. In: E. H. Lenneberg, *Biological foundations of language*, John Wiley and Sons, New York, London, Sidney, 1967, pp. 397-442.
- 99) G. Fant, Distinctive features and phonetic dimensions. In: G. E. Perren and J. L. M. Trim (eds), *Applications of Linguistics, Selected Papers of the Second International Congress of Applied Linguistics* (Cambridge 1969), Cambridge University Press, Cambridge, 1971, pp. 219-239.
- 100) I. Lehiste, Manner of articulation, parallel processing, and the perception of duration. Paper read at the Symposium on Experimental and Theoretical Approaches to the Role of Time in Speech, Language Centre, University of Essex, January 1972.
- 101) P. Ladefoged, *Linguistic Phonetics*. University of California, Los Angeles, Working Papers in Phonetics 6, 1967.
- 102) S. E. G. Öhman, A. Persson and R. Leanderson, Speech production at the neuromuscular level. Paper read at the 6th International Congress of Phonetic Sciences (Prague Sept. 7-13, 1967), Royal Institute of Technology, Stockholm, Speech Transmission Laboratory, Quarterly Progress and Status Report 2-3, 15-19, 1967.
- 103) M. A. A. Tatham and K. Morton, Some electromyographic data towards a model of speech production. *Language and Speech* 12, 39-53, 1969.
- 104) P. Lieberman, Towards a unified phonetic theory. *Linguistic Inquiry* 1, 307-322, 1970.
- 105) R. Collier, From pitch to intonation. Unpublished doctoral dissertation, Leuven 1972.
- 106) J. W. Bresnan, Sentence stress and syntactic transformations. *Language* 47, 257-281, 1971.
- 107) A. M. Liberman, F. S. Cooper, K. S. Harris and P. F. MacNeilage, A motor theory of speech perception. Stockholm Speech Communication Seminar (Aug. 30-31, 1962), Royal Institute of Technology, Stockholm, Speech Transmission Laboratory, 1963, Contributed Papers vol. 2, Paper D3.
- 108) A. M. Liberman, F. S. Cooper, D. P. Shankweiler and M. Studdert-Kennedy, Perception of the speech code. *Psychological Review* 74, 431-461, 1967.
- 109) G. D. Allen, Timing control in speech production: some theoretical and methodological issues. Paper submitted to the Symposium on Experimental and Theoretical Approaches to the Role of Time in Speech. Language Centre, University of Essex, January 1972.
- 110) C. D. Creelman, Human discrimination of auditory duration. *J. acoust. Soc. Amer.* 34, 582-593, 1962.
- 111) M. Treisman, Temporal discrimination and the indifference interval: implications for a model of the "internal clock". *Psychological Monographs, American Psychological*

