ARIPUC VOL. 15.

ISSN 0589-6681

ANNUAL REPORT

of the
Institute of Phonetics
University of Copenhagen

COPENHAGEN 1981

alm ts

ARIPUC VOL. 15.

ISSN 0589-6681

ANNUAL REPORT

of the
Institute of Phonetics
University of Copenhagen

96, Njalsgade
DK 2300 Copenhagen

COPENHAGEN 1981

This annual report is published on the basis of a grant from the Faculty of the Humanities (publications account), University of Copenhagen.

KØBENHAVNS UNIVERSITET
REPRO.AFDELINGEN

CONTENTS

(ARIPUC 15, 1981 p. 1-12)

FUNDAMENTAL FREQUENCY PATTERNING AND SENTENCE INTONATION IN TWO JUTLANDIC DIALECTS - A FIRST APPROACH

NINA THORSEN AND BENT JUL NIELSEN*

In two Jutland areas (Thy and Århus) the fundamental frequency pattern associated with the succession of a stressed plus unstressed syllables can be described as a HIGH plus LOW or FALLING pattern, which is the inverse of the Standard Copenhagen LOW plus HIGH-FALLING one. We suggest that the difference be accounted for in terms of a difference in timing with respect to the peak and troughs of a basically triangular "wave". - Non-statement intonation contours may be rarer in these dialects than in Standard Copenhagen.

I. INTRODUCTION

Recent investigations have laid bare what may be termed the groundworks of (Advanced) Standard Copenhagen Danish (ASC) intonation. This paper is a report on some very preliminary experiments which should be regarded as an extension of the work on ASC Danish, and as a first step on the way to establish the basics of intonation in the main dialects of Danish.

The material is so small that a full scale account of the considerations about choice of words and sentences to be recorded, of recording conditions, and of registrations and measurements would be out of proportion to the account of the results, and the reader is therefore referred to papers published previously about ASC (see the references). Suffice it to say that the focal points of interest were the fundamental frequency (Fo) patterning associated with stressed and unstressed syllables and the sentence intonation contours in

---

* Institute of Danish Dialectology

declarative, non-final and interrogative sentences in two Jutlandic areas, Thy and Århus.

II. PROCEDURES

A. MATERIAL

To establish the relationship between stress and Fo one should look at polysyllabic words - in varying positions in the utterance - where the placement of stress is systematically varied from the first to the last syllable, everything else being equal (stressed monosyllables succeeded by unstressed ones should also be included); to establish the relationship between sentence function and intonation contour one should look at sentences which are as much alike semantically, syntactically and rhythmically as possible, but differing in their function as e.g. statements and questions. The first condition is not easily met in the vocabulary of Danish (although stress placement can be considered to be "free" and a few minimal stress pairs can be set up), so one must resort to nonsense words where syllabic structure can be controlled, see further Thorsen (1978 and 1979). However, since the majority of the subjects were linguistically naïve, we thought it best to avoid nonsense words. As a compromise between the naturalness and the ideal experimental conditions we settled upon:

kufferter 'trunks' [ˈg̊ʰɔfʌd̥ʌ]

kartofler 'potatoes' [g̊ɑˈd̥ʰʌflʌ]

statistik 'statistics' [sd̥ad̥iˈsd̥ig̊][1]

These words were embedded in initial, medial and final position in short meaningful statements (i.e. containing a total of three or four stress groups), which were to some extent adapted morphologically and syntactically to meet the different demands of the two dialects in question.

To look at sentence intonation contours, we chose the following:

Der går mange busser fra Thisted. 'There are many buses out of Thisted.' [d̥ɑ g̊ɒ ˈmɑŋ ˈb̥usʌ fʁ̥ɑ ˈd̥ʰisteð]

Der går mange busser fra Thisted, så vi behøver ikke køre i bil. 'There are many buses out of Thisted, so we need not drive a car.'

Går der mange busser fra Thisted? 'Are there many buses out of Thisted?'

These sentences were embedded in small dialogues. All of the utterances were mixed in the transcript that was presented to the subjects (one page in all).

Apart from the structural difficulties inherent in the material, we expected that the fact that it was to be read (as opposed to a free speech situation) would present a problem with some of the subjects, especially since we wanted to elicit both fluent and natural - and above all neutral - speech from our speakers. We found no other solution to this problem than to get as many recordings from each subject as possible and hope that after critical sorting and discarding of unsuitable items, enough material would remain for analysis.

B. DIALECTS AND SUBJECTS

Two rural speakers from Thy were recorded: KJ, female, about 40 years old and NJ, male, about 70. One subject from Århus, BBA, female, about 30 years of age and one from Randlev (near Århus), ÅF, female, about 50, were recorded. BBA and ÅF speak an Eastern Jutlandic variety of Standard Danish. For a further account, see Nielsen (1959 p. 111-112) and Jensen (1967).

C. RECORDINGS

The tape-recorder was a professional Nagra R 4.2., the microphone a Sennheiser MD214N. One of the authors (BJN) supervised the Thy recordings, and Magda Nyberg (Institute of Danish Dialectology) kindly agreed to supervise the Århus recordings. Three subjects were recorded in their homes, BBA at work. Subjects were given the least possible instruction and were asked to read in as natural a manner as possible.

D. REGISTRATIONS AND MEASUREMENTS

The registrations were made at the Institute of Phonetics, with hard-ware intensity and Fo meters, registered on a mingograph, and the Fo-tracings were measured, - all in a manner completely analogous to that described e.g. in Thorsen (1980a). Mean Fo values were converted to semitones (re 100 Hz) and the mean tracings (slightly stylized) were drawn.

III. RESULTS

Figure 1-4 depict Fo tracings of each of the tri-syllabic words in sentence medial position (including the Fo course from the previous stressed vowel to the succeeding stressed vowel), except that *kartofler* with ÅF is in initial position. The number of items (N) underlying these average tracings is indicated in the top right of each figure. For reasons of space, sentence initial and final position are not shown here, but they have of course been analysed and the conclusions to be drawn below are equally valid for words in both these marginal positions.

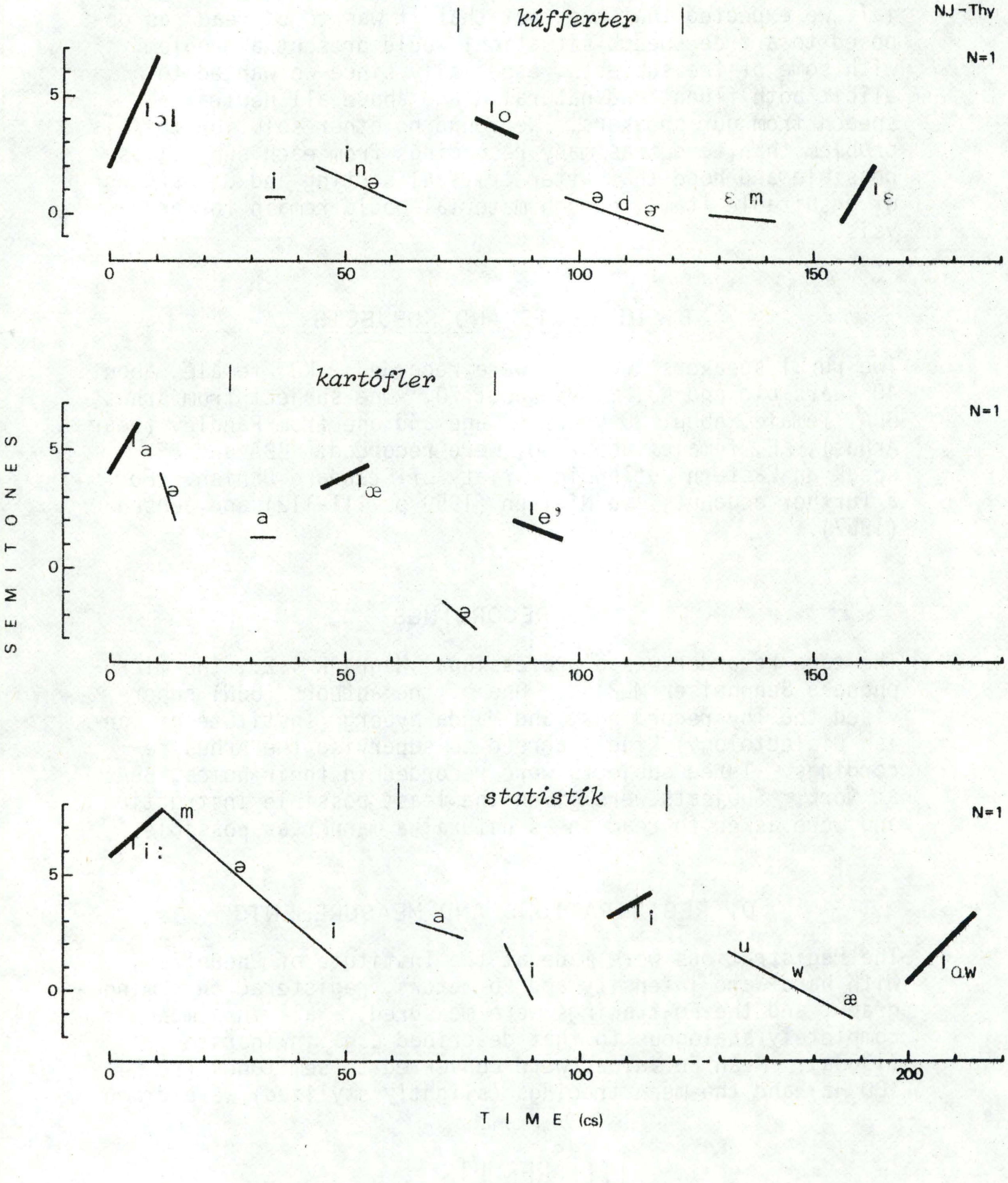


Figure 1

Fundamental frequency tracings (slightly stylized) of *kúfferter*, *kartófler* and *statistík* in sentence medial position, including the Fo course from the preceding to the succeeding stressed vowel. The sentences were: *Hun har áltid hendes kúfferter gemt hén. De har kørt mánge kartófler tíl mejeríet. Et par tímer i statistík kunne være gávnligt.* (The acute accent denotes the stressed vowels.) Zero on the logarithmic frequency scale corresponds to 100 Hz. Subject: NJ-Thy. (NJ paused slightly between *statistík* and *kunne*.)

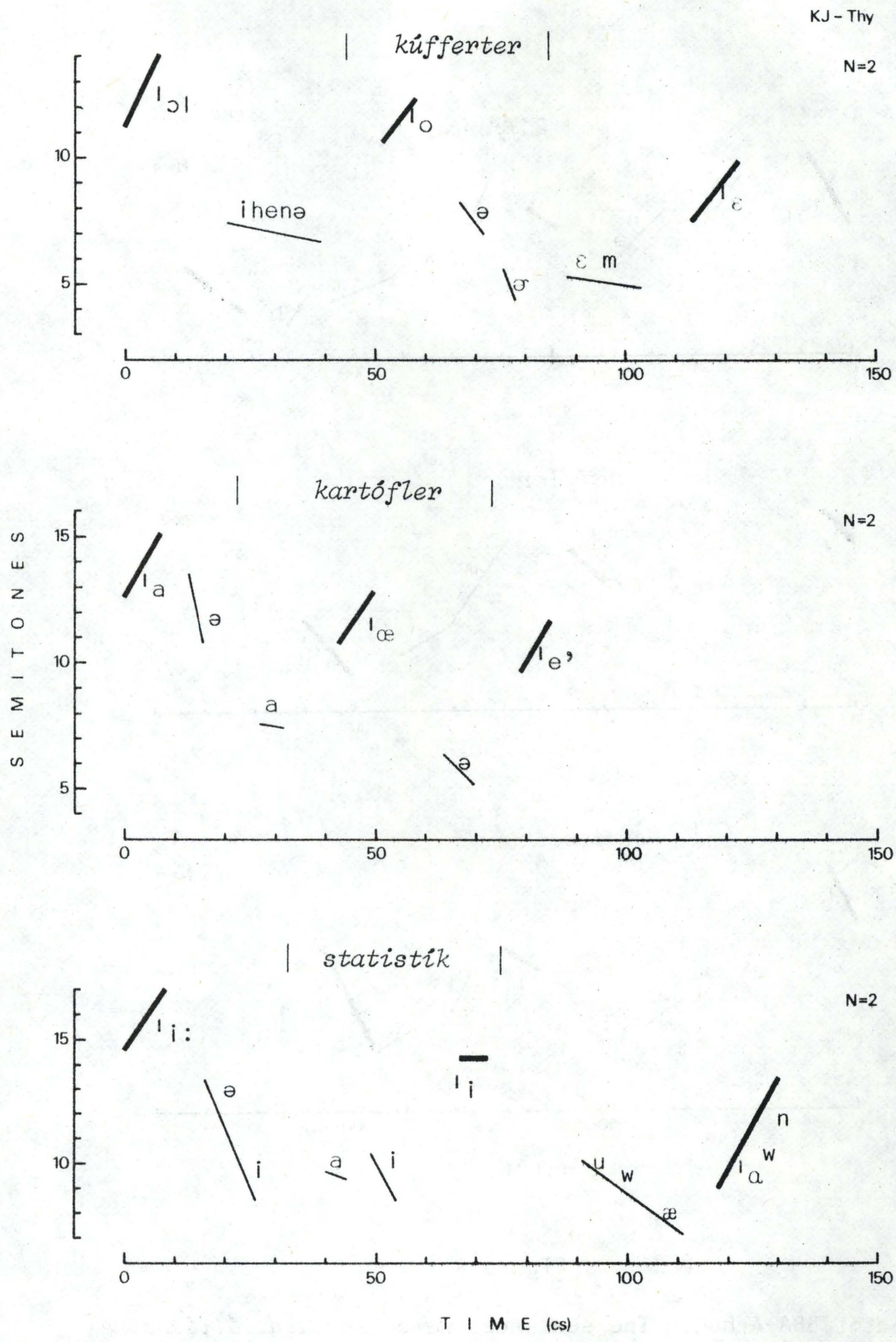


Figure 2

Subject: KJ-Thy. See further the legend to figure 1.

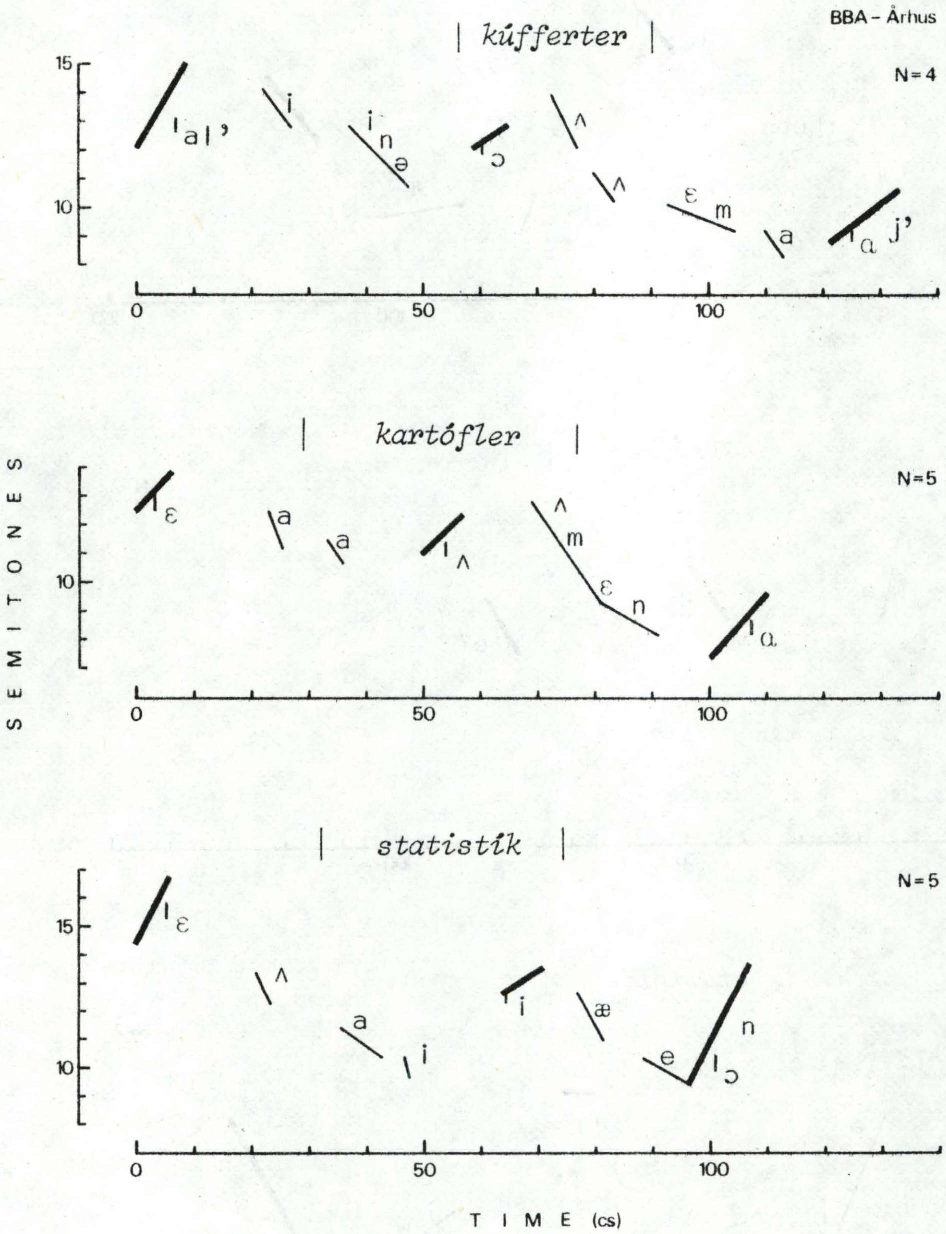


Figure 3

Subject: BBA-Århus. The sentences were *Hun havde áltid sine kúfferter gemt af véjen. Jeg kan bédst tage kartófler med en gáffel. Hans respékt for statistík er beúndringsværdig.* See further the legend to figure 1.

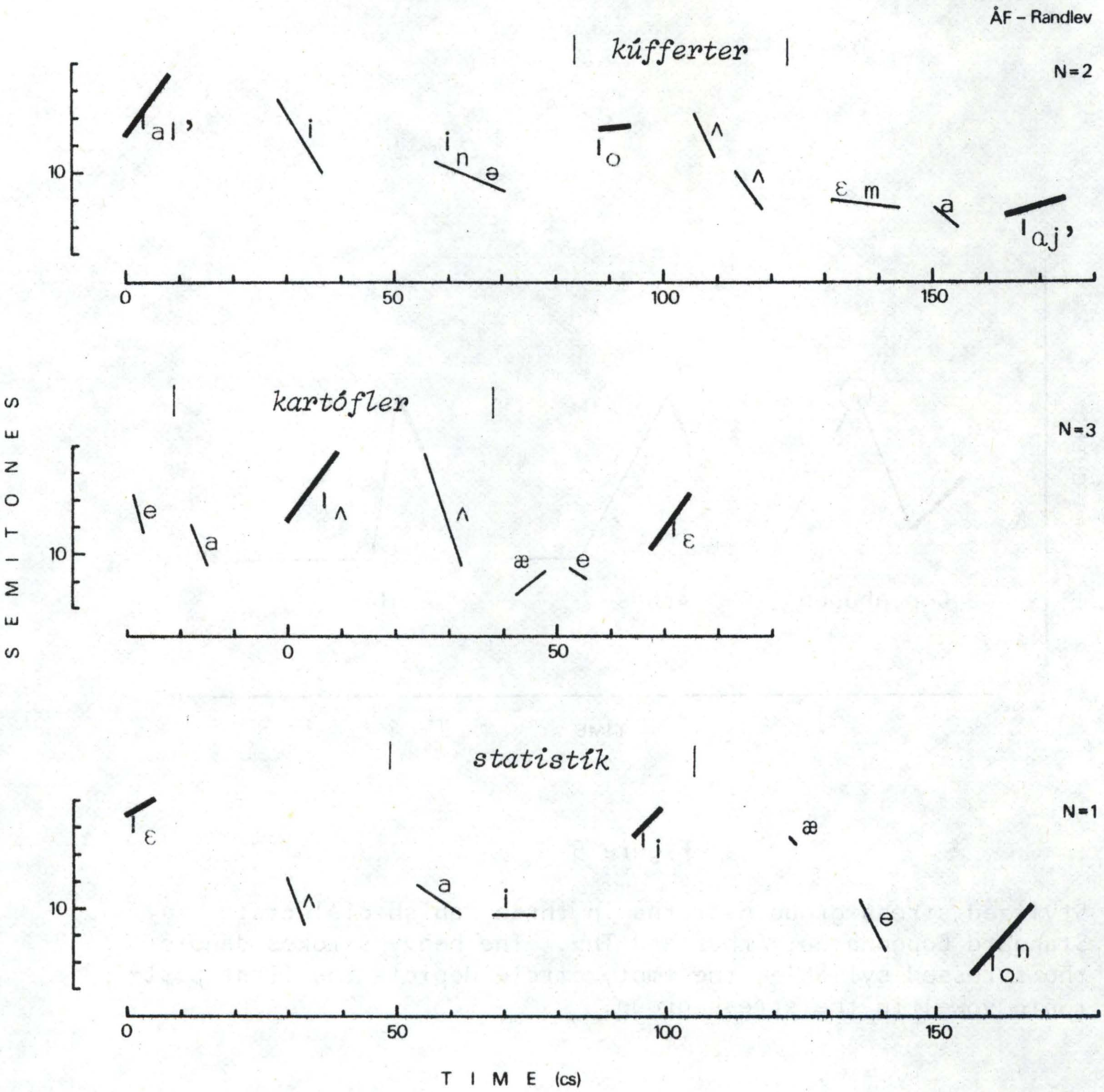


Figure 4

Subject: ÅF-Randlev (near Århus). The sentences were *Hun havde áltid sine kúfferter gemt af véjen. Til kartófler er det bédst med en gáffel. Hans respékt for statistík er beúndringsværdig.* See further the legend to figure 1.

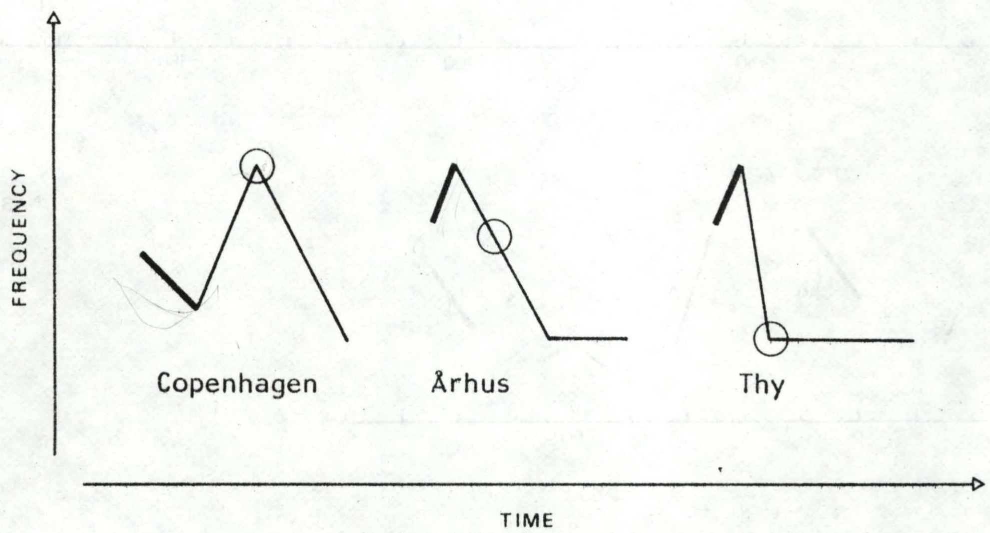


Figure 5

Stylized stress group patterns in three Danish dialects: Standard Copenhagen, Århus and Thy. The heavy strokes denote the stressed syllable, the empty circle depicts the first post-tonic vowel in the stress group.

A. STRESS GROUP PATTERNS AND INTRAVOCALIC FO MOVEMENT

As in ASC Danish the patterning of Fo seems to be independent of word boundaries, i.e. the stress group may be defined as follows: a stress group consists of a stressed syllable plus all succeeding unstressed ones, irrespective of word (or other syntactic) boundaries within the same simple sentence or, rather, on the same intonation contour (see further Thorsen 1980a, 1980b, 1980d). Thus, there is no sign in the Fo traces that e.g. a pre-tonic syllable does not simply attach itself to the tail of unstressed syllables after the preceding stressed syllable, i.e. there is no sign of any particular tonal attachment between pre-tonic and stressed syllable, nor is there any sign of a break between unstressed syllables pertaining to the preceding stressed syllable and unstressed syllables pertaining to the succeeding stressed syllable.

Both dialects generally have high and rising Fo in the stressed vowels, even in absolute sentence final position (see e.g. .. *hen*. in figure 1 and 2, top). The Thy dialect has a steep drop to the first post-tonic (see figure 1 and 2), whereas the Århus speakers have a smoother fall from the high level of the stressed syllable, and even occasionally a very slight rise to the first post-tonic. (The similarity between the Århus model in figure 5 and the stress group patterns of BBA and ÅF (figures 3 and 4) becomes more apparent when the first post-tonic vowel is reduced to a point on the frequency scale corresponding to the frequency at a point in time at 2/3 of the distance from vowel onset (which is the measuring procedure generally adopted in the analyses of ASC Danish, see e.g. Thorsen 1980d).)

The stress group patterning in the two dialects may be characterized roughly as one of HIGH + LOW (Thy) or HIGH + FALLING (Århus) - as opposed to ASC where it is LOW + HIGH-FALLING. It is of course possible to see one as the inverse of the other, but we think that a more advantageous description is to see these variations in terms of differences in timing, as follows: If we regard the archetypical stress determined Fo deflection as a triangular wave, see figure 5, we are in a position to state that in ASC the stressed syllable falls in the very earliest part of this pattern, with the first post-tonic at the peak of the wave; in the two Jutland dialects investigated the stressed syllable occurs relatively later, on the last part of the rising flank. In Thy we should add that the falling slope is extremely steep, to the effect that already with the first post-tonic the trough of the wave is reached, and succeeding post-tonics stay low and level after that. An account along these lines is very reminiscent of the way the tonal word accent differences between Swedish dialects is described, see Bruce and Gårding (1979).

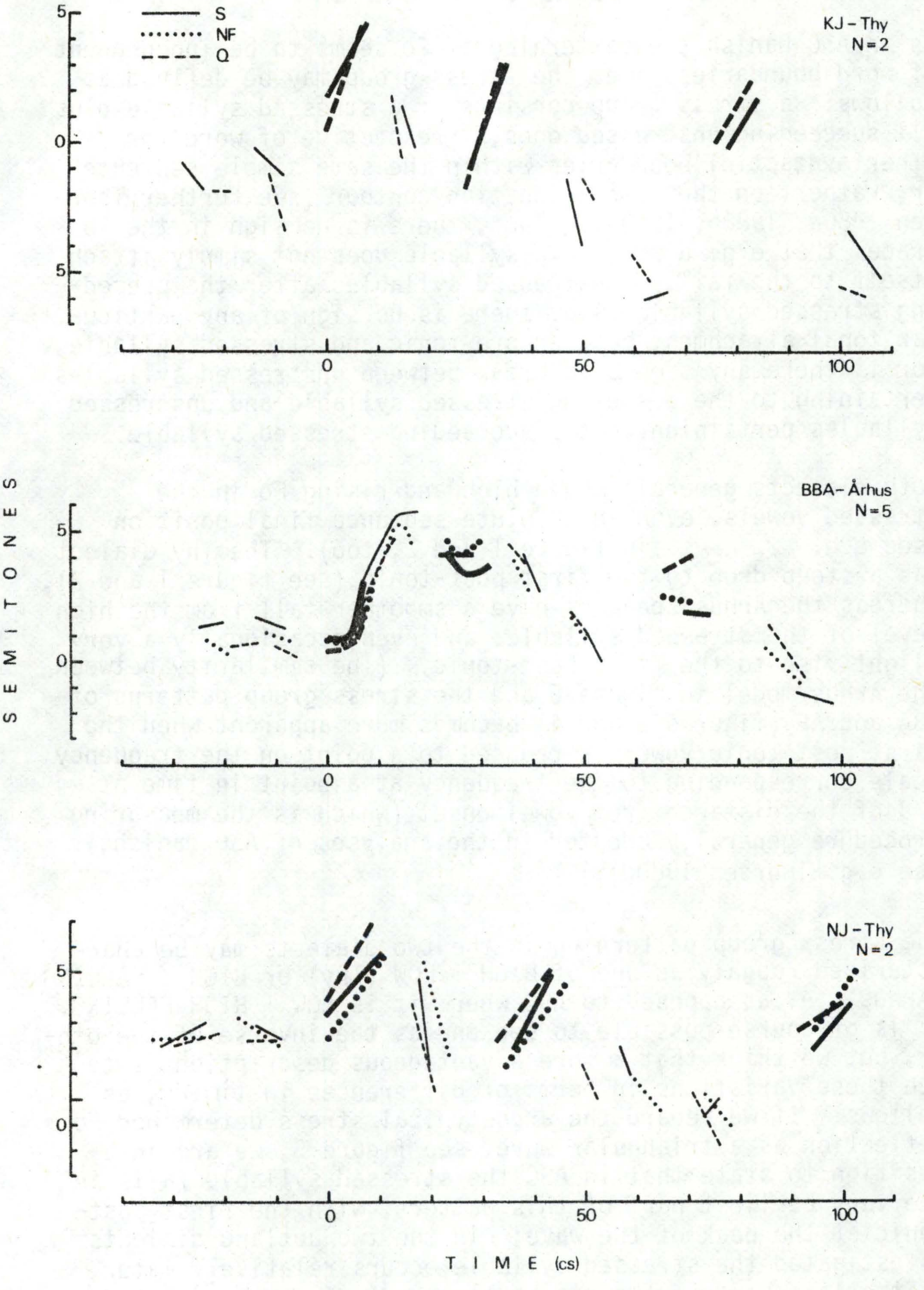


Figure 6

Fundamental frequency tracings of a statement (full lines), a non-final period (dotted lines) and a question (broken lines) by three subjects (means of 2 and 5 readings, respectively). Heavy lines denote stressed vowels, thin lines depict unstressed syllables - see further section II.A. The tracings have been lined up according to the beginning of the first stressed vowel. Zero on the logarithmic frequency scale corresponds to 100 Hz.

Such a description also allows for a very simple account of intrasyllabic Fo movements, which are seen simply as a consequence of the position of the syllable on the Fo pattern, i.e. syllables (or vowels) just "float" on the Fo pattern and do not carry specific movements of their own.

B. INTONATION CONTOURS

Figure 6 depicts Fo tracings (for three subjects only) of a statement, a non-final main clause (which was followed by a coordinate main clause), and a question with word order inversion, cf. section II.A. (Note that there is no non-final sentence with KJ, and that the final post-tonic vowel could not be measured with NJ.)

A declination through the stressed vowels (which defines the intonation contour proper in ASC - a definition which it seems permissible to extend to the Århus and Thy dialects as well) is observed - but we must refrain from making conclusions about the shape of this declination - firstly because the material is so limited, secondly because the data in figures 1 and 2 versus figure 6 is conflicting. (Even with a rough estimate of the correction for differences in intrinsic Fo levels between vowels of different tongue height, the declination would be close to rectilinear with NJ and KJ in figures 1 and 2, but rather asymptotic in figure 6. With BBA, declination would be roughly rectilinear in every instance and thus resemble the declination characteristic of short sentences such as these in ASC Danish.)

BBA's question is the only instance where any clear difference between the three sentence types can be seen. With NJ and KJ all the contours are concurrent, or very nearly so. Now, the question is syntactically marked as such and does not necessarily call for a marked intonation contour (i.e. one which is less falling than in the statement), but the same question was clearly separated from the statement contour with three out of four ASC speakers, cf. Thorsen (1978). The non-final sentence was clearly separated from the statement contour with all four ASC speakers - not so with any of the present subjects. - The material and number of speakers is extremely limited, and the speech situation may not be ideal for the elicitation of intonation contours as they may appear in free speech, but it seems reasonable to state that we should not expect these dialects to exhibit more (neither in number nor in degree) marked intonation contours than does ASC Danish, and future investigations may even prove marked intonation contours to be rarer in Thy and Århus than in Copenhagen.

IV. DISCUSSION

The results should not surprise anyone who is the least bit familiar with Mid and Northern Jutland dialects - which is really why we dare present the results of such a meagre investigation at all: they conform rather well with what is generally assumed about intonation in these parts. But, needless to say, the matter is far from closed, and larger scale investigations should clearly be undertaken.

V. NOTE

1. The vowel qualities vary somewhat with dialects and speakers, see the figures.

REFERENCES

Bruce, G., and Gårding, E. 1979: "A prosodic typology for Swedish dialects", in *Nordic Prosody* (eds.: Gårding, E., Bruce, G. and Bannert, R.), (Department of Linguistics, Lund University), p. 219-228

Jensen, E. 1967: "Om sproget i Århus", *Dialektstudier. 2*, (København), p. 197-270

Nielsen, N.A. 1959: *De jyske dialekter*, (Gyldendal, København)

Thorsen, N. 1978: "An acoustical analysis of Danish intonation", *J.Phonetics 6*, p. 151-175

Thorsen, N. 1979: "Interpreting raw fundamental frequency tracings of Danish", *Phonetica 36*, p. 57-78

Thorsen, N. 1980a: "Neutral stress, emphatic stress, and sentence intonation in Advanced Standard Copenhagen Danish", *Ann.Rep.Inst.Phon., Univ.Cph. 14*, p. 121-205

Thorsen, N. 1980b: "Intonation contours and stress group patterns in declarative sentences of varying length in ASC Danish", *Ann.Rep.Inst.Phon., Univ.Cph. 14*, p. 1-29

Thorsen, N. 1980c: "A study of the perception of sentence intonation - Evidence from Danish", *J.Acoust.Soc.Am. 67*, p. 1014-1030

Thorsen, N. 1980d: "Word boundaries and Fo patterns in Advanced Standard Copenhagen Danish", *Phonetica 37*, p. 121-133.

(ARIPUC 15, 1981 p. 13-47)

INTONATION CONTOURS AND STRESS GROUP PATTERNS IN DECLARATIVE SENTENCES OF VARYING LENGTH IN ASC DANISH - SUPPLEMENTARY DATA

NINA THORSEN

Intonation contours in non-compound declarative sentences containing from one to eight stress groups were analysed with special regard to the manifestation of prosodic boundaries. The results are compared with a previous analysis of a very similar material. The neat and easy match between the prosodic and some of the major syntactic boundaries exhibited by the earlier corpus is not immediately replicated by the present utterances. Resetting of the intonation contour taken rigidly to manifest prosodic phrase group boundaries leads to counterintuitive phrase contours, but on the other hand, the alternative or supplementary boundary criterion suggested in the paper cannot be applied automatically and unambiguously. This strengthens the claim that in syntactically unambiguous non-compound sentences, prosodic structure does not directly reflect or map syntactic structure. However, the four speakers' productions show a high degree of intra- as well as inter subject coherence, and the variation found in the intonation contours with varying sentence length is not random. Nevertheless, the principles governing the intonational properties, especially of the longer utterances, are not easily recovered, and it seems - maybe not surprisingly - that the semantic content of the (constituents of an) utterance must also be considered. The need for further research, also into the perception of prosodic structure, is evident.

I. INTRODUCTION

In the previous volume of ARIPUC (Thorsen 1980a) were published the results of an analysis of eight non-compound declarative sentences, containing from one to eight stress groups, recorded six times by each of four speakers. This is a report of an analysis of an only slightly dissimilar material, recorded by the same speakers, under similar conditions to the 1980 material, and the reader is referred to Thorsen (1980a) for an introduction to the subject with references to the relevant literature, as well as for accounts of recording procedures and measurements. However, I do wish to stress once more the fact that the two materials (forthwith identified as the 1980- and 1981 material) were not originally intended to investigate the relation between prosodic and syntactic boundaries but to test a hypothesis about the equidistant distribution on the frequency scale of the stressed vowels in declarative sentences. It was the refutation of this hypothesis which led to considerations and questions about the syntactic/prosodic interplay. I also wish to point out that temporal relations have not been investigated. Finally: everything that is said in the following pertains exclusively to syntactically unambiguous non-compound sentences.

II. MATERIAL

The material consists of eight non-compound statements, all variations on the same theme:

1. *Til / Tíflis.*
2. *Túkke $ skal til / Tíflis.*
3. *Búster $ skal med / bússen $ til / Tíflis.*
4. *Kísser $ skal med / bússen $ i / nát $ fra / Tíflis.*
5. *Líssi $ skal med / bússen $ klokken / ét = i / nát $ fra / Tíflis.*
6. *Pýtte $ skal med / bússen $ til / Thísted $ klokken / ét = i / nát $ fra / Tíflis.*
7. *Hútters $ skal med / bússen $ fra / kírken = i / Thísted $ klokken / ét = i / nát $ til / Tíflis.*
8. *Knúdsen $ skal med / bússen $ fra / pládsen = ved / kírken = i / Thísted $ klokken / ét = i / nát $ til / Tíflis.*

Sentence no. 8 translates: 'Knudsen is taking a bus from the square by the church in Thisted at one o'clock tonight for Tiflis.' The stressed vowels are indicated with an acute accent. "$" denotes the boundary between major syntactic constituents (noun phrase, verb phrase, complements of place and time); "=" indicates complement internal boundaries; "/" denotes prosodic stress group boundaries (see further page 21). Note that syntactic and prosodic stress group boundaries never coincide. The syntactic boundaries occur after the first post-tonic syllable in the stress group, except that it occurs directly after *ét* and *nát*.

The stressed vowels are all short, but the sequence /ir/ in *kírken* is realized as a diphthong [iɐ̯], which was measured at the midpoint in time of the voiced stretch; they are all high except [e̞] in *ét* and [æ] in *nát*; they are surrounded by unvoiced obstruents except [n] in *nát*, [l] in *Líssi*, and the stressed vowel in *klokken ét* ([-n̩ ˈed̥]); [l] in *plådsen* is devoiced by the preceding aspirated stop.

Sentence no. 1 occurred in a small dialogue, - all the others were naked but were mixed with a material recorded for a different purpose, being evenly distributed over two full pages of recording material, which appeared in three different randomizations, each being read twice on two separate occasions (only once by JR), giving a total of six (three) readings by each subject.

III. RESULTS

Stylized tracings of the eight sentences are depicted in figure 1-4, the grand mean in figure 5. (The last post-tonic vowel could not be measured in JR's sentence 7 and is therefore also lacking in the mean.) As with the 1980 material, the calculation of a mean over all subjects (mean of means) is justified by the rather good qualitative similarity between them. The stressed syllables are connected with full and broken lines, denoting the intonation contour proper. (Note that these lines are "imaginary" - they are not directly present in the course of fundamental frequency (Fo).) Broken lines occur wherever the connection between two stressed vowels is less steep than the preceding as well as succeeding ones; two slopes are considered different if they differ by an (arbitrary) amount of 0.5 semitones/second. (This criterion of slope identity or difference was adopted in order to obliterate slope differences that may occur merely as a consequence of slight differences in the timing of the stressed vowels. 0.5 semitones/second may be too narrow a step, but lacking data on difference limens for the perception of slopes such as these a rather strict criterion is preferable.)

The stressed vowels in the tracings have been corrected for differences in inherent Fo levels, see further Thorsen (1980a p. 6). No correction has been attempted for the unstressed vowels, partly because the difference in inherent Fo level in unstressed vowels and sonorant consonants is smaller than in stressed vowels (Reinholt Petersen 1978), and partly also be-

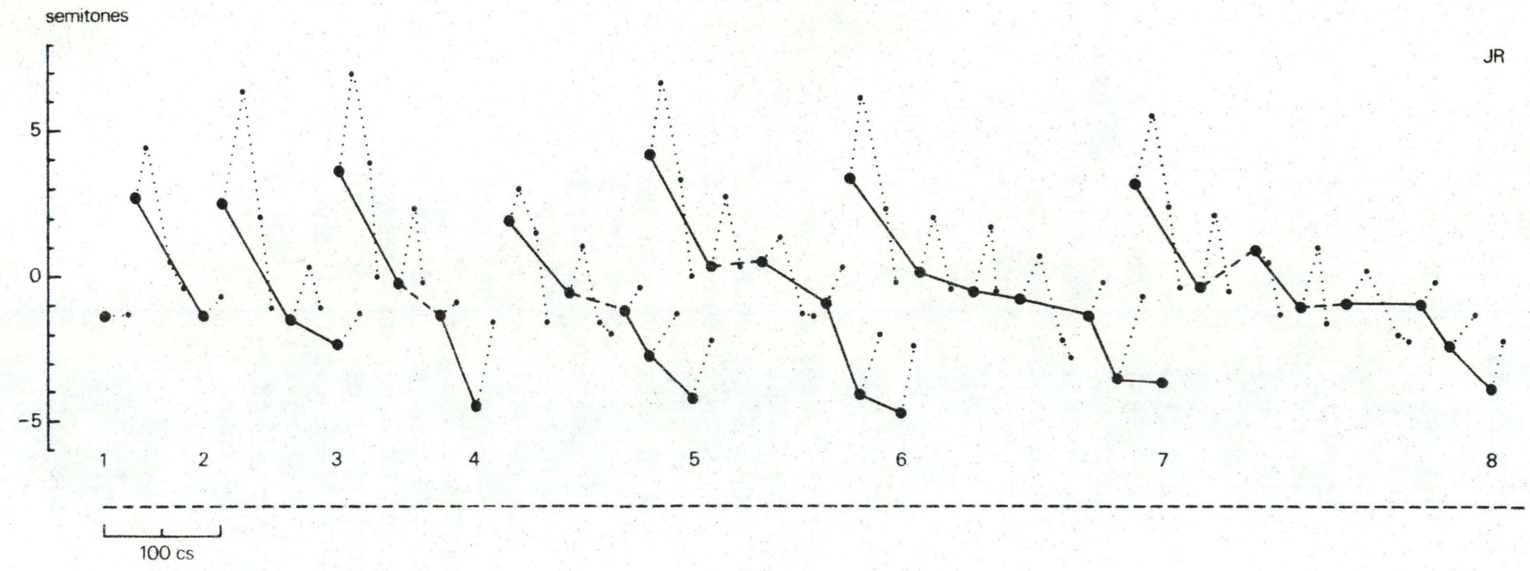


Figure 1

Intonation contours (full and broken lines) and stress group patterns (dotted lines) in declarative sentences containing from one to eight stress groups. Large dots represent stressed syllables, small dots unstressed syllables. Zero on the logarithmic frequency scale corresponds to 100 Hz. Subject: JR.

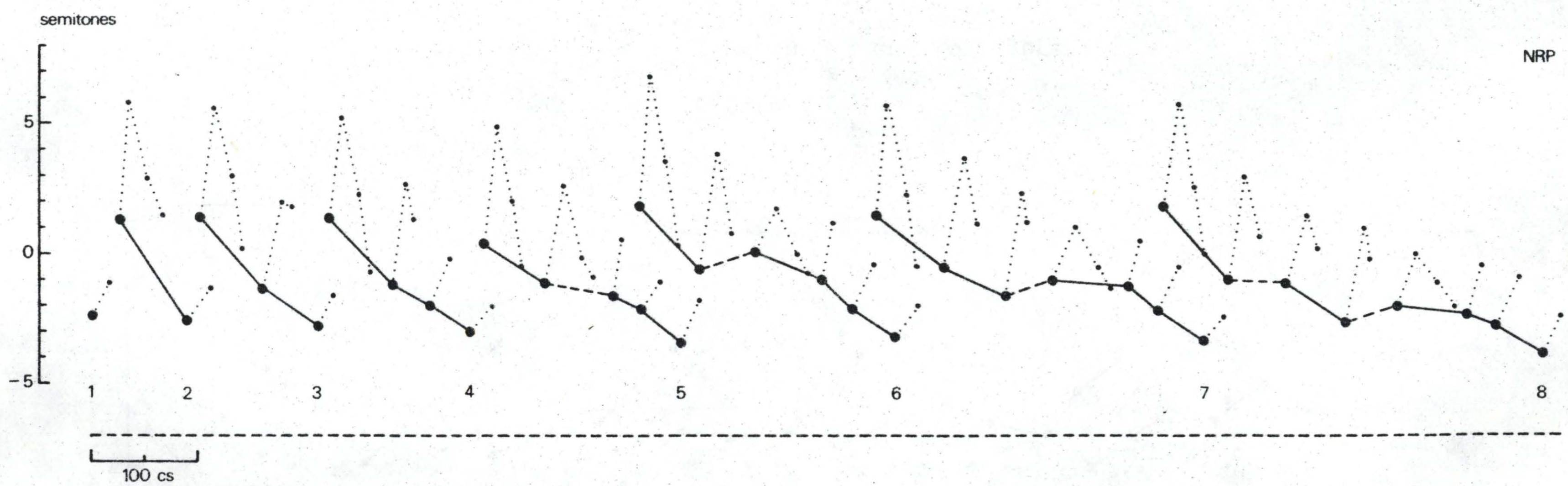


Figure 2

Subject: NRP. See further the legend to figure 1.

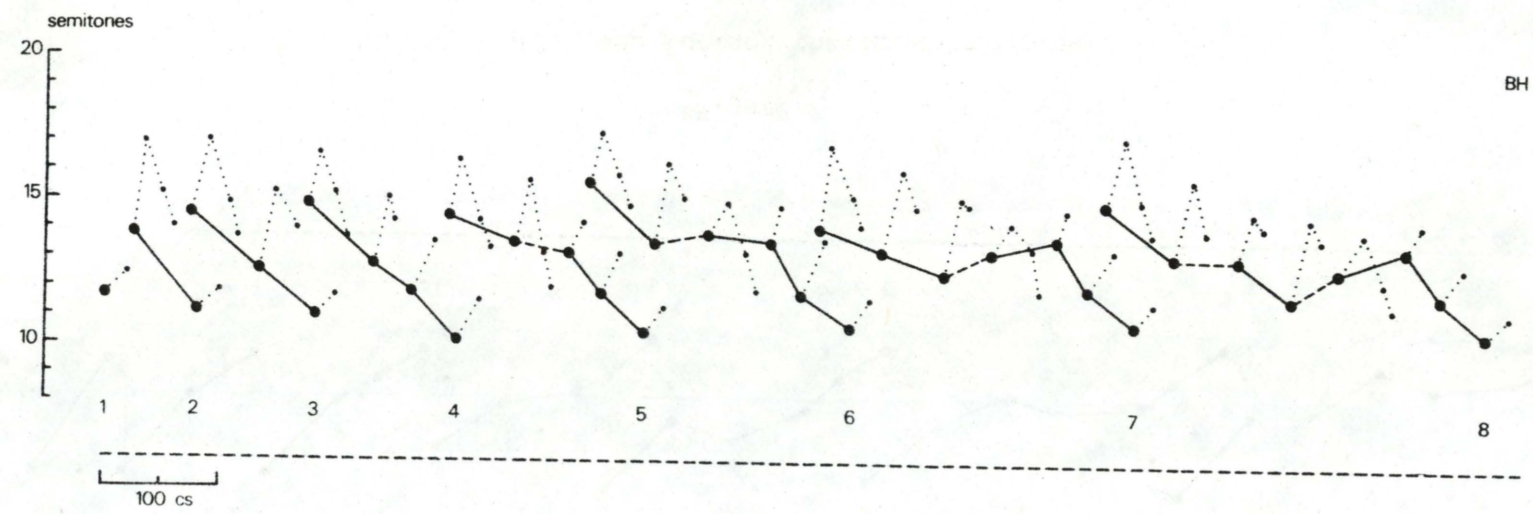


Figure 3

Subject: BH. See further the legend to figure 1.

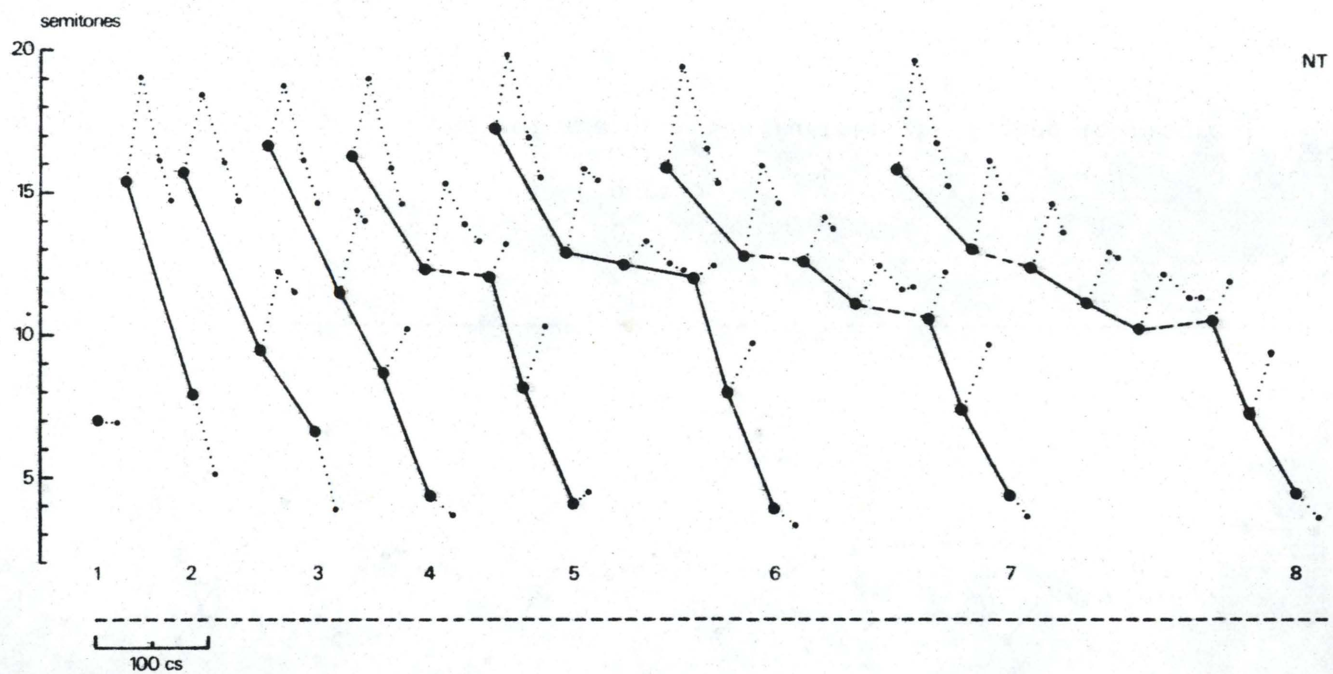


Figure 4

Subject: NT. See further the legend to figure 1.

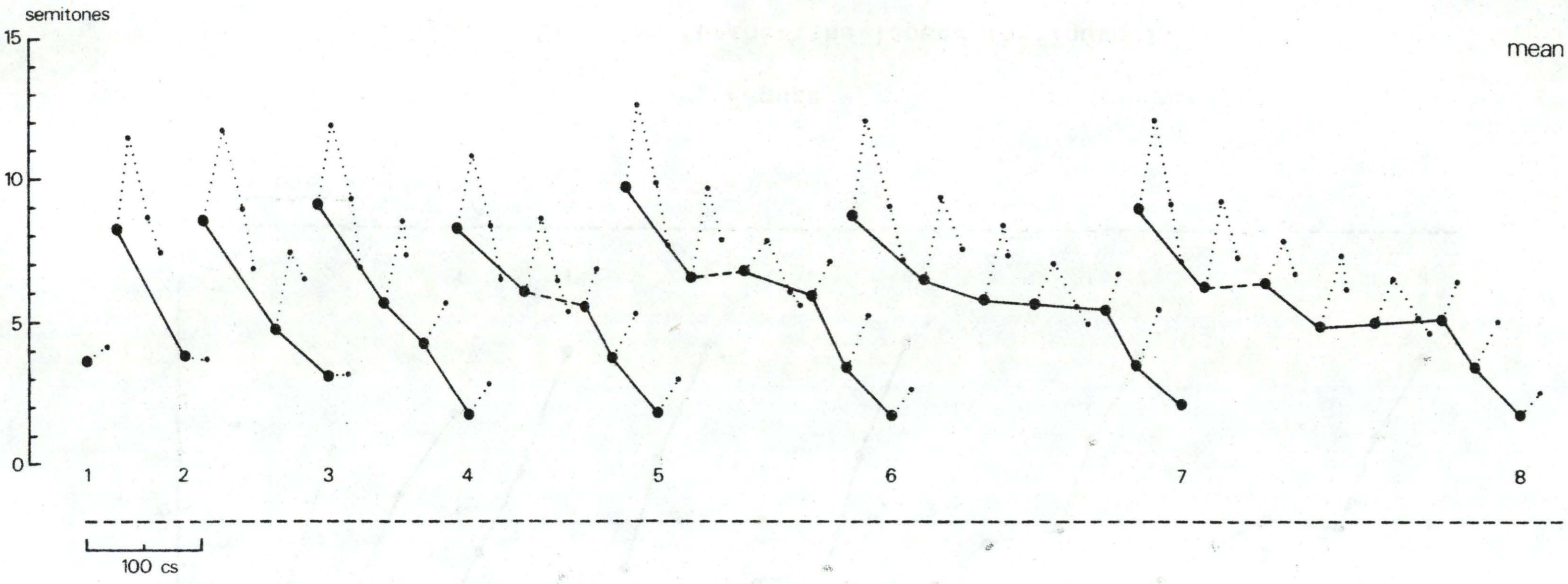


Figure 5

Average over four subjects (mean of means). See further the legend to figure 1.

cause reliable quantifications of these differences are lacking. However, Reinholt Petersen (1980) found that inherent Fo level differences in a stressed vowel are to some extent carried over into the first post-tonic vowel, i.e. a post-tonic vowel will have a higher fundamental frequency after a stressed vowel with a high than after a stressed vowel with a low tongue position, everything else being equal. It might therefore have been appropriate also to raise those post-tonic vowels that occur after non-high stressed vowels, but again the exact magnitude of these corrections is not well established and no such compensation has been attempted.