- Association, Washington D.C., 1963, vol. 77, No. 13, No. 576.
- <sup>112)</sup> J. A. Michon, Timing in temporal tracking. Institute for Perception RVO-TNO, Soesterberg, 1967.
- <sup>113)</sup> S. A. Zimmerman and S. M. Sapon, Note on vowel duration seen cross-linguistically. *J. acoust. Soc. Amer.* **30**, 152-153, 1958.
- <sup>114)</sup> A. S. Abramson, The vowels and tones of standard Thai: acoustical measurements and experiments. *International J. Amer. Linguistics* **28**, no. 2, part 3, 1962.
- <sup>115)</sup> E. A. Meyer and Z. Gombocz, Zur Phonetik der ungarischen Sprache, *Le Monde Oriental*, 122-187, 1907-1908, publ. in 1909.
- <sup>116)</sup> J. von Lazicius, Zur Lautquantität. *Archiv für vergleichende Phonetik* **3**, 245-249, 1939.
- <sup>117)</sup> S. Einarsson, Beiträge zur Phonetik der isländischen Sprache, Oslo, 1927.
- <sup>118)</sup> S. Bergsveinsson, Grundfragen der isländischen Satzphonetik. *Phonometrische Forschungen*, Munksgaard, Copenhagen; Metten, Berlin, 1941, Reihe A, vol. 2.
- <sup>119)</sup> J. Lehtonen, Aspects of quantity in standard Finnish. *Studia philologica Jyväskyläensia* **6**, Jyväskylä, 1970.
- <sup>120)</sup> W. A. Ainsworth, Duration as a cue in the recognition of synthetic vowels. *J. acoust. Soc. Amer.* **51**, 648-651, 1972.
- <sup>121)</sup> F. Äimä, Phonetik und Lautlehre des Inarilappischen 1-2. *Mémoires de la Société Finno-Ougrienne Helsinki* **42**, 1-118; **43**, 1-249, 1914-1918.
- <sup>122)</sup> D. B. Fry, Experiments in the perception of stress. *Language and Speech* **1**, 126-152, 1958.
- <sup>123)</sup> D. L. Bolinger, A theory of pitch accent in English. *Word* **14**, 109-149, 1958.
- <sup>124)</sup> P. Lieberman, On the acoustic basis of the perception of intonation by linguists. *Word* **21**, 40-54, 1965.
- <sup>125)</sup> W. Jassem, J. Morton and M. Steffen-Batóg, The perception of stress in synthetic speech-like stimuli by Polish listeners. In: W. Jassem (ed.), *Speech analysis* **1**, Państwowe, Wydawnictwo Naukowe, Warsaw, 1968, pp. 289-308.
- <sup>126)</sup> A. F. V. van Katwijk and G. A. Govaert, Prominence as a function of the location of pitch movement. Institute for Perception Research, Eindhoven, *Annual Progress Report* **2**, 115-117, 1967.
- <sup>127)</sup> R. Collier and J. 't Hart, A grammar of pitch movements in Dutch intonation. Institute for Perception Research, Eindhoven, *Annual Progress Report* **6**, 17-21, 1971.
- <sup>128)</sup> J. L. Flanagan, C. H. Coker, L. R. Rabiner, R. W. Schaefer and N. Umeda, Synthetic voices for computers. *IEEE Spectrum* **7**, no. 10, 22-45, 1970.

**Author Index** (Alphabetically listed names of authors mentioned in the text, followed by page numbers).

- Abercrombie, D.: 3.  
Abramson, A. S.: 18, 26, 113.  
Åimä, F.: 119.  
Ainsworth, W. A.: 117.  
Allen, G. D.: 109, 110, 111.
- Barnwell, T. P.: 76, 77.  
Berg, B. van den: 47.  
Bergveinsson, S.: 114.  
Blom, J. G.: 79, 122.  
Bolinger, D. L.: 119, 120.  
Boomer, D. S.: 5, 6.  
Bresnan, J. W.: 106.
- Chistovich, L. A.: 19, 49, 118.  
Chomsky, N.: 23, 25, 48, 76, 103, 104, 105, 112, 113.  
Cohen, A.: 5, 13, 18, 24, 26, 39, 41, 45, 47, 48, 49, 51, 78, 79, 86, 104, 112, 120.  
Collier, R.: 79, 106, 120, 122.  
Collinder, B.: 63.  
Cooper, F. S.: 76.  
Creelman, C. D.: 110.
- Davis, H.: 79.  
Delattre, P.: 24, 85, 112, 114.  
Domburg, J.: 9.
- Einarsson, E.: 113.  
Elert, C.-C.: 13, 26, 29, 35, 112, 113, 114.  
Eykman, L. P. H.: 27, 45, 50.
- Fant, G.: 104, 105, 107.  
Fischer-Jørgensen, E.: 29, 30, 40, 41.  
Firestone, F. A.: 79.  
Flanagan, J. L.: 121.  
Fliflet, A. L.: 41.  
Fónagy, I.: 63.  
Fromkin, V. A.: 5, 6, 47.  
Frøkjær-Jensen, B.: 52, 53.  
Fry, D. B.: 119.
- Gaitenby, J.: 63.  
Gay, T.: 47.  
Geiger, P. H.: 79.  
Gombocz, Z.: 113.  
Govaert, G. A.: 119.  
Groot, A. W. de: 26, 39, 45.
- Hadding-Koch, K.: 26.  
Haeringen, C. B. van: 45.  
Halle, M.: 23, 25, 48, 76, 86, 103, 104, 105, 112, 113.  
Hart, J. 't: 18, 41, 51, 78, 79, 120.  
Heike, G.: 50.  
Hockett, Ch. F.: 24.  
Hohenberg, M.: 24, 114.  
House, A. S.: 1, 85, 111.  
Huggins, A. W. F.: 18, 60, 78, 85, 111, 121.
- Jakobson, R.: 40.  
Jassem, W.: 119.