It is worth noting, as with the 1980 material, that standard deviations on Fo and time measurements are small: in sentence no. 8, for instance, they range (with individual subjects) for the stressed vowels between 4.9% and 1.1% of the mean and for the unstressed vowels between 5.2% and 0.5% of the mean in the Fo measurements. The standard deviations on the total duration of sentence no. 8 range between 3.2% and 1.7% of the mean. The figures must therefore be fairly reliable indications of the subjects' behaviour. - Furthermore, there is no tendency for the first recording(s) of each item to be deviant from the following ones (which is not necessarily excluded by the small standard deviations, if only later recordings are sufficiently undispersed) - i.e. there is no apparent sign of subjects having to go through a learning procedure which then made for more or less automated readings of later repetitions of a given item. I note this expressly because sentences no. 7 and 8 are very long indeed and might have caused this kind of behaviour. The small standard deviations, i.e. the good "production stability" across a subject's six (three) renderings of each utterance, are interesting also in the light of the difficulty I have in recovering the rules that the subjects must have employed for their productions, see further below.

The results of range variation and of variation in starting and end points resemble the 1980 material to a point where they are hardly worth accounting for. The appropriate figures and tables and accompanying comments are given in an appendix.

A. STRESS GROUP PATTERNS

A prosodic stress group has been defined previously as a stressed syllable plus all succeeding unstressed syllables, irrespective of intervening syntactic boundaries within the same intonation contour. This definition still holds. The stress group patterns in figure 1-5 look the same, basically, and no trace of either word boundaries or stronger syntactic

Table 1

Least squares regression line slopes of the stressed vowels and of the first post-tonic vowels in each stress group depicted in figure 1-5 and their correlation coefficients.

	Sentence no.	JR Slope/corr.	NRP Slope/corr.	BH Slope/corr.	NT Slope/corr.	mean Slope/corr.
stressed vowels	2	-7.1	-6.0	-4.8	-12.3	-7.5
	3	-4.9/-.97	-3.8/-.99	-3.3/-1.0	-7.8/-1.0	-5.1/-.99
	4	-6.4/-.99	-3.1/-.98	-3.5/-1.0	-8.4/-1.0	-5.4/-1.0
	5	-3.5/-.98	-1.8/-.97	-2.2/-.95	-5.8/-.95	-3.4/-.97
	6	-3.8/-.96	-1.8/-.93	-1.8/-.91	-4.6/-.93	-3.0/-.94
	7	-2.3/-.95	-1.2/-.89	-0.8/-.74	-3.3/-.94	-1.9/-.94
	8	-1.7/-.90	-1.2/-.89	-0.9/-.75	-2.7/-.94	-1.6/-.92

(table 1 continued)

unstressed vowels	2	-7.8	-10.0	-7.7	-20.3	-11.3
	3	-7.6/-.97	-6.2/-1.0	-4.8/-.97	-11.9/-.99	-7.7/-1.0
	4	-7.0/-.98	-5.1/-.99	-3.5/-.98	-9.7/-.98	-6.4/-1.0
	5	-3.0/-1.0	-3.6/-1.0	-2.8/-.97	-6.9/-.96	-4.1/-.99
	6	-3.8/-.98	-3.2/-.99	-2.2/-.96	-5.6/-.95	-3.7/-.98
	7	-2.5/-.94*	-2.3/-.99	-1.7/-.94	-4.3/-.94	-2.3/-.98*
	8	-2.0/-.93	-1.9/-.97	-1.4/-.94	-3.6/-.94	-2.2/-.96

* The last post-tonic could not be measured with JR and is omitted from the calculations.

boundaries (which occur after the first post-tonic syllable in most instances) can be detected, i.e. the prosodic stress group cuts across the syntactic structure (still: in syntactically unambiguous non-compound sentences).

In the 1980 material a trend was found towards higher rises to the post-tonic vowel in a stress group immediately preceding a discontinuity (a resetting) in the intonation contour. A similar trend is not present here, and I am inclined to believe that I over interpreted the 1980 material on this point. If a consistent trend had been found to higher rises to the post-tonics (or any other systematically deviant stress group pattern) in certain places, then that of course might have been added to the list of criteria for prosodic boundary assignment, cf. below.

B. INTONATION CONTOURS

1. OVERALL DOWNDRIFT

a. Degree of downdrift In table 1 are given least squares regression line slopes for the stressed vowel data points and the first post-tonic vowel data points, respectively, in sentences 2 through 8, cf. figure 1-5. These regression lines may be taken as an expression of the degree of overall downdrift. For ease of reference I shall forthwith call the lines connecting the stressed vowel data points (i.e. the intonation contour proper) and the lines that would connect the first post-tonic vowel in each stress group "baselines" and "toplines", respectively, but note that this is not the way these terms are ordinarily understood, see further Thorsen (1980a p 2-3). As with the 1980 material the "topline" is steeper than the "baseline" slope, and "baseline" and "topline" slopes are highly correlated across the seven utterances (r = 0.93 (JR), 1.00 (NRP), 0.94 (BH), 0.97 (NT)). The Pearson product moment correlation coefficients come out with rather high values, i.e. straight lines are fairly good approximations to the data. Exceptions are the "baselines" of BH's sentence 7 and 8, with correlation coefficients of -.74 and -.75, respectively, but note that the jagged "baseline" is somewhat smoothed out in the "topline" which would be rather better fitted to a straight line (cf. the coefficients of -.94). A similar tendency (though slight - due to the generally high "baseline" coefficients) can be observed with the other subjects, but I do not think the smoother "toplines" warrant considerations about their being in some sense primary and "baselines" secondary, for a number of reasons:

The stressed vowel and the first post-tonic in each stress group are highly correlated across a given sentence, cf. table 2. The correlations have been calculated both from data where no compensation for differences in intrinsic Fo level between stressed vowels of different tongue height is performed as well as from the "compensated" data (as they appear

Table 2

Pearson product moment correlations between the stressed vowel and first post-tonic vowel Fo's in six sentences, containing from three to eight stress groups, i.e. three to eight data pairs. The correlations have been calculated on data where no compensation for differences in intrinsic Fo level between stressed vowels of different tongue height is performed ("uncomp.") as well as on data where such a compensation is carried through ("comp."). The second decimal is rounded off to the nearest integer. In JR's sentence no. 7 the correlation is calculated on the first six vowel pairs only, since the last post-tonic could not be measured.

	JR		NRP		BH		NT	
sentence no.	uncomp.	comp.	uncomp.	comp.	uncomp.	comp.	uncomp.	comp.
3	1.00	1.00	0.99	0.98	0.99	0.97	0.97	0.96
4	0.99	0.95	0.93	0.96	0.91	0.99	0.97	0.99
5	0.97	0.99	0.98	0.96	0.92	0.99	0.98	0.99
6	0.97	0.98	0.92	0.94	0.91	0.95	0.97	0.99
7	0.92	0.96	0.92	0.93	0.88	0.89	0.96	0.98
8	0.97	0.91	0.93	0.96	0.88	0.90	0.95	0.99

in the figures): Stressed and post-tonic vowels do not lead separate lives, but I would argue, as on a number of previous occasions, that the post-tonic vowels are the dependent variables in this relation. Firstly, a stress group obviously has to have a stressed vowel in it, but not necessarily an unstressed one, secondly the stressed vowels in an utterance seem to be stronger perceptual cues to the identification of sentence intonation than the post-tonics (see Thorsen 1980b). Finally, one might argue that if prosodic boundaries are indeed signalled via a non-smooth declination, then the slightly more irregular "baseline" is a better carrier of such information than the "topline", and I shall thus consider "baselines" only in the section on prosodic phrase group boundaries. These arguments do not necessarily deprive the "topline" of any perceptual relevance at all - on the contrary, the smooth "topline" may be seen both as a carrier of information about the degree of overall downdrift, and as a reference which sets off the more irregular "baseline".

A tendency appears in table 2 for the correlations on "compensated" data to be (albeit only slightly) higher than on uncompensated data. Given that the correlations are generally very high (excepting BH's sentence 7 and 8) and should be so if post-tonic vowels are predictable from the stressed ones, then the even better correlations that we obtain from data where intrinsic Fo level differences are compensated for is a point in favour of just this procedure. And one might speculate that if a partial compensation had also been performed in the first post-tonic vowel after non-high stressed vowels (cf. above) then correlation coefficients would have been still closer to unity.

That slope variation is not a linear function of utterance length is seen also in figure 6 and 7. Rather, the slope of the overall downdrift decreases asymptotically with utterance length (whether defined in terms of number of stress groups - figure 6 - or in terms of actual duration - figure 7) and reaches a mean saturation value of about -1.5 semitones/second (stressed vowels) and -2.0 semitones/second (post-tonic vowels), respectively. (In the 1980 material the corresponding values were -2.0 and -2.5 semitones/second.) For a brief discussion of this point, see Thorsen (1980a p. 20-21).

b. Shape of the downdrift In the 1980 material a tendency was found (by visual inspection of the tracings) with some of the subjects in some of the utterances towards greater "baseline" declination in the early part of the utterance, i.e. a tendency towards an asymptotic declination throughout the longer sentences. A similar tendency is not manifest in the present material. On the contrary, with two subjects, BH and NT, the final part of the intonation contour of the longer sentences is decidedly more steeply falling than the beginning. With JR

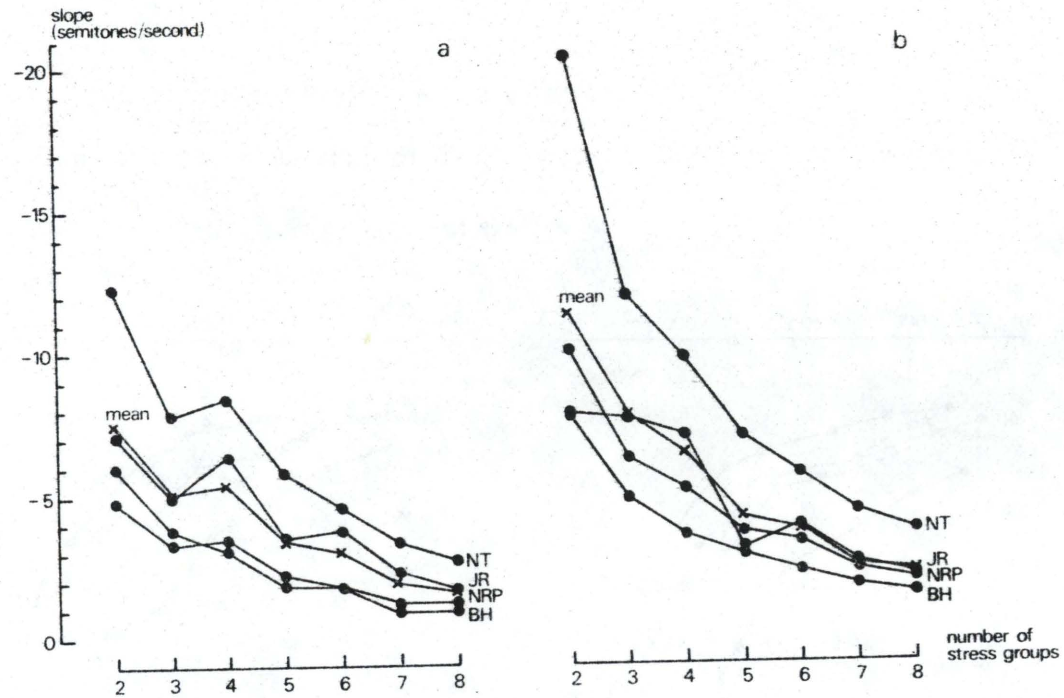


Figure 6

Slope of the overall downdrift of seven utterances, containing from two to eight stress groups, depicted as a function of the length of the utterance (in terms of number of stress groups). Four subjects and their grand mean (crosses). In (a) is depicted the downdrift as determined by the stressed vowels; in (b) downdrift is determined by the first post-tonic vowels. In (b), JR, sentence no. 7, the slope is calculated from the first six vowels only.

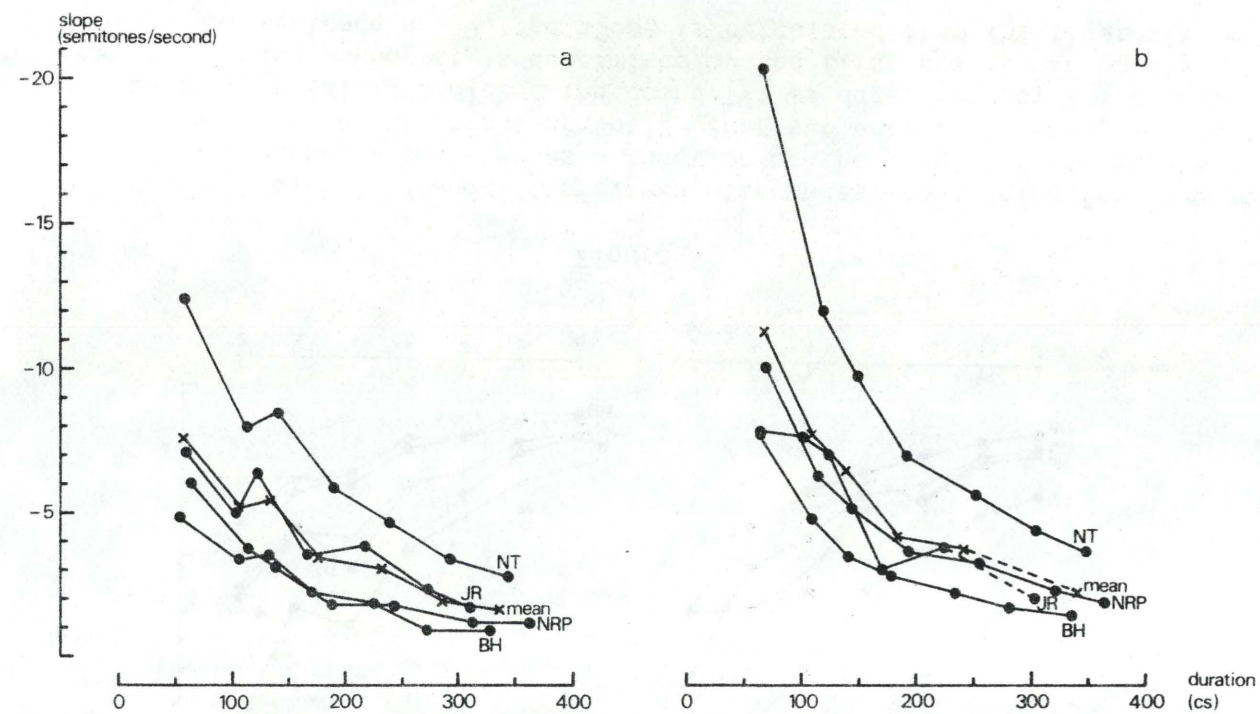


Figure 7

Slope of the overall downdrift of seven utterances depicted as a function of the duration of the utterance. In (b) data on sentence no. 7 is lacking with JR (and the mean). See further the legend to figure 6.

we find a steep beginning, a levelling out, and a steep fall at the end. The subparts of NRP's contours have approximately identical slopes. - The difference in the two materials should probably be ascribed to their different syntactic (and semantic) and prosodic make-up, and I think that this constitutes a good illustration of the precaution needed in the interpretation of the results of isolated analyses.

2. PROSODIC PHRASE GROUP BOUNDARIES

In Thorsen (1980a) a discontinuity in the intonation contour was said to occur at those places in the contour where the slope of the line connecting two stressed vowels is less steep than the preceding as well as succeeding ones (i.e. where a partial "resetting" of the contour takes place), and such discontinuities were taken to be manifestations of boundaries between prosodic phrase groups. The definition seemed a reasonable one, given the shape of the intonation contours and the neat and easy match between the prosodic and some of the major syntactic boundaries thus established.

Although the 1980 and 1981 materials resemble each other rather closely, they do differ in the number and distribution of major syntactic boundaries. Furthermore, the 1980 material had complements of purpose and place, the 1981 material has a complement of time and two complements of place in the longer sentences. The difference between the materials is greatest in sentence 6, 7 and 8 - which are also most evidently different in their intonation contours. The three pairs of sentences are listed here, indicating major syntactic boundaries with "$":

6: (1981) *Pýtte $ skal med bússen $ til Thísted $ klokken ét i nát $ fra Tíflis.*

(1980) *Aníta $ skal med bússen $ til késten for Kísser $ på "Kílden" i Thísted.*

7: (1981) *Hútters $ skal med bússen $ fra kírken i Thísted $ klokken ét i nát $ til Tíflis.*

(1980) *Hútters $ skal med bússen $ til késten for Kísser og Líssi $ på "Kílden" i Thísted.*

8: (1981) *Knúdsen $ skal med bússen $ fra plådsen ved kírken i Thísted $ klokken ét i nát $ til Tíflis.*

(1980) *Knúdsen og Bítten $ skal med bússen $ til késten for Kísser og Líssi $ på "Kílden" i Thísted.*

(Sentence no. 8-1980 translates: 'Knudsen and Bitten are taking a bus to the party for Kisser and Lissi at "Kilden" in Thisted.'.) The major syntactic constituents contain one, two or three stressed syllables (with pertaining unstressed ones) as follows - the actual phrase grouping performed by subjects in the 1980 material is indicated in parentheses:

6:	(1981)	1 + 1 + 1 + 2 + 1	
	(1980)	1 + 1 + 2 + 2	(2+2+2: JR; 2+4: NRP, BH; 6 or 4+2: NT)
7:	(1981)	1 + 1 + 2 + 2 + 1	
	(1980)	1 + 1 + 3 + 2	(2+3+2: all subjects)
8:	(1981)	1 + 1 + 3 + 2 + 1	
	(1980)	2 + 1 + 3 + 2	(3+3+2: all subjects)

This is not precisely the way the contours appeared in Thorsen (1980a figures 1-5) due to the less tolerant criterion of slope identity adopted there, but it corresponds to the tracings in the lower part of figures 8-11 here. With a slightly wider step than 0.5 semitones/second the second discontinuity in JR's sentence 6-1980 would disappear and his grouping would be 2+4 as well.

On the basis of the grouping performed by subjects in 1980 I would venture the following hypotheses, although the risk of making too-far-reaching conclusions is considerable: (a) Four stressed syllables in one prosodic phrase group seems to be the maximum (cf. also the fact that sentence no. 5 was indeed divided prosodically into two groups). NT's exceptionally large range (about 15 semitones) will accommodate even larger prosodic phrases, but it is also possible to postulate a 4+2 grouping in no. 6-1980. (b) A syntactic constituent having only one stressed syllable in it ties up prosodically with a neighbouring constituent. (This point is one that may be disproven by a different material. If the neighbouring constituent(s) is (are) already maximally long (with four stressed syllables in it (them)) then a prosodic phrase with only one stressed syllable may be envisaged, unless a neighbouring constituent is to be cut up internally by a prosodic boundary. However, I do not think that a final syntactic constituent with only one stressed syllable in it will appear as an independent prosodic phrase and thus be preceded by a discontinuity, because such a final "rise" probably would violate the inherent feature of terminal declarative sentence intonation.) With these restrictions the groupings in sentence 6-1980 are all predictable, and the grouping in sentence 7-1980 is the only one possible; in sentence 8-1980 a 2+4+2 grouping is also possible, unless we add a further constraint that (c) prosodic phrase groups be of as nearly as possible equal size - then 3+3+2 is a better candidate than 2+4+2. - The 3+3+2 grouping in sentence 8-1980 might also be due to (d) a tendency for the boundary before the complement (at the second "+") to be stronger than the NP+VP boundary (at the first "+") - and this may be true of this particular utterance but is hardly a general phenomenon (cf. the discussion about the role that semantic content may have, in section 4. below).

If these restrictions on prosodic phrase grouping are carried over to the 1981 sentences, we get the following possibilities:

6-1981: 2 + 4 (hypothesis (d)) or 3 + 3 (hypothesis (c))

7-1981: 2 + 2 + 3 (hypothesis (d)) or 4 + 3

8-1981: 2 + 3 + 3 (the only one possible)

All of the above reasoning naturally rests upon the assumption that a major syntactic constituent will not be cut up internally (at least if it has no more than four stressed syllables in it), or in other words: a prosodic boundary will not occur independently of a major syntactic one (whereas a major syntactic boundary need not be accompanied by a prosodic one).

In figures 8-12 the two sets of intonation contours are shown (omitting the unstressed syllables) with syntactic boundaries indicated, and in the leftmost ("a") edition (if there is more than one) with broken lines at the discontinuities as defined above. Before proceeding any further, I wish to point out two things: Firstly, the "irregularity" of the contours cannot be due to faulty correction for intrinsic Fo level differences between vowels (cf. also A.1.b above). In figure 12 I have indicated in square brackets the lowest vowels in the sentences, and if one looks across the subjects at these particular stretches and compare them with the rest of the contours, it is clear that the raised non-high vowels cannot be made responsible for the breaks and turns in the contours. Secondly, it is also evident that a subject's Fo range will influence the demand for and degree of "resetting" of the contour, compare BH (figure 10) to NT (figure 11), for instance.

The prosodic boundary definition and assignment which worked so well with the 1980 sentences is less satisfactory with the present material, compare the actual "a" groupings to the predictions:

	predicted	JR	NRP	BH	NT
6-1981:	2+4 or 3+3	2+4	2+4	2+4	6
7-1981:	2+2+3 or 4+3	7	3+4	3+4	2+2+3
8-1981:	2+3+3	2+2+4	2+2+4	2+2+4	2+3+3

In sentence 6, three subjects fit one of the predictions, in sentence 7 and 8 this goes for only one subject. That could of course just be a hint that the assumptions on which the predictions are based are false - in other words prosodic boundaries can occur independently of major syntactic ones. - However, some of the phrase contours that arise in the figures ("a" editions) as a result of the "resetting" criterion applied seem distinctly counter-intuitive: (1) No boundary is assigned at the third "$" in JR's and BH's sentence no. 7 and 8, though the contour takes a sharp turn at that place. (2) The final phrase contour is rather sharply rising-falling in BH's sentence 7 and 8. (3) No boundary is assigned at the third "$" in NT's sentence 6, although the resemblance to the final phrase contour in sentence 7 and 8 is striking.

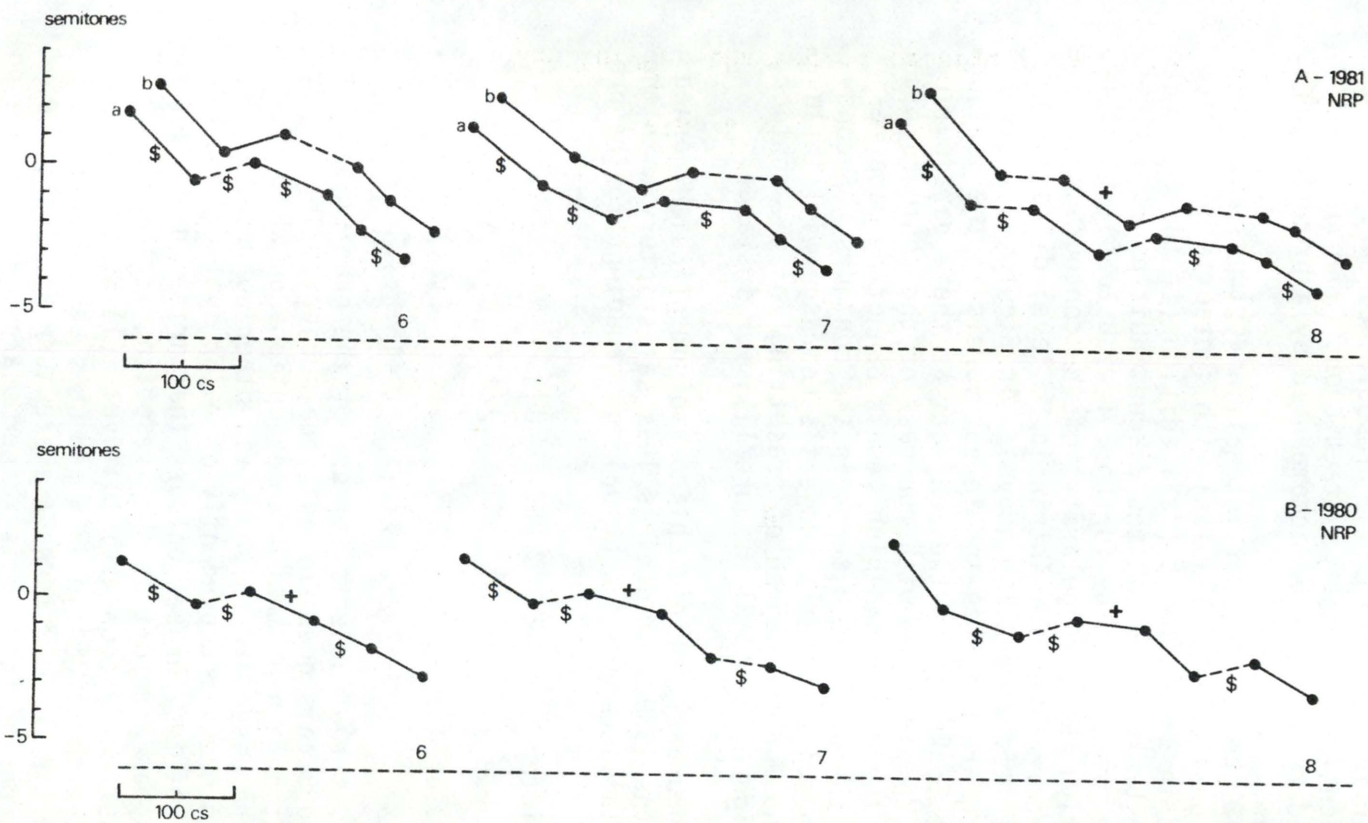


Figure 8

Intonation contours in declarative sentences containing from six to eight stress groups from two different materials: A-1981 at the top and B-1980 at the bottom. The unstressed syllables have been left out. The "a" and "b" editions are identical except for the placement of prosodic phrase group boundaries (indicated with broken lines). The "b" edition has been moved 0.5 semitones upwards and 25 cs to the right with respect to the "a" edition. Subject: NRP.

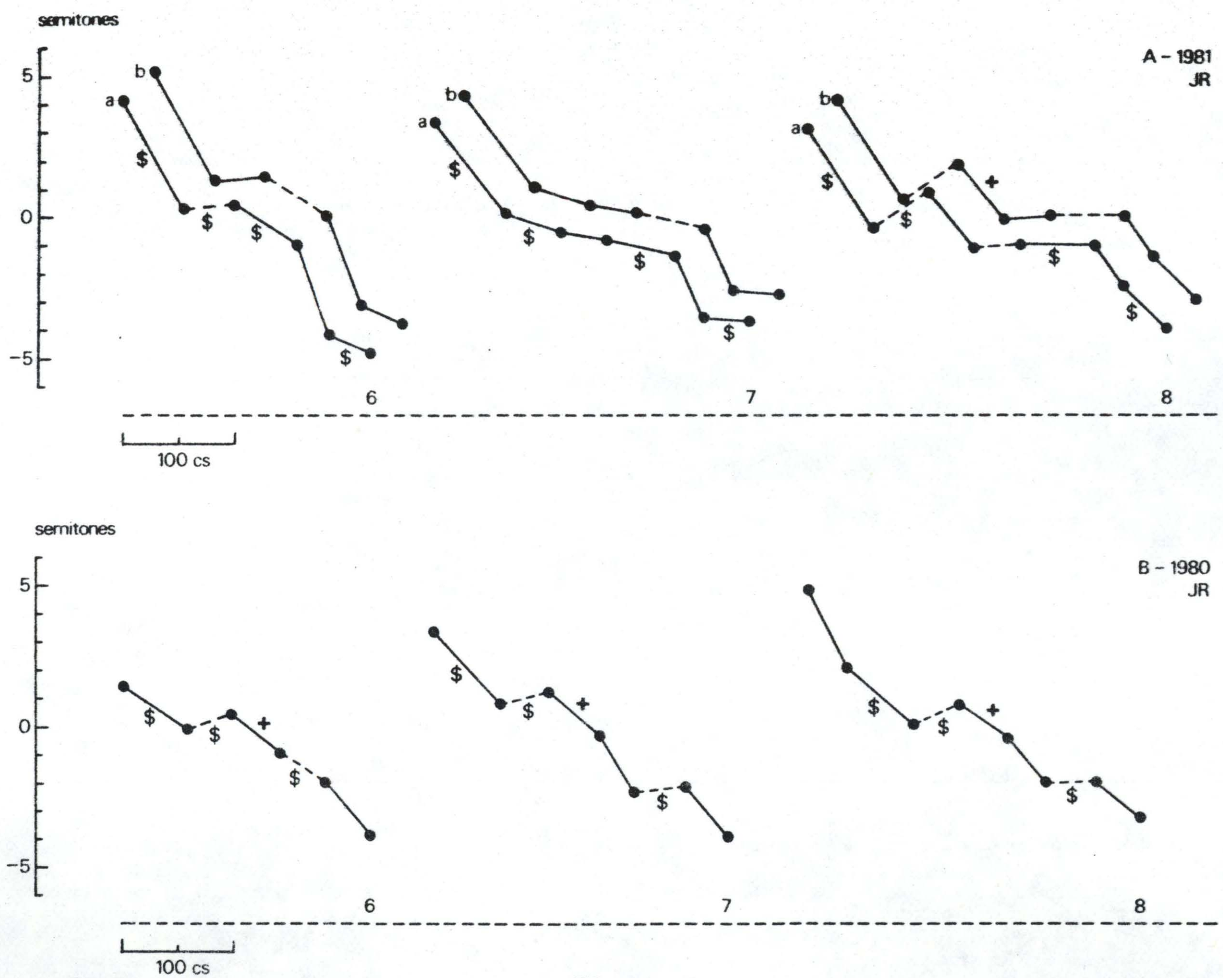


Figure 9

Subject: JR. See further the legend to figure 8.

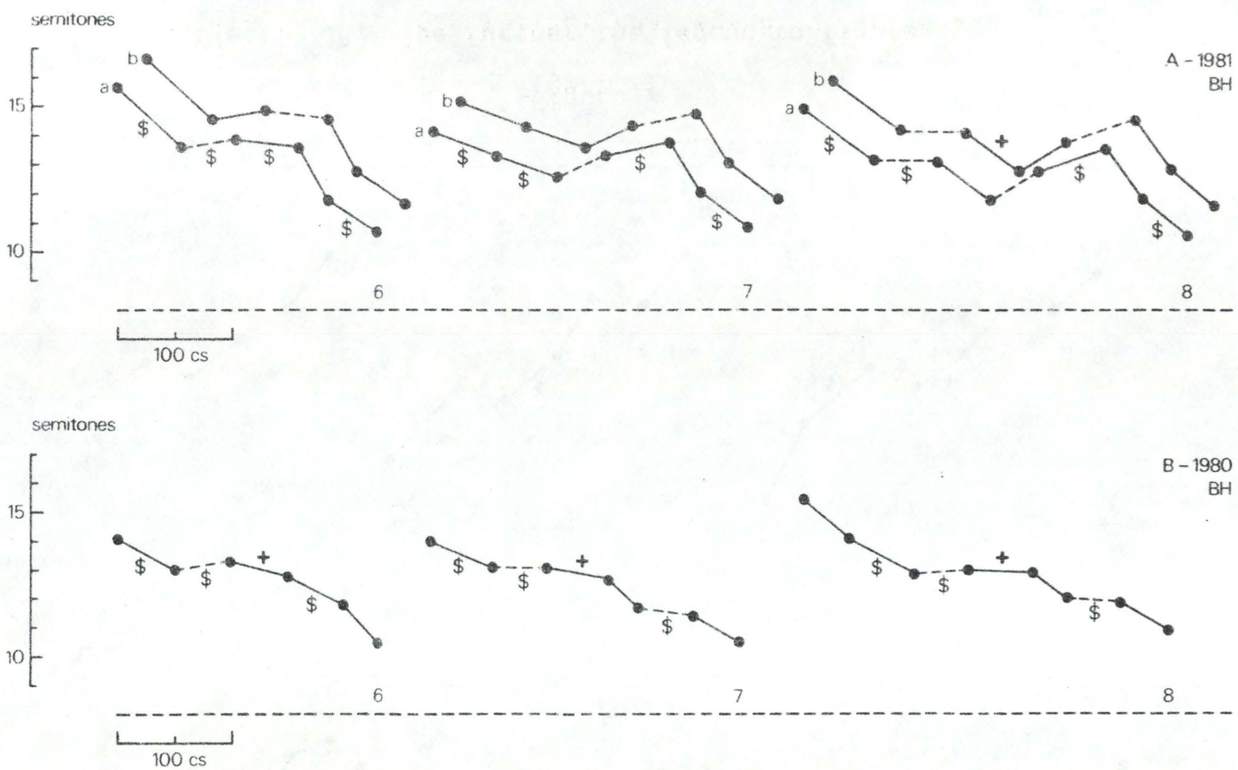


Figure 10

Subject: BH. See further the legend to figure 8.

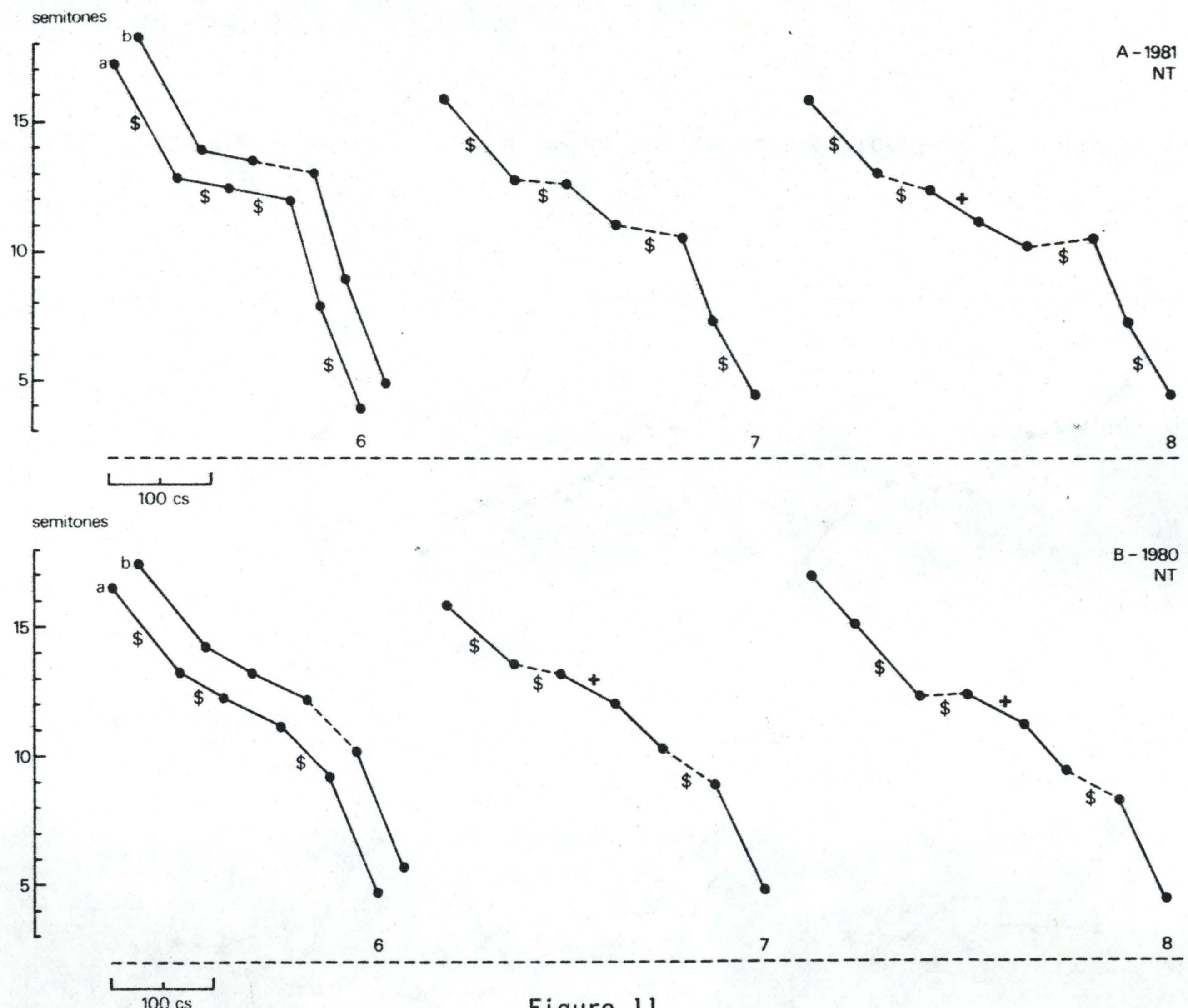


Figure 11

Subject: NT. See further the legend to figure 8.

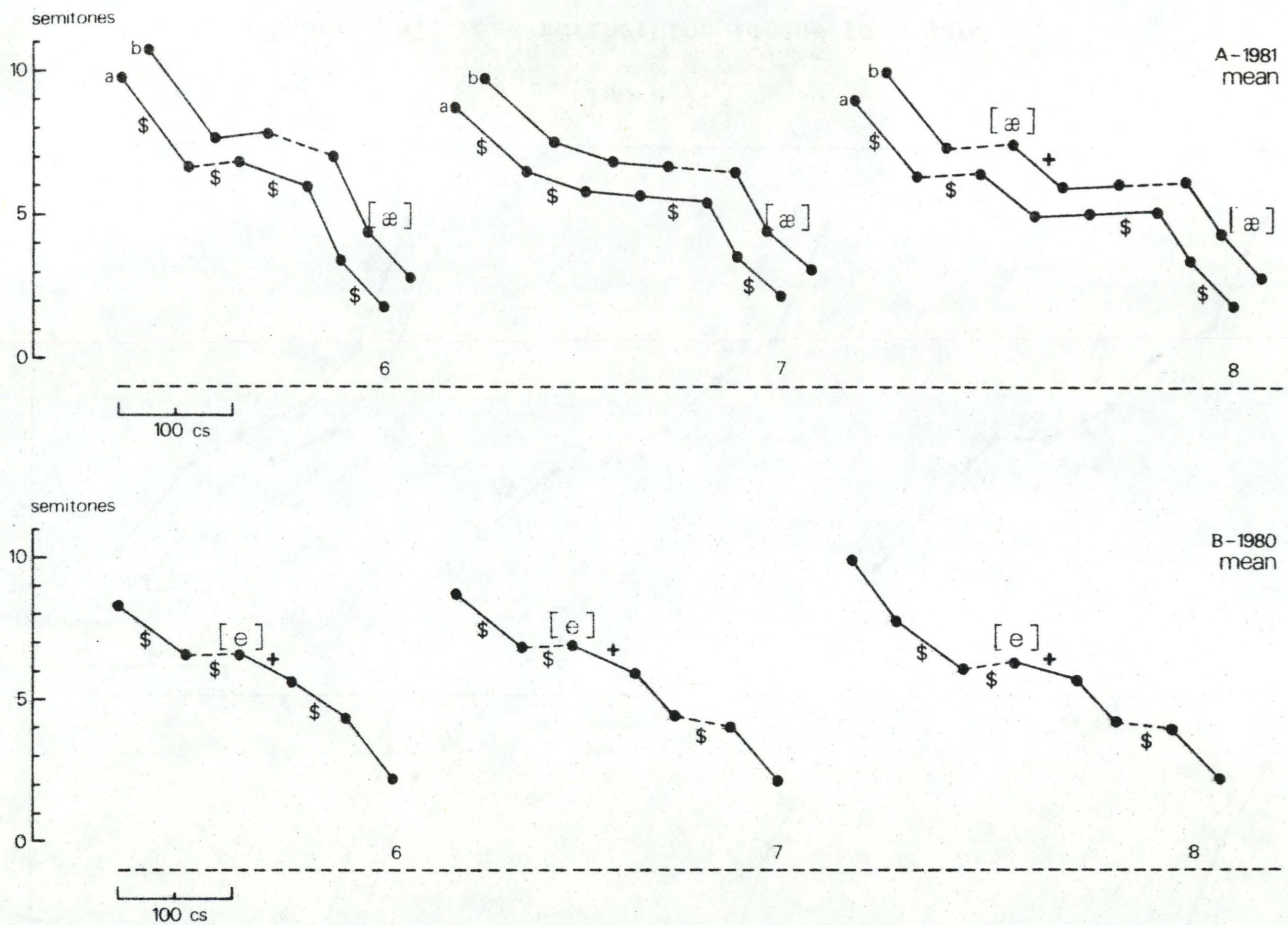


Figure 12

Average over four subjects (mean of means). See further the legend to figure 8.

In the "b" editions of the contours I have suggested an alternative boundary location: Before the antepenultimate stressed syllable of sentence 7 and 8 with JR and BH the contour takes a rather sharp turn, and a boundary is inserted. Since the corresponding final part of sentence 6 bears a striking resemblance to those of 7 and 8, the boundary is also moved (JR and BH) or inserted (NT) here. A similar boundary location can be defended for NRP as well, though his contours are on the whole less jagged-looking. In this way a certain coherence is established across subjects and sentences. Note that the 1980 criterion is not at crosscurrents with the present boundary assignment where the 1980 sentences are concerned, except that a boundary is suggested in NT's sentence 6 at the last "$". - Now the predictions and the "actual" groupings correspond very well indeed:

	predicted	JR	NRP	BH	NT
6-1981:	2+4 or 3+3	3+3	3+3	3+3	3+3
7-1981:	2+2+3 or 4+3	4+3	4+3	4+3	2+2+3
8-1981:	2+3+3	2+3+3	2+3+3	2+3+3	2+3+3

Unfortunately, it is no longer possible to assign prosodic boundaries automatically and unambiguously from the intonation contours alone: with the boundaries suggested in the "b" editions we get phrase contours that describe asymptotic declinations or even rises at the end (see e.g. BH), but when phrase final rises are admitted (and I do not see why, a priori, they should not be) then a number of alternative boundary locations present themselves (which only violate the hypothesis that a prosodic boundary cannot occur unaccompanied by a major syntactic one - a hypothesis that would therefore have to be discarded) as indicated by the crosses in the figures. A further difficulty would arise with an "isolated" sentence, i.e. one that could not be compared to any others in order that a 'coherence across sentences' criterion might be applied.

3. PROSODIC BOUNDARY PERCEPTION

At this point the need to know something about prosodic boundary perception becomes pressing, because if the discontinuities and resettings of these intonation contours have no perceptual significance per se, i.e. as prosodic boundaries properly speaking, then the search for definitions of them and criteria for retrieving them from the intonation contours (without regard to syntactic structure) becomes futile.

A number of studies are concerned with the prosody/syntax relation: Uyeno et al. (1980) find that Fo will disambiguate syntactically ambiguous sentence structures in Japanese (clause structures); Streeter (1978) reports that Fo will disambiguate ambiguous algebraic expressions in American English, whereas duration is a less stable cue; Lehiste et al. (1976) tell us that duration will effectively disambiguate syntactically ambiguous utterances (American English), whereas Lehiste

(1980) mentions an unpublished experiment by Olive and Lehiste where Fo was unsuccessful in disambiguating those same utterances. Although the evidence is slightly conflicting, I will assume - for the sake of the point I want to make below - that Fo can indeed disambiguate otherwise ambiguous utterances. - Cooper and Sorensen (1977) analyse Fo at syntactic (clause) boundaries and find significant differences as a function of syntactic coding but are not concerned with the perception of these differences. - Umeda et al. (1975) and Harris et al. (1981) investigate listeners' perception of (syntactic) boundaries in fluent speech, and Umeda and Quinn (1981) follow up the perceptual experiments with an analysis of word duration, which is shown to be positively correlated with the listeners' perception of boundaries; they also note that some listeners seem to be more sensitive to elongation than others, some are more sensitive to pitch contours than others, and some to a strong initial allophone, but they make no claim that prosody actually cues syntactic boundary perception in general.

Where disambiguating otherwise ambiguous utterances is concerned I think it reasonable to assume that we are actually dealing with perception of prosodic boundaries, which then lead to an interpretation of syntactic structure. But in syntactically unambiguous, non-compound sentences such as the ones under investigation here, I do not know and rather doubt whether prosodic boundaries will be perceived independently of (and thus be able to cue the perception of) syntactic boundaries. Given the complete speech signal a native speaker's prosodic boundary perception may just be a rationalization of his linguistic (and semantic) interpretation of the utterance. - I have listened to the recordings a number of times and tried to listen exclusively for prosodic boundaries (which may be an illusory attempt): I can really detect no boundaries at all with NRP and NT. With BH I can detect a boundary at the third "$" in most instances. With JR I can hear a boundary in sentence 8 at the third "$" and occasionally also at the second "$" in sentence 8. This would be a (weak) support for the boundary assignment in the "b" editions. However, with repeated listening I can induce myself to hear prosodic boundaries practically all over the utterances, and in an erratic fashion. - Evidently, reliable perceptual results should be obtained from low-pass filtered speech or synthetically produced signals.

If it turned out from perception tests with stimuli produced on the basis of intonation contours such as these, retrieved from syntactically unambiguous non-compound sentences, that listeners would not identify and locate consistently prosodic boundaries then it would hardly be appropriate to speak of them as b o u n d a r i e s any longer. It would also mean that we should content ourselves with a description of the intonation contours that arise as a result of syntactic structure (and semantic content, cf. below) - but we would still have to state that certain syntactic boundaries leave no trace in the intonation contour, i.e. certain syntactic constituents will be tied together in one prosodic phrase - and we would also have to describe the conditions under which syntactic boundaries give rise to turns and breaks in the intonation contour and the possible phrase contour shapes that arise as a result. - Clearly, both of the materials investigated are too limited to deal satisfactorily with the last problem; further investigations are called for.

4. THE ROLE OF SEMANTICS?

Let us for the moment assume that the assignment of "boundaries" in the "b" editions in figure 8-12 is an adequate representation of the division into subparts, prosodic phrases, of the intonation contours and let me try and explain the one most striking difference between the intonation contours of the two materials. In 1981 there are a number of instances where the phrase contour rises before the final phrase contour (viz. JR 6 and 8 (only slightly), NRP and BH 6, 7 and 8). If we had had only BH's sentence 7 and 8 to judge from, a likely explanation could have been found in the fact that the final phrase contour must perform a rather steep fall, and since it contains three stress groups, it will have to start rather high up in the frequency scale, so the rise in the preceding contour could be an anticipation of this high start of the final contour. However, NRP and JR contradict this hypothesis. In fact, there is nothing in the intonation contours per se which can explain the phrase contour rises.

The final phrase group in all three sentences (6, 7 and 8) consists in the 1981 material of two complements, one of time followed by one of place (... *klokken ét i nát $ til Tíflis.*). The corresponding final phrase group in the 1980 material was one complement of place (... *på "Kílden" i Thísted.*). I do not think that two versus one complement, or a three-stress versus a two-stress phrase group have anything to do with the difference in intonation contour. I would venture a more semantically oriented explanation: The time complement is less intimately related to the preceding part of the utterance, it is more of a unit apart; that is, the boundary before it is stronger (at least in these particular utterances) and to signal this we get what might appropriately be termed "continuation rises" in the preceding contour. Note that the third "$" would also have been a very likely place for a pause to occur (not that it did). With this analysis we can distinguish two kinds of "rises" in an intonation contour: a phrase final continuation rise, which is a "local" deflection, and a resetting of the intonation contour which separates two phrases (see e.g. JR, sentence 8) and whose purpose is to keep the intonation contour within the speaker's Fo range.

It is tempting to illustrate and support the role of semantics in prosodic structuring by the "boundary" placement in sentence 6 in the two materials (JR, NRP and BH), even though the intonation contours as a whole do not look widely different. In *Pýtte $ skal med bússen $ til Thísted /$/ klokken ét i nát $ fra Tíflis.* (1981) the break ("//") occurs after the first complement (and is preceded by a continuation rise) but in *Aníta $ skal med bússen /$/ til fésten for Kísser $ på "Kílden" i Thísted.* (1980) it occurs before the first complement. In other words, the prosodic phrasing is not governed by surface syntactic structure exclusively, the semantic content of the syntactic constituents is also taken into account, in this case inducing a "boundary" before the "heavier" time complement in the 1981 utterance, which makes the first place complement tie up with the preceding NP and VP. This argument is invalid, however, if *bússen til Thísted* is one syntactic constituent, rather than two (equivalent to *Thísted-bussen* 'the Thisted-bus'). In that case the explanation for the boundary placement in the 1981 utterance is not to be sought in a heavy succeeding time complement but in a weak boundary between *bússen* and *til Thísted.* Note, though, that this syntactic closeness is not accompanied by a stress reduction on the first element which is otherwise characteristic of close-knit syntactic relations in Standard Danish (like *køre bíl* 'to drive a car' and many others).

III. CONCLUSION

In Thorsen (1980a) I concluded that the results presented an argument in favour of a theory put forward in Selkirk (1980) that prosodic categories are distinct entities in the phonology that do not have an isomorphous relation to syntactic structure. This claim is certainly not weakened by the material analysed here. In syntactically unambiguous non-compound sentences the prosodic stress group will cut across any syntactic boundary, and when - in longer utterances - a division of the intonation contour into prosodic phrase groups is necessitated, this phrasing bears no simple relation to surface syntactic structure. Furthermore, I suggest that the matter is rendered even more complicated by the role that semantics may have to play in prosodic structuring.

How and to what extent the results would be applicable to free speech I cannot say. One might speculate that prosody plays a more important role in the production and perception of free speech - which is rarely so syntactically well-formed as the schematized material presented for reading in this investigation. That is: prosodic boundaries may be more evident (also when unaccompanied by pauses) in free speech and may of course take more and different shapes than encountered here.

ACKNOWLEDGEMENT

My sincere thanks are due to Jeanette Holtse for her efficient graphics work.

REFERENCES

Cooper, W.E. and Sorensen, J.M. 1977: "Fundamental frequency contours at syntactic boundaries", *J.Acoust.Soc.Am. 62*, p. 683-692

Harris, M.O., Umeda, N. and Bourne, J. 1981: "Boundary perception in fluent speech", *J.Phonetics 9*, p. 1-18

Lehiste, I. 1980: "Phonetic manifestation of syntactic structure in English", *Ann.Bull.Res.Inst.Log.Phon. 14*, p. 1-27

Lehiste, I., Olive, J.P. and Streeter, L.A. 1976: "Role of duration in disambiguating syntactically ambiguous sentences", *J.Acoust.Soc.Am. 60*, p. 1199-1202

Selkirk, E.O. 1980: "On prosodic structure and its relation to syntactic structure", *Indiana University Linguistics Club*, (Bloomington, Indiana)

Streeter, L.A. 1978: "Acoustic determinants of phrase boundary perception", *J.Acoust.Soc.Am. 64*, p. 1582-1592

Thorsen, N. 1980a: "Intonation contours and stress group patterns in declarative sentences of varying length in ASC Danish", *Ann.Rep.Inst.Phon., Univ.Cph. 14*, p. 1-29

Thorsen, N. 1980b: "A study of the perception of sentence intonation - Evidence from Danish", *J.Acoust.Soc.Am. 67*, p. 1014-1030

Umeda, N., Harris, M. and Forrest, K. 1975: "The placement of auditory boundaries in fluent speech", *J.Phonetics 3*, p. 191-196

Umeda, N. and Quinn, A.M.S. 1981: "Word duration as an acoustic measure of boundary perception", *J.Phonetics 9*, p. 19-28.

Uyeno, T., Hayashibe, H., Imai, K., Imagawa, H. and Kiritani, S. 1980: "Comprehension of relative clause construction and pitch contours in Japanese", *Ann.Bull.Res.Inst. Log.Phon. 14*, p. 225-236.

APPENDIX

1. RANGE

There is an overall tendency for range to increase with increased utterance length, cf. figure 13, but range does not increase monotonically, no matter whether range is determined by the interval between first and last stressed vowel (13a), first and last post-tonic vowel (13b), or absolute Fo maximum (first post-tonic vowel) and absolute Fo minimum (last stressed vowel, except with NT where the minimum is constituted by the last post-tonic vowel in sentences 1-4 and 6-8) (13c).

The range dispersion on the grand mean in figure 13a (stressed vowel interval) is 3.6 semitones (corresponding to 80% of the smallest range), lying between 5.9 (NT) and 1.8 semitones (NRP); in figure 13b (post-tonic vowel interval) it is 2.2 semitones (corresponding to 29% of the smallest range), lying between 3.9 (JR) and 0.9 semitones (BH), and in figure 13c (interval between absolute Fo maximum and minimum) it is 2.6 semitones (corresponding to 31% of the smallest value), lying between 5.7 (JR) and 0.9 semitones (BH).

2. STARTING AND END POINTS

Figure 14 depicts the level of starting points (14a: first stressed vowel, 14b: first post-tonic vowel) and end points (14c: last stressed vowel, 14d: last post-tonic vowel). If we disregard sentence no. 1, which obviously groups itself with the end points, there is only a slight and irregular tendency towards higher starting points with the longer utterances, cf. the slopes of the least squares regression lines on the data points of figure 14a-d and their correlation coefficients in table 3. End points decrease more, at least through sentence 1 to 4 (and with most subjects they also decrease more regularly than starting points increase, the correlation coefficients generally being numerically greater on the end point regressions). Only BH shows a deviant pattern: in the present material starting and end points increase and decrease, respectively, to approximately the same extent, and in the 1980 material (table 3B) starting points increase more than end points decrease. With JR in the 1980 material first and last post-tonic in- and decreased equally. - The near-constancy of the end points from sentence no. 5 and higher was also observed in 1980 and can probably be ascribed to a physiological constraint: the speaker has a lower limit to his Fo range, which he is bound to hit with utterances exceeding a certain length (in terms of number of stress groups).

There are individual differences in the various curves in figure 13 and 14, and no clear-cut pattern in the association between range variation and starting versus end point variation can be found (i.e. even though end points decrease more than starting points increase):

a. Stressed vowel range versus starting and end points — The correlation between stressed vowel range and first stressed vowel level in this material is as follows: JR 0.78, NRP 0.81, BH 0.95, NT 0.85 as opposed to a correlation between range and last stressed vowel level of: JR -0.92, NRP -0.67, BH -0.79, NT -0.97. Thus, the correlation is numerically stronger between range and starting point with NRP and BH. In the 1980 sentences the same calculations yield for range and starting point correlations: JR 0.79. NRP 0.49, BH 0.75, NT 0.86; range and end points: JR -0.67, NRP -0.80, BH -0.35, NT -0.96, i.e. JR and BH have stronger numerical correlations between range and starting points.

b. Post-tonic vowel range versus starting and end points — The correlations between the post-tonic vowel range and level of the first post-tonic in this material are: JR 0.92, NRP 0.85, BH 0.85, NT 0.78; between range and last post-tonic level: JR -0.43, NRP -0.62, BH -0.19, NT -0.88. In the 1980 material the same correlations yield: JR 0.88, NRP 0.62, BH 0.97, NT 0.60 and JR -0.85, NRP -0.56, BH 0.12, NT -0.96. Thus, in both materials range and first post-tonic correlate numerically more strongly than range and last post-tonic with JR, NRP and BH, vice versa with NT.

Concludingly we can say - as for the 1980 material - that fundamental frequency range is not constant over utterances of different length, neither is it a linear function of utterance length. The range variation is brought about by a combination of variation in starting and end points, and at least with some subjects end points lower more than starting points increase, until "saturation" is reached (with utterances of four to five and more stress groups).

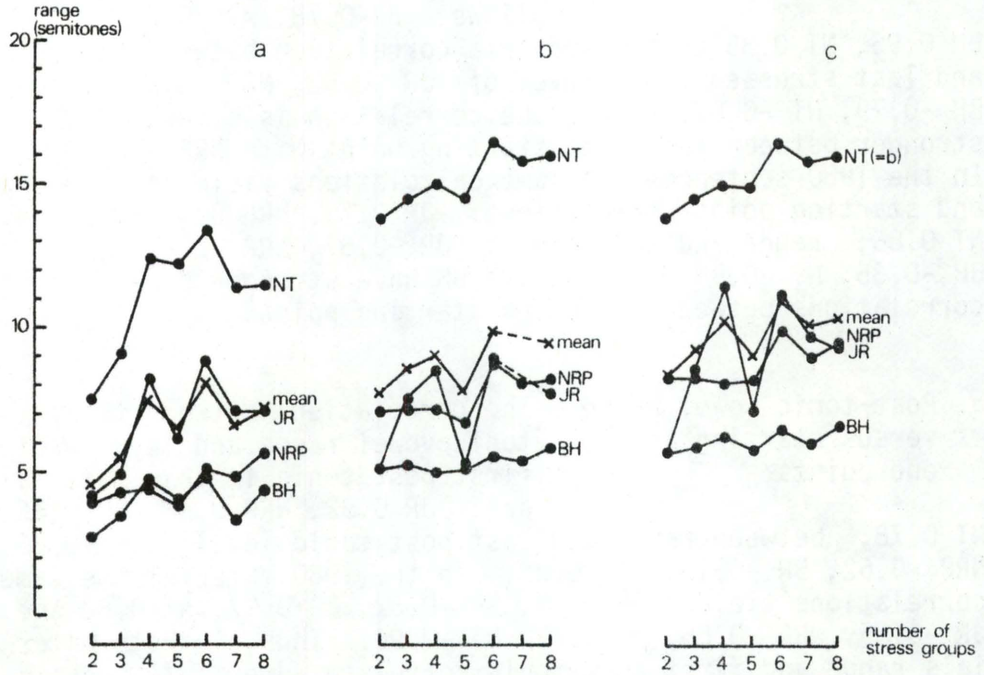


Figure 13

Range of fundamental frequency in seven declarative utterances, containing from two to eight stress groups, depicted as a function of the length of the utterance (in terms of number of stress groups). Four subjects and their grand mean (crosses). In (a) range is defined as the interval between first and last stressed vowel measurement in each utterance; in (b) range is defined as the interval between the first post-tonic vowel in the first and last stress group in each utterance; in (c) range is defined as the interval between the absolute Fo maximum (i.e. the first post-tonic vowel in the first stress group) and the absolute Fo minimum (i.e. the last stressed vowel) in each utterance. (NT's Fo minimum is constituted by the last post-tonic vowel and the curve in (c) is identical to the one in (b). In (b) data on sentence no. 7 is lacking with JR (and the mean).

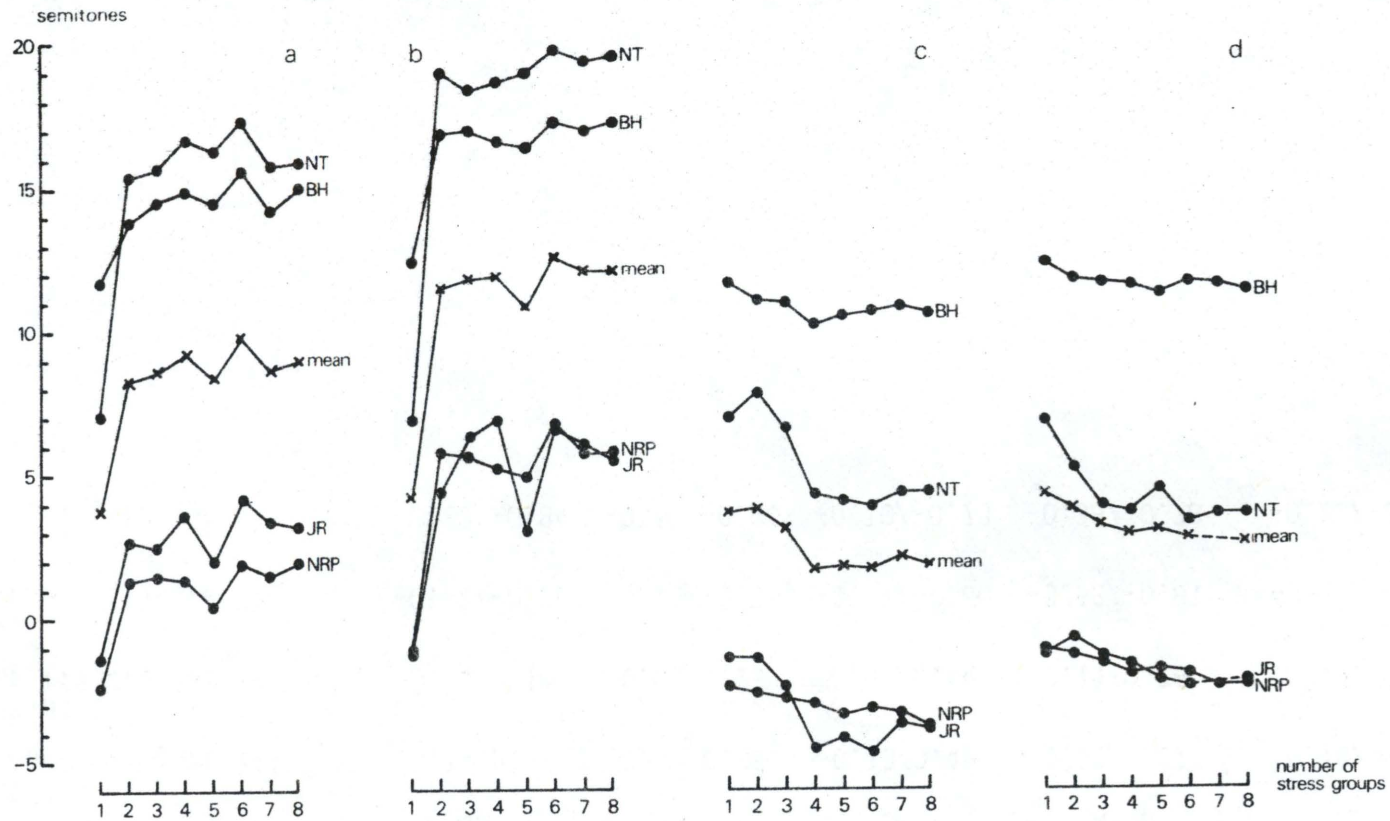


Figure 14

Fundamental frequency starting and end points in eight declarative utterances, containing from one to eight stress groups, depicted as a function of the length of the utterance (in terms of number of stress groups). Four subjects and their grand mean (crosses). In (a) is depicted the frequency of the first stressed vowel in each utterance; in (b) the frequency of the first post-tonic vowel in each utterance. In (c) is depicted the frequency of the last stressed vowel, and in (d) the last post-tonic vowel in each utterance. In (d) data on sentence no. 7 is lacking with JR (and the mean). Zero on the logarithmic frequency scale corresponds to 100 Hz.

Table 3

A: Least squares regression line slopes/Pearson product moment correlation coefficients on the data points in figure 14a-d. Sentence no. 1 is excluded in the calculations on the data in figure 14a and 14b.

	JR	NRP	BH	NT	mean
Starting points, stressed vowels (figure 14a)	0.14/0.18	0.08/0.35	0.13/0.49	0.08/0.27	0.10/0.12
Starting points, post-tonics (figure 14b)	0.09/0.14	0.06/0.23	0.07/0.44	0.11/0.76	0.11/0.42
End points, stressed vowels (figure 14c)	-0.42/-0.76	-0.18/-0.95	-0.11/-0.60	-0.52/-0.81	-0.31/-0.82
End points, post-tonics (figure 14d)	-0.22/-0.84	-0.19/-0.96	-0.10/-0.73	-0.38/-0.78	-0.23/-0.90

(table 3 continued)

B: Same as A on the 1980 material (Thorsen 1980a, figure 8, p. 15)

	JR	NRP	BH	NT	mean
Starting points, stressed vowels (figure 8a)	0.14/0.26	-0.04/-0.21	0.18/0.64	0.11/0.53	0.10/0.42
Starting points, post-tonics (figure 8b)	0.22/0.69	0.05/0.21	0.20/0.65	0.17/0.88	0.15/0.94
End points, stressed vowels (figure 8c)	-0.39/-0.86	-0.23/-0.91	-0.05/-0.31	-0.41/-0.83	-0.27/-0.84
End points, post-tonics (figure 8d)	-0.23/-0.82	-0.20/-0.90	0.02/0.29	-0.52/-0.78	-0.23/-0.82

(ARIPUC 15, 1981, p. 49-60)

THE PRESENT STATE OF THE DESCRIPTION OF DANISH UTTERANCE PROSODY*

NINA THORSEN

I will comment first on *Danish, utterance* and *prosody*: Danish pronunciation varies markedly, depending on which geographical area, which age group, and which social stratum you choose to deal with. Differences due to age and social status may be most pronounced within Standard Danish. - If an utterance is anything that is spoken between two pauses, there are obviously many different kinds of utterances, from single words to long sentences. Sentences may be simple or complex and may be further classified by their status as, e.g., declarative or interrogative. Another choice, which cuts across the choice of utterance type, concerns context. You can choose to investigate free, natural speech or speech in some restricted situation or other: news reading on Radio or Television, a lecture, etc., or you can be compelled to make do with readings of schematized context free utterances. - Prosody is the generic term for several phenomena: primarily rhythm and intonation, but we can also classify such reductions and assimilations as occur in connected speech under *prosody*.

Let me now disclose what I have investigated:

> intonation in (almost exclusively) non-compound periods (with extensive use of nonsense words) read - out of context - by 30 to 40 years old speakers of Standard Copenhagen Danish from the middle classes.

---

* Public lecture with a set title, held on September 23rd, 1981, as the final examination for the philosophical licentiate degree.

Yet another restriction of the material lies in the fact that I have looked at neutral - non-emotional - speech only. Subjects were asked specifically not to give their reading any particular emotional colouring. - It is therefore a very limited part of all that which may come under the heading *Danish utterance prosody* that I have investigated. Add to this other sins of omission: you could expect from an exhaustive description of Danish utterance prosody that it will adopt a historical perspective - to the extent that this is possible - and you would also expect Danish to be compared to our most closely related neighbours - to the extent that that is possible (Swedish is the only Nordic language which is presently adequately described, really); finally, the description could be set off by more general theories of intonation.

The following phenomena have been studied:

- the relationship between stress and fundamental frequency
- intonation in various sentence types
- intonation in utterances with emphasis for contrast
- intonation in utterances of very different length
- the perception of sentence intonation
- the tonal manifestation of words with assimilated or elided schwa
- the relationship between stress and fundamental frequency in two Jutland areas (with Bent Jul Nielsen from the Institute of Danish Dialectology)

The methodology is classic in experimental phonetics:

- systematized material
- several subjects
- several recordings of each subject
- registration of fundamental frequency via pitch meters on mingograms
- segmenting, measuring and averaging - which lead to a
- descriptive model

The complex course of fundamental frequency (Fo) is assumed to be the outcome of a superposition of several components - an assumption which is supported by the results and which is also implicitly or explicitly present in some descriptions of other languages:

the sentence contributes an intonation contour

-	phrase group	- -	phrase contour
-	stress group	- -	stress group pattern
-	stød	- -	stød movement
-	segments	- -	microprosodic modification

To cover more languages, you may have to add to this list a tone or word tone component, a sentence accent, and a terminal contour.

I will go into detail only with the stress group component in the following, because I have greater confidence in the generalizability of the stress group patterns, also to other speech materials and situations, than of any of the other phenomena investigated. - Hans Basbøll has defined a syntactic stress group as a group of words with one main stress, which is on the last word in the group (with certain exceptions - personal pronouns are unstressed, also in stress group final position). The same definition is implicit in Poul Andersen's description. You can say, very roughly, that the syntactic stress group ends with the stress. The prosodic stress group, on the contrary, begins with the stress: it consists of a stressed syllable plus all succeeding unstressed syllables (if there are any), irrespective of intervening word or other syntactic boundaries within the same intonation contour. - The two groupings, syntactic (indicated with slants) and prosodic (indicated with plusses) and the very different parsing of a sentence they lead to can be illustrated thus ('He lay down on the chaiselongue and lit a Caminante' - I have indicated stressed vowels with acute accents):

Han lágde sig / på chaiselóngen / og tǽndte / en Caminánte.

Han + lágde sig på chaise+lóngen og + tǽndte en Cami+nánte.

The interesting thing about the prosodic stress group is not so much the fact that it begins with the stress (in contradistinction to the syntactic stress group), but that it cuts across syntactic boundaries. In other words: at some point in the speech production process the utterance is restructured - some (quite a number, in fact) syntactic boundaries are "deleted". The fundamental frequency pattern shown

(figure 1)

may accordingly depict the course of Fo in any of the heavy passages in the three sentences ('Handball-playing is very tiresome. The bananas in the box are rotten. Allergy is an infamous disease.' - The large dot is the stressed syllable, the small dots are unstressed syllables.) The definition of the prosodic stress group as a stressed syllable plus all succeeding unstressed syllables within the same intonation contour

is based on the observation that this unit is the carrier of a recurrent and rather constant Fo pattern - namely, a low stressed syllable followed by a high-falling tail of unstressed syllables.

The prosodic stress group pattern is subject to variation, according to

position in the utterance

the intonation contour it is superposed upon

emphasis for contrast

syllable reduction

individual

dialect

I shall return to this variation, but first let me show how stress groups combine into whole utterances

(figure 2)

('Mr. Andersen is taking a bus for Thisted.') Since the stress group pattern is - in a sense - constant, i.e. its variation (which mainly concerns the interval between the stressed and the first post-tonic syllable) is predictable, the sentence intonation contour may be defined narrowly as the course described by the stressed syllables alone, disregarding the unstressed syllables. This concept of sentence intonation contour is depicted by the broken lines in the figure

(figure 3)

This is not an illustration of any particular set of utterances. The broken lines which indicate sentence intonation contours have of course no physical reality. The intonation contour is most steeply falling in declarative, terminal utterances (3), least falling in questions that are neither syntactically nor lexically marked as such (1). Between these two extremes you find (2) other question types and non-terminal sentences, with a tendency towards a trade-off relationship between syntax and intonation: the more syntactic or lexical information the sentence contains about its non-declarative, non-terminal status, the more steeply falling, i.e. the more declarative terminal, is its intonation contour, and vice versa. This phenomenon is subject to a certain amount of individual variation, however.

Two points are worth noting about sentence intonation contours in Standard Danish: Firstly, in the type of speech material and -situation I have studied, an intonation contour is never more rising than horizontal, i.e. rising intonation contours, properly speaking, do not occur. Secondly, sentence intonation is a global phenomenon, i.e. information about sentence status is distributed all over the sentence and is not a local feature, e.g. in the shape of a certain Fo movement situated at the end. This analysis has been confirmed by some perceptual tests, i.e. listeners can indeed also utilize the information contained in earlier parts of the utterance in their identification of sentence intonation and -status.

The more or less rectilinear intonation contours in the figure (figure 3) are characteristic only of utterances which are not too long - I shall return to longer utterances later.

STRESS GROUP PATTERN VARIATION

Position and intonation contour dependent variation is apparent from the graph (figure 3): The rise to the first post-tonic syllable is greater early than late in the utterance, which gives the "topline" and "bottomline" of the Fo course together a characteristic wedge shape, - a feature which is also found in other languages. You find the same shrinking of the Fo patterns on less steeply falling contours, but it is not as extensive. In other words: the rise from stressed to post-tonic syllable is greater on less falling intonation contours, everything else being equal. - The shrinking of Fo patterns may have physiological causes, - it may be a voluntary signal of termination, - or it may be a mixture of both - at any rate, it is predictable.

Emphasis for contrast will change the Fo course in the utterance rather drastically

(figure 4)

If you compare this graph to the neutral edition of the same sentence (figure 2) you will note that the three stress group patterns in the neutral case reduce to one, i.e. the rise-fall characteristic of the stress group pattern is reduced or completely eliminated in the stress groups that surround the one which contains the stressed syllable of the emphasized word, and we are left with one low+high-falling pattern. The clumsy formulation is due to the fact that word and other boundaries are still immaterial for the course of Fo. The prominence comes on with the stressed syllable of the emphasized word, and any pre-tonic syllables will again tie up prosodically with the preceding (reduced) stress group. Emphasis may be accompanied by a raising of the stressed syllable in question, but this raising does not in itself appear to be perceptually decisive, and it seems that the tonal reduction of surrounding stress group patterns - which is equivalent to a reduction from main to secondary stress - is the salient feature of emphasis for contrast in Standard Danish, - something which has also been pointed out by Steffen Heger. Therefore, the phenomenon which has been termed *focus, Satzakzent, primary accent, nucleus,* etc. cannot be said to be equivalent to emphasis for contrast in Danish, because *focus* etc. in languages that have such a phenomenon is not accompanied by a similar reduction of surrounding stresses. Danish lacks obligatory sentence accent - and languages that have such a sentence accent of course have emphasis for contrast as well.

On previous occasions I have said that when a first post-tonic schwa is dropped or assimilated to a sonorant neighbouring consonant, it seems that this syllable may also be tonally assimilated (at least partially) to the preceding stressed syllable, to the effect that the Fo maximum in the stress group is reached only in the second post-tonic syllable.

(figure 5)

However, this phenomenon - partial tonal assimilation of a post-tonic syllable with elided or assimilated schwa - is far from being a general one. I am in the process of analysing a rather comprehensive material, recorded by four subjects; - so far I have performed a quantitative two-way description of the stress group patterns, i.e. I have ascertained whether the stress group peak is situated in the first or the second post-tonic syllable after the stressed one. It turns out that across all subjects and all words (with assimilated as well as unassimilated schwa) the stress group Fo maximum lies in the first post-tonic syllable in about 75% of the material. Subjects differ among themselves, however, - with one of them the peak almost invariably lies in the first post-tonic, and with another this is so in a large majority of the words. The remaining two subjects have more instances of late Fo peaks, and it does indeed appear, that the tendency with them is greater towards late Fo maxima in words with assimilated schwa than in words with a vocoid in the first post-tonic syllable. But to say that syllables with assimilated schwa generally behave differently from other unstressed syllables is hardly justified. - A closer qualitative description may still, however, reveal interesting differences.

The conditions in schwa syllables are the only tonal phenomena which have previously been subjected to acoustic (and auditory) analyses. Jørgen Rischel determined how syllable number is distinguished in words of the type *hårde-hårdere* 'hard (pl.)-harder' [ˈhɒːʌ-ˈhɒːʌʌ] and *faldne-faldende* 'fallen (pl.)-falling' [ˈfalnə-ˈfallnə]. He found that, apart from some not quite stable durational differences in the consonants, the peak in the rising-falling tonal pattern is much earlier, relative to the end of the word, in tri- than in dissyllabic words. It is tempting to reformulate this and say that the peak is situated in the first post-tonic syllable; if there is only one post-tonic, the peak will lie very near the end of the word; if there are several post-tonics the peak must be succeeded by a fall. In other words, in Rischel's material a first post-tonic schwa syllable, where schwa is assimilated, did not seem to assimilate tonally to the preceding stressed syllable, at least not to the extent that the Fo peak is "delayed" until the second post-tonic syllable.

The fall through the post-tonics may vary somewhat from one speaker to another

(figure 6)

some have rather steeply falling post-tonics (left) to the effect that the intonation contour will often be transgressed by unstressed syllables (figure 3), but others have only slight falls (right), to the effect that the "bottomline", i. e. a connection of local Fo minima, will be identical to the intonation contour.

In a pilot study I did with Bent Jul Nielsen from the Institute of Danish Dialectology of the relation between stress and Fo in two Jutland areas, Thy and Århus, it turns out that here, too, one can reasonably define a prosodic stress group as a succession of a stressed and all following unstressed syllables within the same intonation contour - but the stress group patterns look somewhat different from Standard Copenhagen

(figure 7)

If you consider the stress determined tonal pattern to be a hat or a triangle or, better, a wave, the difference between these stress group patterns can be accounted for in terms of a difference in the timing of the syllables with respect to the peak and troughs of this wave. In Standard Copenhagen the stressed syllable hits the trough and the rise to the first post-tonic at the peak is fairly steep. Succeeding post-tonics float on a (more or less) falling flank. In Århus, the stressed syllable hits the last part of the rise to the peak, and the unstressed syllables lie on the falling flank. The difference between Århus and Thy is a difference in the steepness of the falling flank - in Thy it is so steep that the first post-tonic will already lie in the next trough. - If this description - in terms of differences in timing with respect to basically the same pattern - holds for other dialects as well, it will be an interesting parallel to the way the two word tones are realised in different Swedish dialects - which is basically a question of how a sequence of low and high syllables are timed relative to the beginning of the word or the stressed syllable in the word. - Another advantage of this description of stress group patterns is that vowel Fo movements need not be separately accounted for. A particular vowel movement is just a consequence of the position of the vowel on the Fo pattern - segments "float" on the Fo pattern - as hinted in the figure by the heavy strokes (stressed syllables).

INTONATION CONTOURS IN LONG UTTERANCES

One might imagine that upper and lower limits for sentence intonation contours are constant, regardless of the length of the utterance, so intonation contours in long and short utterances will differ only in the steepness of their slopes, which will be inversely proportional to the length of the utterances they span

(figure 8)

This is, however, an over simplification. In utterances containing four to five, or more, stress groups the intonation contour is decomposed into smaller phrase contours

(figure 9)

(as indicated by the broken lines in the figure), but an overall fall is preserved. - The question of the location of phrase boundaries relative to the syntax of the utterance is complicated. Generally, one can say that the phrase boundaries will be located near syntactic boundaries, but far from all and every syntactic boundary is thus marked, and the prosodic stress group will still cut right across any syntactic boundary within the simple sentence. Furthermore, surface syntactic structure alone cannot predict the location of phrase boundaries: the semantic content of the phrases must be taken into account. This last point is one which (together with several others) calls for much more comprehensive investigations, however.

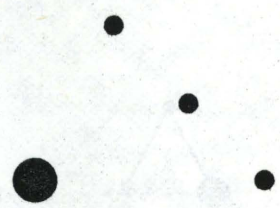
REFERENCES

Andersen, Poul 1954: "Dansk fonetik", Chapter XV of *Nordisk Lærebog for Talepædagoger*, p. 308-354 (Rosenkilde og Bagger, Copenhagen)

Basbøll, Hans 1977: *Dansk fonetik og fonologi - Skitse til en systematisk indføring* (Nordisk Institut, Odense Universitet, Odense)

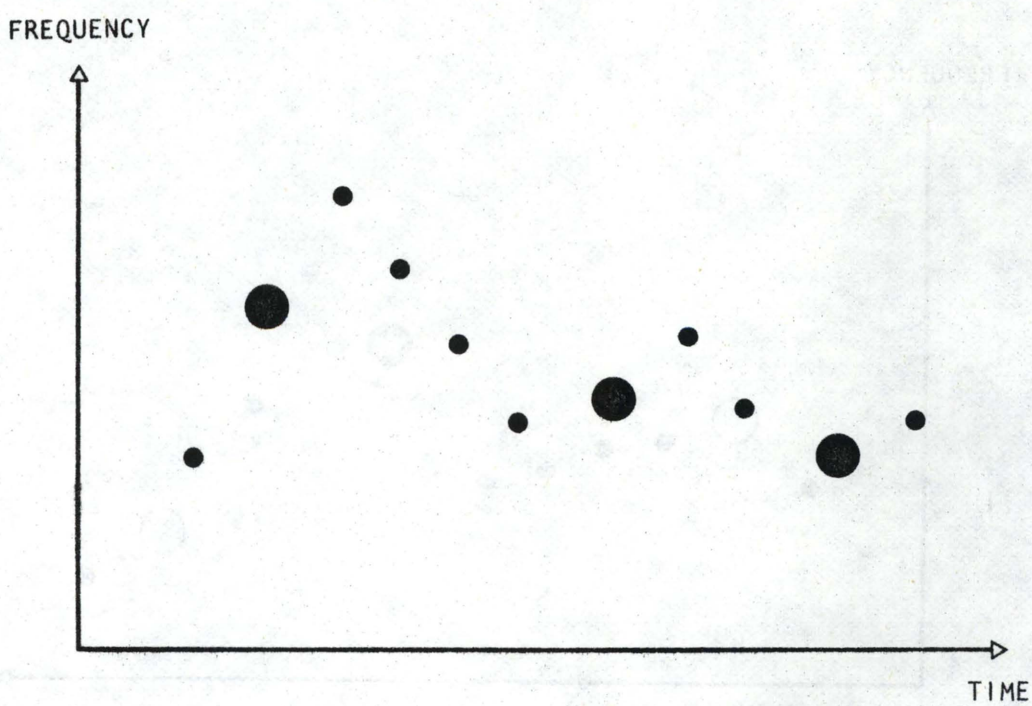
Heger, Steffen 1975: *Tale og tegn, Elementær dansk fonetik 2* (Gjellerup, Copenhagen)

Rischel, Jørgen 1970: "Acoustic features of syllabicity in Danish", *Proc.Phon. 6*, p. 767-770 (Academia, Publishing House of the Czechoslovak Academy of Sciences).



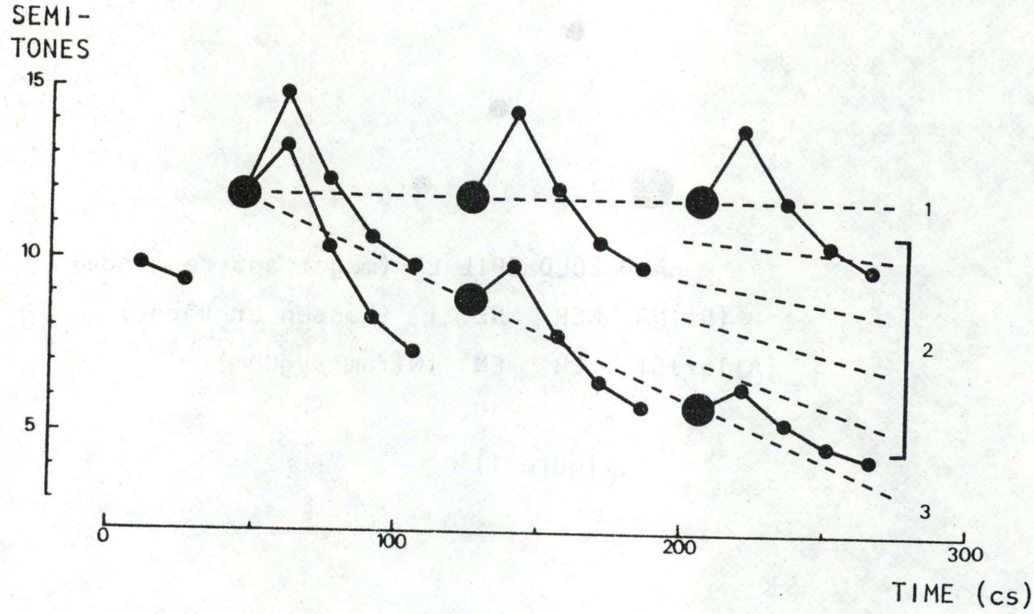
HÅND BOLD SPIL ER (meget anstrengende)
(Ba)NA NER NE I (kassen er rådne)
(Aller)GI ER EN IN(fam sygdom)

(figure 1)

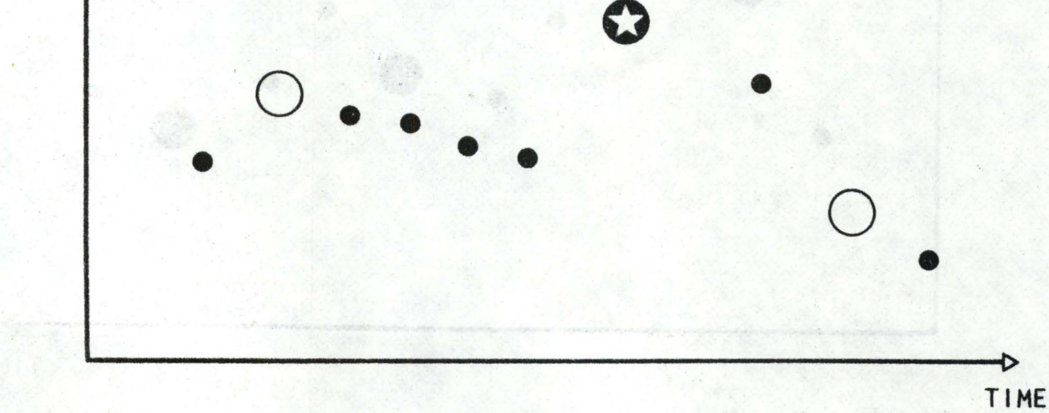


Herr Ándersen skal med bússen til Thísted

(figure 2)

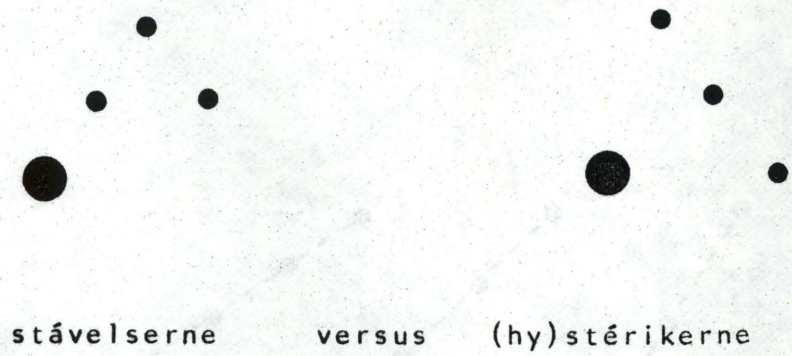


(figure 3)

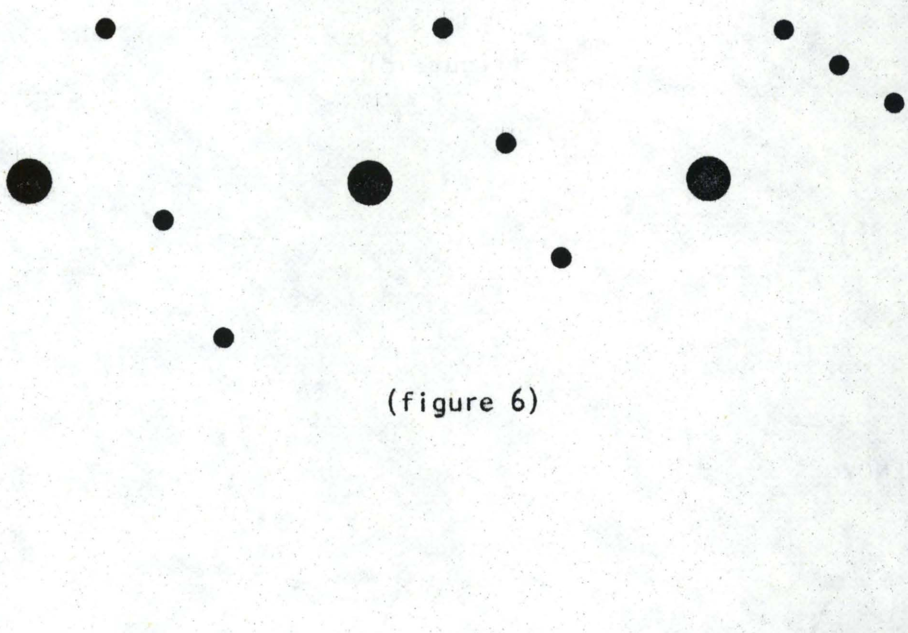


Herr Andersen skal med BÚSSEN til Thisted

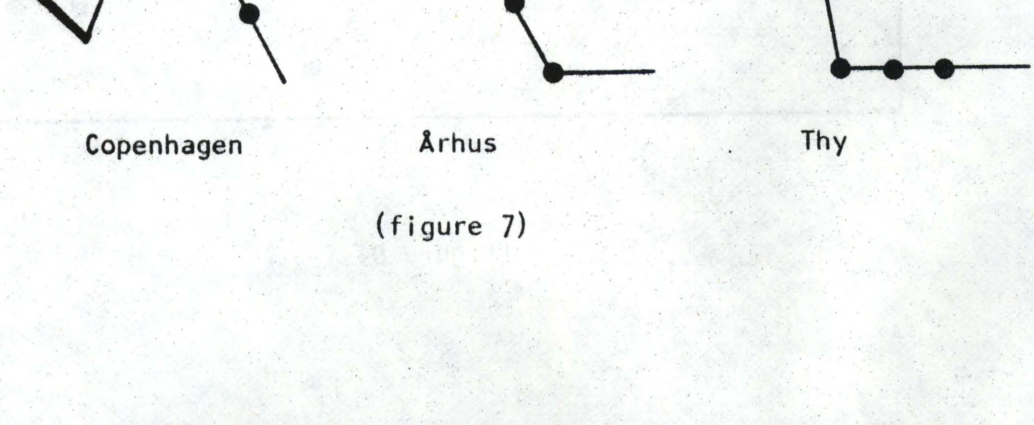
(figure 4)



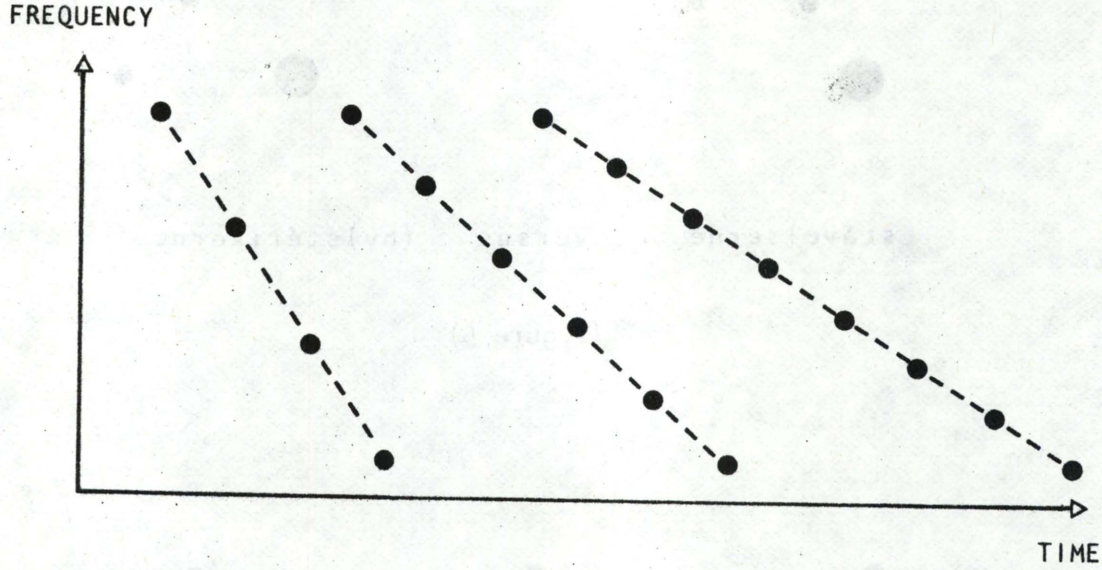
(figure 5)



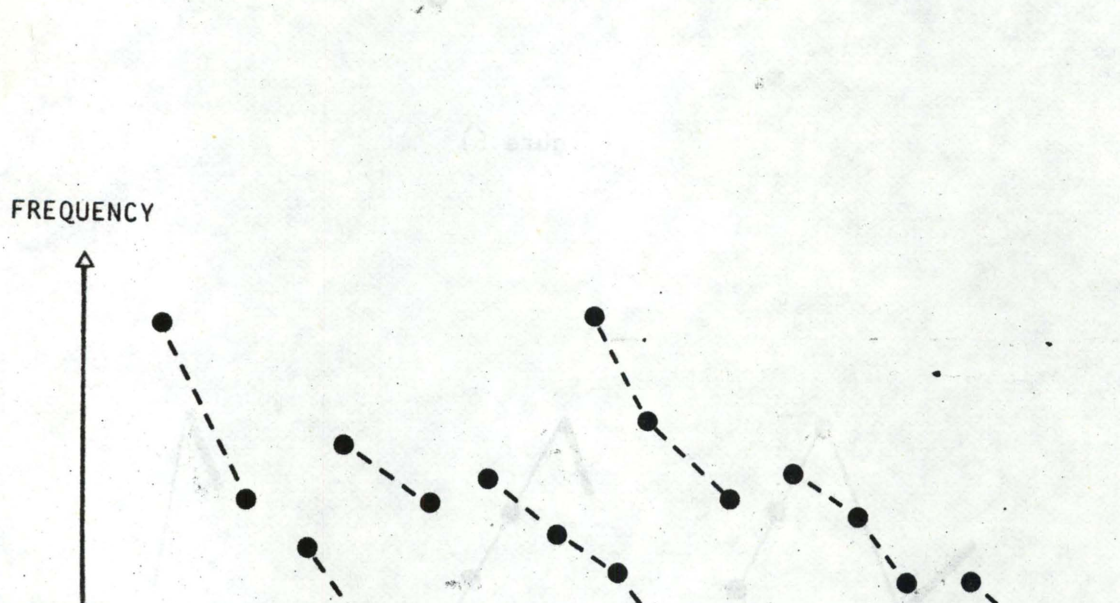
(figure 6)



(figure 7)



(figure 8)



(figure 9)

(ARIPUC 15, 1981 p. 61-75)

FIFTY YEARS WITH PHONETICS AND PHONETICIANS: A CAUSERIE GIVEN AT THE INSTITUTE OF PHONETICS 28/2 1981*

ELI FISCHER-JØRGENSEN

I have called this talk a causerie in order to emphasize that you should not expect any systematic survey of the development of phonetics during the last decades, nor any deep reflections on aims and methods of phonetic research. I only want to present some informal and not too serious, and quite personal memories from a long life as a phonetician.

I have called it "Fifty years with phonetics and phoneticians" but more exactly 52 years have passed since I took my first course in phonetics in 1929, and if I start from my first phonetic observation, it will be 62 years: We had just moved from Lolland to Funen, and it was the first day in my new school. We had gymnastics, and I still remember the comical contrast between what I expected to be a command and the friendly singing Funish dialect of the teacher. I decided that I would not speak this dialect. I would stick to the Standard Danish of my parents. And, as a matter of fact, I never learned to speak genuine Funish, but I came to like the dialect more and more - it became, as it were, a symbol of a very happy childhood; and later, when I went home on holidays from Copenhagen, it was always a big moment when after having passed the Belt I entered the train and heard the conductor speak Funish. Then I felt I had come home.

---

* This is a translation of a talk given at the Institute of Phonetics on the day of my retirement, 28.2.1981, slightly revised, particularly at the end. I am grateful to Betsy Uldall for improving my English style.

However, this phonetic observation was a rather isolated phenomenon. I have always been much more attentive to what I see than to what I hear; and I remember I once amazed my sister by describing a fellow student as a typical Copenhagener. "But have you not noticed that he speaks a pronounced South Jutland dialect?" she exclaimed. I must confess that I had only noticed his behaviour, not his language. Perhaps I should never have been a phonetician!

When I started studying languages it was not because I was interested in phonetics. On the contrary, I had found the phonetic transcriptions in our English and French primers rather irritating, and my first phonetic course did not change this attitude. It was a course in German phonetics, given by an assistant whom we disliked, and it consisted in learning physiological descriptions by heart and making transcriptions from orthographic texts. We never heard nor pronounced a single sound. A course in French phonetics, given by professor Sandfeld, did not interest me very much either. He was a great linguist, but phonetics was not his chief concern.

I was above all interested in general theoretical linguistics and in literature, and somewhat in the history of sounds since our Latin master at school had told us about the Germanic sound shift. However, I had to choose some concrete languages. I first thought of Danish and English, but as I heard that these subjects were rather crowded, I chose German and French instead. This rather accidental choice turned out to be very lucky. There were outstanding professors in both subjects. German was my main subject, and I owe most to Louis Hammerich, from whom I learned linguistic and philological method; from Carl Roos I learned the method of literary research. Kristian Sandfeld impressed me by his fine syntactic observations, and Viggo Brøndal by his bold and original theories and his philosophical perspectives. Brøndal's course in French phonetics was more interesting than Sandfeld's. He had constructed a general vowel system based on a restricted number of abstract features that could be combined, according to universal laws, into more and more complex vowel units. But it was not until later, when I took a course in Danish phonetics with Poul Andersen that I learned solid, phonetic method and observation.

Phonetics was, however, still a secondary interest. In those years I read all the books on general linguistics I could get hold of. I was particularly impressed by Saussure, Meillet, Schuchardt and Jespersen, quite different scholars, but each in his own way very stimulating. Around 1930 the first Travaux of the Prague School appeared; Hammerich drew our attention to them and lent me the books as soon as they arrived, and I read all the articles, with particular enthusiasm those by Jakobson and Trubetzkoy. It is perhaps difficult now to understand how revolutionary these new ideas appeared to us at that time. They meant an enormous widening of the horizon, a completely new way of looking at the sounds of language and the way they were integrated into the functional linguistic system. At the same time I found that the authors passed too lightly over the phonetic substance which at the start was pushed somewhat aside as belonging to natural science. I began to be interested in the interplay between phonological structure and phonetic substance. For my master's thesis, however, I vacillated for some time between a stylistic analysis of the Middle High German epos on Tristan and Isolde and Wilhelm von Humboldt's philosophy of language but ended up by writing about the importance of dialect geography for the conception of sound change.

In the spring of 1933 I became a member of the Linguistic Circle of Copenhagen, and this was decisive for my linguistic development. It was also in other respects a useful experience. I was a quick learner and I had found the studies very easy. In the Linguistic Circle I learned modesty (although not so much that it prevented me from intervening in the discussions). At least I realized the great difference between being able to understand a theory and being able to create a theory. The Circle was at that time dominated by a number of great and original personalities, and the intellectual level was very high. Particularly Brøndal and Hjelmslev were brilliant debaters, and I have rarely - if ever - felt as deep and pure an intellectual joy as when I listened to these discussions. The Linguistic Circle was also characterized by its openness to new ideas from outside. New linguistic literature was discussed at every meeting.

Hjelmslev did not yet have any appointment at the University, but after his thesis had been accepted in 1932 he used his right to lecture as "privatdocent", and I attended his lectures on Rasmus Rask and on Grammont's theories of "phonétique évolutive". Hammerich's textbook on German phonetics, which appeared in 1934, also contained a long chapter on Grammont, and it gave on the whole a good introduction to general phonetics.

After my MA in 1936 I received a scholarship for studies in Germany, and I decided to attempt a phonological description of German dialects. My decision to study phonology more seriously was partly due to my admiration for the Prague phonologists, but it was also negatively motivated by the fact that after having written a prize essay on the definition of the sentence I was fed up with syntax and with all the pseudo-philosophical twaddle I had had to read for this purpose, and I felt a strong need for some hard facts. I studied for two terms at Marburg but did not find much stimulation for my plans. Then I wrote to Trubetzkoy and asked whether I could study with him in Vienna. I received a positive and very kind answer and this is, according to Roman Jakobson, the last existing letter from Trubetzkoy's hand. He died shortly afterwards from a heart attack, and I never met him personally.

Instead I went to Paris in order to study phonology with Martinet and in order to learn experimental phonetics, which was not taught in Copenhagen. Martinet lectured on French phonology. He was an excellent teacher, and he and his Danish wife showed me much hospitality.

At the institute of phonetics I attended a course in experimental phonetics, consisting of exercises in palatography and kymography. There was a rumour that the institute also possessed an oscillograph which, however, was rusting away in the basement. There were two kymographs, one at the disposal of the students and one for research, which I had the privilege of using for a short time. This latter, finer kymograph was driven by means of an old rope, connected to a motor, and in order to get it going one had to pour a pail of water over the rope. To obtain the kymographic tracings we spoke into a mouthpiece connected by means of a rubber tube to a small capsule with a rubber membrane which reacted to the airflow and the vibrations. A small straw attached to the membrane followed its movements and recorded them on a turning drum with a coating of soot. After the recording we checked the speed of the drum by means of a tuning fork in the fond hope that it had been moving at the same speed during the recording. Pitch was calculated by measuring the distance between the peaks of the vibrations and at the speed normally used this distance was 0.5-1 mm, not much more than the thickness of the tracing line, so that even if we used a microscope the measurement was rather inexact.

Nevertheless, I found the course useful. Besides palatography I learned to handle the kymograph and to evaluate the curves. Marguêrite Durand, who as professor Fouchê's assistant was charged with the practical training of the students, was an excellent phonetician who was capable of getting interesting results in spite of the miserable instruments because she knew their restrictions and was able to put interesting questions. In a study of vowel length she showed that the perception of duration is influenced by pitch. If the pitch varies, the vowel is heard as longer. This was a new approach at that time.

From Fouchê himself I did not learn much. He is known to have made some interesting contributions on the subject of the historical development of the French language, but as a phonetician he had no very great capacity, either theoretically or practically (he taught us - in accordance with his own dialect - to pronounce French *r* as a tongue tip-*r* and to pronounce the nasalized vowels as [ɑŋ] and [oŋ]).

The lectures which impressed me most during my stay in Paris were Emile Benveniste's lectures on Indo-European morphology. They were simply brilliant.

The climate of the Institute of Phonetics was rather rough. If a student missed a class or two, even if it was because of illness, he might be excluded from the course. Fouchê was not very nice to Mlle Durand, and she vented her irritation on us. If somebody was butterfingered (and I was), he was cursed up and down. On the whole, studying in Paris required a certain robustness. I did not have much money, so I stayed in a filthy and beastly cold hotel room facing a dark wall, and the French students were fed up with the many foreigners and did not bother to talk to them. For several months I only knew the Martinet family and the two small mice which chased each other up and down the curtains when I had gone to bed and which, after a while, produced a young one which ate out of my hand. - Nevertheless, I enjoyed this stay in Paris immensely. I went to the Louvre every weekend and walked for hours along the quays and through the old streets, absorbing all the new impressions with all my senses alert.

Shortly before my stay in Paris I had taken part in the third international congress of phonetic sciences in Ghent. It was very exciting to meet all the people whose works I had read, above all Roman Jakobson, and some whose works I did not know, among them Eberhard Zwirner. He impressed me very much by his extensive knowledge and sharp intelligence, and we agreed that I should come to Berlin to study with him in April 1939 after my planned studies in Paris.

Zwirner had come from neurology to phonetics, realizing that the study of aphasia had to be built on knowledge of the normal language. He was well acquainted with many aspects of the humanities, e.g. the history of science and arts, including the history of linguistics (and he had taken a degree in the philosophy of history besides his degree in medicine). He was less acquainted with modern linguistics. However, following his own way of reasoning, he had come to a concept of the study of sounds which came very close to the approach of the Prague phonologists and of glossematics, i.e. that the phonetic investigation should be based on the restricted number of sound classes in a given language that are used to distinguish meaning. Their phonetic realization should be described by means of variational statistics. In Ghent he and the structuralists found each other, and Hjelmslev published an enthusiastic article about Zwirner's theories and tried to find a post for him in Copenhagen.

Zwirner was not much impressed by the phonetics he found in Germany, particularly represented by Panconcelli-Calzia, and besides the fundamental requirement of basing the investigation on linguistic classes he set up two further requirements, which were in sharp contrast to the Panconcelli-Calzia tradition, namely (1) one should not use instruments hampering the naturalness of speech, but make acoustic analyses based on gramophone records, and (2) one should not use isolated words or small sentences but connected speech. These are ideal requirements, which are not yet fulfilled in modern phonetics, and which in fact cannot always be met. It is not possible to investigate articulation without inconvenience to the speaker, and it is often necessary to start out with small homogeneous sentences in order to isolate the complicated factors involved in speech. But it is very useful to be reminded of the ideal conditions and not to be satisfied with the results obtained from small isolated sentences but to go on to investigate the interplay of all the complex factors in connected speech. And for Zwirner's original aim, the study of aphasic speech and above all its prosodic features, these claims were really compelling.

Zwirner was also a pioneer in the use of new instruments. Inspired by electro-cardiographic recordings he used an ink writer and an intensity meter which was very similar to the one introduced more than twenty years later by Fant, and he made more reliable pitch measurements based on oscillograms taken at high paper speed. The first automatic pitch meter was constructed at the same time by Grützmacher and Lottermoser, but it was so complicated that it filled almost a whole house and required constant nursing by two engineers.

Zwirner's pioneering work did not have the influence it deserved, mainly because he wrote in German and because his books and papers were published shortly before the war and thus for many years were hardly known outside Germany and Scandinavia. And in the meantime new instruments and methods were developed, particularly by American communication engineers for whom it was natural to take account of the linguistic units of communication and to use statistical methods, thus realizing some of Zwirner's main ideas without knowing his work.