- Jespersen, O.: 40.  
Jones, D.: 63, 119.
- Kamp, L. J. Th. van der: 78.  
Katwijk, A. F. V. van: 47, 119.  
Klatt, D. H.: 60, 76, 77, 121.  
Koopmans-Van Beinum, F. J.: 47.  
Kozhevnikov, V. A.: 19, 49, 118.
- Ladefoged, P.: 105, 107.  
Lammers, C. A.: 82.  
Laver, J. D. M.: 5, 6.  
Lazicsius, J. von: 113.  
Lehiste, I.: 1, 17, 18, 24, 26, 59, 60, 63, 85, 88, 104, 118.  
Lehtonen, J.: 114, 118.  
Lieberman, A. M.: 76, 109.  
Lieberman, P.: 105, 107, 119.  
Lindblom, B. E. F.: 1, 24, 35, 57, 58, 63, 66, 72, 115, 119.  
Lisker, L.: 3, 18, 76.  
Loonen, A. R. M.: 80.
- Mackay, D. G.: 5.  
Magdics, K.: 63.  
Malmberg, B.: 63.  
Mattingly, I. G.: 76, 77, 78.  
McNemar, Q.: 31.  
Melotte, H. E. M.: 9.  
Meyer, E. A.: 63, 113, 119.  
Michon, J. A.: 110.  
Mol, H.: 47.  
Moonen, G. J. J.: 9, 82.  
Morton, J.: 119.  
Morton, K.: 105.  
Moulton, W. G.: 18, 45, 47, 50.  
Muller, H. F.: 9, 82.
- Nooteboom, S. G.: 5, 47, 86.
- Öhman, S. E. G.: 105.
- Paul, A. P.: 1.  
Perkell, J. S.: 24.  
Peterson, G. E.: 1, 88.  
Plomp, R.: 78.  
Pols, L. C. W.: 78.
- Rapp, K.: 63, 66, 72, 73, 115, 119.  
Roudet, L.: 62, 119.
- Sapon, S. M.: 112.  
Schouten, J. F.: 9, 18.  
Sievers, E.: 40.  
Slis, I. H.: 9, 13, 18, 20, 41, 47, 59, 76, 79, 85, 112.  
Steffen-Batóg, M.: 119.  
Stetson, R. H.: 50.  
Stevens, K. N.: 1, 86.  
Stevens, S. S.: 79.
- Tarnóczy, T.: 63, 119.  
Tatham, M. A. A.: 105.  
Terbeek, D.: 78.  
Treisman, M.: 110.  
Trubetzkoy, N. S.: 40, 50.

Uys, J. Z.: 79, 122.

Wang, W. S. Y.: 86.

Ward, D.: 79.

Wijk, N. van: 40, 45, 47.

Willems, L. F.: 9, 10, 78, 80.

Zimmerman, S. A.: 112.

Zwaardemaker, H.: 27, 45, 50.

List of phonetic symbols with key words

a:	Dutch	<u>ma</u> an
ɑ	„	ma <u>n</u>
ɛ	„	he <u>k</u>
ɔ	„	ko <u>p</u>
æ	„	hu <u>t</u>
ɪ	„	ki <u>p</u>
o:	„	bo <u>o</u> t
ø:	„	be <u>u</u> k
e:	„	bee <u>k</u>
u	„	bo <u>e</u> k
y	„	fu <u>u</u> t
i	„	ni <u>e</u> t
au	„	ko <u>u</u> d
ʌy	„	lu <u>i</u> d
ɛi	„	me <u>i</u> d
ə	„	be <u>g</u> in
æ:	„	fre <u>u</u> le
ɔ:	„	ro <u>s</u> e
ɛ:	„	ser <u>r</u> e
t	„	ta <u>a</u> k
p	„	pa <u>a</u> l
k	„	ko <u>e</u> k
m	„	ma <u>a</u> t
n	„	na <u>a</u> d
r	„	pa <u>a</u> rd
l	„	lo <u>o</u> s
z	„	de <u>z</u> e
v	„	ov <u>e</u> r
d	„	di <u>e</u>
θ	Swedish	skuff <u>ʃ</u>

In the cases where the symbols used in the illustrations deviate from those in the above list the normal IPA symbols are used. Stressed vowels are printed in bold face in the text and are marked with accentuation marks in the illustrations.