Moreover, shortly before the war Zwirner's phonetic research was stopped. He was dismissed from the Kaiser Wilhelm Institute, accused of educating his children in a pacifistic spirit. Then he served as an officer during the whole of World War Two, and after the war he had to earn his living as a medical doctor and did not get a chair of phonetics until 1963; and in the meantime he had concentrated almost all his efforts on establishing an extensive collection of recordings of spoken German dialects.

But several of those who read his early works at the end of the thirties were deeply impressed by them, and I for my part feel that it was from him that I learned phonetic method. Therefore I have mentioned his work in more detail. But our personal contact was rather sporadic during my stay in Berlin because he was soon called up for military service. And since all the instruments were packed and stored, I could only study the curves already taken and discuss them with him when he was on leave. But I wrote my first two phonetic articles on the basis of his curves, and this encouraged me to go on with experimental phonetics.

At the same time I attended lectures by the well known specialist on African languages, Dietrich Westermann, who among other things has described the use of sound symbolism in West African languages, and I also took a course in phonetics with Wethlo, who was quite a character, and who had a small laboratory where, for the first time, I saw a wire recorder. I remember that we recorded the word 'Anna', and when we played it back reversed, the Germans still said 'Anna' but I said 'Hanna'.

Berlin was not a very pleasant place at that time. The classes were constantly interrupted by parades or Hitler's birthday or speeches by Goebbels, and, although most Germans still believed in Hitler's promise not to start a war, the imminent danger was evident to everybody else. - I came home two weeks before the war broke out.

After my return I got a job as teaching assistant in German for professor Hammerich. In 1943 I was appointed lecturer in phonetics, a new post attached to the chair of linguistics. In those years Hjelmslev became the leading Danish linguist. His well known book "Prolegomena to a theory of language" appeared in 1943, and there was a growing interest in glossematics among the young generation of linguists. The glossematic theory was discussed in a number of meetings in the Linguistic Circle and later, in the beginning of the fifties, the glossematic committee of the Circle had a long series of very stimulating meetings in which we discussed with enthusiasm for hours, particularly syntactic theory. Besides Hjelmslev, who was the sovereign leader, the most active discussants at that time were Knud Togeby and Paul Diderichsen; I had long arguments with Hjelmslev about form and substance.

The salary of the lectureship was very moderate, but I could just live on it, and there should have been time for research. But I did not write much during those years except for a number of reviews. One reason was that I had constant sinus trouble, and for more than 10 years I woke up every morning with a splitting headache which was not very conducive to productive work. Another reason was that I was more interested in reading and learning than in writing, and for many years after my exam I continued to hold the opinion, which I shared with a number of my fellow students of that generation, that the worst thing that could happen to me was to get a permanent post with a pension and become settled and bourgeois, so I did not aspire to any chair. I would have liked to do some experimental research, though, but we had no instruments. Moreover, there were more important things to do during the war.

However, Hjelmslev made great efforts to procure some instruments, and he worked out a detailed application to the Ministry of Education. He encouraged the director of the Institute of Speech Pathology to send in the same application to the Ministry of Social Affairs, and they agreed that the one who got the grant should put the instruments at the disposal of the other. The grant was given to the Institute of Speech Pathology, and at the end of the war it became possible to start a university course in practical experimental phonetics at that institute.

In the beginning we only had a kymograph. It was an electrokymograph, running at a constant speed and thus better than the one in Paris, but as a matter of fact it was almost out of date when we acquired it. However, it was quite useful for the students to learn kymography since a large number of the older papers they had to read contained kymographic tracings, and there were quite a number of phonetic properties that could be investigated with a reasonable degree of exactitude by means of the kymograph, e.g. duration, voicing, nasality, and - to a certain extent - pitch.

I enjoyed working with an instrument which did not require any deeper technical insight or technical assistance. When the straw broke, I could simply go out into the fields and pick a new one, and I have always loved to pick flowers. The rubber membranes posed a more difficult practical problem: I was too shy to enter a rubber shop, so I used to send one of the young students. It turned out later that he had been just as shy, and he had assured the shop keeper that it was not for himself, but for "a machine". We were very innocent at that time.

Later we got a "Frequenzspektrometer", permitting a frequency analysis of sustained vowels with an accuracy of 1/3 octave, and in 1951 we got our first tape recorder.

The fact that the instruments were situated at the Institute of Speech Pathology in Hellerup brought us in contact with Svend Smith, who had just finished his pioneering thesis on the Danish stød. He was in charge of the laboratory on behalf of the Institute, and neither the students nor I would have liked to miss the amusing memories of our small squabbles. We only had access to the laboratory in the evenings, and at the start we had no key to the door. When the door was locked at 10 P.M., we had generally just got the kymograph working, so when we left around midnight we had to help each other out of a window in the basement leading out into a flowerbed of roses. I suppose the gardener complained, so we got a key.

After the war I received a scholarship from the World Federation of University Women for a year's studies abroad. Hjelmslev advised me to go to London because he found that I needed practical phonetic training badly, in which he was certainly right. So I first went to London, and I enjoyed this stay very much. At University College I attended Daniel Jones's lectures and took courses in English and French phonetics. It was characteristic of the London School that the whole emphasis was on practical phonetics. The important thing was to be able to "make noises". - I can still hear Mlle Coustenoble saying "Encore! Encore! Encore!" until we could pronounce the word exactly as she did. She was known for being able to make strong and sturdy men break down in tears, but she was very efficient. And we always had dear old "Parky", Jones's lovable secretary, to comfort us. Daniel Jones was a quiet, reserved, polite and frail old English gentleman with a weak voice (which is known all over the world from his gramophone records of cardinal vowels). It was a great favour to obtain an interview of a few minutes, and I think it was only due to Hjelmslev's recommendation that he gave me private lessons in clicks. At the School of Oriental Studies I took courses in Yoruba and Chinese phonetics and attended lectures by J. R. Firth, the founder of the prosodic school. He was very different from Jones, a big, noisy man who laughed a great deal at his own jokes; but his lectures were interesting. I spent most evenings mending my only pair of stockings; it was just after the war, and everything was rationed.

From England I went to Holland. There was not much going on in phonetics in Holland at that time, but I gathered material for a book on the accent of compound words in the Germanic languages, and particularly in Danish, inspired by a theory of

Hammerich's concerning the influence of Dutch on the accent of Danish compounds which I did not find quite convincing. Later I put the material aside, mainly because Hjelmslev did not find the topic suitable for a phonetic thesis. In Holland I moved from one city to another, according to where they had sufficient fuel to heat the University library. It was a very cold winter; but spring was beautiful with lots of tulips and fruit trees in blossom. I also enjoyed the art galleries, the beautiful architecture of the towns and the canals where I skated during the winter and went canoeing in the spring.

In the years immediately following the war, acoustical phonetics made rapid progress, mainly due to the technical development which had taken place during the war. The most important new instrument was the sonagraph which made it possible to make a frequency analysis of 2½ seconds of connected speech in five minutes, whereas a few years earlier it could take several hours to analyse a single vowel. But we had no possibility of getting a sonagraph in Copenhagen. In 1952 I received a Rockefeller scholarship for studies in America. I had been looking forward to a six weeks' stay in Cambridge, but unfortunately I was ill most of the time. However, I managed to make a number of spectrograms of Danish words which I could use later for teaching purposes and for research. I was also glad to be able to meet Roman Jakobson, who was very kind and hospitable, and whom I have admired since I was a young student and still admire for his enormous knowledge, his incredibly fertile brain, his temperament, and his personal charm.

One of the last days I gave a talk at MIT. Among the audience I noticed an unusually intelligent face - it was a young research assistant named Chomsky.

From Cambridge I went to New York and worked for a few weeks in the Haskins Laboratories. It was a new institution at that time, comprising a relatively small group of people: Frank Cooper, Alvin Liberman, Pierre Delattre, and a few more. Their teamwork was excellent, and the atmosphere was very stimulating and very pleasant. They had just constructed the first really usable speech synthesizer, the so-called pattern play back, which was of a relatively simple construction. Stylized spectrograms were painted with white paint on cellophane and converted into sound by means of a tone wheel. I learned the technique and took part in some of the tests. In the late evening when the machine was not occupied I painted the words 'Københavns Universitet' and 'Fonetik', which I used for several demonstrations later. It was a very hot summer, and beads of perspiration dribbled down from my face and got mixed up with the paint, which may have added a special accent to the words.

I was very tired when I went from New York to Oklahoma, where Kenneth Pike had his big Linguistic School for missionaries, so I was not very enthusiastic about the very hard schedule he had set up for my visit, starting early in the mornings at 8 A.M. I simply refused to do anything before 9 o'clock. I had read and reviewed various of Pike's works and was very much impressed by them, especially by his excellent book on tone languages, and I was eager to see how the teaching was organized. The day started with prayer, in the hope that God would be interested in the teaching of implosive *b* and voiceless *l*, and it ended in the evening with discussions, e.g. about the structural parallelism between the Trinity and a phoneme with three variants. In between it was very competent linguistic research and teaching. For a heathen sceptic, who believed neither in God nor in the tagmeme, and who had sometimes asked herself whether her work was really of any use to anybody, it was a peculiar and at the same time very positive experience to meet people who believed in both and who were convinced that linguistics was not only useful on this earth but that it could help - through Bible translations - to bring souls to eternal salvation. This extra motivation made the students study with an unusual enthusiasm and assiduity.

You may wonder that I could travel so much. But at that time there was no exam in phonetics in Copenhagen, and no fixed schedule, and the students only came for fun, so it was possible to skip a term.

After my return I had six years with good possibilities for research: still no exams, good health, and support from the Science Foundation.

Inspired by Roman Jakobson I had become very much interested in the auditory dimensions of sounds, including their symbolic value and a possible parallelism to colours, and I started a large number of tests on these problems. However, it turned out that the available statistical methods were not yet very satisfactory, so I put it aside and only used some of it in a short paper for a Roman Jakobson Festschrift. - Instead I took up a more traditional subject, viz. an investigation of Danish stop consonants with special reference to the fortis-lenis problem. For this purpose I needed more spectrograms, and in January 1954 I went to Stockholm for a couple of weeks. This was the start of a close cooperation with Gunnar Fant, who was on his way to becoming an international authority on acoustic phonetics. I remember particularly this first stay in Stockholm. Since the sonagraph was used all day, I had to work at night. I slept in the morning and went skiing in the afternoon and had a very pleasant time.

In this same period I made an extensive perceptual test based on cutting and splicing of tapes in order to throw some light on the importance of formant transitions and explosion noise for the identification of Danish stops. This was before the time of the electrical segmentator, so it had to be done by hand. Since the laboratory in Hellerup was not accessible on weekdays, I had to do the final splicing during the Easter vacation, working day and night and living on canned food with three students helping me in shifts. At one time I had 4000 pieces of numbered tape hanging all over the place, and I was pleasurably surprised when everything had been spliced together and only one piece was running backwards. That night I found that the Science Foundation had not bestowed its money in vain.

Together with Oluf Thorsen I also undertook an investigation of intraoral pressure in obstruents. We used a manometer constructed for medical purposes (a new possibility in 1955) and were permitted to work in the University Hospital and later in the Hospital in Gentofte. We used oscillographic recording and, since the laboratory had to be ready for use next morning, the films had to be developed and dried the same night. In Gentofte they did not have any machine for drying the film, so I got the idea that I could stick them up on the tiled walls of the bathroom (that was the way we dried and ironed our handkerchiefs at college). The result was disastrous: the films stuck to the tiles and had to be torn off in small pieces. When, at three o'clock in the morning, I had finished this sad task and cleaned the walls, all the doors were locked and I had to climb out of a window in the basement. I was very afraid of being taken for a patient escaping from the psychiatric ward.

I managed to write 100 pages of the planned book on stops. Then I was thrown from a horse and got concussion, the effects of which lasted for a long time. During the following five years I could only work for short periods. I could lecture on topics I knew, but for long periods I could not read. When I recovered I got so much administrative work that I gave up finishing the book, but most of the material has later been used in papers.

The sixties were a period of expansion. From 1944 I had given regular courses in elementary phonetics for students of modern languages. I started out with three participants, but the number has grown slowly but constantly, and at the end of the fifties it was around 60-80 each autumn. In the sixties this course was included in the list of recommended courses for the various languages, and at the same time the number of students increased rapidly, so in the beginning of the seventies about 700 took part in the course. They had to be divided up into small groups, and more teachers were needed. From the end of the fifties some courses were given by teaching assistants, but in 1963 a fixed post was established for Jørgen Rischel, and during the following years 5 more posts (lectureships with tenure) were obtained. In 1966 I got a chair of phonetics (during my illness I had in fact become interested in a job with a pension). Since 1960 it had been possible to take a diploma in phonetics, and later this was recognized as a subsidiary subject for the MA. An MA in phonetics as a main subject was established in 1968. During this period our possibilities for instrumental research were also improved. In 1953 we had received a sonagraph from the Rockefeller Foundation, and from 1960 we could place a few other instruments in a small room in town in the Institute of Linguistics and Phonetics. The room could not be heated, however, and it needed repair badly. But in 1966 we got a real laboratory: 4 small rooms in the basement (which, for a change, were overheated), and at the same time a separate institute of phonetics was established and we got a technician and a secretary, and in the following years - in connection with a rapid expansion of our equipment - two engineers. In 1971 we got better rooms for the laboratory and in 1975, as everybody knows, we moved to Amager to premises which had been built specifically for our purposes.

This has been a very rapid development compared to the extremely slow improvements during the first 18 years. The change in conditions for recording can perhaps be used to symbolize the whole development: until 1966 the speaker had to put his head and the microphone into an old sheepskin coat serving as a "soundtreated room". In 1966 staff members built a (remarkably good) room out of old bicycle stands covered with Rockwool and curtains and, finally, now we have a super-modern anechoic room with suspended wire-mesh floor and sound-damping wedges.

The rapid expansion was only possible due to the general economic growth during the sixties. But another necessary condition was that capable young people were ready to take over the new posts and willing to do a hard job helping to build up the institute.

The fact that we now had an institute and a laboratory involved a new way of life. I had been used to working at home, except when I had to make recordings, and I must confess that I still prefer my writing table at home in quiet surroundings when I have to do serious work. But administration and meetings often required my presence at the institute, and I also realized that the rapid development of phonetic techniques and the abundance of new literature made teamwork a necessity in many cases. I have enjoyed discussing with my younger colleagues and learning from their expert knowledge within particular fields. And, on the whole, I think that we have managed to build up together an institute with a good climate, which functions well, and which stood the test when we arranged the Ninth International Congress of Phonetic Sciences in 1979. That was a job which made heavy demands both on our time and on our capacity for cooperation. But I think it was worth while.

Soon after its start the institute began to develop connections with institutions and scholars abroad. - Phonetics has always been a field with close international cooperation. When I started in phonetics there were very few phoneticians, in some countries only one or two, and most of us knew each other and read each other's publications (and so we all became world-famous in a very simple way), and many of us travelled quite a lot. Besides the journeys already mentioned I recall with pleasure a three-months' stay in India and a two months' stay in Japan as very exciting experiences. Now that we had an institute the personal relations could be supplemented by cooperation and exchange of reports with other institutes of phonetics. We have had especially close and enjoyable relations with the Institute of Phonetics in Lund and particularly with Kerstin Hadding and Eva Gårding and, on the whole, with our Swedish colleagues, and we have been able to invite foreign scholars to stay with us for a longer time, e.g. John Ohala from Berkeley and Hirose and Niimi from the Tokyo group, who helped us with our EMG-project. This has prevented us from being too provincial.

Looking back at these 50 years the most striking fact is the fantastic development of phonetics as a discipline. It can hardly be said to be the same subject as the one I became acquainted with for the first time in 1929. At that time the dominating trend was classical phonetics, as represented by Otto Jespersen, i.e. the description of speech sounds in terms of their production by the peripheral speech organs, based on auditory identification, kinesthetic impressions, and perhaps a mirror for control. Experimental phonetics existed, and had existed for some decades and it had obtained valuable results, but its methods were still restricted and classical phonetics, which dominated the University teaching completely, at least in Copenhagen, did not take much notice of it.

The beginning of the period - the thirties - was characterized by the appearance of phonology, i.e. the development of a theoretical framework for the description of the functions of speech sounds. Then - in the forties - technical progress made it possible to undertake an acoustic analysis of speech sounds within a reasonable time. And for a while acoustic phonetics was quite dominant. The acoustic aspect is perhaps not so interesting in itself, but it is a necessary link between sound production and sound perception, and the filling in of this missing link made it possible to proceed to an investigation of the whole speech chain.

The relation between speech production and the acoustic result has been investigated intensively since the fifties, e.g. by Gunnar Fant and K. N. Stevens. And as for the connection between the acoustic stimuli and perception, the main contributions came from the Haskins group. - For a while there was less interest in aerodynamic studies and in the function of the larynx, but new technical possibilities soon revived the interest in these fields. Finally, in the last few years, there has been a growing interest in neurophonetics, including the problem of hemispheric dominance. Our knowledge in this field is still restricted, but it is growing, and we approach a situation where we can add this last link to the speech communication chain. This also means that, whereas fifty years ago the phonetic description of individual languages was the main concern of phonetics, the interest has now shifted to general phonetics (although there is still much to be done in the former field). One now tries to set up models for both speech production and speech perception, and modern computer technique offers new possibilities for testing the models. On the whole, I find that phonetics has become much more interesting and has much wider perspectives than 50 years ago. It has been exciting to follow this development.

If at last I should try to look back briefly on my own research during these years, my dominating impression is that it looks very much like a village shop with smoked hams, clogs, and salt herrings. My relatively numerous papers and my few books deal with the most varied subjects: phonology, articulatory, acoustic, and perceptual phonetics, sound history, and particularities of Danish, German, French, Dutch, and Gujarati phonetics. It does not look as if there has been any plan - and in fact, there hasn't. In the first place I generally do not plan much in advance and, by the way, in a quickly developing discipline long-term planning is not too easy. Moreover, for many years the lack of appropriate instruments set narrow limits for the realization of the projects I was interested in. Finally, my research was often directed by the requirements of the courses taught. For almost twenty years I was the only teacher of general phonetics, except for the elementary courses given by assistants from the end of the fifties. And I have always felt that teaching a rapidly developing subject to a small selected group of intelligent and really interested students should not consist primarily in information about facts but rather in an introduction to scientific method and in discussions of new approaches and unsolved problems. I therefore felt obliged to keep up to date as best I could with the different areas of phonetics, and there has always been a close relation between my research and my teaching. On the other hand, it also happened that I had to interrupt a project because preparation for classes took too much time.

There are, however, a few subjects to which I have returned repeatedly because I could not solve them, e.g. close and open contact and the fortis-lenis difference (I still do not know the precise difference between French *p* and Danish *b*, and that irritates me). - And I have returned to other subjects because they appealed to me, e.g. perceptual phonetics. There has also been an obvious development in my interests from phonology to experimental phonetics. It is true that I have also written about phonology in later years, but this has mainly been because of the requirements of teaching. The growing interest in experimental phonetics may be partly due to a general tendency for old people to get tired of very abstract theories. I have seen too many theories replace each other in the course of time, and each time the adherents think that this is the only adequate approach and that everything that has been said before is nonsense, although it may be nothing but a new fashion, a complementary way of looking at the same facts. The models set up in experimental phonetics are of a different kind, They are hypotheses that can be verified or falsified by later research. There is still much we do not know, new facts, not only new points of view, that can be discovered. But I must confess that my contributions to these discoveries have not been very conspicuous. The results of my research remind me of a village shop also in the sense that the shop is not characterized by new elegant models from Paris but by old homespun cloth, sometimes with a small new detail in the pattern but maybe so small that it could also be an error in the weaving.

My predilection for experimental phonetics may also partly be due to the fact that I am predominantly a visual type - and I like to look at curves. Finally, it is more exciting and dramatic because it often happens that the machine breaks down or you have forgotten to turn a knob, and everything has to be started over again, - I enjoy surmounting difficulties.

I am glad that I have been offered a small office in the institute so that I can continue this exciting job for a few years, although I will miss the inspiration which I have always found in teaching gifted students.

I want to thank my former students and my colleagues for their loyalty, help and inspiration, and for the many merry hours we have spent together. And finally, I want to thank the University of Copenhagen, to which I have been attached as a student and a teacher for more than 50 years, for giving me continually improved working conditions, for having made it possible to build up this institute, and for having preserved the chair of phonetics under difficult circumstances.

May I finally express the wish that the future working conditions for University teachers and students will be better than we fear at the moment.

(ARIPUC 15, 1981 p. 77-102)

ASPIRATED STOP CONSONANTS BEFORE LOW VOWELS, A PROBLEM OF DELIMITATION, - ITS CAUSES AND CONSEQUENCES

ELI FISCHER-JØRGENSEN
AND
BIRGIT HUTTERS

Attention is drawn to the fact that low vowels may show a gradual start after aspirated consonants, so that it is possible to consider the vowel to start at different points: (a) at voicing start, (b) at the start of Formant 1, (c) at the start of higher formants. The choice of delimitation is shown to have serious consequences for the statement of various temporal relations. It is argued that the choice of point (c) is the most rational choice, both from the point of view of production and perception. It also gives the most regular temporal relations.

I. INTRODUCTION

The study of temporal relations in speech has been given much attention lately, one reason being that these relations may give new insight into the way speech is organized from the point of view of motor commands, and perhaps also from the point of view of perception. But a statement of the temporal relations presupposes a relatively unambiguous delimitation of the units studied. This delimitation is often undertaken on the basis of acoustic curves, although in most cases the phonetician is more interested in an interpretation in terms of speech production and speech perception. Particularly in complicated consonant clusters the perceptual segmentation may be at variance with the articulatory boundaries. The delimitation of intervocalic aspirated consonants is relatively simple in this respect, since here articulatory and perceptual boundaries may be expected to coincide. But it raises a problem of a different kind which is caused by the gradual transition between aspiration and vowel, often found in the case of low vowels. The differences between a delimitation made on the basis of start of periodicity and a delimitation made on the basis of start of full formant structure of the vowel may be up to 60 ms in individual cases and more than 40 ms for averages, more than enough to have crucial consequences for the temporal relations set up on this basis. The problem cannot thus be brushed aside as negligible.

II. THE START OF THE CONSONANT

The determination of the start of a stop consonant does not give rise to serious problems. The only difficulty is that the implosion noise is generally too weak to show up on spectrograms except in the case of apicals. In this latter case it is often seen as a thin line; and as it appears immediately after the cessation of the formants of the preceding vowel phoneticians generally agree to define the start of the consonant as the end of the vowel formants in all cases. This means that the (very short) implosion is included in the measurement of the closure phase of the stop.

On oscillograms the corresponding point is the start of an abrupt decrease of the amplitude of the speech wave, coinciding with an abrupt fall of the intensity curve and a sharp drop of the fundamental frequency (Fo) curve. The oscillogram very often shows 10-30 ms of weak oscillations after this point.

Some Swedish phoneticians seem to have a slightly deviating tradition on this point. They often place the delimitation line between vowel and following voiceless stop at the point where voicing ends (e.g. Lindblom and Rapp 1972, Leanderson and Lindblom 1972, Löfqvist 1975a). The reason for this rule seems to be that in Swedish the abduction of the vocal folds often starts before the oral closure, so that there may be a short period of preaspiration, particularly after stressed vowels. In some cases the voicing does continue after the closure, but it is generally too weak to show up in oscillograms, cf. Fant (1973 (1969), p. 120), Karlsson and Nord (1970), Lindqvist (1972), with tracings of glottograms, and Löfqvist (1975b). It is not quite clear how Lindqvist determines the point of occlusion, but Karlsson and Nord have used electrodes on the lips, and they demonstrate that for four Swedish subjects the closure of the lips comes immediately after the cessation of voicing when the preceding vowel is unstressed, and often 10-40 ms later when the preceding vowel is stressed. -- However, in curves of Danish, German, and English stops, and also in Dutch and French, the normal case seems to be a continuation of voicing during a few centiseconds after the oral closure, and in Danish the glottal abduction seems to start immediately after the implosion. The appropriate point of delimitation must thus be the point where the intensity falls abruptly.

In velar consonants the decrease is often more gradual than in labials and dentals, but the area of uncertainty is rarely more than 10 ms.

III. THE ASPIRATION

A. THE PROBLEM

The release of a stop can be located safely on almost all types of curves.

On acoustic curves, particularly spectrograms, it is often possible to distinguish the transient noise of the release from a following fricative phase, which in its turn can be distinguished from the aspiration proper, characterized by a more *h*-like noise (see, e.g., Fant 1973 (1969), p. 111). On mingograms this three-way distinction cannot be made, and generally the three phases of transient noise, fricative noise, and aspiration are taken as one segment, which is sometimes called 'burst' (although this term is also used for transient plus fricative phase alone), or 'open interval', or simply 'aspiration' (thus used here in a wide sense). These finer segmentations will not be treated here. The problem we want to take up is the delimitation between aspiration and the following vowel.

In most cases the start of the vowel after a voiceless consonant does not raise any problems. Generally voicing starts simultaneously and abruptly in the whole spectrum. This is the case after fricative consonants, after unaspirated stops, and in most cases also after aspirated stops followed by high vowels. But in the case of low vowels there often seems to be a more gradual transition between aspiration and vowel, which may give serious problems of delimitation.

B. MATERIAL AND SUBJECTS

In order to throw some light on these problems we have examined a large number of curves of Danish words with stop consonants. In Danish *bdg* are voiceless and *ptk* strongly aspirated, for older speakers around 70 ms, for younger Copenhagen speakers around 90-100 ms (see Fischer-Jørgensen 1980). The investigation is based on recordings by 16 different Danish speakers who had acted as subjects in earlier investigations, some of them in more than one: A, the material used by Hutters (1979) (5 subjects: HU, MF, FJ, LG, and PA); B, part of the material used by Fischer-Jørgensen (1980), particularly lists M2 (6 subjects: HU, MF, LG, BM, JJ, and PM) and S5 (3 subjects: NK, BL, KS); C, the material used by Jeel (1975), which was re-measured for this purpose (6 subjects: LG, BM, VJ, EC, EH, and JJ); D, a material used by Reinholt Petersen for measurements of fundamental frequency and spoken by himself (NR); and finally E, a recording of *pale* and *pile* by PD.

Some subjects from material B were left out because of various problems, e.g. JR, who showed a variable vowel start, often with gradual appearance of F1, and BH, whose higher formants were so weak that their starting point could not be identified. PM has only been included for aspiration measurements, not for vowel durations, because the delimitation from the following consonant was uncertain.

The measurements of material A are based on a comparison between spectrograms and mingograms. As for material B, the measurements of list S5 are based on spectrograms, those of M2 on mingograms controlled by means of a restricted number of spectrograms. For material C, D, and E only mingograms were available. The mingograms contained a duplex oscillogram, a high fidelity intensity curve, an intensity curve highpass filtered at 500 Hz, and a fundamental frequency curve (material A also comprised photo-electrical glottograms). The mingograms did not contain any intensity curve highpass filtered at higher frequencies.

The slow start of the vowel is found more often in Danish *pa* and *ka* than after *t* which is affricated. *t* has therefore been left out, and only words with *pa* and *ka* have been examined and compared to words beginning with *pi*, *ki*, *ba*, *ga*, *bi*, *gi* and *fa*, *fi*. Danish *i* is very high, also when it is short; Danish *a* is a front vowel varying between cardinal vowel No. 4 and 3, for younger speakers closer to No. 3, thus [æ].

C. TYPES OF VOWEL START

An inspection of the curves showed a high degree of individual variation, but four main types could be distinguished: (1) some speakers have an abrupt start of all formants in both high and low vowels (although there is sometimes a strong fundamental and a certain weakening of F1 at the start). The abrupt start is mainly found in older subjects with relatively short aspirations (KS, NK, BL), but in one recording also in the case of JJ (list C), who has long aspirations. (In a different recording (list M2) he shows later start of higher formants.) (2) Some speakers start out with a few centiseconds of very low frequency vibrations (narrow band spectrograms show that the main intensity is in the fundamental), but there may also be some very weak striation in the region of the first formant; then, after 20-40 ms, all formants appear with full intensity (subjects LG and BM). (3) Some speakers start with a relatively strong first formant, but the higher formants do not appear until some centiseconds later (PA and JJ, list M2). (4) Finally, some speakers (in our material only female speakers: HU, MF, and FJ) often have a still more stepwise start: first low frequency (sometimes with very weak striations in the region of the first formant), then a strong F1, and finally the higher formants. In the following we will indicate the start of low frequency vibrations as point (a), the start of F1 as point (b), and the start of higher formants as point (c). There are, however, intermediate cases and it is sometimes difficult to decide when the first formant should be considered to start.

In oscillograms the low frequency start shows up as simple vibrations, often - but not always - of relatively low amplitude. The duplex oscillogram shows lack of higher components. The start of F1 and higher formants cannot be distinguished in the oscillogram. It is, however, possible to identify the start of the higher formants by means of an intensity curve, highpass filtered at a higher frequency, e.g. 1500 Hz. Moreover, in Danish - at least in Copenhagen speech where stressed vowels generally have a falling fundamental frequency contour (which is particularly clear in low vowels) - there will be a turning point from rising to falling fundamental frequency at the point where the higher formants start. This has been controlled for several speakers and seems to be quite consistent. Some speakers do not always have the rising-falling fundamental frequency, but when they have it there is coincidence with the start of higher formants. Thus the turning point in the fundamental frequency curve could be used to identify point (c) in the cases where only mingograms were available.

Figure 1 shows some typical cases, and table 1 gives a survey of the magnitude of the differences in ms between points (a), (b), and (c). The difference a-c could be measured in all cases; (b) may coincide with (a) (PA, JJ) or with (c) (LG, BM), or it may not be measurable on the available curves (list C and HU, MF in list B). The average difference in ms between points (a) and (c) is 30.1 ms in *pa* and 24.5 ms in *ka*.

Apart from individual differences, other factors may influence the start of the vibrations. The force of the speaking voice plays a role. The higher the loudness level, the higher the chance of finding an early start of low frequency vibrations, and in this case they may even show up before high vowels (this has been shown by Preben Andersen (1981)). Some speakers often have a slightly later start of F2 and rising fundamental frequency at the start in high vowels. The preceding vowel may also be of some influence: we have, e.g., found a slightly earlier start of the vibrations in *api* than in *ipi*. Moreover, a quicker tempo favours voicing, and in quick connected speech an unstressed aspirated Danish consonant may be completely voiced. Finally, it cannot be excluded that the distance to the microphone may have a certain influence on the intensity of higher formants. It cannot, however, explain the different types of vowel start found in the present investigation. - The difference we have found between high and low vowels is thus fully valid only in the usual types of test sentences like "they said *pile*, they said *pale*", etc. - i.e., in distinct speech at a moderate tempo and loudness level.

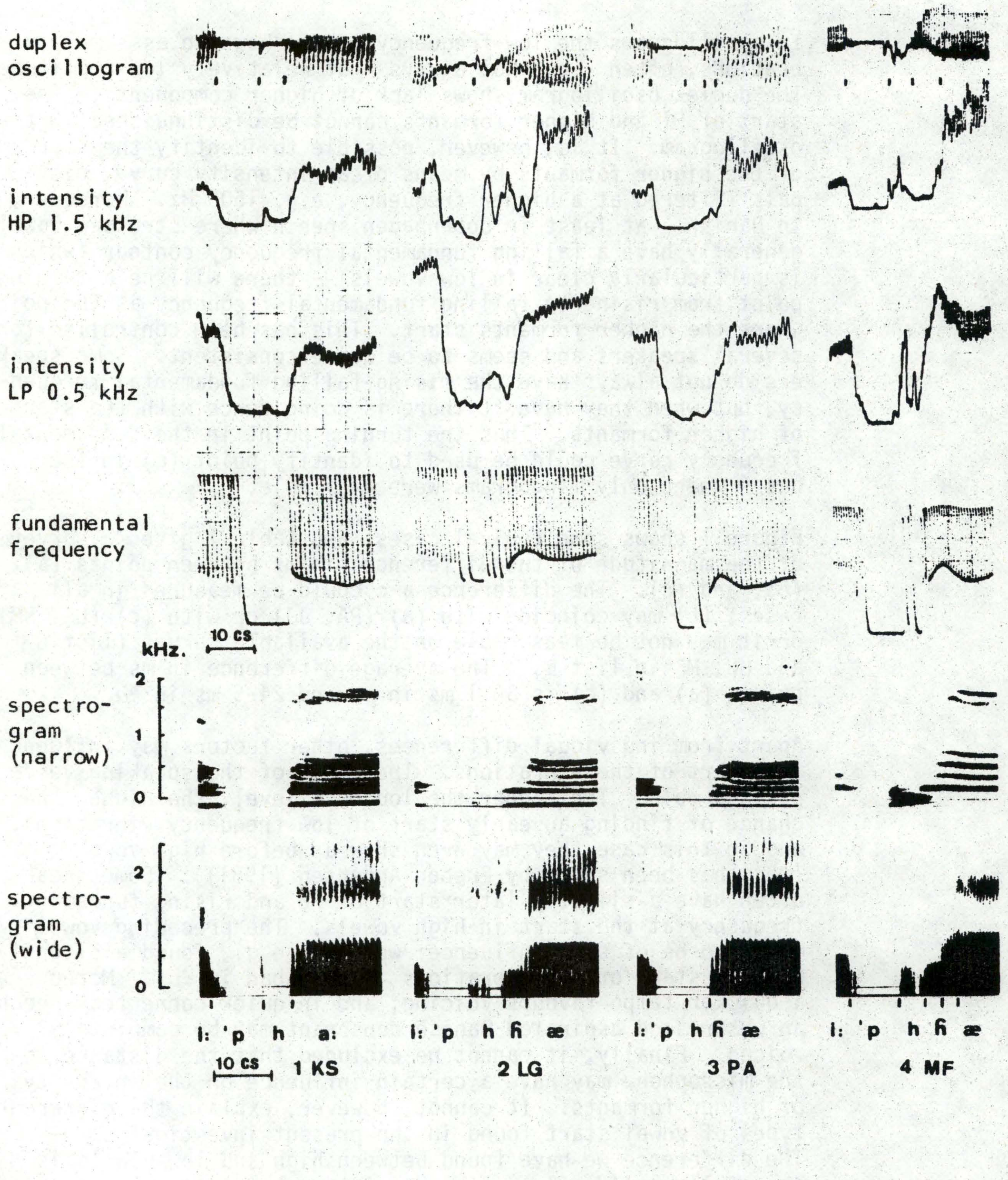


Figure 1

Four types of vowel start in Danish *pa*: 1 (KS) abrupt start; 2 (LG) start Fo-F1,F2 (a-bc); 3 (PA) start Fo,F1-F2 (ab-c); 4 (MF) start Fo-F1-F2 (a-b-c). (In MF's mingogram there is a click from a camera simultaneously with *p*'s open interval.)

Table 1

Differences between measurements in ms according to voicing start (point a), start of F1 (point b), and start of higher formants (point c) for speakers with gradual vowel start in *pa* and *ka*, separated according to the types mentioned in the text: a-b-c (type 4), a-bc (type 2), ab-c (type 3), a-c (cases in which point b could not be identified). (N = number of averages, each including 9 tokens, on the average. In the other tables, N = number of tokens.)

Type	Speaker	Material	N	*pa* Difference a-b	*pa* Difference b-c	*pa* Difference a-c	*ka* Difference a-b	*ka* Difference b-c	*ka* Difference a-c
a-b-c (4)	HU	A	1	17	29	46	7	28	35
	MF	A	1	17	16	33	19	10	29
	FJ	A	1	21	14	35	17	11	28
a-bc (2)	LG	B	3	34		34	36		36
	BM	B	2	32		32	24		24
ab-c (3)	PA	A	1		34	34		21	21
	JJ	B	1		21	21		30	30
a-c	HU	B	1			31			27
	MF	B	1			22			12
	EC	C	1			31			19
	EH	C	1			21			9
	VJ	C	1			12			16
	NR	D	1			34			32
	PD	E	1			36			
Grand means						30.1			24.5

D. CHOICE OF CRITERIA IN THE PHONETIC LITERATURE

The gradual start of the vowel may be particularly prominent in Danish with its long aspirations, but it is not restricted to Danish. We have found similar gradual transitions between aspiration and vowel in curves of Indian aspirated consonants (although not as a regular phenomenon) and in curves of German and English aspirated stops.

It is, therefore, not surprising that phoneticians who have measured vowel duration or duration of aspiration have used different criteria of delimitation. A few phoneticians consider the vowel to start at the release of the stop consonant (or at the end of the fricative phase). The main argument for this delimitation is that the supraglottal part of the vocal tract may be in the position for the vowel already at this time, and the formants - at least the higher formants - can often be traced in the aspiration noise; thus the vowel may be considered to have started, it is only voicing that begins later. This might be an acceptable point of view if the voiced part of the vowel were correspondingly shortened so that the sum of the aspiration (= voiceless vowel) and the voiced vowel would be of approximately the same duration as a vowel after unaspirated consonants. It is true that there is, e.g. in Danish, a certain shortening of the (voiced) vowel after aspirated consonants compared to unaspirated consonants, but this shortening is of only about 15-20 ms (sometimes in the case of speakers with very long aspirations 25 ms), and this modest shortening does not by any means compensate for aspirations of about 100 ms. If the aspiration were included as part of the vowel, short vowels after *ptk* might be acoustically more than twice as long as short vowels after *bdg*, and they would often be longer than long vowels after these latter consonants. This would give very irregular vowel durations. It can thus be concluded that considering the aspiration as part of the vowel in languages with long aspirations will lead to absurd temporal relations. (These arguments may not be valid for sonorant consonants following aspirated stops, but this problem will not be treated here.) Moreover, neither the speaker nor the hearer is aware of any difference in vowel duration after *ptk* and *bdg*. The vowels seem to be intended and perceived as being of the same duration. Obviously, the vowel, as the syllabic peak, has to be voiced in order to fill its function. - The remaining possibilities are thus points (a), (b), and (c).

Those who have measured oscillograms have generally considered the start of the voicing (point a) as the start of the vowel (e.g. Löfqvist (1975a), Kent and Moll (1969), Zue (1976)). This is the point which can be located with the highest degree of precision on oscillograms. The term "voice onset time", introduced by Lisker and Abramson, also at first glance seems to indicate a delimitation according to voicing, but it appears quite clearly from their discussion of segmentation problems (1964, p. 416-418, and 1967), that they do not include what they call weak "edge vibrations", appearing near the base line of the spectrogram, which they have found in a small number of utterances (their example is *k + a*). These edge vibrations are excluded because perceptual experiments have shown them to be inaudible. This means that in the cases of gradual transitions between aspiration and vowel they would not place the end of the voicing lag period at point (a) but rather at point (b) (start of the first formant) where voicing will probably be fully audible, but hardly at point (c). -- On the whole, those who base their measurements on spectrograms will hardly include these "edge vibrations" since they do not always show up very clearly in spectrograms (they may be partly covered by the base line and sometimes by the effect of puffs of air in the microphone after an aspirated *p*).

Lehiste and Peterson (1960) also indicate that they consider the vowel to start "at the moment in time at which periodic striations are stable in the first formant frequency" (thus point (b)). Still others choose point (c), thus Klatt (1975, p. 687) who states that "the VOT is indicated by the sudden onset of vertical striations in the second and higher formants" and Weismer (1979). Finally, it should be said expressly that placing the VOT limit e.g. at point (b) or (c) does not necessarily imply that one considers the vowel to start at this point (VOT means "(audible) voice onset", not "vowel onset"). Abramson (personal communication) is more inclined to consider the aspiration as a (voiceless) part of the vowel.

In earlier writings by one of the authors (FJ) there has been a certain vacillation between point (b) and point (c), whereas Hutters (1979) has chosen point (c).

E. DISCUSSION OF THE PRODUCTION AND PERCEPTION ASPECTS

As long as the only aim is acoustic measurements, the problem is not so important. One may simply measure the different acoustic segments. However, as mentioned above, one generally wants to make measurements which are meaningful from the point of view of production and/or perception. Moreover, when the purpose is to set up rules of temporal relations, one may prefer a delimitation of the units involved which permits a maximal generalization of rules, covering, e.g., both high and low vowels by the same rule. There is finally the practical problem of choosing a point which can be located with a sufficient degree of certainty on different types of curves.

Point (a) is the easiest one to locate on oscillograms. But this choice can be criticized both from the point of view of perception and production. From the point of view of production it is evident that vibrations starting in point (a) are produced before the glottis is closed for the vowel. That the vocal folds can vibrate in a relatively open position could be seen in the early high-speed films of glottal movements, e.g.

the Bell film and the film by Elizabeth Uldall, "Vocal Cord Action in Speech: A High Speed Study", 1957. That this is also the case in Danish *pa*- and *ka*-syllables has been demonstrated by means of a large number of simultaneous acoustic and glottographic recordings of Danish stop consonants (Hutters 1979). In *pi* and *ki* voicing normally starts very shortly before or at the end of the glottal gesture, whereas in *pa* and *ka* it starts at some point during the closing phase, sometimes even at the top of the glottographic curve. This difference can be explained by the fact that the open vocal tract in *pa* [phæ] and *ka* [khæ] causes a quick pressure drop above the glottis which permits voicing to start early, whereas the more constricted vocal tract in *pi* and *ki* will delay the intraoral pressure drop, and more so in *ipi* than in *api* because the tongue will be raised all through *ipi*. After Danish *t* the pressure drop is delayed by the affrication both before *i* and *a*. The earlier start of the voicing in very loud speech can be explained by the higher subglottal pressure.

When voicing starts before the glottis is closed the glottal spectrum will contain only relatively low components, and the formants will not appear until later, or only in the form of noise. As long as the vocal folds are separated the fundamental will also be relatively low, and it will be rising until the glottis is (almost) closed. This means that point (a) may occur at any moment of the glottal closing movement and before the tongue is in position for the vowel, depending on subglottal pressure, tempo, and other factors which do not have a direct relation to the segmentation of the speech chain in consonants and vowels but belong to the feature of voicing. Voicing is not a sufficient criterion for vowel start. Point (c), on the other hand, i.e. full formant structure and start of the Fo movement characteristic of the vowel, can be shown to coincide with the end of the glottal movement (see Hutters 1979). This will, moreover, give the same criterion for high and low vowels. It is thus a well defined point (see figure 2).

Point (a) can also be criticized from the perceptual point of view. As mentioned above, Lisker and Abramson found that these low frequency "edge vibrations" were not perceived as voicing, and in an experiment with tape cutting and splicing of Danish stop consonants (Fischer-Jørgensen 1972a) it was shown that the inclusion or removal of a segment of 10-15 ms containing such low frequency vibrations (and probably some weak high frequency noise which was not, however, visible on the spectrogram although the vibrations were) was decisive for the perception of the preceding stop as unaspirated or aspirated, thus this segment was heard as belonging to the consonant. Moreover, it can be stated more informally (we have not made any tests) that aspirations which, according to criterion (a) are of about 30-40 ms but according to criterion (c) about 80-90 ms, are heard as normal Danish aspirations.

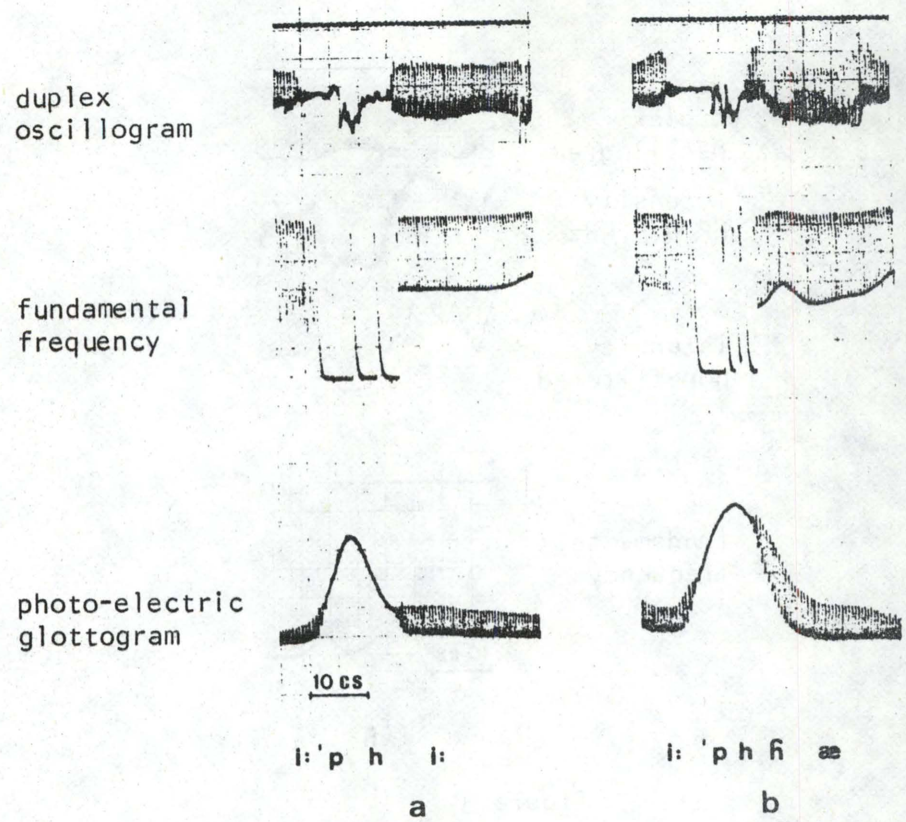


Figure 2

Danish *ipi* and *ipa*, subject MF

The rejection of point (a), i.e. start of periodicity, is corroborated by an investigation of Darwin and Pearson (1980). They have tested the *p-b* boundary on various continua of synthetic sounds and found that periodicity (in this case weak periodicity of formants) is of little importance for the perceptual boundary between *p* and *b*. Perceptually aspiration seems to end at a point where the formants of the vowel have reached a certain level of intensity. Probably this point coincides more or less with our point (c): start of higher formants. Further experiments are required to test whether overall intensity or intensity of particular formants is the determining factor.

Point (b) is less well defined than (c). From the point of view of production it is situated somewhere during the closing movement of the glottis, and acoustically the first formant has appeared as voiced, but the higher formants only as noise. Perceptual tests are needed on this point. Moreover, point (b) is rather difficult to locate on the oscillograms because there is often a gradual transition from low frequency vibrations to

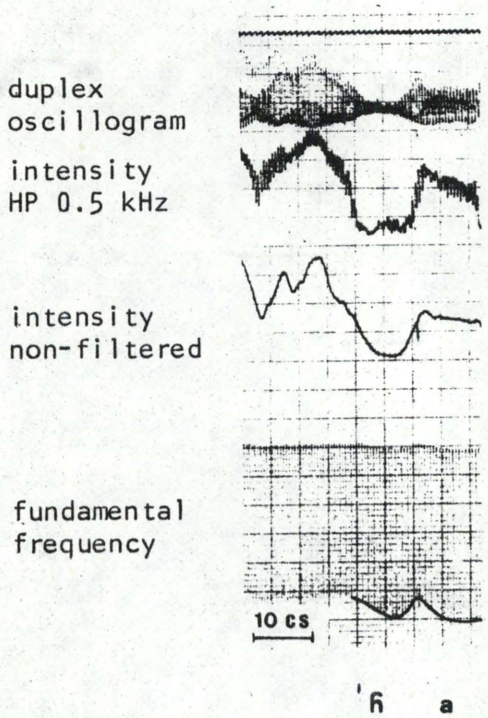


Figure 3

Danish voiced *h* in *-ha-*,
subject EH

full first formant, and the intensity of the vibrations may be gradually increasing.

An argument supporting the choice of point (c) is that this point corresponds to the most appropriate point of delimitation between voiced *h* and a following vowel. Voiced *h* has a dip in intensity and fundamental frequency, and if the following vowel has a falling fundamental frequency, the turning point is very clear. Point (c) will give a common criterion for vowel start after aspirated voiceless stops and *h* (see figure 3).

F. CONSEQUENCES FOR THE DESCRIPTION OF TEMPORAL RELATIONS

Important arguments against point (a) and for point (c) can also be found by looking at the consequences of the choice for the regularity of rules concerning temporal relations.

The relations examined are: (1) differences of vowel duration after *ptk* compared to *bdg* and *fs*; (2) differences of duration between the vowels *a* and *i*; (3) differences of duration of aspiration before *a* and *i*.

1. DIFFERENCES OF VOWEL DURATION AFTER *ptk* COMPARED TO *bdg*

As mentioned above, Danish vowels are generally slightly longer after *bdg* than after *ptk*. This can be shown for high vowels where no problem of delimitation arises. Since the materials were not originally recorded for this purpose, there are not always examples of all consonants, but the number is sufficient for a thorough examination.

The difference between the durations of *i* after *ptk* and *bdg* could be measured for 15 different speakers, some of them taking part in various recordings. There were 54 averages, each based on about 9 tokens, and they all showed the vowel to be longer after *bdg* than after *ptk*, on the average 16.4 ms. This is thus a very well established relation. Finally, at least for some subjects, the choice of point (c) gives a smaller range of variation.

The relations for the vowel *i* can now serve as a standard of comparison for measurements of the vowel *a* after *ptk* and *bdg* according to different criteria of delimitation. In table 2a and figure 4a the averages of *i* and *a* are compared for four speakers who have an abrupt start of both vowels. The grand mean of the difference between the durations of the vowels after *b* and *p* is 22.0 ms for *i* and 21.6 ms for *a*. For *g*-*k* it is 16.2 ms for *a* (there are no comparable cases with *i*). Table 2b and 2c and figure 4b-c show the differences for speakers with gradual vowel start. In table 2b and figure 4b the difference between the durations of *a* after *pk* and *bg*, delimitated after two different criteria, (a) and (c), is compared to the durations of *i* after *pk* and *bg* for the same speakers. The grand mean for *i* is 17.4 ms; for *a* measured from the start of the vibrations (criterion (a)) it is -0.4 ms, whereas measured from the start of the higher formants (corresponding to the turning point of the pitch curve) it is 24.1 ms. Table 2c and figure 4c show the difference between the duration of *a* after *pk* and *bg* according to criteria (a) and (c) for those speakers who did not speak words with *i*. The grand mean for point (a) is -6.8 ms, for point (c) 18.3 ms.

These comparisons show quite clearly that there is full agreement between the measures for *a* and *i* if the vowel *a* is defined as starting at point (c), but absolute disagreement if it is defined as starting at the beginning of voicing (point (a)). In this latter case the vowel *a* will be of approximately the same duration after *pk* and *bg* or, for some speakers, clearly shorter after *bg*, in contradistinction to the consistently longer duration after *bg* for the vowel *i*, which was also found for *a* in the case of the speakers with abrupt vowel start. The relations are also valid for the individual speakers except that in table 2b criterion (a) is preferable to (c) for JJ (*b*-*p*), and (a) and (c) are approximately equivalent for BM (*b*-*p*). For LG (a) is preferable both in the case of *b*-*p* and *g*-*k* when compared to the values for *i* of the same speaker but not when compared to the differences for other speakers, and in table 2c criterion (c) is obviously preferable for LG.

Table 2

Differences of vowel duration after *pk* and *bg* (in ms)

a. Speakers with abrupt start of all vowels

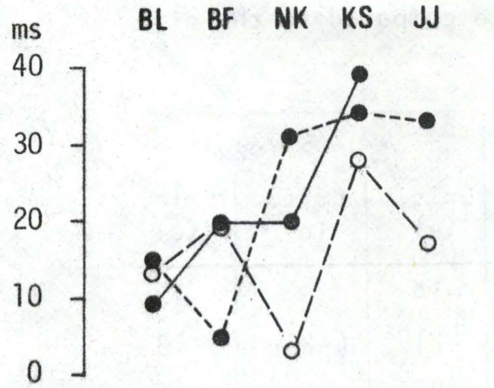
Speaker	Material	N	*(b)i>* *(p)i*	N	*(b)a>* *(p)a*	N	*(g)a>* *(k)a*
BL	B	2	9	6	15	6	13
BF	B	2	20	2	5	2	20
NK	B	2	20	6	21	6	3
KS	B	2	39	6	34	6	28
JJ	C			20	33	20	17
Grand means			22.0		21.6 ↘ 18.9		↙ 16.2

b. Speakers with gradual vowel start in *pk* + *a*

					Crit. (a) (PA and JJ: ab)		(Crit. (c)	
Speaker	Material	N	*(b)i>* *(p)i>*	*(g)i>* *(k)i>*	*(b)a>* *(p)a*	*(g)a>* *(k)a*	*(b)a>* *(p)a*	*(g)a>* *(k)a*
LG	B	10	10	7	4	6	26	28
BM	B	10	9	17	-3	1	24	27
HU	B	6	8	24	-20	-7	11	20
PA	A	5-15	27	15	-19	-12	15	9
MF	B	10	26	23	12	0	33	12
JJ	B	10	11	31	24	9	45	39
Grand means			15.2	19.5	-0.3	-0.5	25.7	22.5
			17.4		-0.4		24.1	

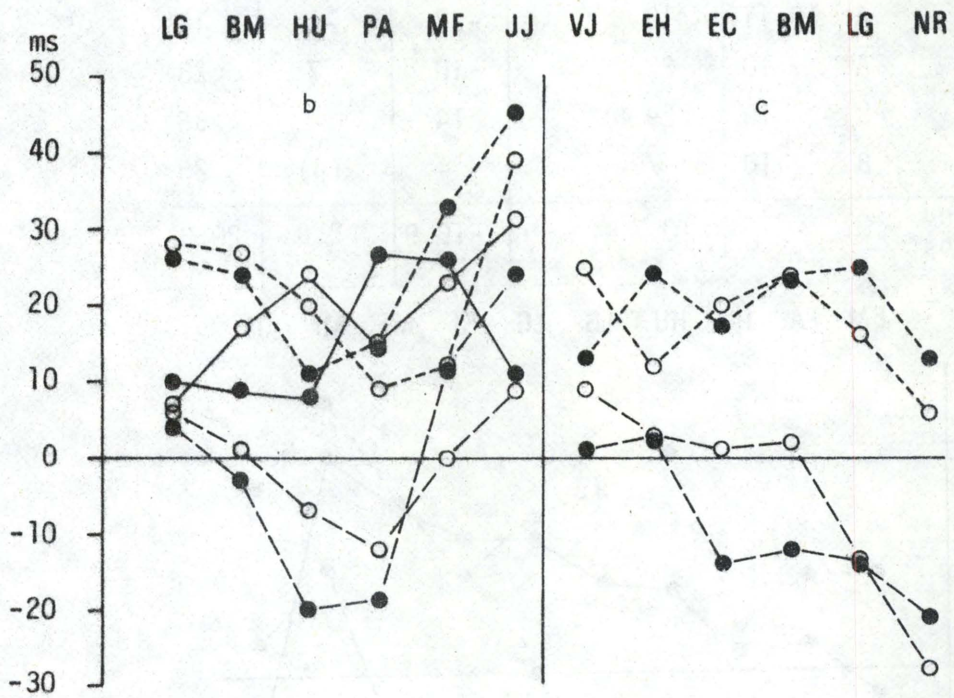
c. Speakers with gradual vowel start in *pk* + *a* (no examples with *i*)

		N		Crit. (a)		Crit. (c)	
Speaker	Material	*b-p*	*g-k*	*(b)a>* *(p)a*	*(g)a>* *(k)a*	*(b)a>* *(p)a*	*(g)a>* *(k)a*
VJ	C	10	5	1	9	13	25
EH	C	12	10	3	3	24	12
EC	C	12	8	-14	1	17	20
BM	C	11	10	-12	2	24	24
LG	C	8	9	-13	-13	25	16
NR	D	9	9	-21	-28	13	6
Grand means				-9.3	-4.3	19.3	17.3
				-6.8		18.3	



a. Speakers with abrupt vowel start in *a*.

●——● *(b)i* > *(p)i* ●- - -● *(b)a* > *(p)a* O——O *(g)a* > *(k)a*



b - c. Speakers with gradual vowel start in *a*. (LG and BM have spoken two different materials.)

●——● *(b)i* > *(p)i* O——O *(g)i* > *(k)i*

●— —● *(b)a* > *(p)a* (crit.a) ●- - - -● *(b)a* > *(p)a* (crit.c)

O— —O *(g)a* > *(k)a* (crit.a) O- - - -O *(g)a* > *(k)a* (crit.c)

Figure 4

Differences of vowel duration after *pk* and *bg*.

Table 3

Differences of vowel duration (in ms) after *p* and *f*. Speakers with gradual vowel start in *pa*, delimitation according to different criteria compared to the differences for the vowel *i*

Speaker	Material	N	*(f)i > (p)i*	*(f)a > (p)a* Crit. (a)	Crit. (b)	Crit. (c)
BM	B	10	-3	-14		13
PA	A	5-15	4	-21	(-21)	13
HU	B	6	3	-25		6
HU	A	10-11	11	-17	0	29
LG	A	10-13	12	-18		24
LG	B	10	18	0		21
FJ	A	10-11	18	-22	-1	13
MF	A	10	22	-10	7	23
MF	B	10	29	14		36
JJ	B	10	-7	4	(4)	25
Grand means			10.7	-10.9	2.0	20.3

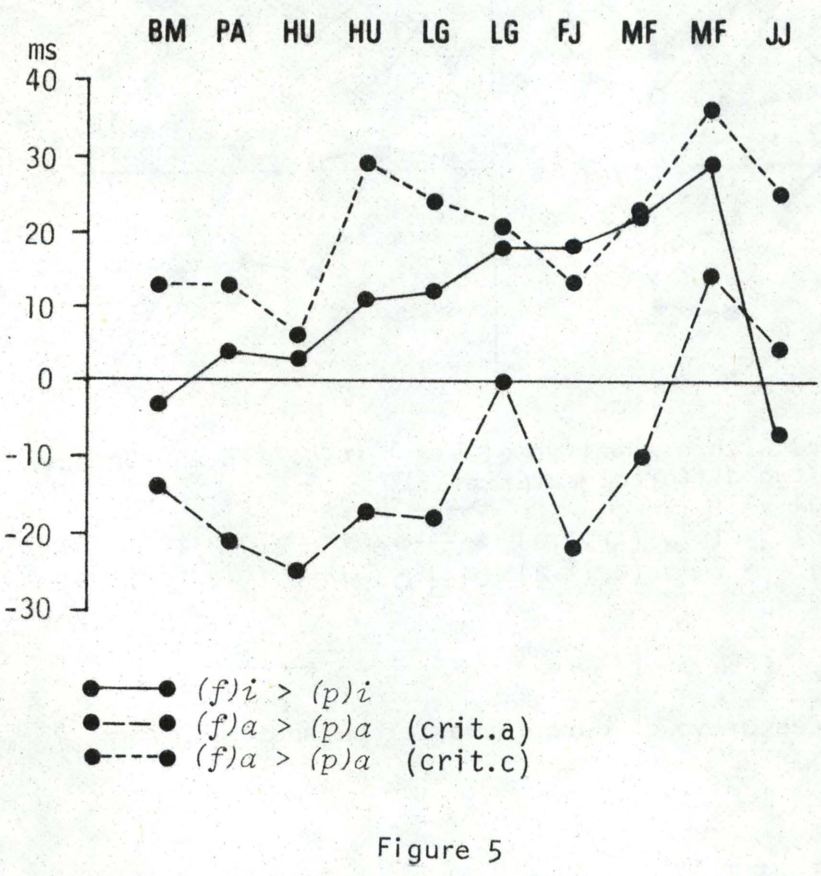


Figure 5

Differences of vowel duration after *f* and *p*. (LG, MF, and HU have spoken two different materials.)

Since the measurements of material B, C and D were made on the basis of mingograms (with a few spectrograms for control), it was not possible to distinguish the start of F1. However, in the case of PA and JJ, spectrograms were taken of all examples. They showed, as mentioned above, that for these two subjects F1 starts at once, i.e. point (a) and (b) coincide. For PA it is evident that (c) is preferable to (b) as defining the start of the vowel, whereas for JJ this is only obvious for *g-k*. Subject BH, who was left out because her higher formants were so weak that point (c) could not be identified with certainty, had a large difference between (a) and (b), and the differences according to (b) were in agreement with her differences for the vowel *i*.

The results concerning the relations between vowels after *ptk* and *bdg* can be supported by a comparison of vowel durations after *p* and *f*. High vowels are longer after *f* than after *p*, although the differences are not quite as regular as in the case of vowels after *ptk* and *bdg*. There are no examples for the speakers with abrupt vowel start. Table 3 and figure 5 show that for the speakers with gradual vowel start the differences between the durations of *a* after *f* and *p* is in better agreement with the differences for *i* when the measurements are made according to criterion (c) than when they are made according to criterion (a), the grand mean for *i* being 10.7 ms and for *a* 20.3 ms according to criterion (c) but -10.9 ms according to criterion (a) (thus giving a shorter vowel after *f* than after *p*). This is also valid for all individual speakers except BM and JJ. It has been possible to distinguish points (a), (b), and (c) for subjects HU, MF and FJ. Point (c) is obviously preferable when the values are compared to those for the vowel *i*. For PA (a) and (b) coincide, and the measurements have been listed both under criterion (a) and (b). In his case (c) is also in better agreement with the measurements for *i*.

2. DIFFERENCES OF DURATIONS BETWEEN THE VOWELS *a* AND *i*

It is a universal finding that the vowel *a* is longer than *i*. This is also true of the present material, and the difference is so large that it exists irrespective of the criteria of delimitation chosen. But it is possible to compare the magnitude of the difference according to different criteria with the magnitude of the difference in the positions where the delimitation does not give any problems. Table 4 shows that the difference between *a* and *i* is around 50 ms in these latter cases, i.e. for speakers who have abrupt vowel start in *a* (table 4a, grand mean 50.7 ms), similarly in vowels after *g* and *s* for all speakers (table 4b, grand mean 59.2 and 51.4 ms after *g* and *s*, respectively), and finally after *b d g f s* for PA (table 4b, grand mean 48.0 ms). After the consonant *k* criterion (c) gives approximately the same result (55.8 ms, for PA 56.0 ms), but with a relatively large range of variation, whereas criteria (a) and (b) give considerably higher values (86.7 and 78.7 ms, respectively, and criteria (ab) for PA gives 74 ms).

Table 4

Differences between the duration of the vowels *i* and *a*

a. Speakers with abrupt vowel start in both vowels

Speaker	Material	N	*a* > *i* after *p*	after *t*	after *k*
BF	B	2			45
NK	B	12	46	48	40
BL	B	12	55	54	47
KS	B	12	67	59	46
Grand mean				50.7	

b. Speakers with gradual vowel start in *ka*

Speaker	Material	N	*a* > *i* after *g*	after *s*	after *k* Crit. (a)	Crit. (b)	Crit. (c)
HU	A	10-11		47	96	89	61
MF	A	10-12		52	77	58	48
FJ	A	10-11		57	106	89	78
LG	A	10		65	112		63
HU	B	6	82	39	113		66
MF	B	10	62	34	85		73
LG	B	10	62	50	63		41
BM	B	10	43	52	59		33
JJ	B	10	47	67	69	(69)	39
Grand means			59.2	51.4	86.7	78.7	55.8

				a > *i* Crit. (ab)	Crit. (c)
PA	A	5-15	after *b:* 43 *d:* 46 *g:* 32 *f:* 65 *s:* 54	after *p:* 89 *t:* 73 *k:* 60	after *p:* 59 *t:* 70 *k:* 39
Grand means			48.0	74.0	56.0

This is also valid for the individual speakers, except for BM and LG in material B (but in material A, LG shows the same relations as the other subjects). The results for JJ are ambiguous. On the whole, this relation thus supports the choice of criterion (c).

3. DIFFERENCES IN THE DURATION OF THE OPEN INTERVAL OF THE STOPS BEFORE *i* AND *a*

There is a universal tendency for the open interval after stop consonants to be longer before high vowels than before low vowels. This may be explained by the slower escape of air after the release because of the narrower constriction (but in a low back [ɑ] the constriction in the pharynx may be expected to have a similar effect). The consonant *t* (which is not included here) always has a longer open interval before *i* than before *a* [æ], whereas for *p* and *k* the open interval is only consistently longer before *i* if the start of voicing (point (a)) is taken as the start of the vowel. However, as shown in table 5 and figure 6, the differences obtained by choosing this point (30.4 ms for *p* and 30.5 ms for *k*) are, on the average, considerably higher than the difference found for those speakers who have abrupt vowel start (9.0 ms for *p* and 12.2 ms for *k*). Point (b), too, gives higher values (25.3 and 22.0 ms). On the other hand, point (c) results in a deviation in the opposite direction so that there is, on the average, practically no difference before *i* and *a*. As shown clearly in figure 6, point (c) gives negative values (i.e., the aspiration is shorter before *i* than before *a*) for half the subjects (here point (a) is better) and positive values for the other half (here point (a) gives too large differences compared to the subjects with abrupt vowel start).

The question is now how consistent one should expect these relations to be in the case of aspirated *p* and *k* before *i* and *a*. If the tendency is due to the slower escape of air after the release because of a more constricted vocal tract, one should expect the differences to be particularly clear in cases of short open intervals, e.g. for Danish voiceless *bdg* and French unaspirated *ptk*, and this is indeed the case (see, e.g., Fischer-Jørgensen 1972b and 1980). In Danish *bdg* and, according to various investigations, e.g. Kagaya and Hirose (1975), in Indian unaspirated *ptk*, the glottis is almost closed at the release and ready for voicing, and the duration of the open interval is thus exclusively dependent on the time it takes for the pressure to decrease in the oral cavity. Vowel start is equal to voicing start. But in aspirated stops the constriction evidently influences the start of voicing (cf. the large differences between the open intervals before *a* and *i* when the delimitation is made at voicing start (point (a)). But the intraoral pressure may have decreased to zero when the adduction movement of the vocal cords is finished and all formants appear, so that one should not necessarily expect a difference at point (c). In Fischer-Jørgensen 1980 a consistent difference was found for *ki*-*ka*, but in some cases the delimitation

Table 5

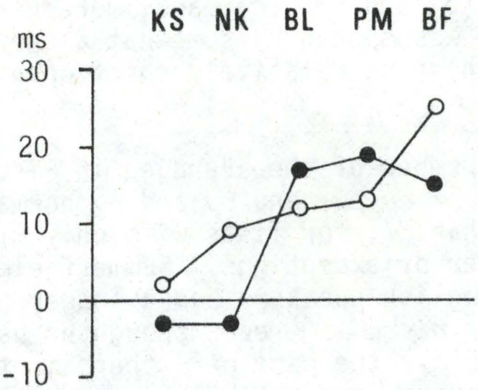
Differences between the open intervals of *p* and *k* before *a* and *i* (in ms)

a. Speakers with abrupt vowel start

Speaker	Material	N	*(p)h(i)>* *(p)h(a)*	*(k)h(i)>* *(k)h(a)*
KS	B	12	-3	2
NK	B	12	-3	9
BL	B	12	17	12
PM	B	10	19	13
BF	B	2	15	25
Grand means			9.0	12.2

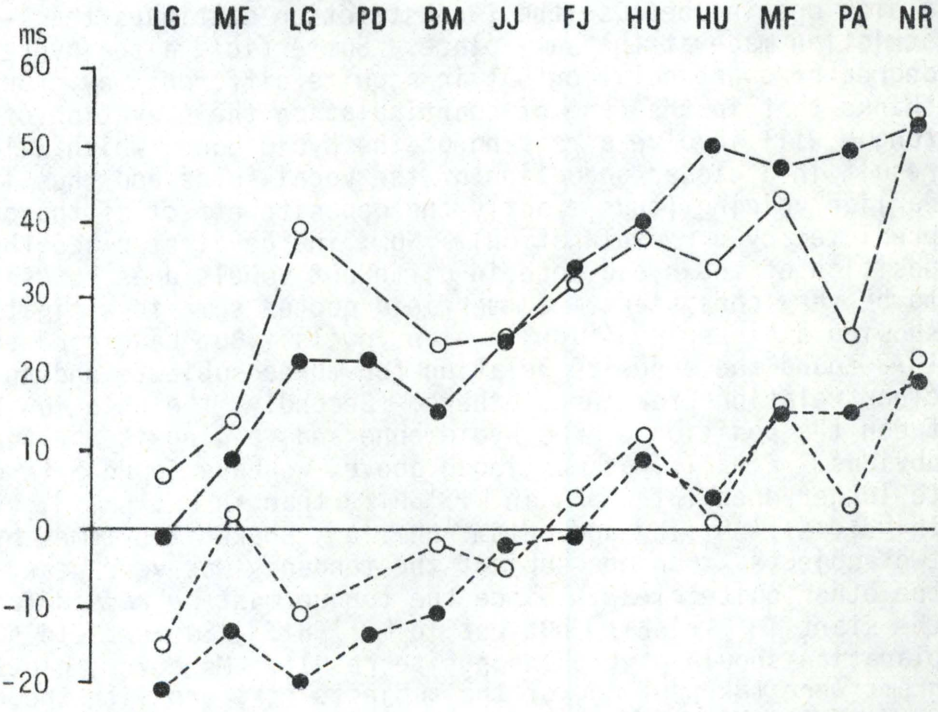
b. Speakers with gradual vowel start

			(p)h(i)> *(p)h(a)*			*(k)h(i)>* *(k)h(a)*		
Speaker	Material	N	Crit. (a)	Crit. (b)	Crit. (c)	Crit. (a)	Crit. (b)	Crit. (c)
LG	B	10	-1		-21	7		-15
MF	B	10	9		-13	14		2
LG	A	10-13	22		-20	39		-11
PD	E	10	22		-14			
BM	B	10	15		-11	24		-2
JJ	B	10	25	(25)	-2	25	(25)	-5
FJ	A	10	34	13	-1	32	15	4
HU	B	6	40		9	38		12
HU	A	10-11	50	33	4	34	27	1
MF	A	10	47	30	15	43	24	16
PA	A	5-15	49	(49)	15	25	(25)	3
NR	D	9	53		19	54		22
Grand means			30.4	25.3	-1.6	30.5	22.0	2.5



a. Speakers with abrupt vowel start.

●——● *(p)h(i) > (p)h(a)* ○——○ *(k)h(i) > (k)h(a)*



b. Speakers with gradual vowel start - according to criteria (a) and (c). (LG, MF, and HU have spoken two different materials.)

●— — ● *(p)h(i) > (p)h(a)* (crit.a)
○— — ○ *(k)h(i) > (k)h(a)* (crit.a)
●- - - -● *(p)h(i) > (p)h(a)* (crit.c)
○- - - -○ *(k)h(i) > (k)h(a)* (crit.c)

Figure 6

Differences between the open intervals of *p* and *k* before *i* and *a*.

may have been made at point (b) and not (c), and, moreover, a large part of the material was spoken by somewhat older subjects or non-Copenhageners who have relatively short open intervals in aspirated stops.

As far as *p* is concerned, a number of the averages in Fischer-Jørgensen 1980 were based on *pi py pu*, and *py* and *pu* normally have longer open intervals than *pi*; for lists with only *pi* (e.g. M2), there were a number of exceptions. Summerfield 1975 found that 5 out of 6 English speakers had a longer open interval in *pa* [pʰɑː] than in *pi* [pʰiː], even though he used voicing start as a criterion. In the case of *p* there is the special condition that the tongue need not be raised completely at the release (particularly not in the case of a diphthongized British [ɩi] and, on the other hand, a British [ɑː] may have a very narrow constriction in the pharynx). The open interval in *pi* (and also in [pʰɑː] and [kʰɑː]) may thus depend partly on the degree of coarticulation, whereas *ki* always has a slow opening because the *i*-constriction continues the *k*-constriction made at the same place. Summerfield also invokes the degree of coarticulation but in a quite different way. He thinks that in the case of coarticulation the elevation of the tongue will involve a raising of the hyoid bone, which will result in a closer adduction of the vocal folds and thus in earlier voicing (thus exactly the opposite effect of the one predicted by our explanation). Now, in the first place the position of the hyoid bone in different vowels does not seem to be very consistent. Summerfield quotes some investigations showing a higher position in high vowels. But Ladefoged et al. 1972 found the opposite relation for three subjects and no clear relations for three others. Secondly, the relation between the position of the hyoid bone and voicing is not very obvious. Finally, as mentioned above, we have found a tendency to longer open intervals in [iːˈphiː] than in [æːˈphiː], viz. in [divisiːˈphiːlə] and [disæːˈphiːlə], spoken ten times by two subjects. For one subject the tendency was very weak, for the other quite clear. Since the tongue must be raised from the start in [iːˈphiː] but not in [æːˈphiː], Summerfield's explanation should give the opposite result. Moreover, glottograms were taken of one of the subjects (the one with the weak tendency); they showed that generally voicing started earlier during the adduction of the vocal folds after *a* than after *i*. This has strengthened our belief that the main factor in the delay of voicing is the constriction in the vocal tract. This is also supported by the fact that Danish *pu* and *py*, in which the lip opening remains small in the transition from *p* to the vowel, have a longer open interval than *pi*. And it is further corroborated by inspection of intraoral pressure curves. Generally, voicing does not start until the pressure curve has reached zero level. In *pa* this may happen almost instantaneously, in *pi* the decay is slower, but variable, and in *pu* it often takes up to 70-80 ms, which means that the slow decay may really delay voicing in aspirated stops in some surroundings, and even sometimes vowel start according to criterion (c), which presupposes both start of voicing and start of higher formants. But as it need not do so in *pi*, nor, in the case of long aspirations in *ki*, we do not think that the lack of difference between the open intervals in *pi-pa* and *ki-ka* can be used as an argument against the choice of criterion (c).

4. DIFFERENCES IN INTRINSIC Fo

It should finally be mentioned briefly that the choice of delimitation criterion may also influence the relations of intrinsic Fo. Generally, the Fo curve starts at a higher frequency value after Danish *ptk* than after *bdg*, but since the curve rises from (a) to (c) in *pa-*, *ta-*, *ka*-syllables, the difference will be larger in point (c) than in point (a). Jeel 1975 used point (c), and for one of her subjects the relations got reversed if point (a) was chosen.

IV. CONCLUSION

The purpose of the present paper was in the first place to draw attention to a problem of delimitation which is not of a negligible magnitude (up to 45 ms for averages, and up to 60 ms in individual cases). All three criteria ((a) start of periodicity, (b) start of F1, and (c) start of higher formants) have been used by different phoneticians, and we should like to emphasize that the so-called "weak edge vibrations" may show up with quite considerable amplitude in oscillograms so that point (a) may seem to be a natural point of delimitation both for vowel start and (still more obviously) for VOT, when oscillograms are used as a basis. We have further shown that different points of delimitation result in rather different temporal relations, and we have argued for using point (c). It seems quite evident to us that point (c) is preferable to point (a), (1) because it is physiologically better determined as coinciding approximately with the closure of the glottis, (2) because the low frequency vibrations starting in point (a) are perceptually weak and are not heard as part of the vowel, and (3) because point (c) gives much more regular temporal relations. The evidence for preferring point (c) (start of higher formants) to point (b) (start of F1) does not carry the same weight because it is based on a much more restricted material, but in almost all cases point (b) gives less regular temporal relations than point (c) and, moreover, it is rather difficult to locate on oscillograms.

REFERENCES

Andersen, P. 1981: "The effect of increased speaking rate and (intended) loudness on the glottal behaviour in stop consonant production as exemplified by Danish *p*", *Ann. Rep. Inst. Phon. Univ. Cph.*, this issue

Darwin, C.J. and Pearson, M. 1980: *What tells us that voicing has onset?*, unpublished paper given to the Institute of Acoustics Speech Group, York, December 1980

Fant, G. 1969: "Stops in CV-syllables", *Speech Transm. Lab., Quart. Prog. and Status Rep., Royal Inst. Techn., Stockholm 4/1969*, and *Speech Sounds and Features, 1973*, p. 110-139

Fischer-Jørgensen, E. 1972a: "Perceptual studies of Danish stop consonants. Tape cutting experiments with Danish stop consonants in initial position", *Ann. Rep. Inst. Phon. Univ. Cph. 6*, p. 104-168

Fischer-Jørgensen, E. 1972b: "PTK et BDG français en position intervocalique accentuée", *Papers in Linguistics and Phonetics for the Memory of Pierre Delattre*, (The Hague), p. 143-200

Fischer-Jørgensen, E. 1979: "Temporal relations in consonant-vowel syllables with stop consonants based on Danish material", in *Frontiers of Speech Communication Research* (Lindblom, B. and Öhman, S. eds.), (London)

Fischer-Jørgensen, E. 1980: "Temporal relations in Danish tautosyllabic CV sequences with stop consonants", *Ann. Rep. Inst. Phon. Univ. Cph. 14*, p. 207-262 (revised and enlarged version of 1979)

Hutters, B. 1979: *Glottisfunktionen ved produktion af ustemte obstruenter i dansk, en glottografisk og fiberoptisk undersøgelse*, unpublished report (Institute of Phonetics, University of Copenhagen)

Jeel, V. 1975: "An investigation of the fundamental frequency of vowels after various consonants, in particular stop consonants", *Ann. Rep. Inst. Phon. Univ. Cph. 9*, p. 191-212

Kagaya, R.A. and Hirose, H. 1975: "Fiberoptic, electromyographic, and acoustic analyses of Hindi stop consonants", *Annual Bulletin, University of Tokyo, Logopedics and Phoniatrics 9*, p. 27-46

Karlsson, I. and Nord, L. 1970: "A new method of recording occlusion applied to the study of Swedish stops", *Speech Transm. Lab., Quart. Prog. and Status Rep., Royal Inst. Techn., Stockholm 2-3/1970*, p. 8-18

Kent, R.D. and Moll, K.L. 1969: "Vocal tract characteristics of stop cognates", *J. Acoust. Soc. Am. 46*, p. 1549-1555

Klatt, D.H. 1975: "Voice onset time, frication and aspiration in word initial consonant clusters", *J. Speech and Hearing Research 18*, p. 686-706

Ladefoged, P., DeClerk, J., Lindau, M., and Papçun, G. 1972: "An auditory-motor theory of speech production", *Working Papers in Phonetics, UCLA 22*, p. 48-75

Leanderson, R. and Lindblom, B.E.F. 1972: "Muscle activation for labial speech gestures", *Acta Otolaryngologica 73*, p. 362-373

Lehiste, I. and Peterson, G. 1960: "Studies of syllable nuclei", *Speech Research Laboratory Report 4*, (Ann Arbor)

Lindblom, B. and Rapp, K. 1972: "Reexamining the compensatory adjustment of vowel duration in Swedish words", *Speech Transm. Lab., Quart. Prog. and Status Rep., Royal Inst. Techn., Stockholm 4/1971*, p. 19-25

Lindqvist, J. 1972: "Laryngeal articulations studied on Swedish subjects", *Speech Transm. Lab., Quart. Prog. and Status Rep., Royal Inst. Techn., Stockholm 2-3/1972*, p. 10-27

Lisker, L. and Abramson, A.S. 1964: "A cross-language study of voicing in initial stops: acoustical measurements", *Word 20*, p. 384-422

Lisker, L. and Abramson, A.S. 1967: "Some effects of context on voice onset time in English stops", *Language and Speech 10*, p. 1-29

Löfqvist, A. 1975a: "On intrinsic and extrinsic Fo variations in Swedish tonal accents", *Phonetica 31*, p. 228-247

Löfqvist, A. 1975b: "A study of subglottal pressure during the production of Swedish stops", *J. Phonetics 3*, p. 175-189

Summerfield, Q. 1975: "Aerodynamics versus mechanics in the control of voicing onset in consonant-vowel syllables", *Speech Perception, Series 2, Number 4*, (Belfast), p. 61-72

Uldall, E. 1957: *Vocal cord action in speech: a high-speed study*, (film), University of Edinburgh and Swiss Federal Institute of Technology

Weismer, G. 1979: "Sensitivity of voice-onset-time (VOT) measures to certain segmental features in speech production", *J. Phonetics 7*, p. 197-204

Zue, V.W. 1976: "Acoustic characteristics of stop consonants: a controlled study", *Technical Report 523*, Lincoln Laboratory, Massachusetts Institute of Technology

(ARIPUC 15, 1981 p. 103-146)

THE EFFECT OF INCREASED SPEAKING RATE AND (INTENDED) LOUDNESS ON THE GLOTTAL BEHAVIOUR IN STOP CONSONANT PRODUCTION AS EXEMPLIFIED BY DANISH P.

PREBEN ANDERSEN

The present combined fiberoptic-glottographic experiment has the purpose of studying the glottal behaviour in Danish stop consonant production at increased speaking rate and intended loudness, respectively. At increased speaking rate the degree and duration of the glottal abduction is found to decrease proportionally, i.e. the rate of the glottal abduction is constant, irrespective of speaking rate. Different individual strategies seem to be applied in the temporal organization of the glottal adduction. At increased intended loudness level no effect on the temporal organization of the gesture is observed, but the rate of glottal abduction is found to be faster. A model is proposed for the glottal behaviour under each of the two speech conditions. A very limited EMG-material (1 subject) shows changes in INT-activity under both conditions, whereas changes in PCA-activity occur only in connection with increased intended loudness level. The present findings concerning the speaking rate condition are compared to results from research on supraglottal articulatory organization at varying speaking rates.

I. INTRODUCTION

The present study originated from observations made in a fiberoptic-glottographic pilot study. Glottograms of repetitive productions of stops in conversational speech displayed a considerable variation in maximum amplitude, and supplementary acoustic curves displayed variation also in the duration as well as in the intensity of the open interval of the stops.

The present investigation can be considered a preliminary contribution to the determination of a potential systematic relationship between degree of maximum glottal abduction and (1) intensity and (2) speaking rate (as reflected by the duration) of stops in conversational speech. Its specific purpose is to clarify the effect of deliberate variations of intensity level and speaking rate on the glottal behaviour of Danish stops. The application of the results, presented here, to the normal, conversational speech condition will be the subject of a later paper.

This paper has another, perhaps more important purpose, namely to contribute to the debate on timing in speech production. It is commonly known that allophonic variation due to the suprasegmental features of stress and speaking rate among other things includes durational differences. The majority of research in this area involves an attitude to a model, proposed by Björn Lindblom in 1963 (Lindblom 1963). This model assumes that the allophonic variation in destressed vowels or vowels produced with faster speaking rate is merely a consequence of the shorter duration and can be accounted for as a pure timing phenomenon. For vowels subject to variation in stress and speaking rate, Lindblom found a high correlation between the duration of the vowel and an expression of its degree of spectral reduction, i.e. the shorter the vowel, the closer its formant pattern to that of schwa. Lindblom concludes that this neutralization is a consequence of articulatory and acoustic undershoot, caused by a temporal overlap of the muscle commands.

The essential question raised in connection with Lindblom's model has been, whether a model of destressed speech and speaking rate control can be based on timing changes alone. Later experiments (Harris 1971, 1978, Kent and Netsell 1971, Gay 1978) have given evidence that production of vowels under varying stress conditions seems to involve a change in the intensity of the muscle commands as well and, at the peripheral level, a change in the rate of articulator displacement. At the muscular level, Sussman and MacNeilage (1978) and Sussman (1979) have observed an increase in the number of active motor units during the production of stressed syllables and, at the individual motor unit level, a higher instantaneous discharge rate. A conclusion from the results of recent research has been (Harris 1978), that control of vowels under different stress conditions seems to be explicable in terms of an "extra energy model" (Öhman 1967) rather than of Lindblom's "undershoot model".

For vowels of different duration caused by varying speaking rate, the mechanism of the underlying control seems less evident. Gay et al. (1974) observe for vowels produced at faster speaking rate an undershoot in tongue displacement, a decrease in tongue EMG-activity, but no concurrent decrease in articulator velocity. They conclude that although an articulatory undershoot is stated, one cannot fit in the findings with Lindblom's model, as this presupposes the undershoot to be a mere consequence of temporal overlap of muscle commands with no changes in driving force. As regards velocity of articulator displacement, Kuehn (quoted from Gay et al. 1974) has shown that different speakers use different strategies to increase speaking rate, e.g. by trade-offs in displacement vs. velocity. Harris (1978) found longer and greater EMG-activity for vowels in stressed position (see above) but no consistent effects on EMG-activity from speaking rate. In a spectrographic study Gay (1978) found variations of vowel-target frequency as a function of stress, but not as a function of rate. He concluded that control of stress and speaking rate implies different mechanisms.

As Lindblom's model accounts for vowels only, an essential question is whether the complex of problems connected with the model can be applied to consonant production as well, i.e. does consonant production, when influenced by changes in stress and speaking rate, imply articulatory reorganization comparable with that observed for vowels? The effect of speaking rate on labial consonant production has been studied by Gay et al. (1973, 1974). Besides differences in the timing of labial muscular events they found an increase in labial muscle activity and in rate of labial movement by increased speaking rate. Accordingly, their conclusion was that a general model of speaking rate control, based solely on timing changes, is too simple. The concept of undershoot was reconsidered, as vowel production involves articulator movements towards spatial targets, whereas consonant production most often implies movements towards occlusive or constrictive targets. Consequently, consonant production is supposed not to allow undershoot, and the authors find it reasonable to assume that consonant production, when constrained by increased speaking rate, implies faster movement of the articulators towards the target.

As for consonant production under different stress conditions, Harris et al. (1968) and data from other investigations (Lubker and Parris 1970) suggest no consistency as concerns differences in driving force, although consonants in pre- and post-stressed position have different durations.

From the above it is evident that research on mechanisms of stress and rate changes has been primarily concerned with the control of the supraglottal articulation. The second purpose of the present paper is to contribute further to the complex of problems connected with articulatory organization under different stress and tempo conditions by shedding light also on the glottal articulation. The study is concerned specifically with the glottal behaviour during stop consonant production under different speech tempo conditions. The essential question raised in connection with the experiment is, if an increase of speaking rate also implies a reorganization of the glottal gesture in terms of faster rate of ab- and adduction and hence, underlyingly, a change in the timing and intensity of the commands to the laryngeal muscles? This is part of a more general question: Are supraglottal and glottal articulations during consonant production under different speaking rate conditions controlled by one and the same underlying mechanism? An alternative would be that supraglottal and glottal articulations in this respect are independently controlled and work as two different systems. The latter possibility seems plausible if the concept of target undershoot is taken into consideration. Whereas the supraglottal articulation implies articulator movement towards an occlusive target, the glottal articulation might be perceived as a movement towards a far less constrained spatial target, the nature of which would be expected to allow articulatory undershoot in contradistinction to that of the supraglottal target.

Experimentally the problems are approached here through extensive fiberoptic-glottographic recordings in order to account for the dynamic properties of the glottal gesture, supplemented by a very limited EMG-material, supposed to yield some information about the mechanisms causing the dynamic variation.

II. METHOD

A. SUBJECTS AND LINGUISTIC MATERIAL

The speakers were two adults, a female (subject HU) and a male (subject JR).

The linguistic material for the investigation was limited to the Danish stop *p* in covered initial position. The test sound was embedded in the phrase *de vil sige pige* [d̥i ve si: ˈb̥ʰi:] 'they will say (the word): girl'. The articulatory course of the phrase was supposed to minimize alterations in the position of the fiberoptic cable.

The material to be recorded consisted of repetitions of the test phrase under the following two speech conditions:
a) in speech where the subject, repetition by repetition, in subjectively determined steps, was told to gradually speak more loudly, starting the series with the phrase being produced with weak "loudness". (The change in the subject's performance as a consequence of this instruction is here referred to as a change in "loudness".) Each subject spoke 9 such series (in the following referred to as '"loudness" series').
b) in speech where the subject, following the same principle as described above in a), produced series of repetitions in which the speaking rate of the phrase was gradually increased. Subject HU spoke 9 such series, subject JR 8 series (in the following referred to as 'rate series').

B. THE INSTRUMENTAL SET-UP

The fiberscope (Olympos VF, type 4A) and the photo-electric glottograph (a slightly reconstructed version of the one described by Frøkjær-Jensen 1967) are the central components of the instrumental set-up shown in figure 1. The light guide of

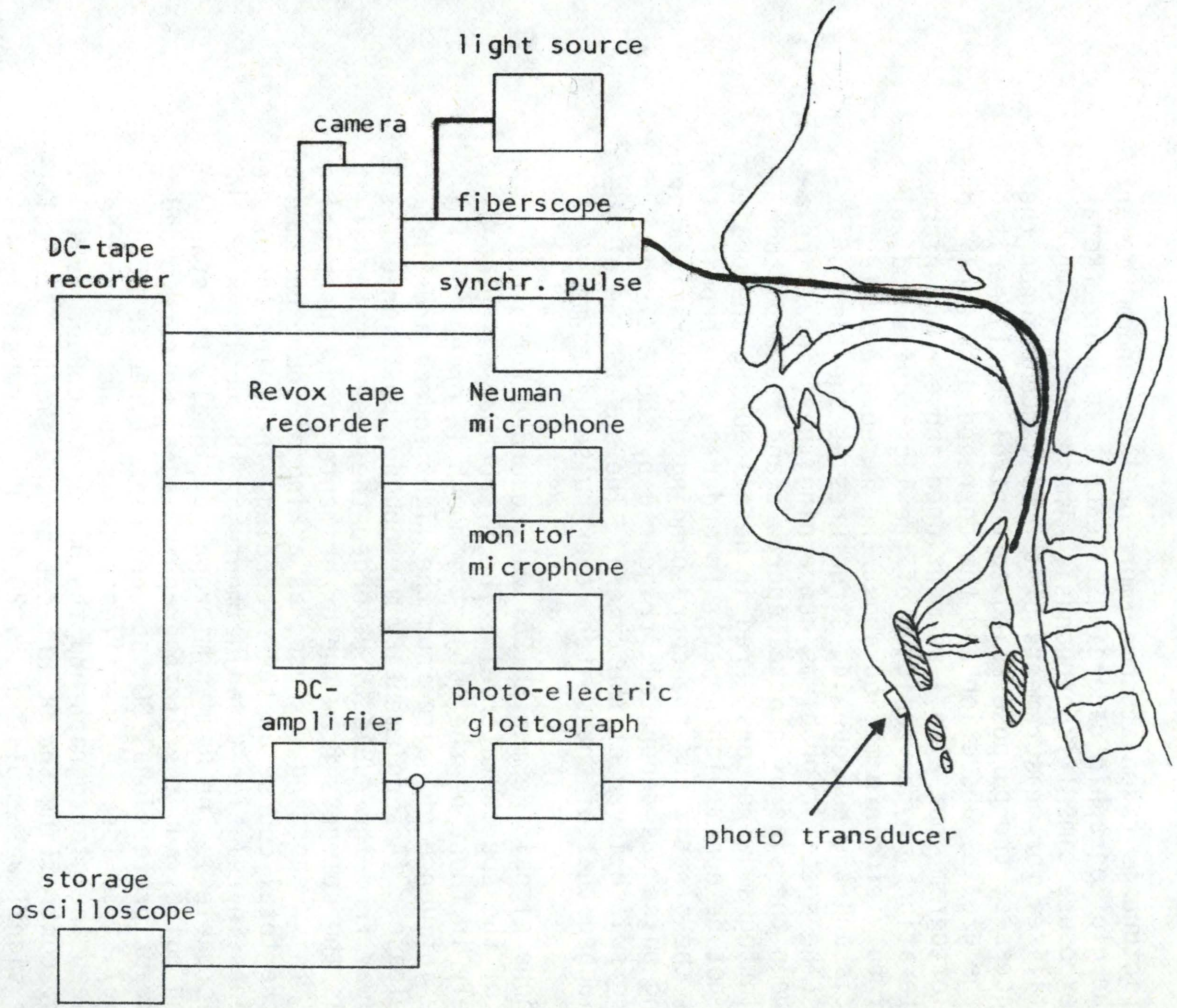


Figure 1

The instrumental set-up.

the fiberscope, placed in the subject's pharynx, serves as the light source for the photo-electric glottograph. This technique makes it easier to control the position of the light source by means of the fiberoptic pictures transmitted through the image guide. A photo transducer, placed on the frontal part of the subject's neck in a position approximately between the thyroid and cricoid cartilages, turns it into an electric voltage variation which, after amplification, can be recorded graphically.

A still-picture is taken at some (more or less random) moment during the glottographic recording of each test sound (Kodak TX 135, exposure time 1/125 seconds). These pictures were mainly utilized for measurements of the distance between the vocal processes, the purpose being to control the linearity between the glottal abduction and the corresponding amplitude of the glottogram, but also - as mentioned above - to inform about alterations in the position of the light guide and, altogether, to yield information about the immediate state of the glottis and the surrounding structures at the time of exposure. (The specific problems concerning the (linear) relation between the size of glottal aperture and the amplitude of the glottogram were considered in detail during the project, but will not be dealt with here.) In order to synchronize the stills of the glottis and the corresponding glottograms, a synchronizing pulse was generated, triggered by the M-X synchronizing switch lever of the camera. The onset time of the shutter lagged that of the switch lever pulse by 15 ms.

A microphone signal (Neuman KH 56) was recorded for acoustic analysis of the speech material. The microphone was placed immediately in front of the subject's mouth in order to eliminate the influence of noise from the light source and the camera release. (A supplementary microphone, positioned farther away from the subject, recorded the speech material as well as the comments from the experimenter and others being present. This recording was used as a kind of sound documentation of the total course of the recording, including the pauses and comments in relation to adjustments of the position of the fiberoptic cable.) The Neuman microphone signal, the synchronizing pulse, and the glottographic signal were recorded on an 8-track professional DC tape recorder (Lyrec TR 86, Agfa PE 36) at 30 ips. An oscilloscope was connected to the line between the glottograph and the DC tape recorder as a device for controlling the DC off-set of the signal. The glottographic signal was amplified by 20 dB (PT Universal Amplifier) for both subjects.

C. GRAPHICAL REGISTRATION OF THE RECORDED MATERIAL

The speech material was registered graphically on the mingograph (paper speed: 100 mm/sec.). From the Neuman microphone recording the following curves were extracted for further analysis: Duplex oscillogram (BFJ Transpitchmeter); 2 intensity curves, one logarithmic HP-filtered at 500 cps with an integration time of 2.5 ms, and one linear Hi-Fi with an integration time of 10 ms (BFJ Intensitymeter). Further, the synchronizing pulse and glottograms were registered.

D. CHOICE OF PARAMETERS REPRESENTING "LOUDNESS" AND SPEAKING RATE, RESPECTIVELY, IN THE PRODUCTION OF THE SPEECH MATERIAL

The goal of the present experiment was to test the correlation between variations in glottal parameters (spatial as well as temporal ones) and variations either in speaking rate or in "loudness", whereby one might get an indication of the presence or absence of systematic changes in the glottal behaviour under the two speech conditions. This procedure, however, implies the existence of a parameter representing the speaking rate and "loudness", the values of which can be correlated with the data of the glottal parameters. Such a parameter was established in the shape of a sequence of natural numbers functioning as a rank scale assigned to the rate and "loudness" series, respectively. Each of the subjectively determined steps in the single rate or "loudness" series was ranked according to the intention of the subject in such a way that the first step in a series (i.e. the token with the intended slowest speaking rate or intended lowest "loudness" level) was given rank 1, the second step was given rank 2, and so forth throughout the series.

The correspondence between the subjective ranking of the steps in a series and the physical facts was to be tested by correlating the established 'intention' parameter, consisting of ranks, with physical parameters consisting of intensity and duration measures. To test the relation between the intention of the subject and the physical facts of the "loudness" series, the intensity level of the first 40 ms after the release of the closure of the test sound *p* was used. (This intensity level was measured with the level of the most forceful vowel in the material as reference.) To test the corresponding relations as regards the rate series, the acoustic duration of the stop (closure + open interval) was measured. The intensity and duration data were converted to ranks, and for each subject the Spearman's coefficient of rank correlation between the 'intention' parameter and each of the physical parameters was computed. The results of the computation is shown in figure 2. As regards the "loudness" series, the rank correlation coefficient for the 'intention' parameter and the average intensity level of the vowel *i* following *p* in the the test phrase is included in the figure in order to give an impression of the intensity variations not only in the test consonant. It is evident that there is - within each series - a highly significant, positive correlation between the intentions of the subjects and the physical facts. As the aim of the present experiment was to state, for the test sound, the relations between its "loudness" and rate variations, respectively, and its glottal behaviour, not only in terms of correlation coefficients but, as far as possible, also in terms of physical,

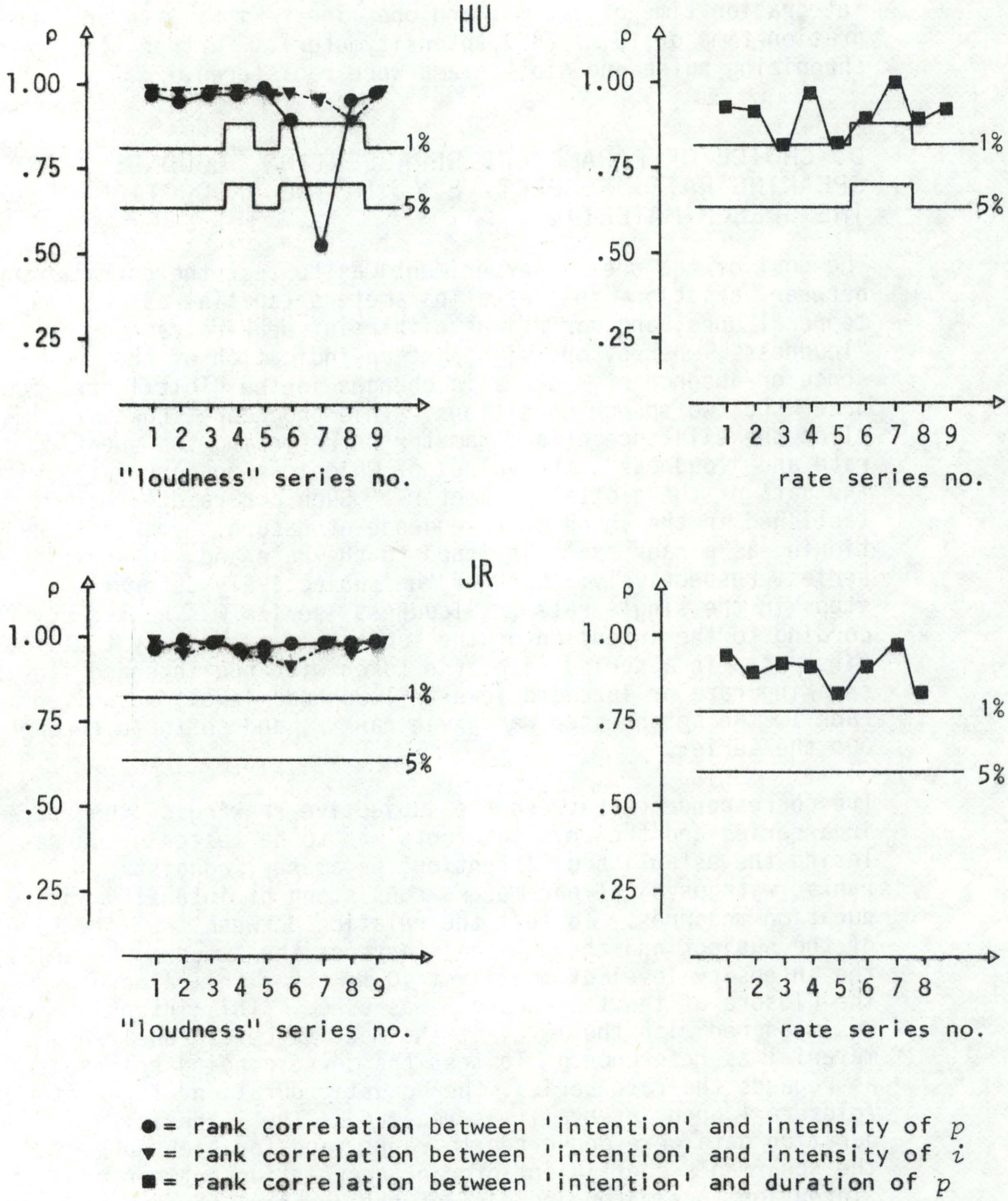


Figure 2

Spearman's coefficient of rank correlation (ρ) for the 'intention' parameter and (a) the intensity of *p* (intensity level of the first 40 ms. after the release of the closure) ,(b) the intensity level of the *i* following *p*,(c) the duration of *p* (closure + open interval). The first two coefficients are calculated for each of the "loudness" series, and the third one for each of the rate series. Two levels of significance are included.

numerical relations, the two physical parameters were chosen to represent the rate of performance of the sound and thereby the variations in respectively the speaking tempo and the variations in "loudness".

1. AVERAGE "LOUDNESS" AND RATE SERIES

Figures 3a and 3b illustrate graphically the measured intensity of the explosion and aspiration of *p* through each of the 9 recorded "loudness" series and the measured segment duration of *p* through each of the 9 (subject HU) or 8 (subject JR) recorded rate series. Though there is some dispersion of the values on each step which causes a certain amount of overlapping between consecutive steps, an average increase in intensity and decrease in duration through the series is obvious for both subjects. One might suspect that the dispersion of the values on each single step is caused by systematic differences in the course of intensity increase or duration decrease between the series, for instance in the sense that one series was dislocated one step in relation to another. However, no such clear parallelism in the course of any two series is observed, and one must conclude that the dispersion of data is random for each step. This fact makes it possible to pool the series and, thus in the following, to operate with only one average "loudness" and average rate series. (The average value of each step is indicated on the figure by a small horizontal line.)

In the rate series, for subject JR each one spanning 9 steps and for subject HU only 8 steps, it can be observed that through the series subject JR shortened the segment duration by 100 ms, on the average, subject HU by only 73 ms on the average. The total reduction, however, differed for the two subjects with a value just about equal to the range between the first and second step in subject JR's average series, and as the second step in the average series of subject JR had a value corresponding to that of the first step in the average series of subject HU, one might tentatively conclude that subject JR started his rate series one step slower than did subject HU. In any case, for the sake of comparison, the first step in the material of subject JR was left out in the further analysis.

The values of the steps in the average series were used as the parameters representing "loudness" and rate, respectively, in the further analysis. In the following, the two parameters constitute the standard of reference in the testing of the glottal and partly the supraglottal behaviour of *p* under the influence of varying "loudness" and speaking rate. The parameters to be compared (i.e. correlated) with the "loudness" and tempo parameter will be described in the following section. In accordance with the "loudness" and tempo parameter, each of these spatial and temporal parameters will also be represented by only one value per step, namely the average value of the step in question for all the series recorded.

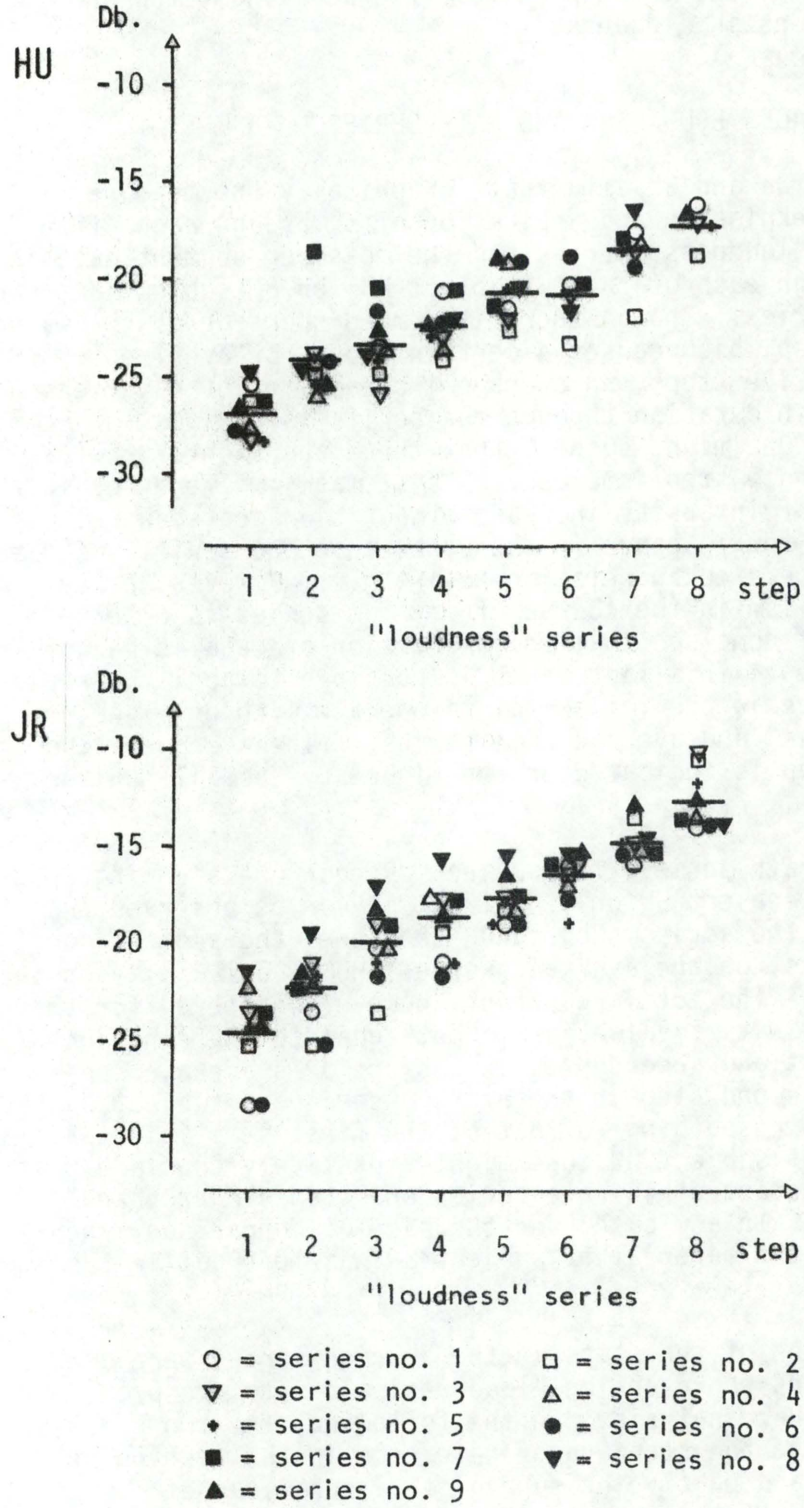


Figure 3a

The measured values of the intensity level of *p* (intensity level of the first 40 ms. after the closure release) throughout the "loudness" series.