## Samenvatting

De bedoeling van deze studie was te zoeken naar regelmatigheden in de meetbare duren van vocalen en regels te formuleren die kunnen helpen bij het voorstellen van de correcte duuropbouw van woorden en zinnen in het Nederlands. We hoopten dat de resultaten een bijdrage zouden kunnen leveren in het verhelderen van de relatie tussen de discrete linguïstische specificatie van spraak en een fonetische specificatie van de min of meer continue eigenschappen van het spraakgebeuren (hoofdstuk 1).

In hoofdstuk 2, gewijd aan de beschrijving en discussie van een aantal articulatorische metingen aan segmentduren in nonsenswoorden, werd ruim aandacht gegeven aan de rol van fonologische vocaalkwantiteit als een onderliggende determinant van meetbare vocaalduren. We konden laten zien dat de vocalen van het Nederlands in te delen zijn in twee fonetische categorieën, een categorie van lange en een categorie van korte vocalen. De vocalen die geen deel hebben aan een fonologische kwantiteitsoppositie, passen zich fonetisch aan bij een van de kwantiteitscategorieën.

Op het niveau van de fonetische representatie kunnen vocalen een specificatie krijgen van het kwantiteitskenmerk. Voor het Nederlands hoeven maar twee mogelijke waarden van deze specificatie verondersteld worden, één voor de lange en één voor de korte vocalen. Deze waarden kan men zien als optimale, ideale, vocaalduren, of als doelwaarden op een interne representatie van de tijd. De realisaties van deze optimale duren zijn sterk afhankelijk van zulke factoren als klemtoon en positie. De effecten van klemtoon en positie kunnen beschreven worden met behulp van enkele eenvoudige kwantitatieve regels. Een van deze regels beschrijft het effect van woordlengte op de duur van beklemtoonde vocalen. Dit effect bleek hoofdzakelijk beperkt te blijven tot het effect van het aantal syllaben dat later in het woord komt. De andere regels genereren de duren van onbeklemtoonde vocalen. Dezelfde regels zijn in principe van toepassing op woorden met een toonhoogteaccent en op woorden zonder een toonhoogteaccent. Tussen geïsoleerd gesproken woorden en woorden die ingebed zijn in een woordgroep bestaan kwantitatieve verschillen die makkelijk te beschrijven zijn (hoofdstuk 2).

Het soort van regelmatigheden dat beschreven wordt door deze regels kan men zich niet makkelijk bewust maken door introspectieve reflexie op het klankaspect van woorden en zinnen. Toch leek het van belang om de perceptieve realiteit van de beschreven regelmatigheden aan te tonen. Dit kon met behulp van de methode van gelijkstelling aan een intern criterium. Aan proefpersonen werd gevraagd de duur van vocalen die ingebed waren in gesynthetiseerde woorden in te stellen. We konden laten zien dat de regelmatigheden die voorspeld waren door de duurregels ook tot uiting kwamen in de instellingsproeven,

hoewel in veel gevallen de proefpersonen zich van de betreffende duurverschillen niet bewust waren (hoofdstuk 3).