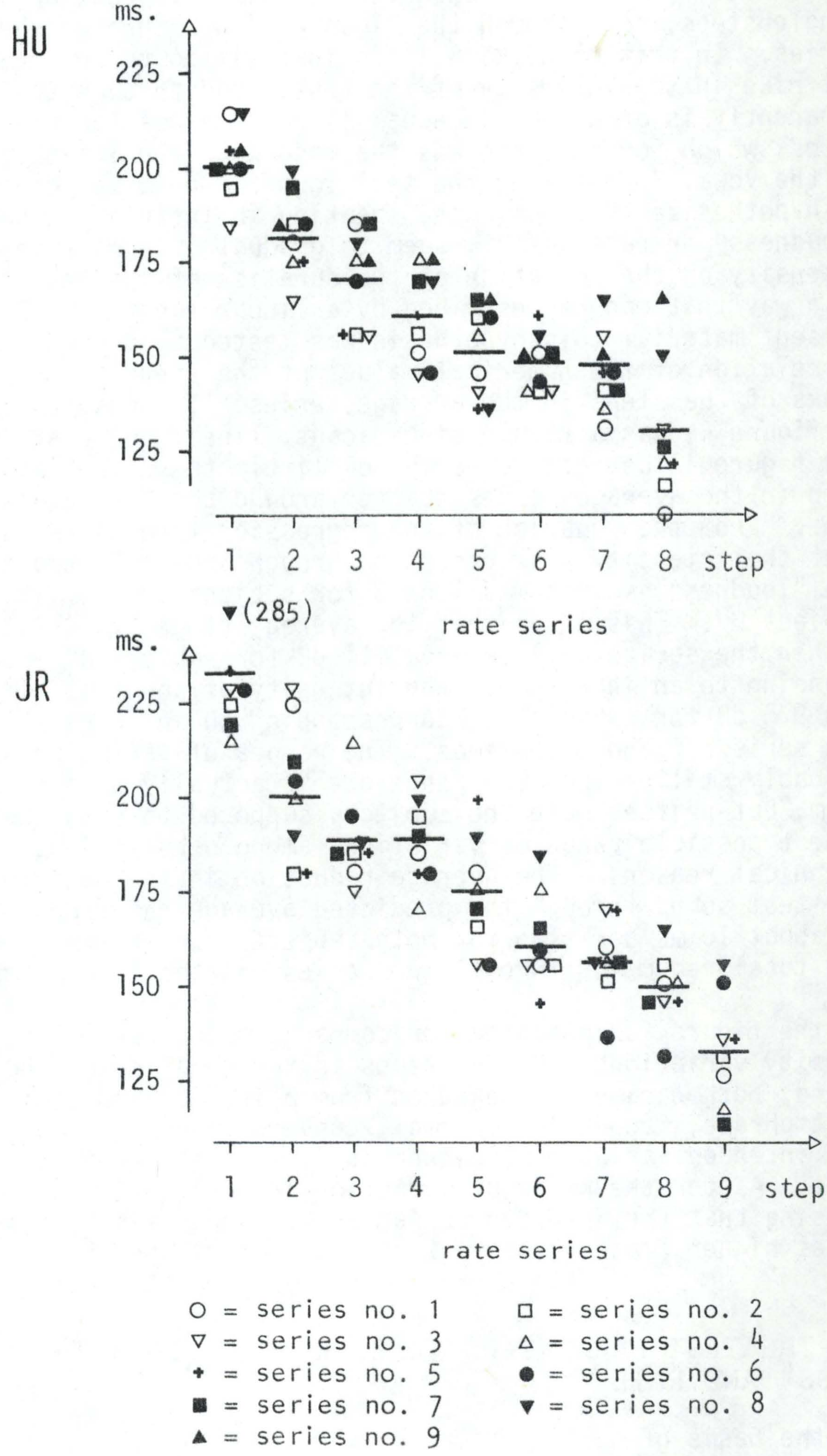


Figure 3b

The measured values of the duration of *p* (duration of closure + open interval) throughout the rate series.

2. PREDICTED AVERAGE "LOUDNESS" AND RATE SERIES

For neither subjects is there any obvious tendency towards a systematic difference in the average intervals between the single steps, not through the "loudness" nor through the rate series. In this connection it is interesting to observe that the rise in the intensity of the test sound through the series apparently is graduated in equal linear and not logarithmic steps, which for instance was the case with the intensity rise of the vowel *i* following the test sound. One might consequently hypothesize that subjects, speaking an infinite number of "loudness" or rate series, seem to graduate the rise in the intensity or the reduction of the duration of the test sound in a way that can be described by a linear function. In the present material this hypothesis was tested by product-moment correlation of the numerical values of the steps with the ranks of the steps in the average series. The result, shown in figure 4, was a highly significant, linear correlation. The figure illustrates clearly how little the values of each step in the average series scatter around the Y'-regression line. From the equation of the regression line we can derive that the intensity rise per step through the predicted average "loudness" series was 1.6 dB for subject JR and 1.3 dB for subject HU. This means that the average range of variation within the series will be about 11 dB for subject JR, corresponding to an increase in the intensity of about 13 times, and 9.5 dB for subject HU, corresponding to an increase through the series of about 10 times. The ranges of variation are probably smaller than the range one is actually able to perform, but neither were the subjects supposed to span their widest possible range of variation, among other things for technical reasons. The average reduction in the duration of the test sound through the predicted average rate series will be about 10 ms per step for both subjects, which means that the total reduction through the series was about 70 ms or 50%.

On the figure is indicated for comparison the range of intensity variations and variations in the duration of the test sound, both parameters measured from *p* in 150 samples of the test phrase, repeated in normal, conversational speech without intended variation, neither in "loudness" nor in speaking rate. As for the range of variation in duration, it seems surprising that the intended variation mainly consists in speaking rates higher (rather than lower) than normal speech.

E. THE TEST PARAMETERS TO BE CORRELATED WITH "LOUDNESS" AND RATE

On the basis of delimitations on the acoustic curves and the glottograms, 5 temporal parameters, all measured in milliseconds, and one spatial parameter were extracted and measured for correlation with the rate parameter and the "loudness" parameter. Of the extracted parameters, 2 were purely acoustic, and 4 purely physiological (glottographic). The extracted parameters were the following:

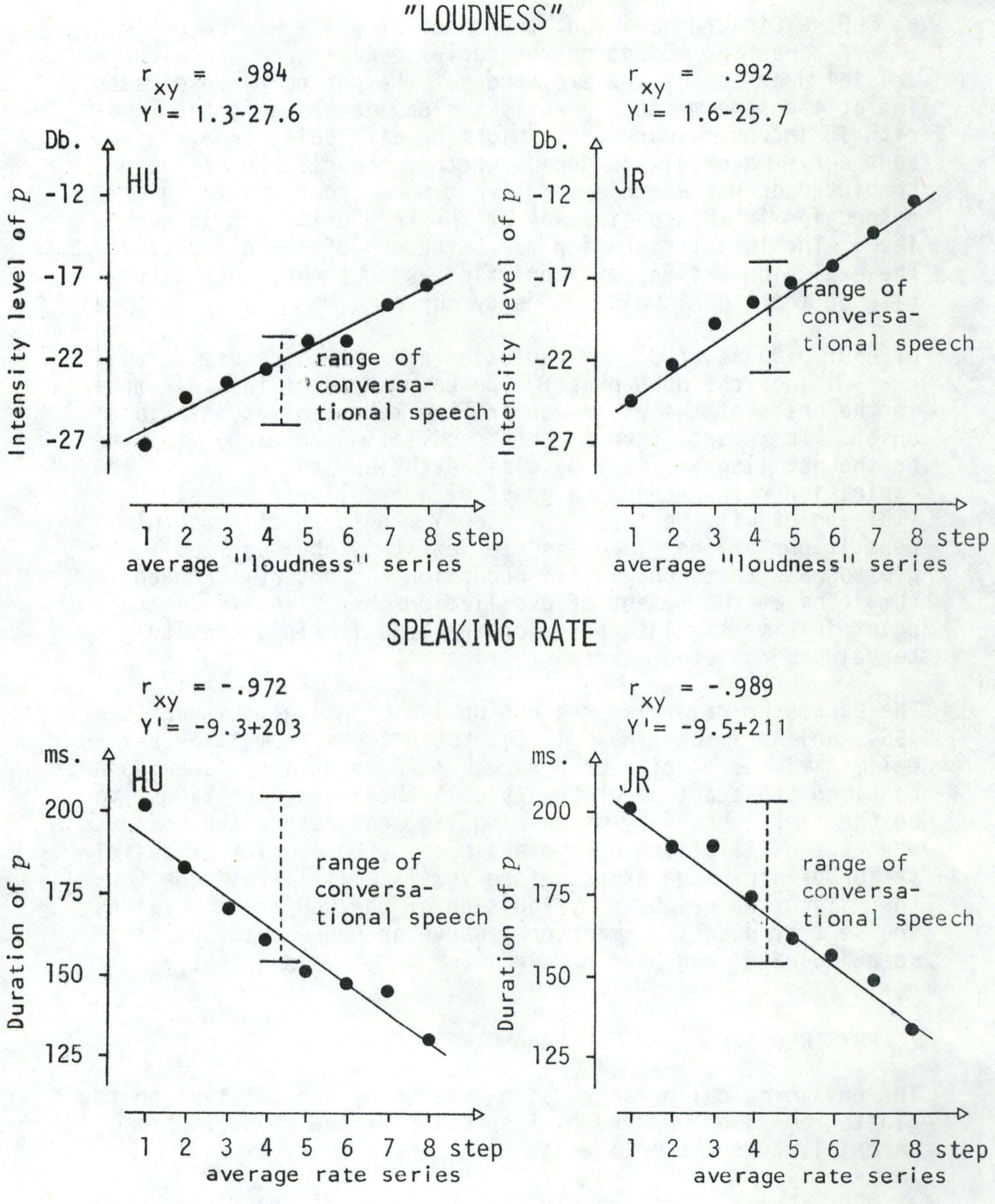


Figure 4

Intensity level of *p* on each step in the average "loudness" series as a function of the rank of the step (upper half) and duration of *p* on each step in the average rate series as a function of the rank of the step (lower half).

1. ACOUSTIC TEMPORAL PARAMETERS:

a. CLD: Closure duration, the onset of which was determined as the moment on the duplex oscillogram where the crucial decrease of the amplitude of the preceding vowel sets in; at the same moment there is a clear decrease in the logarithmic intensity curve. It might be difficult, from the present curve material, to decide whether the closure has been completed or not at this point, but more important in this connection is the application of the same criterion in each item. The labial explosion marks the end of the closure. The explosion was easily identified as an abrupt intensity rise on the logarithmic intensity curve.

b. ASD: Duration of the explosion and the voiceless period of the open phase. The boundaries of this parameter is the onset of the explosion and the moment of oscillation onset. The start of vocal cord oscillation, clearly visible on the oscillograms, may be claimed to mark the offset of the aspiration rather than the onset of the following vowel. (This point will be taken up later.) In the most forceful repetitions of the "loudness" series it is obvious from the glottograms that the glottal abduction has not by far been completed at the moment of oscillation onset. At the same point in time the intensity (on the logarithmic intensity curve) has not even started to rise.

The parameter resembles the notion 'voicing lag' (Lisker et al. 1964, and Abramson et al. 1965), but differs from it by the delimitation criterion used here, as Abramson and Lisker do not consider the start of what they call the 'edge vibrations' to be the 'left' boundary of voicing lag, but rather the moment when the oscillations become audible. (The problem of delimitation of aspirated stops before vowels will likewise be further discussed below.) By addition of the CLD and ASD values the segment duration, mentioned above as representative of speaking rate, had been derived.

2. PHYSIOLOGICAL TEMPORAL PARAMETERS

The physiological parameters, extracted by segmentation on the glottograms supplemented by inspection of the photographic material, were the following:

a. ABD: Duration of the glottal abduction of the test sound. The moment of completion of the abduction is defined as the moment of maximum amplitude on the glottogram, which was rather unproblematic to identify. As for the definition of the start of the abduction, this moment was taken to be the moment of initiation of the apparent movement of the arytenoid cartilages (deduced from fiberoptic pictures and glottograms) and the moment of initiation of the glottal gesture. A preceding check of the stability of the minimum level of the glottograms showed that this level to a high degree seemed to reflect no movement of the arytenoid cartilages and, translucency and incomplete closure excepted, approximately full adduction. On the glottograms the initiation of the glottal abduction was defined as the moment when the minimum level of the glottogram started to rise. An inspection of those stills that happened to have been taken about the moment of change in the minimum level of the glottogram, confirmed that the moment in question was a reasonable criterion for the determination of the start of the glottal abduction. Around this point in time the stills displayed a transition from a state of glottal vibration to the initiation of an active glottal abduction, starting posteriorly.

It shall be mentioned that, for the glottograms of the fastest repetitions in the rate series, it was not always unproblematic to identify the start of the glottal gesture as the minimum level of the glottogram which, because of the fast speaking rate, never reached a steady state level but displayed a kind of undershoot effect. In such cases the initiation of the glottal gesture of the test sound was defined as the moment when the minimum level of the glottogram of the preceding vowel was at its lowest.

b. ADD: Duration of the glottal adduction of the test sound. The onset is defined as the moment of maximum glottogram amplitude, and the completion as the moment when the falling flank of the glottogram has reached the minimum level. The determination of criteria for identifying the completion of the adduction were, as mentioned above, concerning the moment of start of the glottal abduction, also here supplemented by an inspection of the stills, taken "at random" close to the point in time of the moment in question.

c. OADD: Duration of the oscillating phase of the glottal adduction. The onset of the oscillations is clearly visible on the decreasing flank of the glottogram as well as on the duplex oscillogram, and it coincides with the offset of the acoustic parameter ASD, mentioned above.

3. THE SPATIAL PARAMETER

Besides the temporal parameters the maximum amplitude of the glottogram of the test sound was included as a spatial parameter. The maximum amplitude of the glottogram (MAG) was measured with the minimum level of the glottogram as a reference. It shall be mentioned that the results, presented in the following, are based on MAG-values, which are not all identical with the values actually measured on the glottograms. A common problem during fiberoptic-glottographic recordings are disturbances in the transmission line between the light source (i.e. the fiberoptic cable) and the photo transducer. The disturbances are caused by the subjects occasionally having to cough or swallow, or by the speech condition itself (e.g. during forceful speech the tip of the fiberoptic cable often changes position in relation to the glottis), etc. These sources of error will cause variation in the glottogram amplitudes that are not reflections of variation in the size of the glottal aperture. Although a highly positive correlation had been found between the size of glottal aperture, measured on the stills, and the glottogram amplitude, measured at the moment when the still was taken, sources of error might nevertheless have affected the recording, as the errors are not necessarily reflected in the degree of correlation between size of glottal aperture and glottogram amplitude but often in the slope of the regression line, describing the relationship between the two parameters. For instance, an approximation of the light source to the glottis (this can be ascertained from the stills) will intensify the exposure of the glottal opening and the photo transducer and thereby cause an increase in the slope of the regression line describing the relationship between glottal aperture and glottogram amplitude.

A normalization of the glottogram amplitude data was attempted on the basis of a closer inspection of the changes in the slope of the regression line throughout the recording procedure, compared with changes in the experimental conditions, established from the still picture material and from listening to a tape recording of the total recording process, including comments from experimenter, subjects, and others being present. Sections of the recordings, where the slope of the computed regression line was observed to be deviant from that of an arbitrarily chosen 'normal' section of the recording, were taken as a starting point in the normalization procedure. To the extent that the change to a deviant slope of the regression line could be plausibly accounted for by the still picture material and the control tape, the deviant amplitude data were normalized, i.e. their position in relation to the regression line about which they would have scattered if the sources of error had not affected the recording, was calculated.

The procedure of normalizing glottogram amplitude data, as applied here, should be considered only as a preliminary approach to a problem of methodology involving quite a number of uncontrolled factors. Naturally, the normalization procedure cannot account for all sources of error, and hypercorrections have probably been made (see below). However, the general impression is that this attempt to eliminate the influence from sources of error showed plausible results in most respects, based on reasonable assumptions about the nature of the sources of error and their expected influence, supplemented by physical evidence. On the whole, it seems warranted to assume that the normalized amplitude values reflect the physical facts to a higher degree than the originally measured ones, and thus they offer a better basis for pooling glottogram amplitude values of the repetitive recordings.

In the following, the average maximum amplitude through the "loudness" and rate series is represented by the median, as the values of each step displayed no underlying normal distribution. The 6 parameters are illustrated by examples in figure 5.

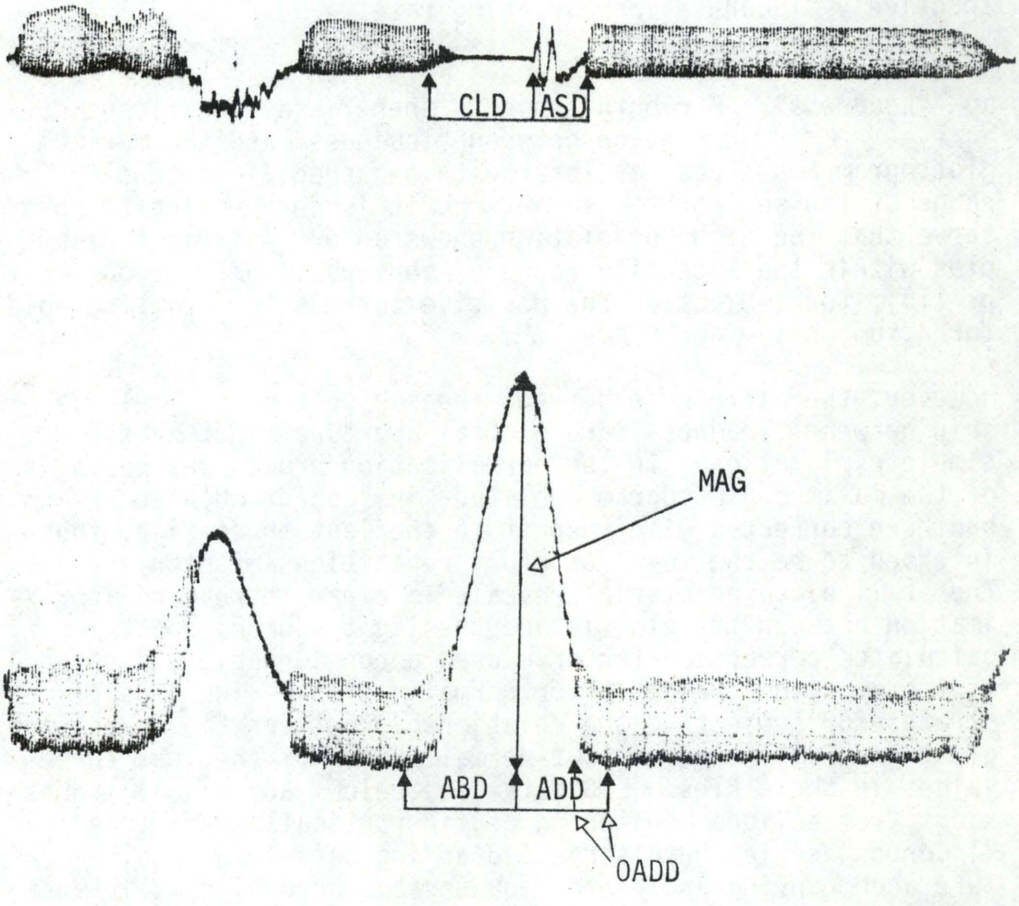


Figure 5

The 6 test parameters to be correlated with "loudness" and rate. The 2 acoustic parameters are illustrated in the upper curve (duplex oscillogram), and the 4 fysiological parameters are illustrated in the lower (photo-electric glottogram).

III. RESULTS

A. THE EFFECT OF "LOUDNESS" AND SPEAKING RATE ON THE SPATIAL REALIZATION OF THE GLOTTAL GESTURE

Figure 6 illustrates graphically for each subject the correlation between the maximum amplitude of the glottogram and respectively "loudness" and speaking rate.

a. "Loudness" For both subjects there is a significant correlation between "loudness" and the maximum glottogram amplitude, at least within a specific intensity range of the series; for subject HU it is interesting to observe that the high correlation shows up exclusively for samples within the intensity range of conversational speech (cf. p. 115); for subject JR the positive correlation seems to hold for a somewhat wider range.

However, the difference between the subjects in the relationship between "loudness" and glottal aperture might have a simple explanation. In the normalization procedure the values of the maximum glottogram amplitude in subject HU's recording had been corrected with respect to the last three (i.e. those intended to be the most forceful) repetitions of each series. The still picture material revealed a clear increasing approximation between the glottis and the light source, and the calculated correction factor caused a considerable reduction of the amplitude values after normalization. Figure 6a illustrates for subject HU the relationship between "loudness" and glottogram amplitude without normalization of the last three values in the series. EMG data (see below) and visual impressions from a Video monitoring of fiberoptically registered "loudness" series both suggested an increase in glottal aperture accompanying increased "loudness". Accordingly, it must be assumed that figure 6a in fact reflects the actual relationship between size of glottal aperture and "loudness" better than the normalized data. The increasing approximation between the glottis and the light source, ascertained from the still picture material, and the observed radical increase in the slope of the regression line taken into consideration, it seems plausible that the calculated small amplitude values in question are due to the presence of an additional source of error rather than the use of too high a correction factor.

It is evident from the figure that the relationship between "loudness" and glottogram amplitude of the non-corrected values for subject HU is much more like that of subject JR. For this subject no correction was made for differences in the distance between the glottis and the light source, caused by the "loudness" condition itself, although a normalization of the value of the last step in the series was considered.

Although the normalization procedure used here seems to warrant a more adequate description of the physical facts (see Note 1),

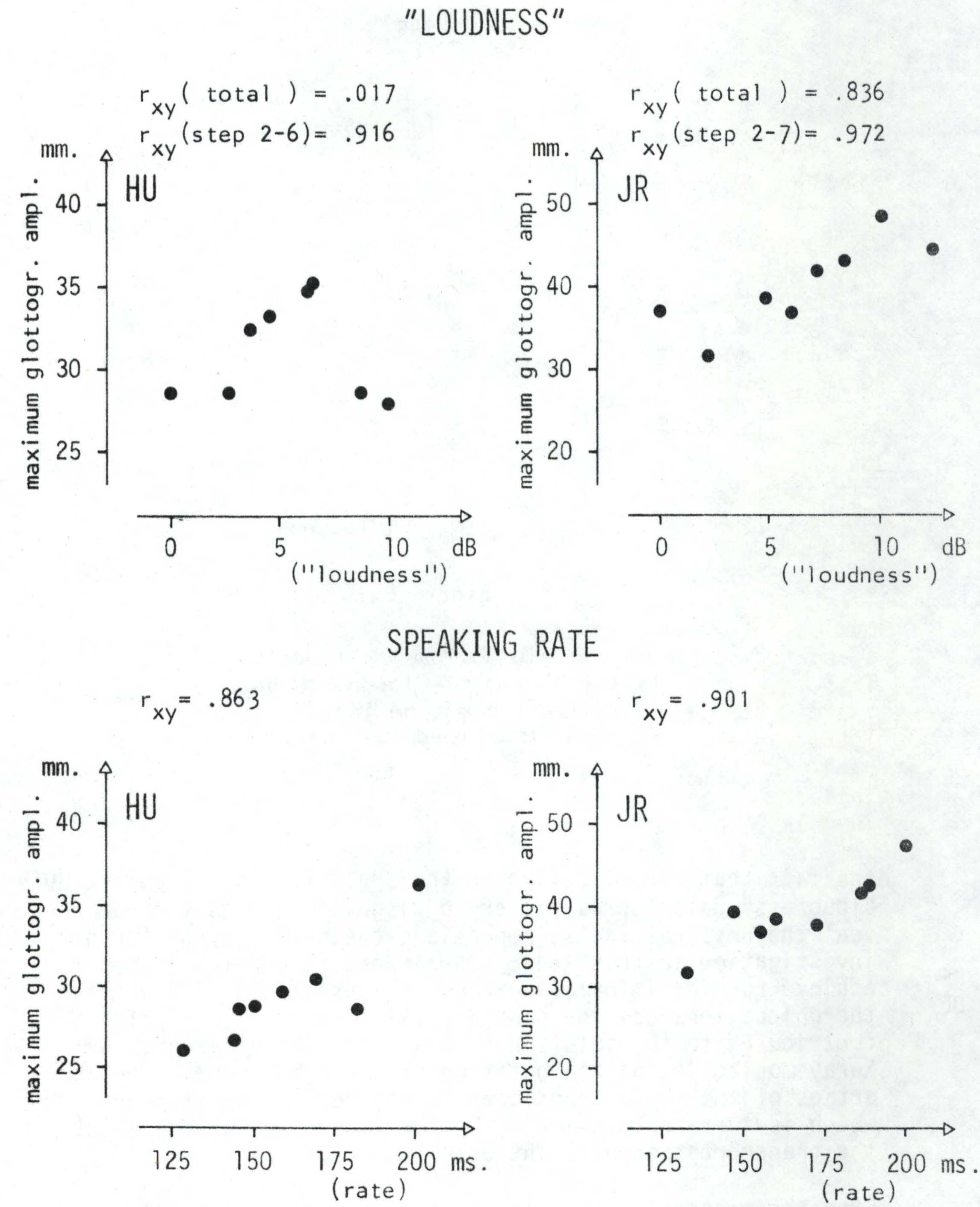


Figure 6

Maximum glottogram amplitude of p as a function of "loudness" (intensity of p) and speaking rate (duration of p).

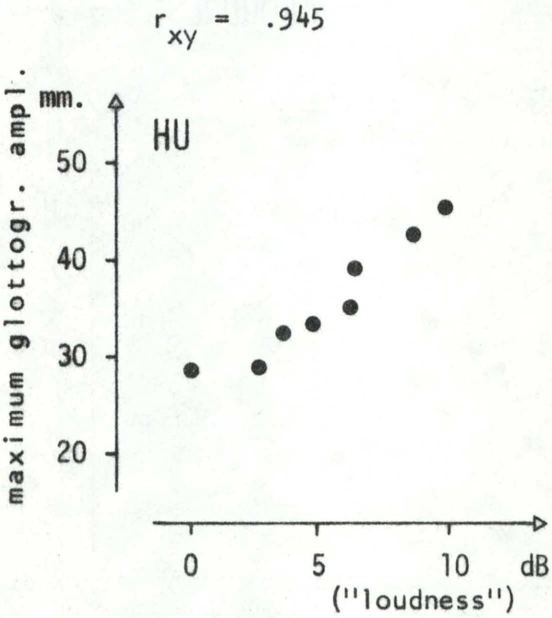


Figure 6a

Maximum glottogram amplitude of *p* as a function of "loudness" before normalization of the last three steps of the "loudness" series.

the fact that normalization in the special case of subject HU's "loudness" data seemed rather to disguise than to further reveal the physical facts, emphasizes the necessity of further investigation in this area. One important problem is how to achieve running information about the exact distance between the object lens and the glottis. Kiritani (1971, 1972) has contributed to the solution of this problem by research on X-ray monitoring of the position of the fiberscope. The properties of the photo transducer is another fundamental problem which is further complicated by the lack of knowledge about the transportation of light below the glottis.

From the present results the following tentative conclusion might be drawn concerning the effect of "loudness" on the spatial control of the glottal gesture: repetitive productions of *p* with gradually increased "loudness" are accompanied by a systematic increase of the maximum glottal aperture. It cannot be generally stated whether there is a difference in the principle of control depending on the "loudness" being weak or not. It seems, however, to be a common feature for the subjects that the principle of spatial control is different when the "loudness" level is low and, for subject JR, perhaps also when the speech assumes a shouting character, but these assumptions must be supported by further investigation.

It is interesting that, within a certain range of "loudness" variation, a linear correlation can be observed between intensity (i.e. dB levels) and the size of the glottal aperture. This must lead to the hypothesis that the control of the degree of glottal abduction in connection with "loudness" variations takes a logarithmic scale as point of departure.

b. Speaking rate: For both subjects speaking rate and size of maximum glottogram amplitude also show a significant correlation. Figure 6 seems to display a close, approximately linear relationship between the degree of maximum abduction and segment duration all through the series.

Later in this paper the results of the spatial realization of the glottal gesture, for the "loudness" as well as for the rate series, will be reconsidered and incorporated as part of a larger framework together with the temporal relationships of the gesture.

B. THE EFFECT OF "LOUDNESS" AND SPEAKING RATE ON THE TEMPORAL RELATIONS OF THE GLOTTAL GESTURE

1. THE GLOTTAL ABDUCTION

The variation in the duration of the glottal abduction through the "loudness" and the rate series is shown graphically in figure 7.

a. "Loudness" It is clear from the figure that the range of variation in the duration of the glottal abduction is very small, about 10 ms for subject JR and less for subject HU. For subject HU the variation seems rather random, the correlation coefficient being zero, whereas the variation estimated from the correlation coefficient seems to display some systematism in relation to "loudness" for subject JR. However, the significant correlation seems to depend entirely upon the values of the peripheral steps in the series; the correlation coefficient for the steps 2-7, the most interesting range as far as size of glottal aperture is concerned, is close to zero. Consequently, we may conclude that increased "loudness" has virtually no effect on the duration of the glottal abduction; irrespectively of the intensity of the test sound the glottal abduction lasts about 100 ms for subject JR and 120 ms for subject HU.

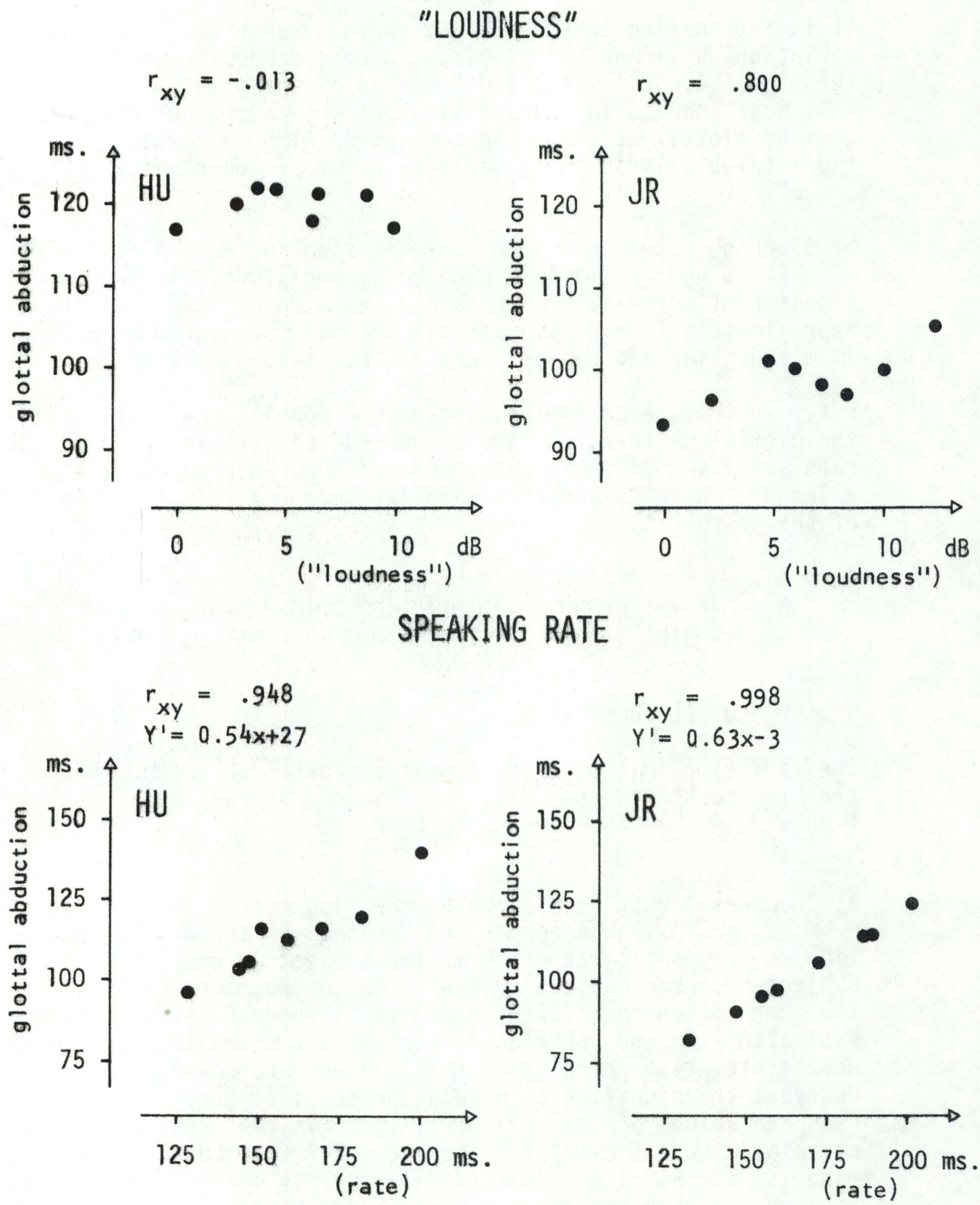


Figure 7

Duration of glottal abduction of *p* as a function of "loudness" (intensity of *p*) and speaking rate (duration of *p*).

b. Speaking rate: In connection with varying speaking rate the range of variation in the duration of the glottal abduction is much wider. Quite contrary to the relations at increased "loudness" - but not unexpectedly - the figure shows a clear, systematically shorter duration of glottal abduction concurrently with increased speaking tempo. The correlation between the segment duration and the duration of glottal abduction is linear, positive, and highly significant (1%).

Since the speaking rate of the test sound is represented by the segment duration, delimited on the acoustic curves, it can be concluded that a faster completion of the oral articulation of the test sound implies a proportionally (one-to-one) faster completion of its glottal abduction. The duration of the abduction is more reduced per step in the series with subject JR than with subject HU. The amount of reduction from slowest to fastest tempo in the series, predicted from the regression line, is on the average 45 ms for subject JR and about 38 ms for subject HU, which means that the duration of the glottal abduction is reduced by 55% through the series for subject JR and only about 40% for subject HU. For the sake of comparison, it is interesting to recall that the reduction of the segment duration from slowest to fastest speaking rate was 70 ms or 50% on the average for both subjects.

c. The relation between duration and degree of glottal abduction: In the previous section it was stated that there is a significant correlation between the degree of maximum glottal abduction and "loudness" and speaking rate, respectively, of the test sound. The duration of the glottal abduction was shown to be of constant value, irrespective of "loudness", but to correlate highly with speaking rate, the duration of the abduction being reduced proportionally with the segment duration. These relationships, or lack of same, might lead to interesting conclusions about the average rate of glottal abduction under each of the two extra-linguistically controlled speech conditions. As the intensity of the test sound (i.e. "loudness") has been shown to correlate highly with degree, but not with duration of glottal abduction, one might not expect to find a correlation between duration and degree of glottal abduction by increased "loudness". Contrariwise, as segment duration (i.e. speaking rate) correlates highly with both degree and duration of the glottal abduction, a high correlation between duration and degree of glottal abduction would be expected by increased speaking rate. These assumptions are confirmed by figure 8, which illustrates these correlations graphically. The two parameters, degree and duration of the glottal abduction, do not correlate in the "loudness" series but show for both subjects a highly significant (1%), positive, linear correlation in the rate series. As the degree and duration of the glottal abduction can be proved to decrease proportionally through the rate series, it might consequently be concluded that the average velocity of the abduction must be constant in the produc-

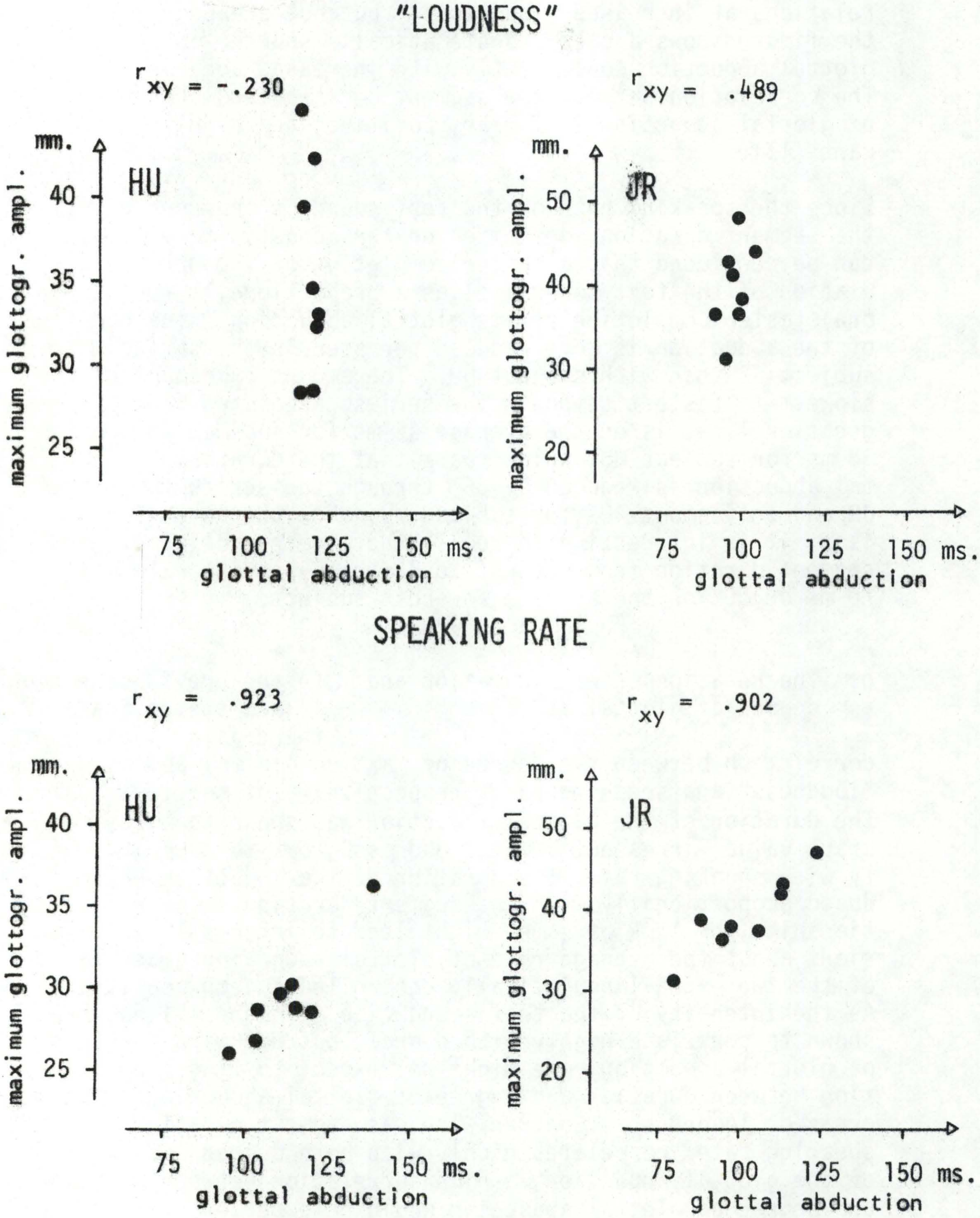


Figure 8

Maximum glottogram amplitude of *p* as a function of the duration of the glottal abduction. The relation is shown for the "loudness" series (upper half) and the rate series (lower half).

tion of the test sound repeated with increased speaking rate. Concerning the average of the glottal abduction in the production of the test sound repeated with increasing "loudness", it was previously stated that the maximum glottal abduction increases proportionally with "loudness" within a certain range of variation. As the time to reach the maximum abduction is constant, irrespective of "loudness", it must be concluded that the velocity of the glottal abduction is greater with increased "loudness".

2. THE GLOTTAL ADDUCTION

a. "Loudness" As shown in figure 9 the variation in the duration of the glottal adduction of the test sound is rather small and obviously random in relation to "loudness" for both subjects. The duration of the adduction averages 108 ms for subject HU, a bit shorter than that of the glottal abduction (120 ms), and 105 ms for subject JR, a little longer than that of the abduction (100 ms). On the whole, one might conclude about the relation between the duration of abduction and adduction that they each make up about 50% of the total duration of the gesture, irrespective of the degree of "loudness".

b. Speaking rate As opposed to "loudness", the effect of speaking rate on the duration of the glottal adduction seems rather different for the two subjects. Figure 9 indicates only a small effect of speaking rate on the duration of the glottal adduction for subject JR, i.e. only at very high and partly at very low rates does there seem to be a positive correlation, and this relationship is probably responsible for the significant (at the 5% level) correlation coefficient for the whole series, the correlation coefficient for the steps 1-7 is but .607, and for the steps 2-7 only .441. The marginal steps excepted, it can be concluded that the duration varies randomly about an average of 105 ms. For subject HU, however, there is a highly significant correlation between the two parameters and indications of a systematic reduction all through the series; the numerical reduction in the duration of the adduction is on the average 4-5 ms per rate step, which is a little less than that for the abduction. For this subject it is interesting to notice that the duration of the abduction and the adduction, respectively, expressed as percentage of the total duration of the gesture, makes up the same percentage, irrespective of speaking rate, namely 52-53% for the abduction and consequently 47-48% for the adduction.

During the adduction phase the onset of vocal cord vibration can be observed, from 20 up to about 75 ms ahead of the moment of completion of the adduction. The earliest start of oscillation is observed in the most forceful samples of the "loudness" series. This is not unexpected, as it is a well known fact that "loudness" implicates an increased transglottal flow,

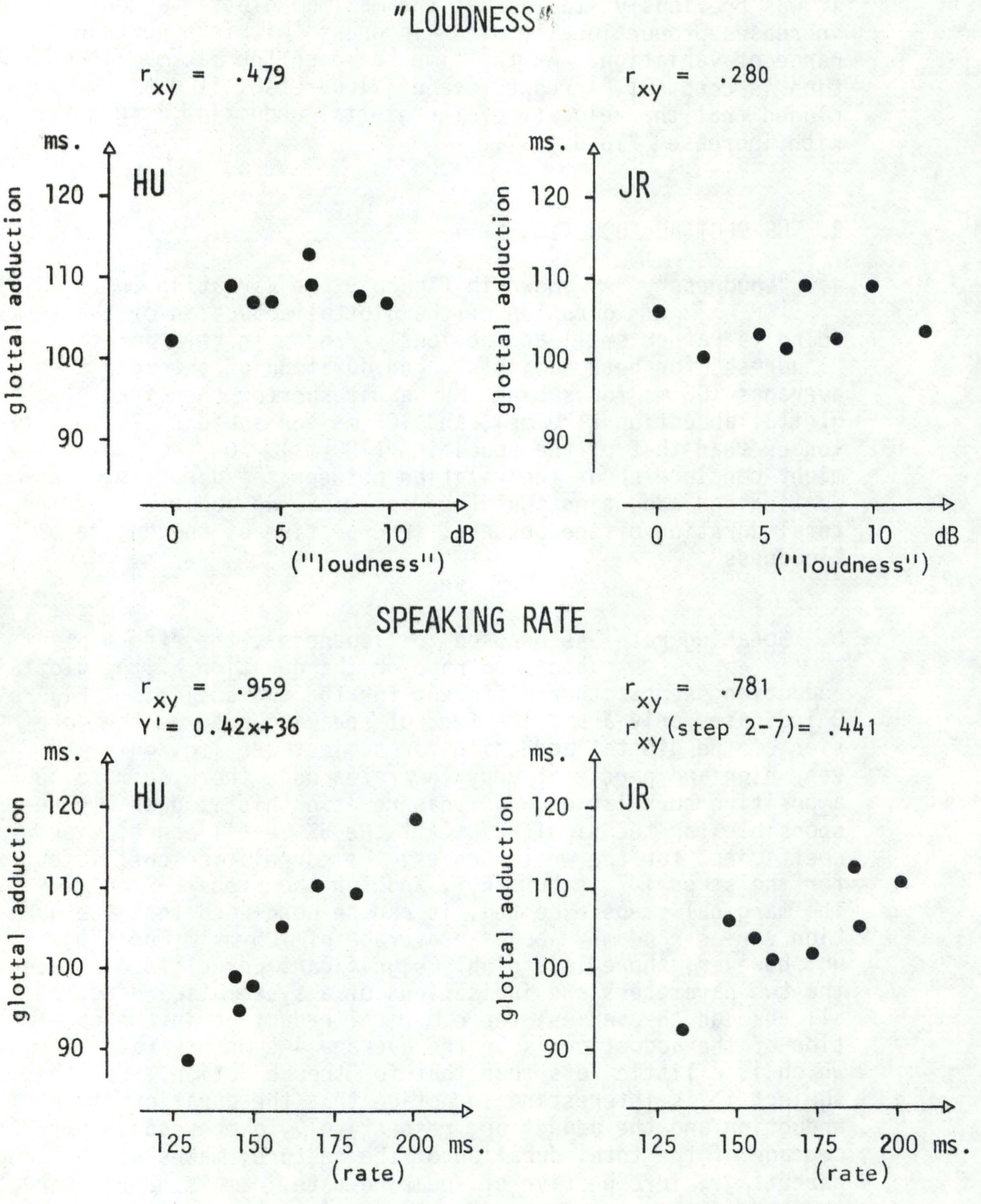


Figure 9

Duration of glottal adduction of *p* as a function of "loudness" (intensity of *p*) and speaking rate (duration of *p*) and the faster the flow after the release, the earlier the conditions for the Bernouilli effect to take place will be met. The highly significant correlation between extension of the vibration period through the adduction phase and "loudness" is illustrated in figure 10.

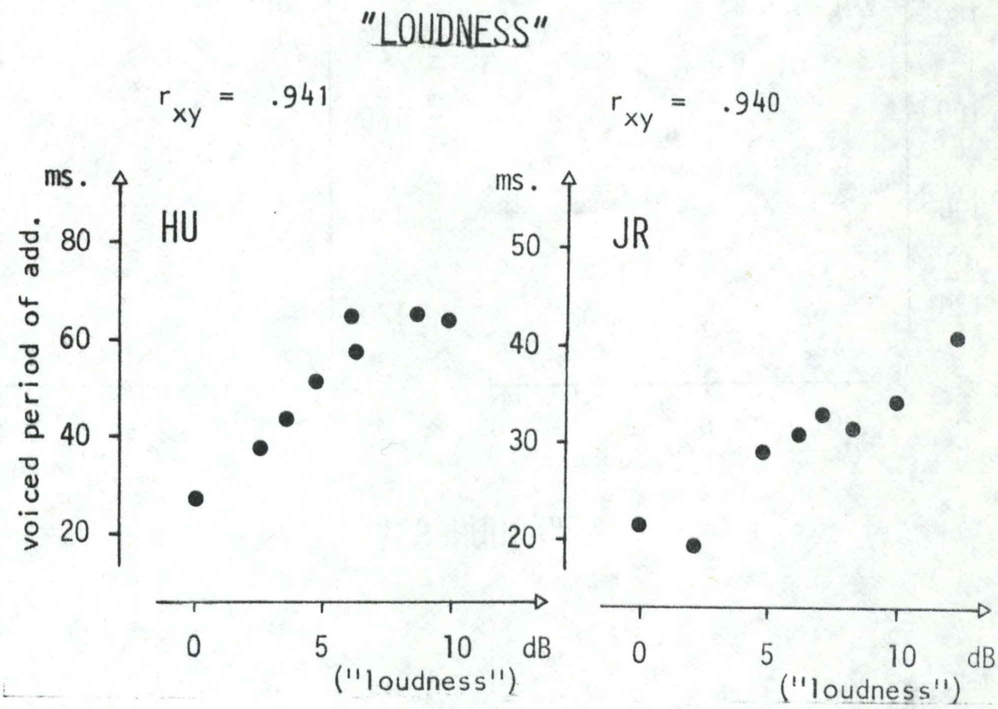


Figure 10

Duration of the voiced period during the glottal adduction of *p* as a function of "loudness" (intensity of *p*).

3. SEGMENT DURATION VS. THE TOTAL DURATION OF THE GLOTTAL GESTURE AND IMPLICATIONS OF THEIR DIFFERENCE FOR DELIMITATION CRITERIA

In the previous section it was stated that the duration of the glottal abduction and adduction, respectively, remained largely unchanged, irrespective of "loudness". Accordingly, it is not surprising that the total duration of the glottal gesture also remains largely unchanged through the "loudness" series, the duration averaging 225-230 ms for subject HU and 205 ms for subject JR. Figure 11 (upper half) shows clearly that segment duration, on the other hand, is systematically reduced when "loudness" is increased. This systematic reduction is almost exclusively due to the criteria used for delimitation of the aspiration of the stop and the onset of the following vowel. Defining the moment of oscillation onset as the start of the following vowel might actually lead to fatal generalizations

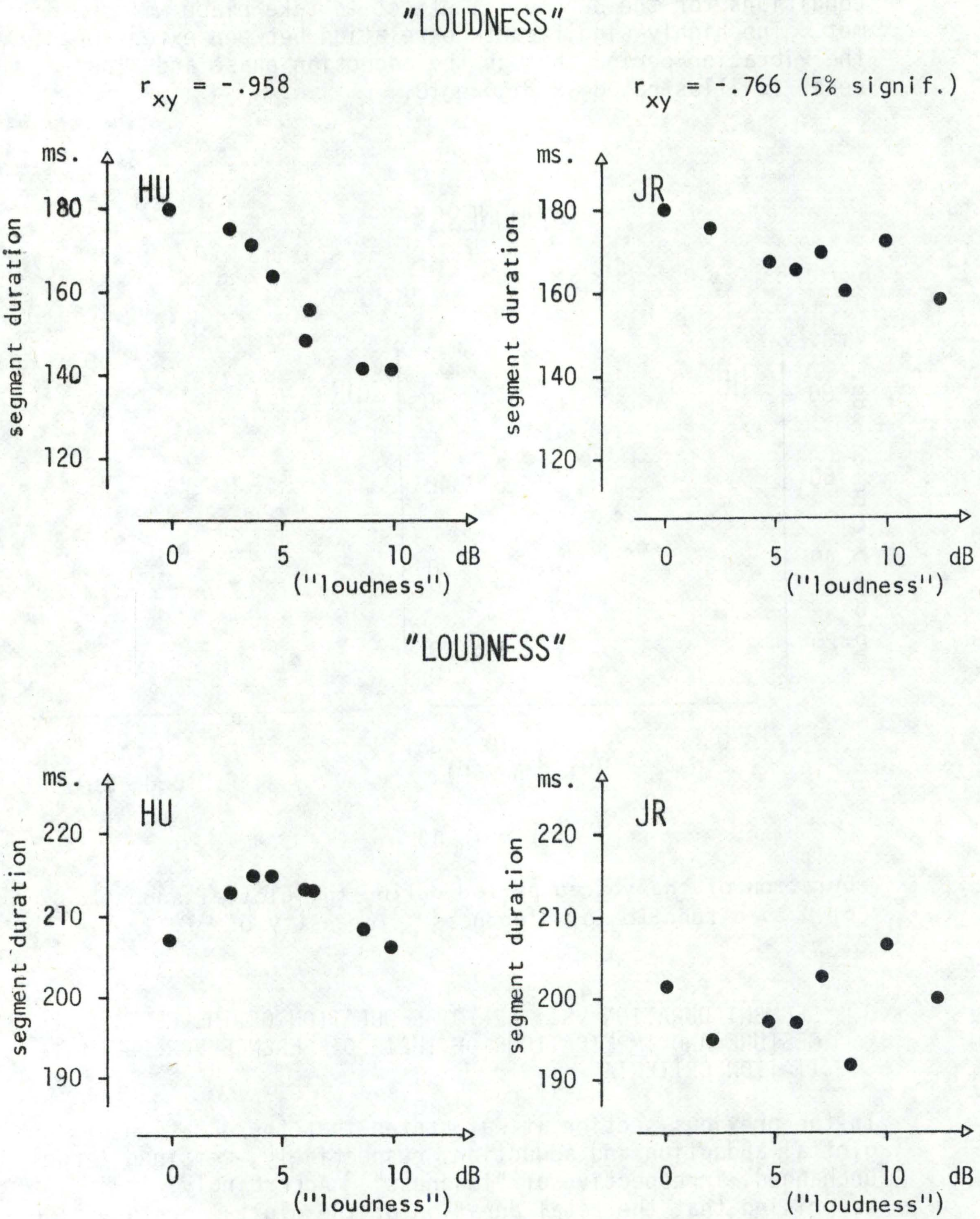


Figure 11

Duration of *p* as a function of "loudness" (intensity of *p*). In the upper half of the figure the relation is shown for duration of *p* being defined as the interval from oral implosion to the onset of vocal cord vibration. In the lower half the relation is shown for duration of *p* being defined as the interval from oral implosion to the moment of completion of the glottal adduction.

about the temporal organization of speech from the point of view of production, as the ever earlier onset of vocal cord vibration during the adductory phase at increased "loudness" may be considered to be a passive consequence of aerodynamic conditions (i.e. heightened subglottal pressure) rather than a programmed reduction in the duration of the aspiration and thereby in the duration of the stop. Seen from that point of view, the chosen criterion for delimitation implies ambiguities of up to ±40 ms for average consonant and vowel durations (in individual cases up to ±60 ms) if "loudness" is not taken into consideration.

An alternative delimitation criterion might have been preferred. If, for the moment, the start of the following vowel is defined as the moment of completion of the glottal adduction, then no systematic reduction in segment duration will occur (cf. figure 11, lower half). Using this criterion for delimitation makes the aspiration of the stop consist of a voiceless plus a voiced interval, where the voiceless interval takes up between 80% and 40% (depending on the degree of "loudness") of the total duration. The moment of completion of the glottal adduction, as determined by means of a physiological curve, seems to appear in the acoustic curves as the moment when the complete formant pattern of the following vowel has been stabilized (as regards the presence of upper formants and their intensity). For a more thorough treatment of the problems of delimitation, the reader is referred to Fischer-Jørgensen and Hutters (1981).

As one of the major purposes of the present paper is to shed light on the temporal organization of speech, it seemed preferable to redefine the concept of segment duration and, in the following, it will be defined as the period from the labial implosion to the moment of completion of the glottal gesture. It shall be emphasized that application of the redefined segment duration as the rate parameter would not appreciably have changed the results and conclusions reported so far concerning the effect of speaking rate.

C. SUPRAGLOTTAL AND GLOTTAL TIMING

This section describes the temporal relationship between oral closure and aspiration, glottal ab- and adduction, and the interarticulatory timing of the glottal and supraglottal events under the two speech conditions.

1. SUPRAGLOTTAL TIMING

It has been stated previously that increased "loudness" has no effect on segment duration, while increased speaking rate is characterized by a systematic, linear reduction in segment duration. Concerning general timing strategies of the closure and aspiration components under the two speech conditions, it might be preliminarily assumed that the shorter segment dura-

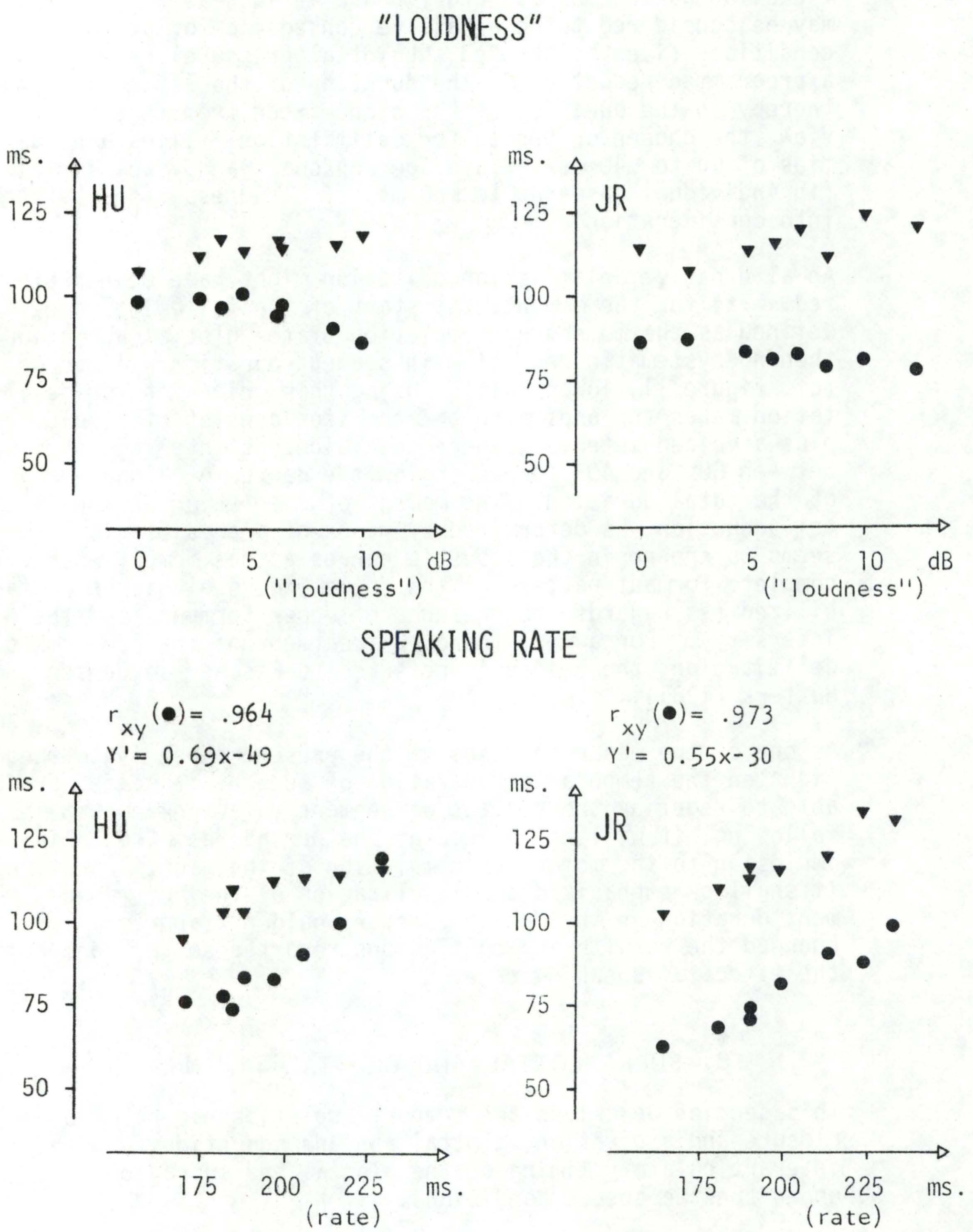


Figure 12

Closure duration (the dots) and duration of aspiration (the triangles) of *p* as a function of "loudness" (intensity of *p*) and speaking rate (duration of *p*) tion at increased rate is a consequence of reduction in the duration of the closure as well as the duration of the aspiration. As for the "loudness" condition, variation (if any) in the duration of closure and aspiration must be inversely correlated.

Figure 12 confirms that increased speaking rate involves a systematic reduction of both closure and aspiration. Though far less radical or systematic, the effect of increased "loudness" is a shorter closure as well but, as expected, the aspiration is prolonged. As for the rate condition, it seems to be a general strategy that the shorter duration is primarily a consequence of a reduction of closure duration, the tendency being more radical for subject HU, where closure reduction accounts for almost 70% of the total segment reduction; for subject JR the figure is only about 55%.

It appears from figure 12a that, for subject HU, shorter segment duration is primarily reflected in closure duration also under different "loudness" conditions. Accordingly, it seems as if the two subjects have different individual timing strate-

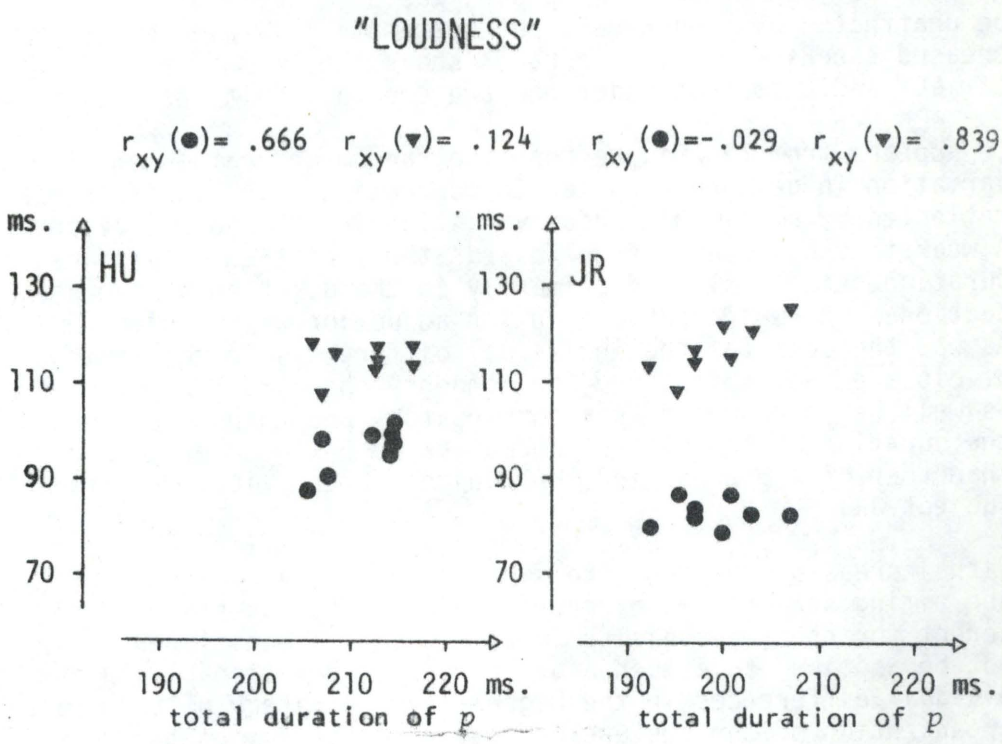


Figure 12a

Closure duration (the dots) and duration of aspiration (the triangles) of *p* as a function of the total duration of *p*. The relations are shown for the "loudness" series.

gies applied to the common constraints laid upon the temporal organization of closure and aspiration under the two different speech conditions. This might be explained as a mere compensatory phenomenon, since closure duration makes up a greater percentage of the segment duration for subject HU than for subject JR, i.e. subject JR has in general a relatively short closure, less susceptible to reduction than that of subject HU. The tendency to compensation is clearly illustrated by the fact that the reduction in closure duration from slow to fast rate amounted to 45 ms for subject HU and 38 ms for subject JR, while reduction in aspiration was 23 ms for subject HU and 30 ms for subject JR. However, it is interesting to observe that the reduction in duration and aspiration, expressed as a percentage of the duration at the slowest speaking rate, is almost identical for the two subjects, namely about 37.5% for the closure; for the aspiration it is 19.5% for subject HU and 22.5% for subject JR.

2. GLOTTAL TIMING

Like the segment duration (intended to express the supraglottal timing), the duration of the glottal gesture, in the abductory as well as in the adductory phase, was shown by and large to be unaffected by "loudness", but systematically reduced at increased speaking rate. Figure 13 shows the timing of the glottal ab- and adduction under the two speech conditions.

It appears from the figure that the random and rather small variation in gesture duration in connection with "loudness" is reflected by small and random variation in ab- and adduction. A weak tendency can be seen towards the variation in gesture duration being reflected primarily in the duration of the abduction with subject JR and in the adduction with subject HU. As was the case for the individual differences in the timing of closure and aspiration, the tendency observed here might as well be interpreted as a compensatory phenomenon, because the duration of the glottal abduction is on the average shorter than that of the adduction for subject HU and vice versa for subject JR.

With increased speaking rate it is clear from the figure that the timing strategy with regard to the glottal gesture is different for the two subjects in this case. The difference cannot be ascribed to compensatory processes but probably to individual differences in the higher-level strategy of temporal organization of the gesture. However, it is common to the two subjects that varying gesture duration is primarily reflected in the duration of the abduction. The primary difference in their timing strategy is that the shorter gesture duration seems to be a consequence of a systematic reduction of both ab- and adduction with subject HU, but to a much higher degree reflected exclusively in a systematically reduced abduction duration with subject JR. From the slowest to the fastest speaking rate the amount of reduction in abduction duration

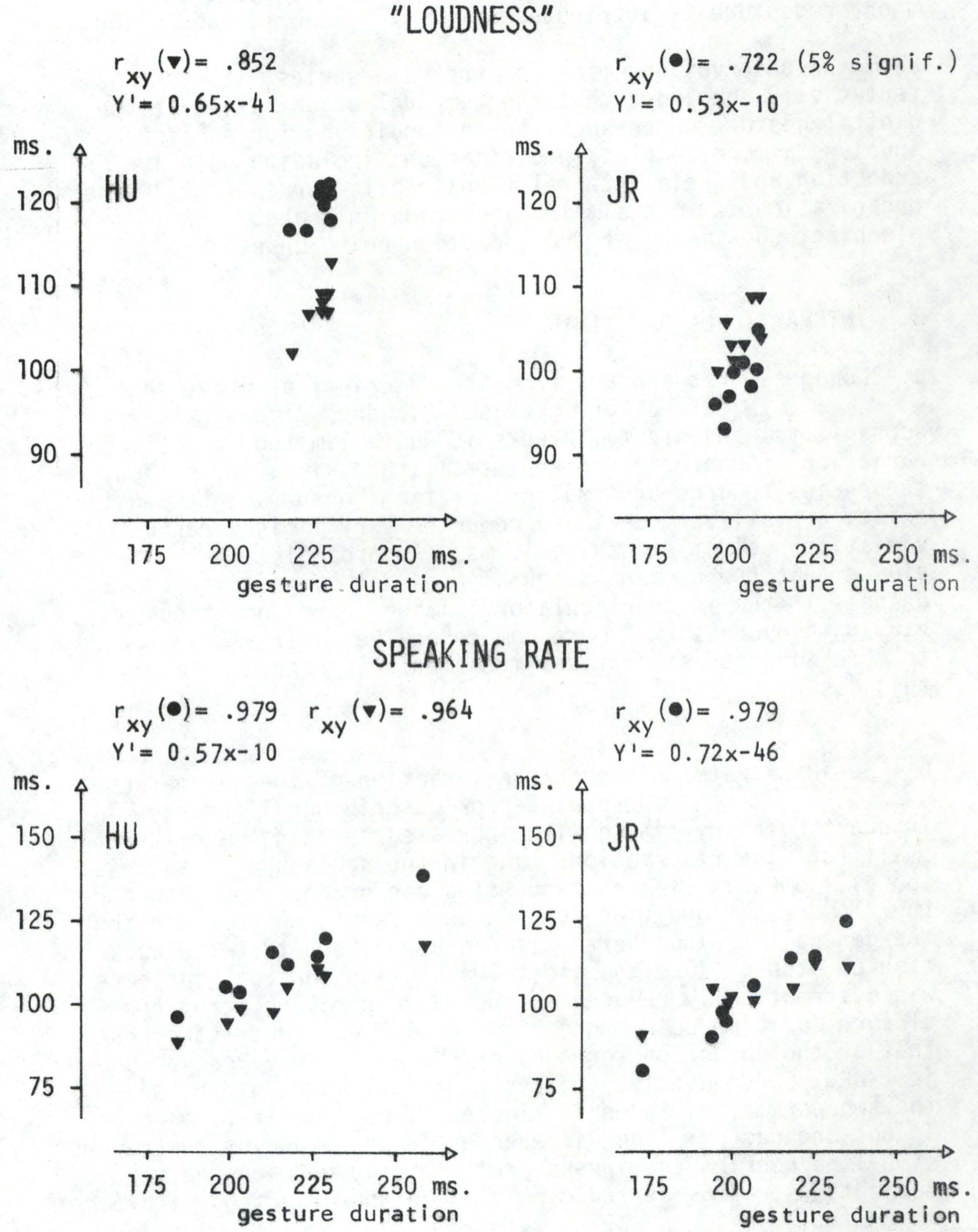


Figure 13

Duration of glottal abduction (the dots) and adduction (the triangles) of *p* as a function of the total duration of the glottal gesture. The relations are shown for the "loudness" series (upper half) and the rate series (lower half).

constitutes about 57% of the total reduction in gesture duration for subject HU and about 72% for subject JR. If the marginal steps (1,2 and 8) are omitted with subject JR, the shorter gesture is attributed solely to a shorter abduction.

From the observations made for the rate series, it might be tentatively concluded that the temporal organization of the glottal gesture under such speech conditions for subject HU involves a more complex specification, including both ab- and adduction and their interrelations, while for subject JR the organization is more simple, involving only abduction in the planning, the timing of the adduction being unspecified.

3. INTERARTICULATORY TIMING

a. "Loudness" It appears from the descriptions above that the effect of increased "loudness" on the timing of supraglottal and glottal events is quite limited. The temporal variations found here are very small, and virtually random; tendencies towards systematic variation, when such tendencies exist, are all very weak and comprise very limited ranges of variation, seldom exceeding 10 ms. Accordingly, it may be concluded that the present material does not suggest any clear changes in the interarticulatory timing under conditions of varying "loudness". A more comprehensive material, subjected to more advanced statistical treatment, is needed to clarify this issue.

b. Speaking rate Although the variation of the parameters appropriate for describing differences in interarticulatory timing with increased speaking rate is rather small (cf. the reservations made in the section on "loudness" above), somewhat clearer tendencies can be observed than for the "loudness" condition, and it seems worth-while to present the general findings here. It has been shown that the duration of stop closure and glottal abduction for both subjects was systematically reduced at increased speaking rate, the closure duration being more reduced with each rate step than that of the abduction for subject HU, and vice versa for subject JR. Consequently, the systematic relationship between the two parameters shown in figure 14, upper half (the dots), is not unexpected. The differences in the slope of the regression line confirm the inverse relationship between the two subjects' strategy of reducing closure and abduction with each rate step.

From the relationships established one might expect to find a change in the timing of the explosion and the moment of maximum abduction at increased speaking rate, the time span between the two parameters being increased with subject HU and decreased with subject JR. Further, a priori assumptions can be made if the relationship between closure duration and the period from oral implosion to the moment of maximum abduction

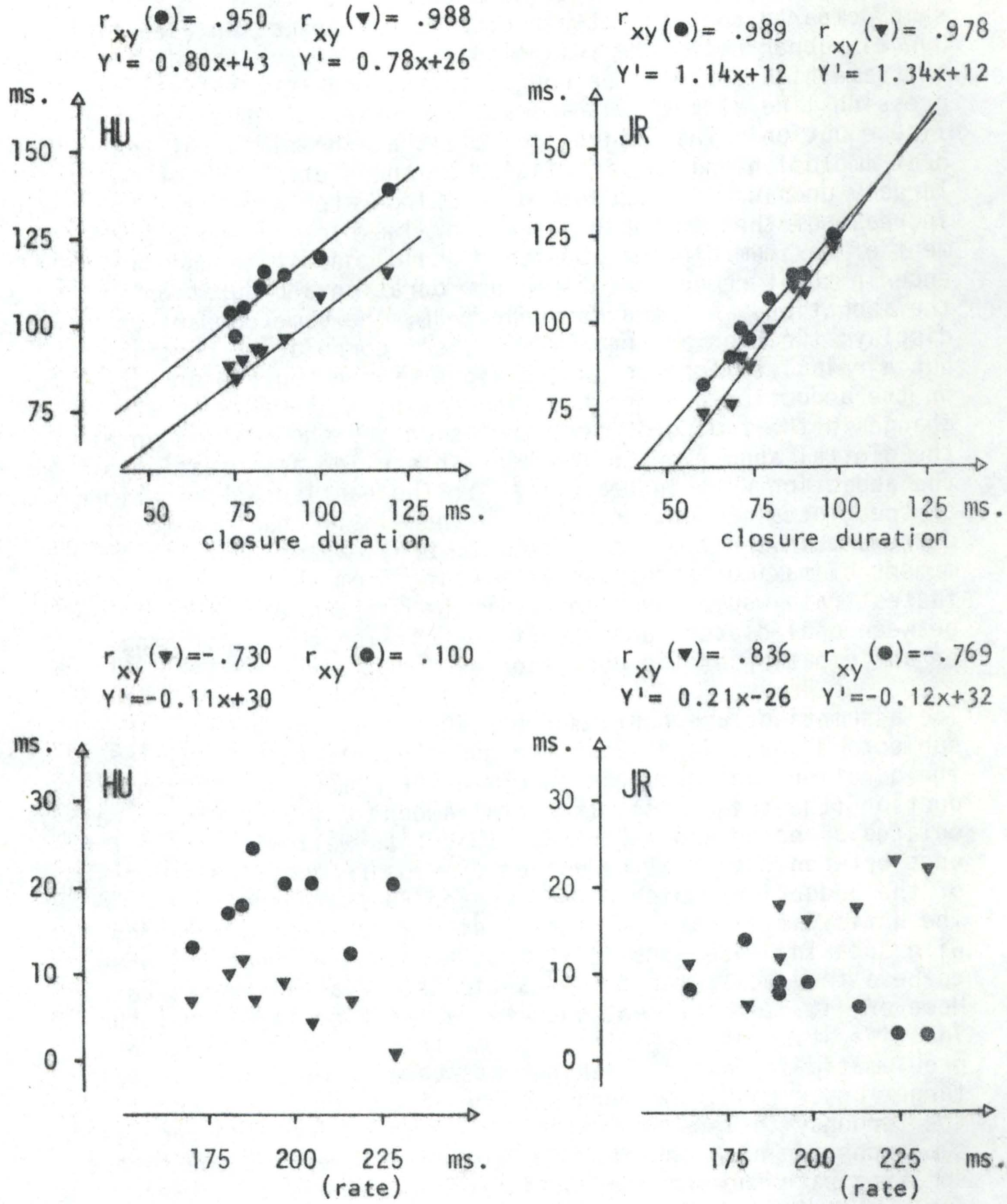


Figure 14

Upper half: Total duration of glottal abduction of *p* (the dots) and duration of glottal abduction of *p* with oral implosion as line-up point (the triangles) as a function of closure duration of *p* at increasing speaking rate. Lower half: The interval from the initiation of the glottal abduction of *p* to the oral implosion (the dots) and the interval from the oral explosion of *p* to the moment of maximum glottal abduction as a function of speaking rate (duration of *p*).

(i.e. duration of glottal abduction with oral implosion as line-up point) is taken into consideration. This relationship compared to that between closure and abduction (see figure 14, upper half, the triangles) shows for subject HU almost the same high correlation coefficients, nearly identical regression line slopes, but a smaller constant in the regression line equation. This might indicate that the timing of the oral implosion and the initiation of the glottal abduction is largely unchanged by increased speaking rate, and the total increase of the time span between oral explosion and the moment of maximum glottal abduction would constitute the difference in total reduction of closure duration and duration of the abduction. As regards subject JR, the same comparison displays two high and nearly identical correlation coefficients but a relative increase in the slope of the regression line on the abduction data points, indicating a tendency towards a change in the timing of oral implosion and the initiation of the glottal abduction in terms of an earlier anticipation of the abduction with increased rate. The earlier anticipation of the glottal abduction would further imply that the total decrement in the time span from the oral explosion to the moment of maximum abduction when going from the slowest to the fastest rate would cover not only the difference in reduction between oral closure and abduction but also the total amount of anticipation of the abduction.

The assumptions are confirmed by figure 14, lower half. For subject HU there is a clear tendency towards an increase of the duration from oral explosion to the moment of maximum abduction (the triangles), the total amount of which can be calculated as being about 7.5 ms. This fits well with the difference between the total reduction of closure duration and that of the abduction, which was 47 ms and 39 ms, respectively. The anticipation data (the dots) display a zero correlation with speaking rate, though it must be admitted that there seems to be a tendency towards less anticipation at increased rate. However, it has been mentioned above (section II,E) that the identification of the initiation of the gesture was not unproblematic at faster speaking rate because of glottal vowel target undershoot, and errors of measurement might account for the tendency to less anticipation. With subject JR, the assumptions are also confirmed. From the slowest to the fastest speaking rate the total reduction of the time span between oral explosion and the moment of maximum glottal abduction averages 14.5 ms; the difference between the total reduction of oral closure duration and abduction added to the total increase of anticipation constitutes 14.5 ms as well.

IV. CONCLUSION AND DISCUSSION

From the results described in the previous section the following can be concluded about the effect of "loudness" and speaking rate on the glottal behaviour of Danish *p*:

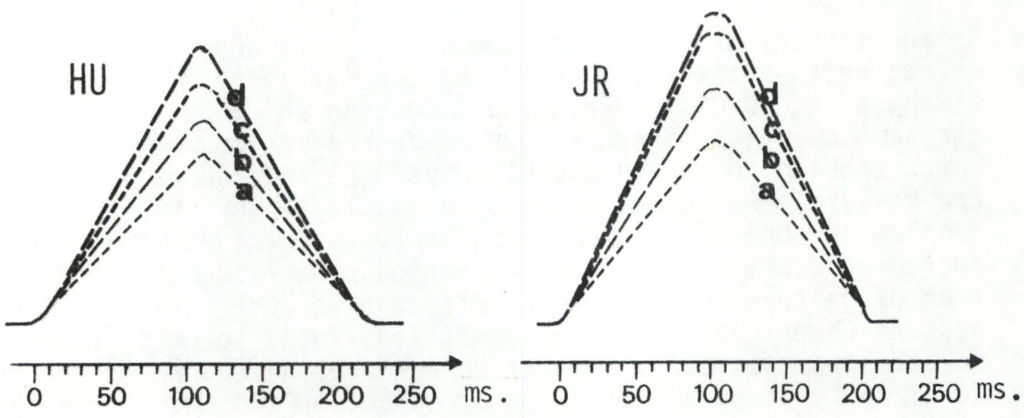
The spatial realization is affected systematically by "loudness" as well as by speaking rate. An increase of "loudness" is accompanied by an increase of glottal abduction, at least for a specific range of "loudness", whereas an increase in speaking rate causes a decrease of glottal abduction.

As for the temporal organization, varying "loudness" seems to affect neither the duration of the glottal gesture, nor the timing of the O-C gesture. The abduction and adduction have the same duration, irrespective of degree of "loudness", each corresponding to an average of 50% of the total duration of the gesture, the abduction being a little longer than the adduction for one of the subjects, and vice versa for the other. Increased speaking rate is accompanied by a systematic reduction of gesture duration. The strategy for the timing of the gesture under different rate conditions seems to be different for the two subjects. Subject HU achieves a reduced gesture duration by an equal reduction of abduction and adduction, whereas subject JR mainly reduces the gesture duration by shortening the duration of the abduction; only in the margins of the rate variation, i.e. from slow rate to the lower end of normal, conversational rate and from the higher end of normal, conversational rate to very fast speaking rate does there seem to be a reduction in glottal adduction duration as well.

On the basis of the present data, a model of the glottal behaviour under the two speech conditions can be set up for each subject. The models are schematized in figure 15. From the two sets of models it might be deduced that the control of the glottal behaviour under the "loudness" conditions is exclusively concerned with the spatial realization of the gesture, while for the rate conditions it seems as if the control mechanism involves spatial as well as temporal organization. However, it has been shown above that the degree of glottal aperture under the rate conditions can be expressed as a function of the abduction duration. On the basis of this relationship the following hypothesis shall be advanced concerning strategies in the planning of the glottal gesture under the two speech conditions:

The strategy applied in the control of the gesture by increased "loudness" aims exclusively at a reorganization of the spatial realization, resulting in a wider glottal aperture at increased "loudness". As the timing of the gesture remains unchanged at increased "loudness", the rate of the glottal abduction is increased in order to achieve a greater glottal aperture. As for the rate condition, the strategy implies primarily a reorganization of the temporal realization, resulting in a reduced gesture duration at faster speaking rates. This result can be achieved in different ways by different subjects. The spatial realization of the gesture is controlled by the duration of the abduction; the O-C gesture is under these circumstances to be considered merely as a ballistic movement of constant velocity, the size of maximum glottal aperture being but a passive consequence of the timing of the ab- and adduction.

"LOUDNESS"



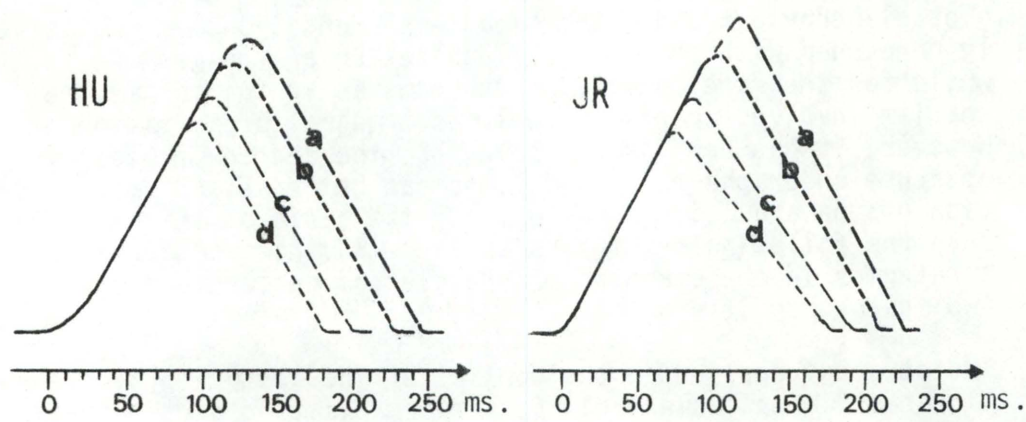
a = low "loudness" level

b = conversational speech low "loudness" level

c = conversational speech high "loudness" level

d = high "loudness" level

SPEAKING RATE



a = slow speaking rate

b = conversational speech slow speaking rate

c = conversational speech fast speaking rate

d = fast speaking rate

Figure 15

Schematized glottograms illustrating the O-C gesture of *p* at varying "loudness" and speaking rate (as for range of conversational speech cf. figure 4)

It is evident that the hypothesis advanced about the planning of the gesture must have interesting implications for the motor control of the glottal gesture. On the motoric level the increased rate of glottal abduction at increased "loudness" must presuppose an increase of the frequency of pulse firing to the posterior crico-arytenoid (henceforth PCA) muscle and/or a greater number of active motor units, i.e. it implies a change in the intensity but not in the timing of the motor commands to the abductor and probably the adductor muscles. At varying speaking rate the hypothesis presupposes an unchanged intensity of the motor commands to the PCA muscle and departs from the pattern of commands for the PCA-activity at slow rates. The relatively wide open glottis at the slowest rate must be a consequence of a rather strong PCA-activity, i.e. a high frequency pulse firing for the motor unit activation and/or a considerable number of active motor units; and an essential point in the hypothesis is that this motor program for PCA-behaviour is unchanged also at increased speaking rate, which means that the accumulated PCA-activity is the same and would have caused the same degree of glottal aperture as at the slowest rate; however, the size of the maximum aperture is decreased because the command to the interarytenoid muscle (henceforth INT) to an even higher degree will overlap the PCA-command and thereby the initiation of adduction will be triggered still earlier at faster tempo. Accordingly, the decrease of maximum aperture at increased tempo could be considered a mere consequence of the timing of the muscle command. As regards speaking rate control, the hypothesis assumes that the principles of Lindblom's (1963) model can be applied to the glottal behaviour of consonants as well.

It would have been useful to have EMG recordings of the laryngeal muscle activity of the present material in order to test at least that part of the hypothesis which is concerned with the motor control of the gesture. Unfortunately, only a very limited EMG material was available for this purpose. The material consisted of recordings from only one subject (HU) having spoken only a few repetitions of the phrase *"det er pile, de siger"* [d̥e: ˈb̥ʰiːlə d̥i ˈsiːʌ] with increased "loudness" and speaking rate. However, the tendencies that could be extracted from the material give some support to the hypothesis advanced above. Figure 16 illustrates the interaction between PCA- and INT-activity of *p* under the two speech conditions. It shall be emphasized that the values on the vertical axes are all quite arbitrary. It appears from the figure that increased "loudness" obviously causes a reorganization of the motor program as regards the intensity of the commands but not their timing. At increased speaking rate the PCA-activity remains essentially unchanged. However, the decrease of glottal aperture through the rate series seems not to be exclusively a consequence of an increased overlapping of the PCA- and INT-commands, since the INT-activity, besides the change in timing, is also increased through the rate series. Accordingly, this aspect of the hypothesis is only partly supported by the results.

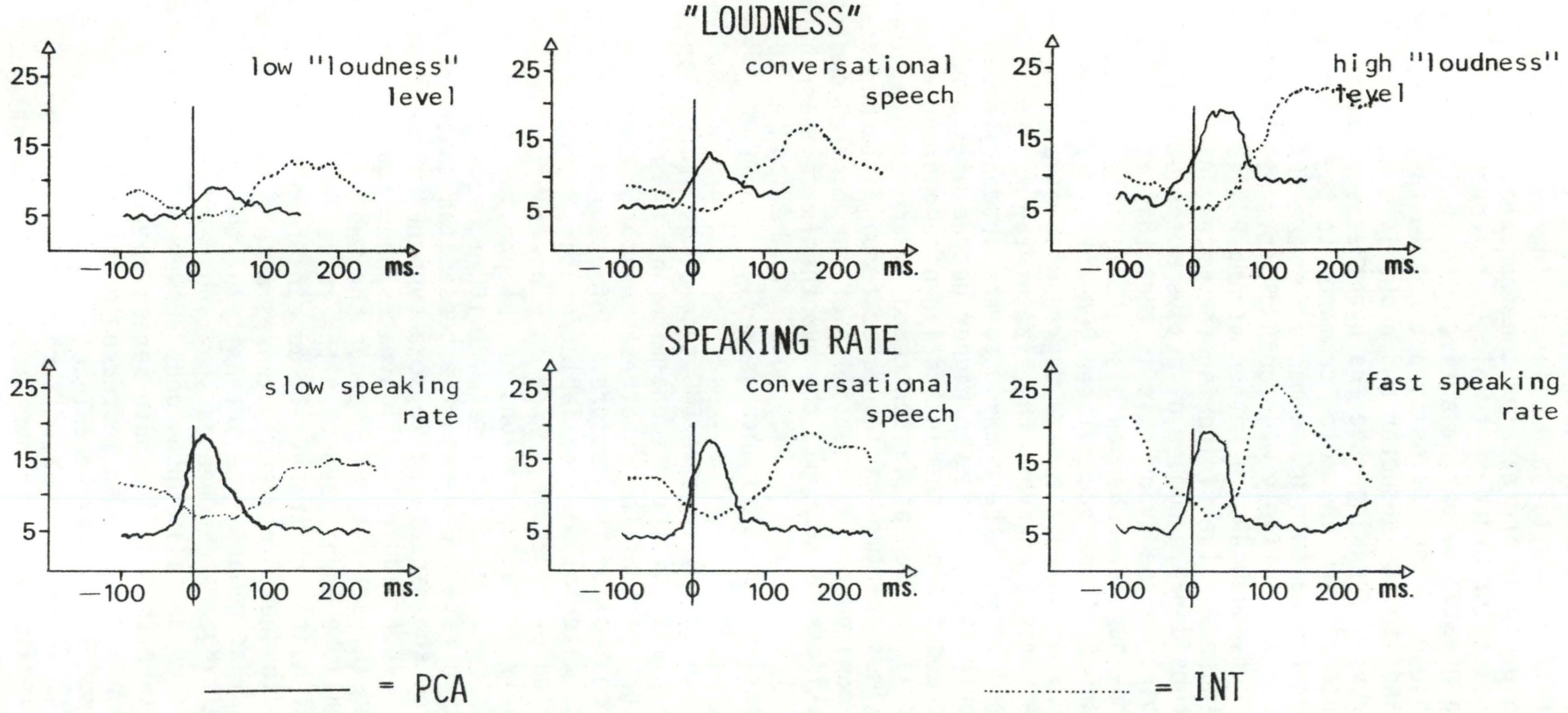


Figure 16

Tracings of PCA- and INT-activity of *p* in the phrase *det er pile, de siger* [ˌd̥e: ˈb̥ʰi:lə d̥i ˌsi:ɔ] at varying "loudness" levels and speaking rates (in this figure no distinction is made between different levels or rates of conversational speech). The implosion of *p* is used as line-up point. The scale used for the vertical axis is arbitrarily chosen.

The results of the present investigation do not give an unambiguous answer to the question raised in the introduction: whether an increase in speaking rate during stop consonant production implies a reorganization of glottal as well as supraglottal articulations in terms of faster rate of articulator movements and a change in the timing and intensity of the muscle commands. It appears, however, that the question of the existence of a common speaking rate control mechanism can be answered in the negative, if the production of a speech gesture is considered in terms of a coordinatory organization of two components: a movement towards and a movement away from a target, and if the comparison of supraglottal and glottal articulatory organization is made on this assumption. From the present investigation it appears that control of laryngeal behaviour implies an articulatory reorganization in terms of an increase or decrease in the rate of the glottal adduction only, while Gay et al. (1974) showed faster rate of both lip closing and lip opening gestures at increased speaking rate. At the muscular level it appears here as if only the command pattern of the muscle responsible for the movement away from the target is changed. In this connection, Gay et al. (1973) established a reorganization in timing as well as in intensity of the commands to the muscles responsible for both lip closure and lip opening gestures. If, on the other hand, the conclusion about the absence or presence of a reorganization of the driving force is to be based on changes in the amount of EMG activity of the total gesture (i.e. the movement both towards and away from the target), then the answer to the question is affirmative, as regards subject HU.

Since no EMG recordings of subject JR were included in the material, the question of PCA- and INT-interaction cannot be settled for this subject, but the different strategies employed by the two subjects in order to achieve a shorter gesture duration might indicate a difference between them in their organization of the laryngeal muscle activity. Whatever turns out to be subject JR's strategy of laryngeal control at increased speaking rate: one of pure timing changes, or one including also reorganization in terms of a decrease in muscle activity, neither of these alternatives corroborates the strategy of labial articulation proposed by Gay et al. (1974).

As mentioned in the introduction, it would not be surprising to reach a result like the one presented here, leaving the general impression that the supraglottal and glottal articulation of stops at increased speaking rate to a high degree work as two systems controlled by different mechanisms. Compared to the supraglottal gesture, the glottal gesture is assumed to be a simpler mechanism, less constrained as regards the necessity of reaching a fixed target, and with more degrees of freedom in its manner of contributing to the adequate aerodynamic characteristics of the stop.

Anyhow, most of the above considerations cannot be considered but as modestly supported hypotheses, and further investigations are needed, especially involving a more comprehensive study of the EMG-activity of the laryngeal muscles under different speaking rate conditions. Crucial points to be clarified seem to be: 1) Whether PCA-activity generally remains unchanged irrespective of the speaking rate (duration) of the stop. 2) The role of INT as a control mechanism by increased rate, i.e. can changes in the timing of the INT commands alone trigger the moment of maximum abduction, inferring that changes in the intensity of the commands have the separate function of regulating the rate of glottal adduction, or does an intention of earlier initiation of glottal adduction imply a change in timing as well as an increase of the INT-activity?

A problem connected to the questions posed above especially emphasizes the necessity of comprehensive EMG-studies. Generalizations about laryngeal control by varying speaking rate are often complicated, not only because different speakers might use different strategies, but perhaps even more by the fact that the different strategies are not necessarily different strategies of speaking rate control but can be assigned to general individual variations in laryngeal control. Sawashima et al. (1978) found individual variations in the sense that one subject was more dependent on the INT-muscle for laryngeal control, while the other relied more on the PCA-muscle. Bearing this in mind, it may not be safe to interpret the results presented concerning PCA-control of subject HU as being indicative of a special mode of laryngeal control at varying speaking rates, nor might we necessarily expect an EMG-recording of subject JR at increased speaking rate to display a PCA-behaviour equal to that of subject HU. Conclusions about strategies for speaking rate control with regard to laryngeal articulation presuppose a clarification of the general mode of laryngeal control for individual subjects.

V. NOTES

1. For example normalization of recording sections which display different linearity between glottogram amplitude and glottal aperture, the difference having arisen from alterations in the position of the fiberoptic cable caused by coughing, swallowing, or other reflex reactions of the subject.

ACKNOWLEDGEMENTS

The present paper is based on the author's thesis for the cand. phil. degree in phonetics. I wish to express my gratitude to Jørgen Rischel for his very constructive suggestions in connection with the present work. I am also indebted to Birgit Hutters and Peter Holtse for useful discussions, and to Svend-Erik Lystlund for technical assistance.

REFERENCES

Abramson, A.S., Lisker, L. and Cooper, F.C. 1965: "Laryngeal activity in stop consonants", *Status Rep. Speech Res., Haskins Labs. 4*, p. 6.1 - 6.13

Fischer-Jørgensen, E. and Hutters, B. 1981: "Aspirated stop consonants before low vowels, a problem of delimitation, its causes and consequences", *Ann. Rep. Inst. Phon. Univ. Cph., this issue*

Frøkjær-Jensen, B. 1967: "A photo-electric glottograph", *Ann. Rep. Inst. Phon. Univ. Cph. 2*, p. 5-19

Gay, T. and Hirose, H. 1973: "Effect of speaking rate on labial consonant production. A combined electromyographic/high-speed motion picture study", *Phonetica 27,1*, p. 44-56

Gay, T., Ushijima, H., Hirose, H. and Cooper, F.S. 1974: "Effect of speaking rate on labial consonant-vowel articulation", *J. Phonetics 2,1*, p. 47-63

Gay, T. 1978: "Physiological and acoustic correlates of perceived stress", *Language and Speech 21,4*, p. 347-353

Gay, T. 1978: "Effect of speaking rate on vowel formant movements", *J. Acoust. Soc. Am. 63*, p. 223-230

Harris, K.S., Gay, T., Sholes, G.N. and Lieberman, P. 1968: "Some stress effects on electromyographic measures of consonant articulation", *Status Rep. Speech Res., Haskins Labs. 13/14*, p. 137-152

Harris, K.S. 1971: "Vowel stress and articulatory reorganization", *Status Rep. Speech Res., Haskins Labs. 28*, p. 167-177

Harris, K.S. 1978: "Vowel duration change and its underlying physiological mechanisms", *Language and Speech 21,4*, p. 354-361

Kent, R. and Netsell, R. 1971: "Effects of stress contrasts on certain articulatory parameters", *Phonetica 24,1*, p. 23-44

Kiritani, S. 1971: "X-ray monitoring of the position of the fiberscope by means of computer controlled radiography", *Ann. Bull. Res. Inst. Logopedics and Phoniatrics, Tokyo, 5*, p. 35-39

Kiritani, S. 1972: "X-ray monitoring of the position of the fiberscope in the pharynx", *Ann. Bull. Res. Inst. Logopedics and Phoniatrics, Tokyo, 6*, p. 1-8

Kuehn, D.P. 1973: *A cinefluorographic investigation of articulatory velocities*, (Ph.D. thesis, University of Iowa)

Lindblom, B. 1963: "Spectrographic study of vowel reduction", *J. Acoust. Soc. Am. 34*, p. 1773-1781

Lisker, L. and Abramson, A.S. 1964: "A cross-language study of voicing in initial stops: acoustical measurements", *Word 20*, p. 384-422

Lubker, J.F. and Parris, P.J. 1970: "Simultaneous measurements of intraoral pressure, force of labial contact, and labial electromyographic activity during production of the stop consonant cognates /p/ and /b/", *J. Acoust. Soc. Am. 47,2*, p. 625-633

Öhman, S.E.G. 1967: "Word and sentence intonation: a quantitative model", *Speech Transm. Lab., Quart. Prog. and Status Rep., Royal Inst. Techn., Stockholm, 2-3*, p. 20-54

Sawashima, M., Hirose, H. and Yoshioka, H. 1978: "Abductor (PCA) and adductor (INT) muscles of the larynx in voiceless sound production", *Ann. Bull. Res. Inst. Logopedics and Phoniatrics, Tokyo, 12*, p. 53-60

Sussman, H. and MacNeilage, P. 1978: "Motor unit correlates of stress: preliminary observations", *J. Acoust. Soc. Am. 64,1*, p. 338-340

Sussman, H. 1979: "Motor unit discharge patterns during speech: temporal reorganization due to coarticulatory and prosodic events", *Proc. 9 Int. Congr. Phon. Sc. II, Copenhagen*, p. 365-371

(ARIPUC 15, 1981 p. 147-166)

THE ROLE OF INTRINSIC Fo AND DURATION IN THE PERCEPTION OF STRESS*

EVA ROSENVOLD

It has been found previously that the intrinsic differences in Fo level and duration, existing between vowels of different tongue height, are compensated for perceptually. The present experiment aims at examining the role of the intrinsic differences in connection with the perception of stress. It is found that the Fo step required to identify a certain stress pattern is of a different physical magnitude for a low than for a high vowel. Similarly, the duration required to identify a vowel as being stressed is not the same for low and high vowels. It appears from the results that the hypothesized perceptual compensation for the intrinsic differences functions in the perception of isolated vowels as well as in the perception and identification of linguistic category stress, which in its turn is dependent on Fo and duration.

I. INTRODUCTION

The present experiment is a contribution to the debate about the perceptual relevance of the so-called microprosodic phenomena. The object is, more specifically, to see whether the intrinsic differences in duration and fundamental frequency level between vowels of different tongue height influence the perception of stress.

---

* This is a condensed version of a thesis for the exam.art. degree in Phonetics. I am indebted to Niels Reinholt Petersen, Peter Holtse, and Nina Thorsen for useful guidance and helpful suggestions.

A. PERCEPTION OF STRESS

Various perceptual tests have shown that of the three acoustic correlates to stress (Fo, intensity, and duration) fundamental frequency seems to be the factor of primary importance. Fry (1958) found that Fo had an all-or-none effect for English, and Morton and Jassem (1965) report similar results for Polish, adding that the greatest effect was achieved by varying the Fo level of the second syllable in bisyllabic words. The authors note that the different vowel qualities used in the test words seemed to influence the results, and that further experiments taking these differences into account are called for. Thorsen (1978, 1979) likewise regards fundamental frequency as the primary cue to stress in Danish; duration may be a secondary cue.

B. INTRINSIC DIFFERENCES IN FO AND DURATION

Many experiments have proved the existence of a tendency for high vowels to have a higher Fo level and a shorter duration than low vowels, other things being equal. This tendency is manifest also in Danish, see Reinholt Petersen (1978), Fischer-Jørgensen (1955), and Bundgård (1980).

It is possible that the perception of Fo and duration proceeds regardless of those acoustic properties of the signal that convey information on vowel quality, and hence without compensation for intrinsic differences resulting from differences in tongue height. It is, however, also conceivable that the perception of Fo and duration is influenced by properties conveying information on vowel quality, thus involving a compensation. If the latter is true, we might expect this same compensatory process to be functioning in the case of stress perception cued by Fo and duration.

Two experiments have investigated the perceptual relevance of intrinsic differences. Reinholt Petersen (1974) had subjects categorize as phonemically long or short vowels of identical quality which occurred in a series of stimuli in which duration was increased in equal steps. He found that perceptual boundary between long and short vowel was dependent on vowel quality, the subjects requiring low vowels to be longer than high vowels in order for them to be classified as phonemically long. This implies that two vowels of different tongue height but of identical duration may be perceived as being different in duration (and hence in length, if the perceived durations are on either side of a phoneme boundary), and that two vowels of different tongue height and of different duration may be perceived as being of equal duration (and length), both kinds of results indicating a compensatory perceptual process. J. M. Hombert (1976) compared the intrinsic differences occurring in natural speech to the differences in cross-over points emerging from his test, and found that they were within the same range. From his experiment with isolated vowels of level Fo he concludes that there is a tendency for [ɑ+] to be judged higher in pitch than [i] or [u], but the perceived difference is smaller than the intrinsic difference. Both experiments thus point towards a compensatory process.

In a recent experiment Reinholt Petersen (1980) has found a certain degree of coarticulation between syllables with regard to inherent fundamental frequency level. In words of the type [ˈfimɑ+] and [ˈfimi] he first of all finds a general tendency for the Fo jump from the first syllable to the second to be of a greater magnitude when the second vowel is [i], attesting the expected intrinsic difference. Secondly, [ɑ+] in the second syllable may exhibit an upward shift in Fo as a coarticulatory effect from [i] of the first syllable, which indicates an articulatory compensatory process helping to maintain the difference in Fo level between the two syllables. This coarticulation is not, however, present for all subjects, and hence Reinholt Petersen concludes that the perceptual system is able to do without this articulatory compensation (the stress pattern: ˈ ̥ being identified in spite of intrinsic differences). It thus seems reasonable to assume that a perceptual compensatory process is at hand both in the case of isolated vowels and in the case of stress perception based on fundamental frequency.

On the basis of Hombert's and Reinholt Petersen's results we may hypothesize that a perceptual compensatory process will make up for the "deficiencies" of Fo and duration as cues to stress which are due to intrinsic differences. Such a hypothesis would involve that the Fo interval between two syllables required for the identification of a given stress pattern must be larger when the second vowel is a high vowel than when it is a low vowel (everything else being equal), and similarly, that low vowels must be longer than high vowels to be perceived as exceeding the limit for unstressed vowels.

The above hypothesis may be verified if the shift in identification from stress on the first syllable to stress on the second syllable occurs at different magnitudes of the physical Fo interval between the two syllables, and/or at different physical durations for [i] and [ɑ+], respectively, in the second syllable.

II. EXPERIMENTAL DESIGN

A. TEST MATERIAL

1. TEST WORDS: *bidbig* [b̥id̥b̥ig̊], *bidbag* [b̥id̥b̥ɑ+g̊]

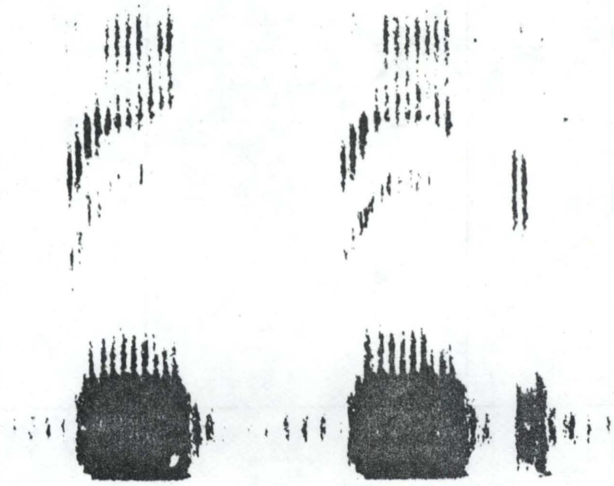
[i] and [ɑ+] represent the extremes of the intrinsic Fo and duration continuum. It is important that the vowels be perceived as speech sounds, and they were therefore embedded in disyllabic nonsense words. These nonsense words are structurally possible

in Danish. The surrounding consonants are unaspirated plosives (this type of consonant was preferred because with the synthesis procedure employed such consonants are easily given a natural sounding quality). The words were presented in isolation to avoid any influence on their perception from, say, a surrounding carrier sentence. They were synthesized on the parallel speech synthesizer of the Institute (cf. Rischel and Lystlund 1972). The synthesis was based on spectrograms of the two words as spoken by Danish listeners and on tables of Danish formant frequencies (Fischer-Jørgensen 1972, Holtse 1973). Figure 1 shows spectrograms of the resulting basic stimuli.

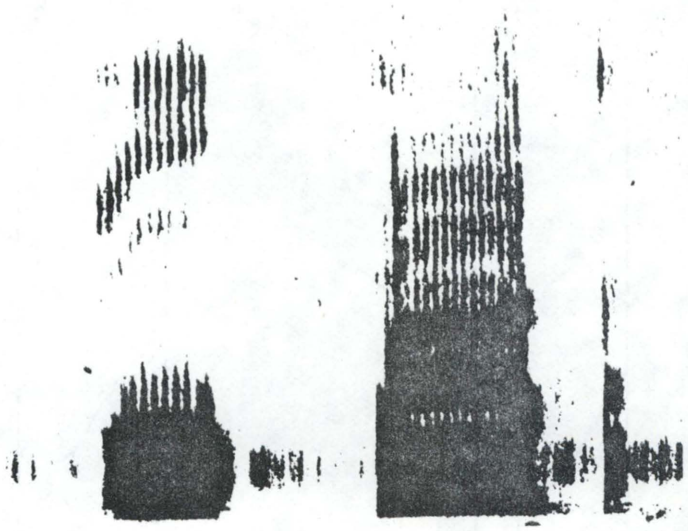
2. VARIATION OF FUNDAMENTAL FREQUENCY AND DURATION

a. Fundamental frequency. Figure 2a and 2b show Fo tracings of natural recordings of nonsense words similar to the test words of this experiment. In these recordings of real speech the two stress patterns depicted (ˈ ˳ = stress 1, and ˳ ˈ = stress 2) exhibit a difference in the direction as well as the slope of the Fo movement which, however, seems to be confined to the second syllable. The resemblance of the two stress patterns may appear puzzling but can be ascribed to the fact that we are dealing with one-word-utterances rendering any unstressed first syllable utterance initial, which accounts for the jump from such a syllable up to the stressed syllable. Stress 1 shows a jump to a higher level on the second syllable, as opposed to stress 2 which has an upward glide and a lower overall level in the second syllable. Due to its short average duration (10 cs) and small average range (20 Hz), this glide is probably below the threshold for audible glides (cf. Rossi 1971 and 't Hart 1977), i.e. it will be perceived as a level tone at a pitch corresponding to a frequency at 2/3 of the distance in time from vowel onset. The magnitude of the Fo jump can, therefore, be expected to be decisive for the identification of either stress 1 or stress 2 in this particular context.