De resultaten van ons onderzoek laten zien dat er veel details in de duuropbouw van spraak zijn die hun neerslag vinden in de impliciete kennis van taalgebruikers over het klankaspect van taal en die als zodanig een legitiem object vormen voor linguïstische beschrijving, maar die niet gemakkelijk beschreven kunnen worden in de discrete linguïstische specificatie van spraak. Wat we gewoonlijk vocaalduren noemen zouden beter duren van syllabekernen genoemd kunnen worden. Syllabekernen lijken te corresponderen met stukken in het spraakcontinuum die luider zijn dan hun omgeving en die op hun beurt weer corresponderen met periodes dat er fonatie is en geen occlusie in het spraakkanaal. De regels voor vocaalduren lijken aan te grijpen op deze syllabekernen en kunnen als zodanig geen deel zijn van de fonologische component van taal. Aansluitend bij suggesties van verschillende fonetici stellen we voor dat de regels voor vocaalduren samen met regels voor implementatie van kenmerken, regels voor coarticulatie en regels voor intonatie, een fonetische component van taal uitmaken. De status van deze fonetische component is vergelijkbaar met die van de fonologische component. De input wordt gevormd door de discrete fonetische representatie van een zin, een representatie van de syntactische structuur en een representatie van de semantische structuur. De output wordt gevormd door een semicontinue representatie van spraak. Deze laatste staat veel dichterbij het akoestisch signaal dan de discrete fonetische representatie maar is daar niet identiek mee. De uitgang van de fonetische component is een mentale representatie. Een optimale specificatie van de uitgangsparameters van een spraaksynthesator moet voldoen aan de perceptieve eisen die gesteld worden door de regels van de fonetische component van taal.

Deze regels kunnen ook een basis verschaffen voor vergelijkende studies van de fonetische eigenschappen van verschillende talen. Bijvoorbeeld de traditionele wijze van vergelijking van talen met betrekking tot vocaalkwantiteit met behulp van de V/V: verhouding is niet bevredigend. Het zou bevredigender zijn als het gebruik van kwantiteit vergeleken werd op basis van de beschrijving hiervan voor iedere taal in termen van een paar abstracte optimale duren plus een verzameling duurregels die laten zien hoe deze optimale duren gerealiseerd worden onder verschillende condities van klemtoon en positie. Deze beschrijving zou zo moeten zijn dat er kwantitatieve voorspellingen uit kunnen worden afgeleid.

De regels voor vocaalduren zoals voorgesteld in deze studie kunnen een passend uitgangspunt vormen voor verder onderzoek naar systematiek in vocaalduren (hoofdstuk 4).

### **Curriculum vitae**

Sibout Govert Nooteboom werd geboren op 19 April 1939 te Makassar (Nederlands Indië). Daarop volgde het eindexamen gymnasium alpha te Rotterdam in 1958, een kleine twee jaar militaire dienst tot 1960, een kandidaatsexamen Nederlandse taal- en letterkunde aan de Rijksuniversiteit te Leiden in 1963 en een doktoraal examen algemene taalwetenschap aan dezelfde universiteit in 1966. Sinds 1 September 1966 is hij medewerker van het Instituut voor Perceptie Onderzoek te Eindhoven en sinds 1 september 1968 tevens docent in de fonetiek aan de Rijksuniversiteit te Leiden.

## STELLINGEN

## STELLINGEN

### I

Een generatief-fonologische beschrijving van een taal volgens de ideeën van Chomsky en Halle houdt zich bezig met een manier waarop woorden en woordgroepen van die taal schriftelijk kunnen worden weergegeven, niet met de wijze waarop zij klinken.

Chomsky and Halle, *The Sound Pattern of English*, 1968.

### II

De hypothese dat ieder mens in aanleg beschikt over dezelfde vaste, niet uitbreidbare, verzameling fonetische kenmerken is niet falsifieerbaar en onnodig in een theorie van taalverwerving.

Chomsky, *The formal nature of language*, in E. H. Lenneberg, *Biological Foundations of Language*, 1967, 397-442, speciaal 403-404.  
Chomsky and Halle, *The Sound Pattern of English*, 1968, 4-5, 293-295.

### III

Het bezwaar dat Chomsky en Halle aanvoeren tegen hun eigen benadering van kenmerken en evaluatie, namelijk dat deze al te formeel is, geldt in dezelfde mate voor de methode die zij kiezen om dit bezwaar te bestrijden, t.w. de markeringsconventies.

Chomsky and Halle, *The Sound Pattern of English*, 1968, 400.

### IV

De perceptieve werkelijkheid van het klankaspect van een taal verschilt sterk van wat „de zorgvuldige en niet-naïeve impressionistische fonetikus, die de taal kent” zich bewust kan maken.