For synthesizing the Fo contour, a line segment (a monotonous Fo function) of 20 cs's duration was superposed upon each syllable separately. Thus the first syllable could be kept constant and the second varied by moving its entire Fo line segment upwards. Both line segments had a slope of approximately 0.6 Hz per centisecond, far below the threshold for audible glides. This slope (falling in the first, and rising in the second syllable) was introduced to add naturalness to the stimuli. Subjective auditive criteria decided on the choice of absolute frequencies ranging from step 1: 85 Hz to step 6: 135 Hz. The first syllable was kept constant at 85 Hz. For a graphical display of the Fo variation and the frequency steps, see figure 3a.



[b̥id̥b̥iɡ̊]



[b̥id̥b̥ɑ+ɡ̊]

Figure 1

Spectrograms of synthesized testwords.

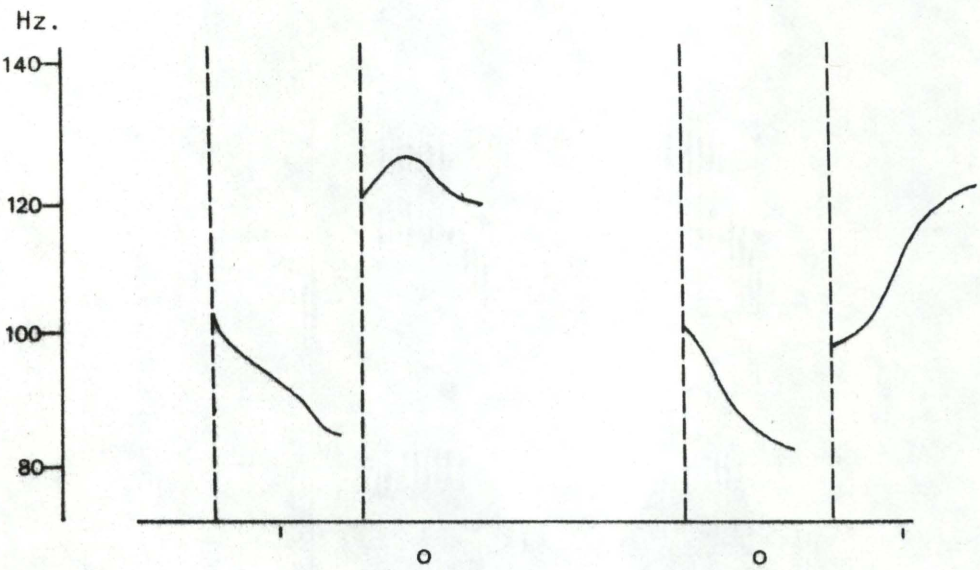


Figure 2a.

Avaraged tracings of 20 recordings of [b̥id̥b̥ig̊] and 20 of [b̥id̥b̥ɑ+g̊] , with the stresspatterns ' o (stress1) and o ' (stress 2). Line up point vowelstart.

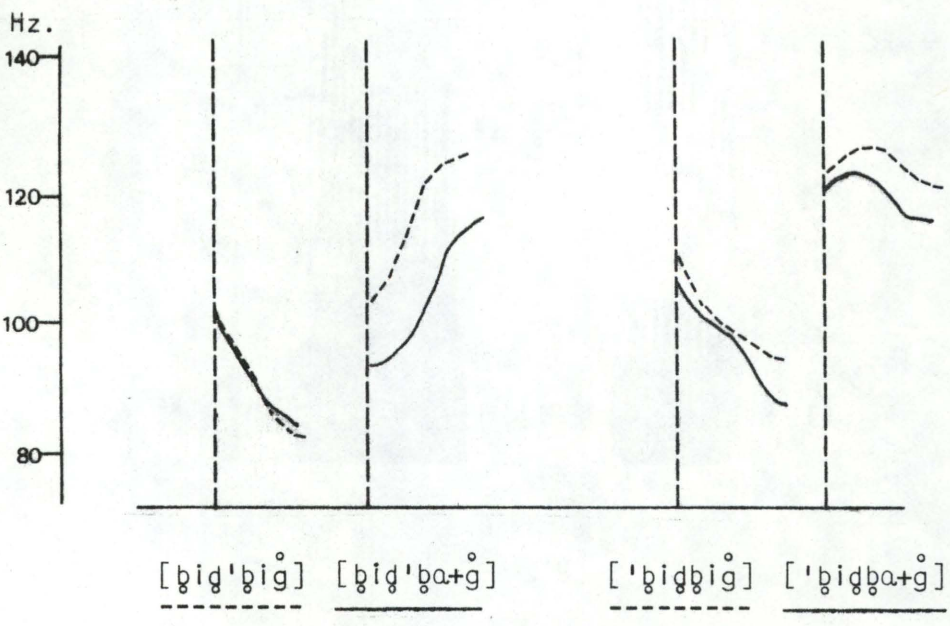


Figure 2b.

Averaged tracings of 20 recordings of the words [b̥id̥b̥ig̊] and [b̥id̥b̥ɑ+g] with either stress 1 or stress 2.

b. Duration The variation range of vowel duration was determined on the basis of average durations of the two vowels. It turned out that a stimulus with an [i] of a duration of 10 cs and an [ɑ+] of a duration of 13 cs had the desired natural effect both with regard to the intrinsic durational difference between the two vowels and the overall duration of the test word. 6 steps were placed around these values to provide a reasonable dispersion. The vowel of the second syllable could thus be varied in 8 steps, whereas [i] of the first syllable was kept constant at 10 cs. For graphical display of the variation of duration and the durational steps, see figure 3b.

In synthesizing the stimuli, the definition of "vowel duration" was based on the control of the voice source amplitude gate. The final stimuli were subject to auditive evaluation, according to which they were of an adequate duration without giving the impression of phonemically long vowels.

It is an essential limitation of this experiment that different criteria were applied to decide the Fo and duration steps. The variation of Fo was established on the basis of actual tracings of the two stress patterns in normal speech, whereas the variation of duration was intended only to ensure a reasonable dispersion around the normal average durations of the two vowels involved. There is no reason to assume that the Fo steps and the duration steps chosen are psychologically equidistant, and thus one cannot arrive at any final conclusions concerning the relative importance of various factors as a cue to stress in Danish on the basis of this experiment. The steps chosen do, however, make it possible to discuss the perceptual relevance of the intrinsic differences, which is the purpose of this test.

For the test every possible duration step could be combined with every possible Fo step for either [i] or [ɑ+], thus forming 96 stimuli types in all (8 x 6 x 2).

III. THE TEST

A. PREPARATIONS

Each stimulus was to be presented three times to the subjects. The three responses given per stimulus were subsequently averaged and thus reduced to one response (to be considered as a sort of random sampling). The three repetitions ought to ensure a reasonable degree of reliability in the results. A tape was recorded on a semiprofessional Revox recorder, 7½ ips. The stimuli were arranged in three randomized series, each comprising 96 stimuli (8x6 *bidbig* and 8x6 *bidbag*). On track No. 2 a pulse matching the start of every stimulus on track No. 1 was inserted in order to trigger a device for displaying numbers visually.

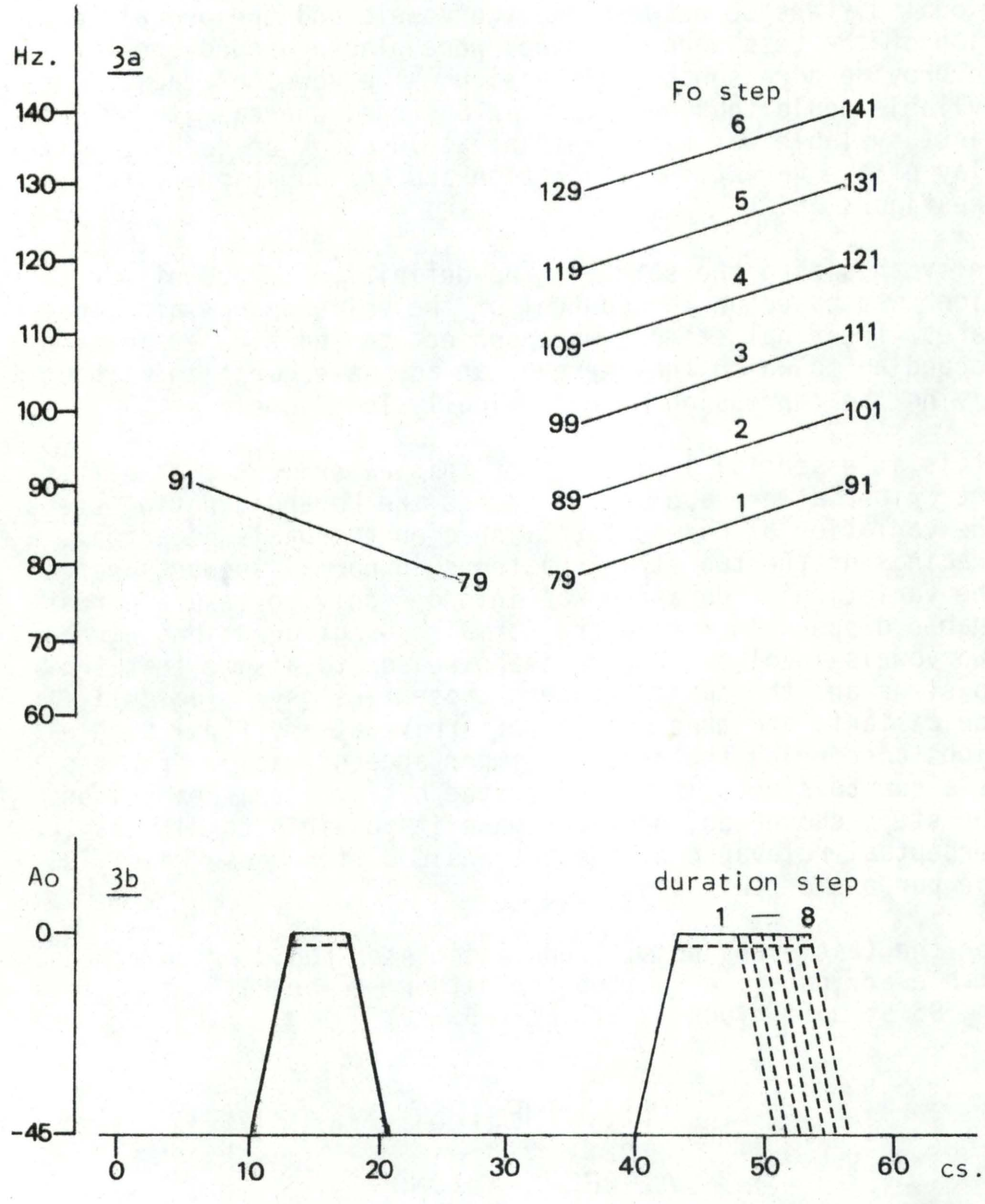


Figure 3a+b.

Variations of fundamental frequency and duration for the vowel of the second syllable, that of the first being kept constant. [ɑ+] 's intensity was slightly lowered to counteract the perceptual effect of the intrinsic intensity differences.

B. THE TEST

The test was run over three afternoon sessions holding 4, 6 and 10 subjects, respectively. The subjects were seated in a room covered with sound absorbing boards, and the stimuli were played back to them via loud speakers. The listeners were informed that they were going to hear the words *bidbig* and *bidbag* (presented in ortography), and they were instructed to note on their answer sheets whether they heard stress on the first or on the second syllable.

A TV monitor in front of the subjects displayed the stimulus numbers successively, enabling the subjects to check their progress.

All three tests were run in the same session with a 15 minutes break in between.

C. SUBJECTS

20 subjects took part in the test. As it was considered desirable to include naive listeners as well as trained phoneticians, the final group consisted of 10 phoneticians and 6 quasi-naive listeners (students attending a course in general phonetics for language students), all of whom were Danish, and in addition 4 foreign subjects who at that time attended the above mentioned course.

IV. RESULTS AND DISCUSSION

When first going through the answers it was found that 4 subjects had failed to perform the task, either because they did not follow the instructions or simply by failing to give answers. Each of the remaining 16 subjects had responded three times per stimulus, and simple majority of responses was used to decide on the preferred stress position for each stimulus (e.g. 2,2,1 = 2, or 1,1,2 = 2). This preferred stress position was from then on considered the answer. Any potential system in the responses could be read off immediately from the answer sheet. Nine out of the 16 subjects used the same criterion, while one subject showed a pattern that was very similar to that of the just-mentioned group, but reversed (see below). Two subjects (No. 11 + 12) displayed an apparently deviating and more complex pattern of response. The four foreign subjects had quite different linguistic backgrounds, and as the results for these subjects proved to be unsystematic between subjects as well as individually, this group is disregarded in the discussion of the results. The results for the "main group" (nine subjects) were pooled, while the results for the two subjects (No. 11 and 12) were left for separate treatment.

A. RESULTS FOR THE MAIN GROUP

The responses for the main group revealed a common principle for the nine subjects in question. They tended to switch from the identification of one stress pattern to the other at the same Fo step, and the influence of increased duration seems to have been of the same nature for all nine. The subjects showed agreement as to which stress pattern they associated with high and low Fo, respectively, on the second syllable. They agreed upon stress 1 at a comparatively high Fo on the second syllable, and on stress 2 at a comparatively low Fo. As already mentioned, subject No. 10 proved to have used the same differential criteria as the main group but favoured stress 1 responses at a low Fo on the second syllable, and stress 2 at a high Fo. Apart from this reversal, the results of this subject corresponded to those of the main group concerning the effect of the intrinsic differences. Figure 4a+b displays graphically the distribution of responses with the two vowel qualities pooled. The figure clearly reveals that a higher Fo in the second syllable causes a shift in identification, whereas increased duration has no such effect. As stated earlier, the steps in the two parameters, Fo and duration, cannot be said to be equidistant, and hence a definite answer to the question of primary cue to stress cannot be given; but we can conclude that variations of Fo in the second syllable can effect a shift in identification, and that increased duration of this syllable causes a decline in the number of stress 1 answers. (It might be argued that a further increase of duration could have led to a further decline and eventually to a shift in identification.)

In figures 5 and 6, vowel qualities are kept apart. Figure 5 shows the number of stress 1 responses as a function of increased fundamental frequency for each durational step, and conversely, figure 6 displays the number of stress 1 responses as a function of increased duration for each Fo step. For the sake of clarity, only stress 1 responses are plotted, and the 50% cross-over points are indicated by a vertical line. As an example, figure 5, graph 1 (dur 0) shows on the vertical axis how many stress 1 responses (out of a maximum of 9) that were given to stimuli with a vowel duration (second syllable) of step 0. The number of responses for each fundamental frequency step can be read off the horizontal axis (1-6). Thus it appears from this graph that when presented with e.g. *bidbag* dur. step 0 (the shortest stimulus) combined with Fo step 1 (the lowest Fo), 5 subjects judged the stimulus to have stress on the first syllable.

1. INTRINSIC Fo

To further investigate the difference between [i] and [ɑ+] with regard to the threshold value for the identification of one or the other stress pattern as revealed in figure 5, the 50% cross-over points were calculated for all graphs by linear interpolation. The values are plotted in figure 7. From this

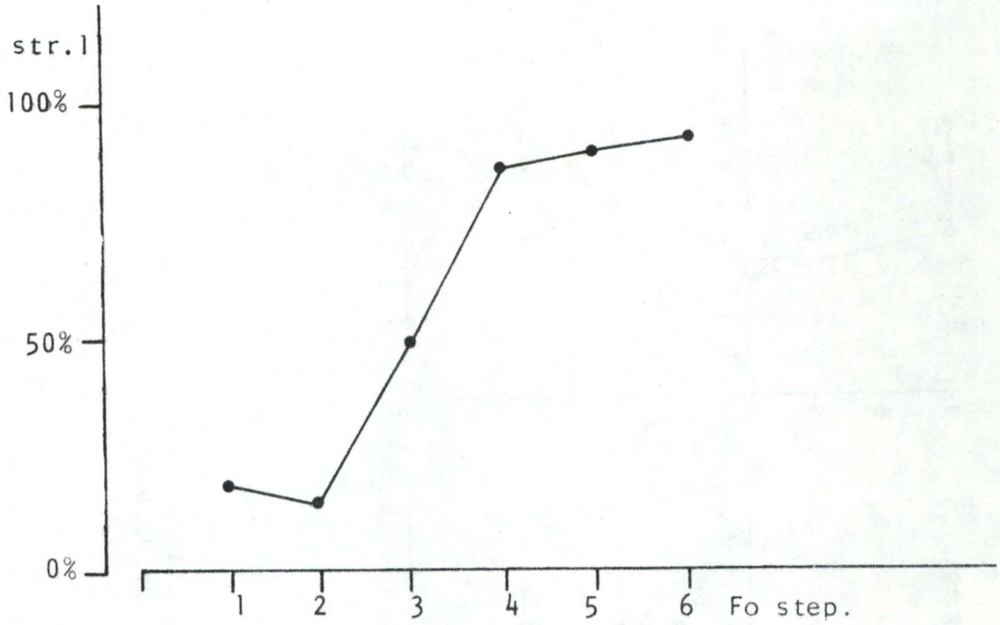


Figure 4a.

Per cent stress 1 responses given by the main group (9 subjects) as a function of increased Fo. The two vowels and all possible duration steps pooled.

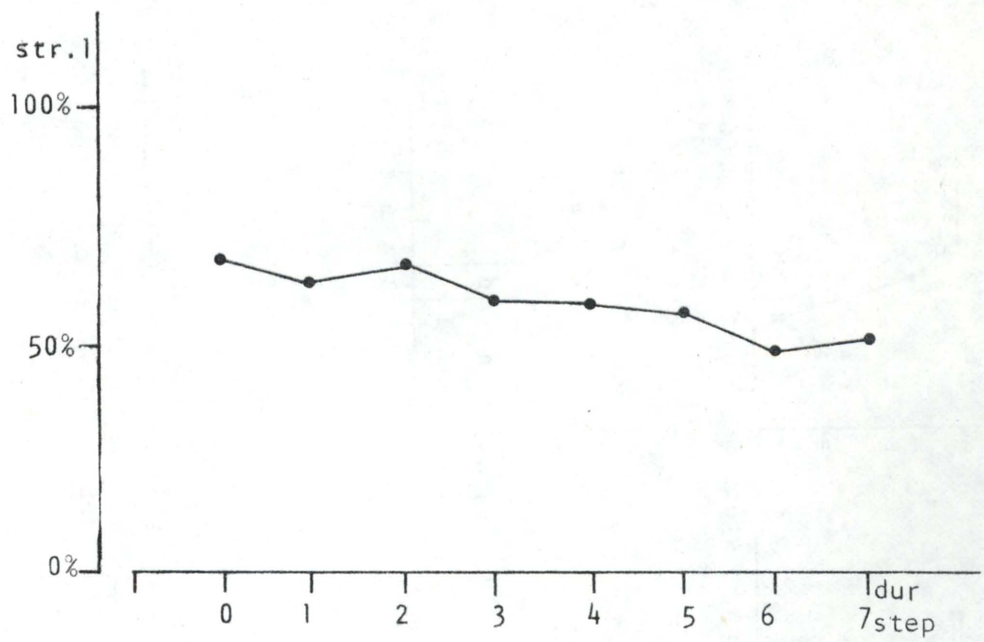


Figure 4b.

Per cent stress 1 responses given by the main group (9 subjects) as a function of increased duration. The two vowels and all possible Fo steps pooled.

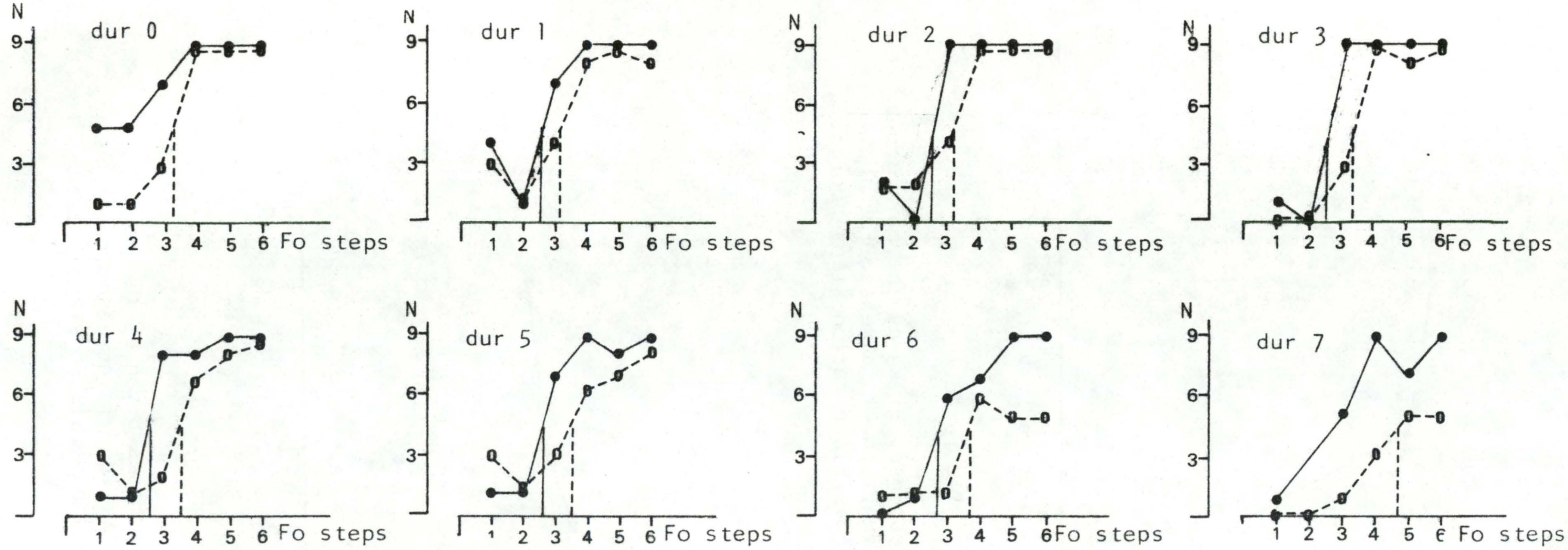


Figure 5.

Number of stress 1 responses for the main group (9 subjects) as a function of increasing Fo at each of the eight duration steps,------ *bidbig* ———— *bidbag*. In the case of duration step 7 the number of stress 1 responses for Fo step 2 is left out due to an error in the test material.

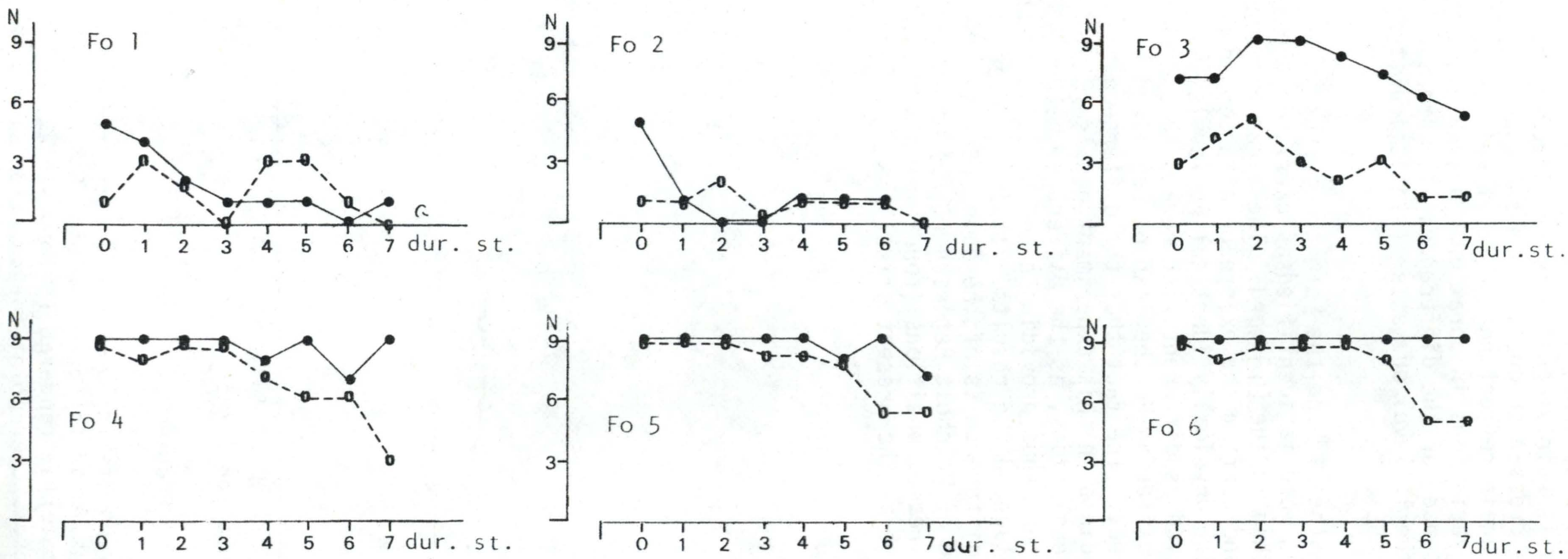


Figure 6.

Number of stress 1 responses for the main group (9 subjects) as a function of increasing duration at each of the six Fo steps, -------- *bidbig* ———— *bidbag*. In the case of Fo step 2 the number of stress 1 responses for duration step 7 is left out due to an error in the test material.

figure it appears that the distance between the [i] and the [ɑ+] cross-over points is fairly constant over different vowel durations and that the cross-over points form a horizontal line for both vowels, again reflecting the relatively insignificant influence of variations in duration. A one-tailed T-test proved significance for the difference, its direction, and its magnitude at the 1% level or better (statistical treatment based on Siegel 1956).

The difference between the [i] and the [ɑ+] cross-over points indicates that the [ɑ+] word is identified as carrying stress 1 at lower Fo levels than is the [i] word. These results conform to expectations and confirm the existence of a compensatory process. Even with physically identical Fo levels, [i] and [ɑ+] can be perceived as different with regard to stress type; with physically differing Fo they can be perceived as identical in this respect. The fact that the difference between the cross-over points in this experiment equals 10 Hz does not warrant any conclusions, as the difference between stimuli is 10 Hz per step, thus providing a quantization which obviously may have influenced the results. It is, however, interesting that this difference is of the same order of magnitude as the one found by Reinholt Petersen (1979) as an average for stressed/unstressed positions (for a voice within the same frequency range as the present stimuli).

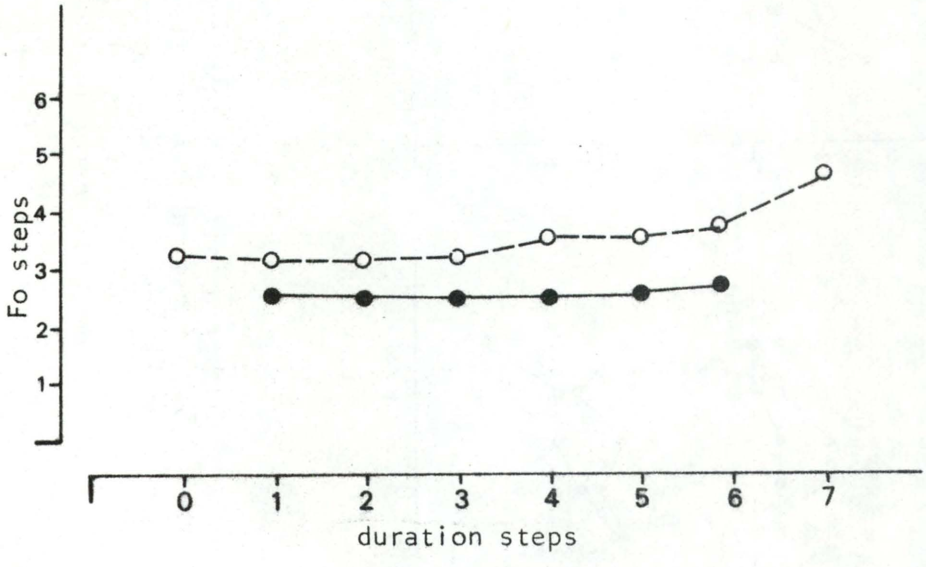


Figure 7.

50 per cent cross-over-points for *a*——— and *i* -------
No cross-over-point was found for duration step 0, and an error in the test material rendered it impossible to calculate the 50% cross-over-point for duration step 7.

2. INTRINSIC DURATION

As the graphs in figure 6 do not display a shift of identification but merely a decline in the number of stress 1 responses as a function of increased duration, no cross-over points can be calculated. Instead we must search for the potential [i]/[ɑ+] difference in the graphs displaying the lowest and the highest fundamental frequency steps. If one examines the low Fo steps which strongly favour stress 2 responses, the graphs show that the shortest durations cause more stress 1 responses for [ɑ+] than for [i]. Some subjects seem to have found these short durations altogether too short for the second syllable to carry stress, more often so with [ɑ+] than with [i]. In other words, a perceptual compensatory process is observable also with regard to duration.

An [i]/[ɑ+] difference appears again in the graphs for the highest Fo steps, which strongly favour stress 1 responses: whereas increased duration with Fo step 4 and higher causes a decline in the number of stress 1 responses for [i], such a change in duration does not influence the number of stress 1 responses for [ɑ+]. This suggests that since an [i] is perceived as being longer than an [ɑ+] of the same physical duration, it does not take as long a duration to make [i] appear too long to be unstressed as is the case with [ɑ+]. Thus these data also give support to the expected compensatory process.

The absence of cross-over points precludes a comparison between physical and psychological difference in the case of duration. It was found that the duration steps chosen for this experiment were too few and covered too small a range to exceed the duration acceptable for an unstressed vowel, which would otherwise have caused a decline in the number of stress 1 responses for [ɑ+] as well. As it is, one might imagine that a further increase of duration would have brought about a downward slope in the [ɑ+] curve, perhaps starting, as it does for [i], about 3 steps from the "normal duration" of the vowel of the second syllable relative to the [i] of the first syllable, i.e. three steps from step 5. It might, however, prove futile to search for durational limits for unstressed versus stressed vowels, considering that not all subjects judged [i] at duration step 4 and 5 too long to be unstressed. The limits we are looking for may turn out to be of a highly individual character, so that one cannot expect to reach unanimity, as was the case for the Fo jump. It suffices to conclude that the perceptual compensation for intrinsic durational differences is clearly detectable even in this rather restricted material.

B. RESULTS FOR SUBJECTS 11 AND 12

The results for these subjects are shown in figures 8 and 9. The subjects appear to have used duration as a cue to stress when presented with short durations in the second syllable (favouring stress 1 responses), while the Fo jump becomes dominant as a stress cue (favouring either stress 1 or stress 2 responses depending on its magnitude) as duration reaches a

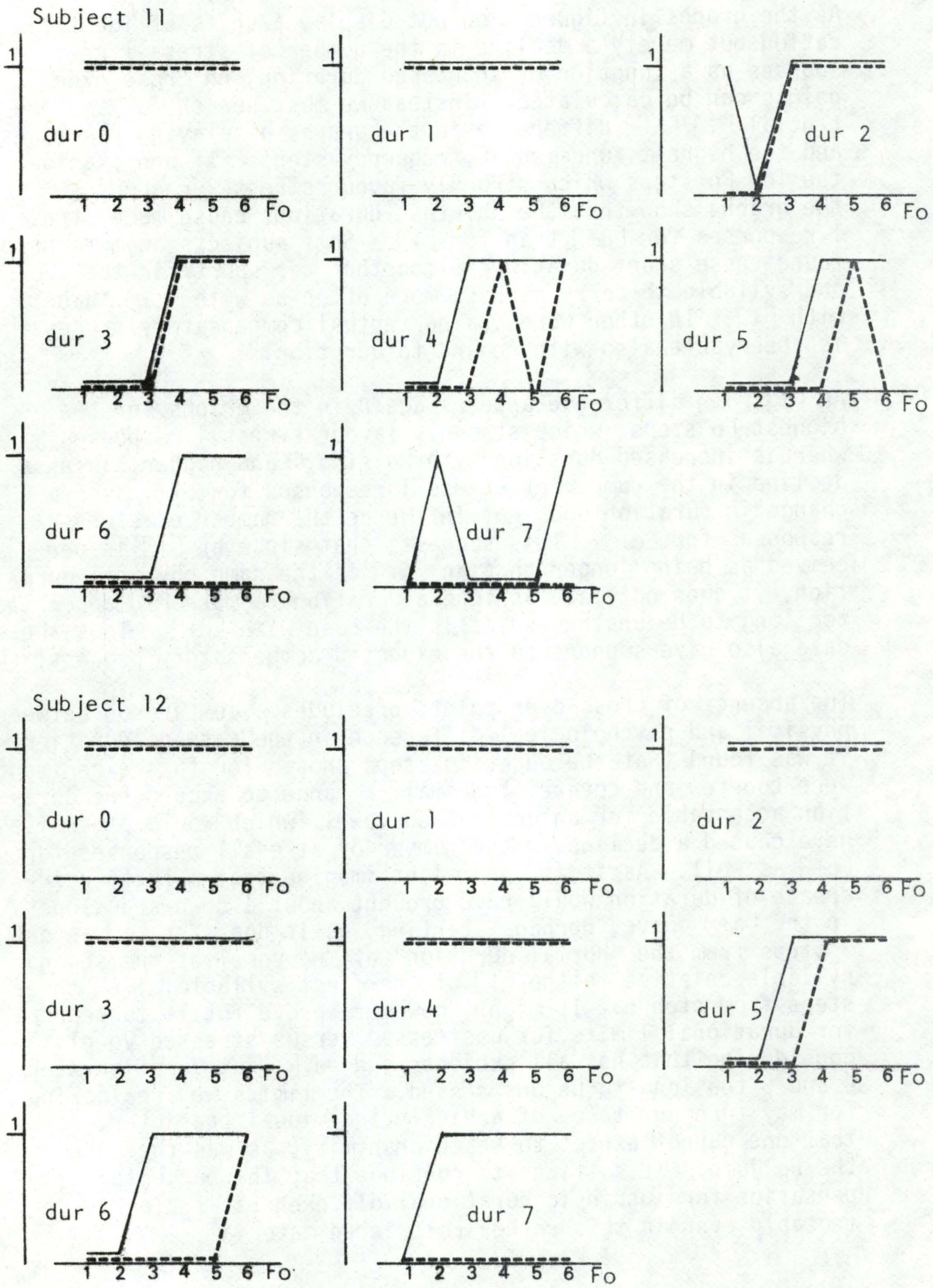


Figure 8.

Stress 1 responses for subjects no. 11 and 12 as a function of increasing Fo at each of the eight duration steps, ------ *bidbig* ——— *bidbag*.

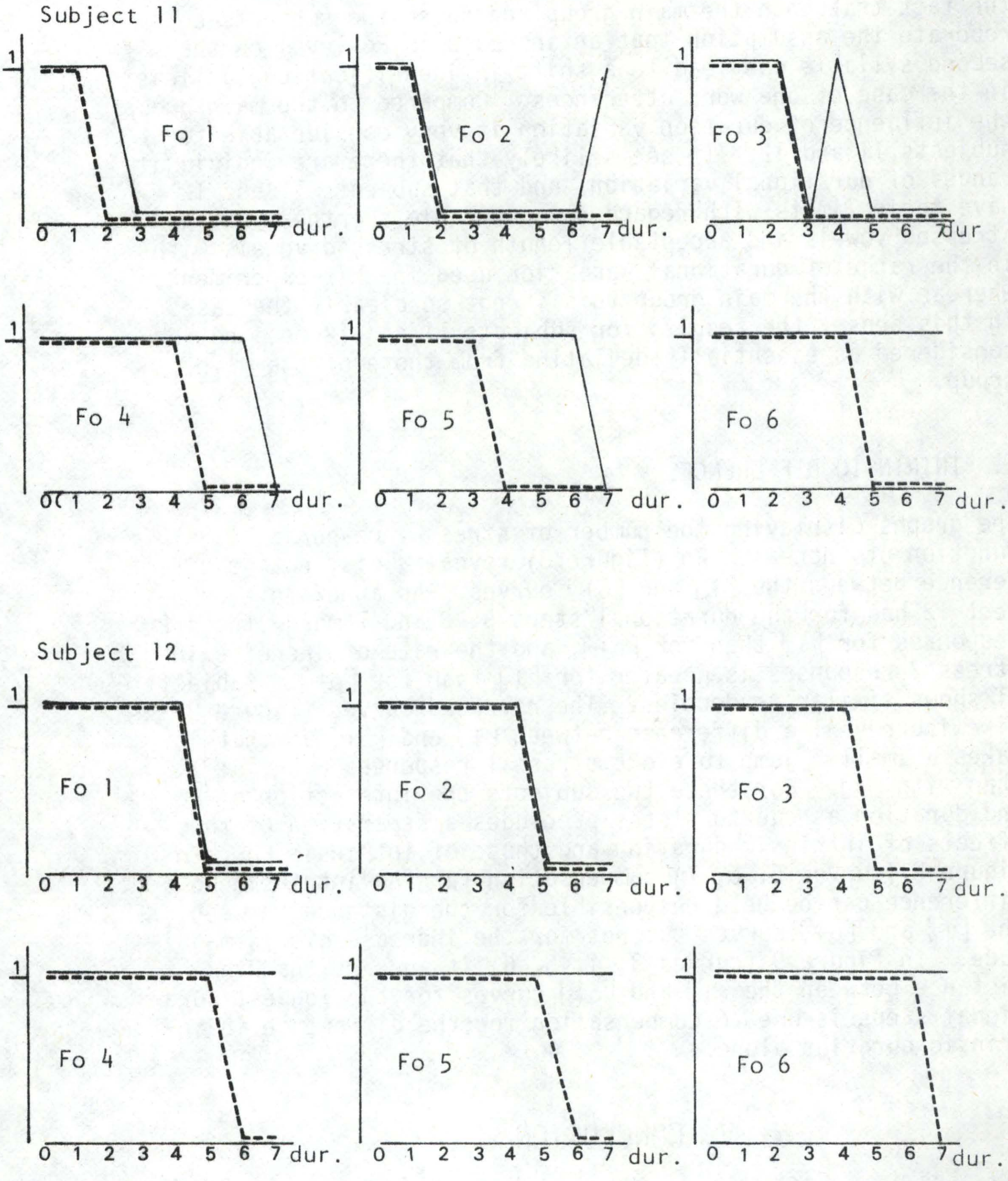


Figure 9.

Stress 1 responses for the subjects no. 11 and 12 as a function of increasing duration at each of the six Fo steps, -------- *bidbig* ———— *bidbag*.

certain critical point. It seems, however, worth pointing to the fact that both the main group and these two subjects corroborate the assumption that an increase in Fo level on the second syllable can lead to a shift in identification, at least in the case of one word utterances. Compared to the main group the influence of duration variation is very considerable for subjects 11 and 12. It seems likely that there are individual ranges of durational variation, and that subjects 11 and 12 have their limits with regard to acceptable shortness of unstressed vowels and acceptable length of stressed vowels within the range of durational variation used in this experiment, whereas with the main group this is not so clearly the case. In this sense, the results for subjects 11 and 12 need not be considered as essentially deviating from those of the main group.

1. INTRINSIC DIFFERENCES

The graphs displaying the number of stress 1 responses as a function of increased Fo (figure 8) reveal a distinctive difference between the [i] and [ɑ+] curves. As an example, subject 12 has for the durational steps 5, 6 and 7 more stress 2 responses for [i] than for [ɑ+], and the rate of increase in stress 2 responses is greater for [i] than for [ɑ+]. Subject 11 shows similar tendencies. The reversed curves (figure 9) likewise reveal a difference between [i] and [ɑ+] in that it takes a smaller jump to evoke stress 1 responses with [ɑ+] than with [i]. For these two subjects the interaction of Fo and duration as cue to stress precludes a separation of the effects of intrinsic duration and those of intrinsic Fo. In figure 8 (curves 5, 6, 7) compensation for the intrinsic Fo difference can be held responsible for the distance between the [i] and [ɑ+] curves but not for the increase of its magnitude. In figure 9 (curves 3, 4, 5, 6) it appears that the distance between the [i] and [ɑ+] curves for the longest durational steps is due to compensation for the difference in intrinsic duration alone.

V. CONCLUSION

This pilot experiment has yielded results which offer probable answers to the questions posed in the introduction, and thus it seems worth while (with relevant extensions of the material) to perform future experiments of a similar nature.

The results from the main group as well as from subjects 11 and 12 strongly suggest the existence of a perceptual compensation for intrinsic differences, as was hypothesized. A shift in identification from stress on the first syllable to stress on the second syllable occurs at different magnitudes of the physical Fo interval. [ɑ+] is perceived as being said on a higher pitch than [i] at physically the same Fo level, and consequently the shift in identification from stress on the first syllable to stress on the second syllable (which depends upon an

Fo jump of a certain magnitude) occurs at a smaller Fo step for [ɑ+] than for [i]. The range of durational variation applied in this experiment did not suffice to bring about an actual shift of identification. The results do, however, clearly indicate a perceptual compensation for the intrinsic duration difference as well. [i] is being perceived as longer than [ɑ+] of physically identical duration, and accordingly the limit at which the vowel of the second syllable is judged too long to be unstressed occurs at a shorter duration step for [i] than for [ɑ+], and [i] is judged long enough to carry stress at an earlier duration step than is [ɑ+]. These limits of duration for unstressed/stressed vowels seem to be of a highly individual character, but more experiments are called for, comprising a wider range of variation in duration.

The results of this experiment agree with Reinholt Petersen's (1974) findings concerning intrinsic duration and Hombert's (1976) results concerning intrinsic fundamental frequency, in affirming the existence of perceptual compensation. With reservations made for a certain bias due to the steps chosen, the difference between the cross-over points (the perceptual compensation) seems to be of a magnitude similar to the one found by Reinholt Petersen. The fact that Reinholt Petersen found instances in which a clear identification was made in spite of absence of a coarticulation (in terms of intrinsic Fo levels) between syllables, can be explained from the perceptual compensation suggested by this experiment.

It may then be regarded as an established fact that the perception of fundamental frequency and duration (at least at syllable and word level), and thereby the perception and identification of linguistic categories pertaining to these factors, is influenced by properties of the acoustic signal which convey information on vowel quality.

REFERENCES

Bundgaard, M. 1980: "An acoustic investigation of intrinsic vowel duration in Danish", *Ann. Rep. Inst. Phon. Univ. Cph. 14*, p. 99-119

Fischer-Jørgensen, E. 1955: "Om vokallængde i dansk rigsmål", *Nordisk Tidsskrift for Tale og Stemme 15*, p. 33-56

Fischer-Jørgensen, E. 1972: "Formant frequencies of long and short Danish vowels", *Ann. Rep. Inst. Phon. Univ. Cph. 6*, p. 189-213

Fry, D. B. 1958: "Experiments in the perception of stress", *Language and Speech 1*, p. 126-152

Holtse, P. 1973: "Identification and discrimination of closely spaced synthetic vowels", *Ann. Rep. Inst. Phon. Univ. Cph. 7*, p. 235-264

Hombert, J.-M. 1976: "Development of tones from vowel height", *Working Papers in Phonetics, UCLA, 33,* p. 55-66

Morton, J. and Jassem, W. 1965: "Acoustic correlates of stress", *Language and Speech 8,3,* p. 159-181

Reinholt Petersen, N. 1974: "The influence of tongue height on the perception of vowel duration in Danish", *Ann. Rep. Inst. Phon. Univ. Cph. 8,* p. 1-10

Reinholt Petersen, N. 1978: "Intrinsic fundamental frequency of Danish vowels", *J. Phonetics 6,* p. 118-127

Reinholt Petersen, N. 1980: "Coarticulation of inherent fundamental frequency levels between syllables", *Ann. Rep. Inst. Phon. Univ. Cph. 14,* p. 317-354

Rischel, J. and Lystlund, S. E. 1972: "A formant-coded speech synthesizer", *Ann. Rep. Inst. Phon. Univ. Cph. 6,* p. IX

Rossi, M. 1971: "Le seuil de glissando ou seuil de perception des variations tonales pour les sons de la parole", *Phonetica 23,* p. 1-33

Siegel, S. 1956: *Nonparametric statistics for the behavioral sciences,*

't Hart, J. 1977: "Vers une base psychophonētique de la stylisation intonative", *8emes journées d'études du groupe de la Communication parlée,* p. 167-174

Thorsen, N. 1978: "An acoustical investigation of Danish intonation", *J. Phonetics 6,* p. 177-189

Thorsen, N. 1979: "Lexical stress, emphasis for contrast, and sentence intonation in Advanced Standard Copenhagen Danish", *Ann. Rep. Inst. Phon. Univ. Cph. 13,* p. 59-86

INSTITUTE OF PHONETICS
JULY 1, 1980 - JUNE 30, 1981

I. PERSONNEL OF THE INSTITUTE

PROFESSOR:
Eli Fischer-Jørgensen, dr.phil.h.c. (until February 28, 1981)

ASSOCIATE PROFESSORS:
Børge Frøkjær-Jensen, cand.mag. (seconded to the Audiologopedics Research Group)
Peter Holtse, cand.phil.
Birgit Hutters, cand.mag.
Nina Thorsen, cand.phil.
Oluf Thorsen, cand.mag.

ASSISTANT PROFESSOR:
Niels Reinholt Petersen, cand.phil.

TEACHING ASSISTANTS:
Michael Bundgaard, stud.mag.
John Jørgensen, stud.mag.
Peter Molbæk Hansen, cand.mag.
Eva Rosenvold, cand.mag.
Lisbeth Strøjer, cand.phil., MA

ENGINEERS:
Otto Bertelsen, M.Sc. (from November 1, 1980)
Preben Dømler, B.Sc.
Mogens Hald Kristensen, M.Sc. (until September 30, 1980)
Mogens Møller, M.Sc. (on leave) (until September 30, 1980)

TECHNICIAN:
Svend-Erik Lystlund

SECRETARY:
Else Parkmann

TEACHERS FROM OTHER INSTITUTES:
Una Canger, Ph.D. (Institute of Linguistics)

GUEST RESEARCHERS:
Eli Fischer-Jørgensen, dr.phil.h.c. (from March 1, 1981)
Dr. Dieter Mehnert (from June 11, 1981 to June 25, 1981)

II. PUBLICATIONS BY STAFF MEMBERS

Eli Fischer-Jørgensen "Fonetik, studiet af sproglyde", Grundvidenskaben idag, 27, 1981

Nina Thorsen "Intonation contours and stress group patterns in declarative sentences of varying length in ASC Danish", Nordic Prosody II (ed. Thorstein Fretheim), TAPIR 1981, p. 75-89

Nina Thorsen "Word boundaries and Fo patterns in Advanced Standard Copenhagen Danish", Phonetica 37, 1980, p. 121-133

Nina Thorsen and Oluf Thorsen Fonetik for Sprogstuderende, 3rd edition, 4th printing, Copenhagen 1981, 169 pp.

Oluf Thorsen, Ole Kongsdal Jensen, and Karen Landschultz: Fransk Fonetik, 10th revised edition, Copenhagen 1981, 272 pp.

III. LECTURES AND COURSES

1. ELEMENTARY COURSES IN GENERAL PHONETICS AND LINGUISTICS

One semester courses (two hours a week) in elementary general phonetics (intended for students of phonetics and linguistics) were given by Birgit Hutters in the autumn semester 1980 and by Peter Holtse in the spring semester 1981.

A course in general and French phonetics (two hours a week) was given by Oluf Thorsen in the autumn semester 1980 and in the spring semester 1981.

A course in general and German phonetics (two hours a week) was given in the spring semester 1981 by Michael Bundgaard.

A course in general phonetics for students of Modern Greek (two hours a week) was given in the autumn semester 1980 by Eva Rosenvold.

A course in general phonetics for students of Italian (two hours a week) was given in the autumn semester 1980 by John Jørgensen.

Courses in general and Russian phonetics (two hours a week) were given by Peter Molbæk Hansen both in the autumn semester 1980 and in the spring semester 1981.

An introductory course in general linguistics (two hours a week) was given in the autumn semester 1980 by Una Canger.

2. PRACTICAL EXERCISES IN EAR-TRAINING AND PHONETIC TRANSCRIPTION

Based on tape recordings as well as work with informants, three semester courses of two hours a week were given as follows:

Beginners, in the autumn semester 1980 by Niels Reinholt Petersen.

Intermediate level, in the spring semester 1981 by Niels Reinholt Petersen and Peter Holtse.

Advanced students, in the autumn semester 1980 by Oluf Thorsen.

3. PHONOLOGY

Courses in phonology (two hours a week) were given by Lisbeth Strøjer for beginners in the spring semester 1980, and for more advanced students in the autumn semester 1981.

4. PHYSIOLOGY AND ACOUSTICS OF SPEECH

A course in the physiology of speech (two hours a week) was given in the autumn semester 1980 by Birgit Hutters.

A course in instrumental physiological phonetics (two hours a week plus individual exercises) was given by Birgit Hutters in the spring semester 1981.

A course in the acoustics of speech (two hours a week) was given by Nina Thorsen in the spring semester 1981.

A course in instrumental acoustic phonetics (two hours a week plus individual exercises) was given in the autumn semester 1980 by Peter Holtse and Niels Reinholt Petersen.

A course in elementary mathematics and electronics (two hours a week) was given by Nina Thorsen in the autumn semester 1980.

5. OTHER COURSES

A course in statistics (two hours a week) was given by Niels Reinholt Petersen in the spring semester 1981.

A course in the theory and practice of the language laboratory (one hour a week) was given by Oluf Thorsen in the spring semester 1981.

Birgit Hutters presided over a seminar for advanced students on normal and pathological functions of the velum in the spring semester 1981.

Courses in the phonetics and the phonology of particular languages (each two hours a week) were held as follows:

English, autumn semester 1980 by Peter Holtse.

Danish, autumn semester 1980 by Nina Thorsen.

French, spring semester 1981 by Oluf Thorsen.

German, spring semester 1981 by Nina Thorsen.

A course in sound typology (two hours a week) was given by Eli Fischer-Jørgensen in the autumn semester 1980.

6. SEMINARS

Professor Gunnar Fant (Royal Institute of Technology, Stockholm) lectured on the voice source in speech.

Professor Peter MacNeilage (University of Texas at Austin) gave a lecture titled: "The production of speech: Development and dissolution of motoric and premotoric processes".

Dr. Mona Lindau (University of California, Los Angeles) gave a lecture titled: "Phonetic differences between languages".

Professor Paul Kiparsky (Massachusetts Institute of Technology, Cambridge) gave three lectures titled: "Lexical phonology", "Analogy", and "Vowel harmony".

Dr. James Lubker (University of Stockholm) gave a lecture titled: "Normal and pathological velo-pharyngeal function".

Professor Ronald Netsell (Boys Town Institute for Communication Disorders in Children, Omaha) lectured on: "Speech production as a special case of motor skill".

Professor Peter MacNeilage (University of Texas at Austin) lectured on speech production.

Professor Alan Bell (University of Colorado, Boulder) gave a lecture titled: "Perception of rhythmic structures in speech".

Dr. Dieter Mehnert (Humboldt-Universität zu Berlin) lectured on microintonation applied to technical speech synthesis.

IV. PARTICIPATION IN CONGRESSES ETC.

Nina Thorsen participated in a meeting on "Prosody and synthesis by rule" in Stockholm January 16-17, 1981.

Nina Thorsen and Oluf Thorsen participated in a meeting on "Phonetic Description" in Lund May 15, 1981.

Nina Thorsen participated in the "Sixth Scandinavian Conference of Linguistics" in Røros, June 19-21, 1981, and gave a paper on: "Phonetic evidence for the autonomy of prosodic categories in Danish".

Nina Thorsen visited the phonetics institutes in Cologne and Munich on July 14-15, 1981, and lectured on "Stress, pitch, and intonation in Danish".

V. INSTRUMENTAL EQUIPMENT OF THE LABORATORY

The following is a list of instruments that have been purchased or built during the period July 1, 1980 - June 30, 1981.

1. AMPLIFIERS

1 Power amplifier, Nikko, type NA 590.

2. EQUIPMENT FOR EDP

1 Graphics tablet, Tektronix, type 4953.

1 Joystick, Tektronix, type 4952.

3. INSTRUMENTATION FOR VIDEO

1 Video monitor, Barco, type CRM 2032.

1 Video monitor, JAI, type JAI 810.

1 Video camera, JAI, type JAI 710A.