Chomsky and Halle, *The Sound Pattern of English*, 1968, 25.

### V

A. W. F. Huggins heeft in zijn studie „Just Noticeable Differences for Segment Duration in Natural Speech” geen juist waarneembare verschillen, maar wel perceptieve toleranties bepaald.

A. W. F. Huggins, *Journal of the acoustical Society of America* 51, 1270-1278, 1972.



## VI

De grote verschillen in gemeten duur tussen de klinkers in de woorden *muur*, *tien* en *hoed*, gevonden door mevrouw F. J. Koopmans-van Beinum, moeten niet aan inherente verschillen tussen deze klinkers worden toegeschreven, maar aan de ene kant aan een fonologisch effect van de *r* in *muur* op de klinkerkwantiteit, waardoor de *uu* lang wordt, en aan de andere kant aan een fonetisch effect van de volgkonsonanten in *tien* en *hoed* op de klinkerduur, waardoor de korte *ie* een grotere duur krijgt dan de korte *oe*.

F. J. Koopmans-van Beinum, *Vergelijkend Fonetisch Klinkeronderzoek*, Instituut voor Fonetische Wetenschappen, Universiteit van Amsterdam, publicatie No 32, Amsterdam 1971.

## VII

De interne representatie van een klinkerduur kan nauwkeuriger zijn dan de spectrografische meting van zijn akoestisch korrelaat.

## VIII

In de titel van het artikel „Measuring duration in Dutch” wordt ten onrechte de indruk gewekt dat het artikel handelt over het meten van duur in het Nederlands.

W. G. Klooster and H. J. Verkuyl, *Foundations of Language* 8, 62-96.

## IX

Bij discussies over de spelling van het Nederlands neemt de kwaliteit van de argumenten af naarmate de extremiteit van de standpunten toeneemt.

## X

De Nederlandse staat is medeverantwoordelijk voor de gevolgen van het Amerikaanse optreden in Zuid-Oost Azië, door haar vrijwel kritiekloze aanvaarding van de politiek die de Verenigde Staten van Amerika in Zuid-Oost Azië voeren.

The first part of the paper is devoted to a general discussion of the problem. It is shown that the problem is equivalent to the problem of finding a function  $f(x)$  which satisfies the conditions

(1)  $f(x) > 0$  for all  $x$  in the interval  $(0, \infty)$ , and

(2)  $f(x) = 0$  for all  $x$  in the interval  $(-\infty, 0)$ .

It is shown that such a function exists and is unique. The function is given by

$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x > 0 \end{cases}$

The second part of the paper is devoted to a detailed study of the properties of the function  $f(x)$ . It is shown that the function is continuous at  $x=0$  and that it is differentiable at  $x=0$ . The derivative of the function at  $x=0$  is given by

$f'(0) = 1$

The third part of the paper is devoted to a study of the asymptotic behavior of the function  $f(x)$  as  $x \rightarrow \infty$ . It is shown that the function approaches the value 1 as  $x \rightarrow \infty$ .

The fourth part of the paper is devoted to a study of the asymptotic behavior of the function  $f(x)$  as  $x \rightarrow -\infty$ . It is shown that the function approaches the value 0 as  $x \rightarrow -\infty$ .

The fifth part of the paper is devoted to a study of the asymptotic behavior of the function  $f(x)$  as  $x \rightarrow 0$ . It is shown that the function approaches the value 1 as  $x \rightarrow 0$ .

The sixth part of the paper is devoted to a study of the asymptotic behavior of the function  $f(x)$  as  $x \rightarrow \infty$ . It is shown that the function approaches the value 1 as  $x \rightarrow \infty$ .

The seventh part of the paper is devoted to a study of the asymptotic behavior of the function  $f(x)$  as  $x \rightarrow -\infty$ . It is shown that the function approaches the value 0 as  $x \rightarrow -\infty$ .

The eighth part of the paper is devoted to a study of the asymptotic behavior of the function  $f(x)$  as  $x \rightarrow 0$ . It is shown that the function approaches the value 1 as  $x \rightarrow 0$ .





