ARIPUC VOL. 18

ISSN 0589-6681

ANNUAL REPORT

of the
Institute of Phonetics
University of Copenhagen

96, Njalsgade
DK 2300 Copenhagen

COPENHAGEN 1984

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

HUMANISTISK FAKULTET
REPRO AFDELINGEN

CONTENTS

Global and local fundamental frequency variation and larynx height: Some preliminary observations (Niels Reinholt Petersen) 1

Variability and invariance in Danish stress group patterns (Nina Grønnum Thorsen) 13

The acoustic manifestation of stress in Danish with particular reference to the reduction of stress in compounds (Eli Fischer-Jørgensen) 45

The perception of voice onset time: A cross language study of American English and Danish (Jens B. Christensen) 163

Intonation and text in Standard Danish (Nina Grønnum Thorsen) 185

A note on work in progress: secondary articulation in Thai stops (Jørgen Rischel and Amon Thavisak) 243

Some basic vowel features, their articulatory correlates and their explanatory power in phonology (Eli Fischer-Jørgensen) 255

On the relation between vowel height and front back: A comment on Eli Fischer-Jørgensen's paper "Some basic vowel features, their articulatory correlates and their explanatory power in phonology" (Hans Basbøll) 277

A note on vowel triangels and quadrangles - an answer to Hans Basbøll (Eli Fischer-Jørgensen) 285

Vocal fold adjustment in Danish voiceless obstruent production (Birgit Hutters) 293

Editorial note

A change in grant policy has necessitated a change in publication dates. The present volume only covers the fall term of 1983, and the account of personnel, publications, courses etc. is postponed to the next volume of ARIPUC (vol. 19, 1985) which will cover the calendar year 1984.

ARIPUC 18, 1984, 1-12

GLOBAL AND LOCAL FUNDAMENTAL FREQUENCY VARIATION AND LARYNX HEIGHT: SOME PRELIMINARY OBSERVATIONS

NIELS REINHOLT PETERSEN

The paper presents data on the relation between fundamental frequency declination and larynx height, and between the stress determined fundamental frequency variation and larynx height. The results can be summarized as follows: a) Fo declination seems to be accompanied by a concomitant declination of larynx height, b) the Fo rise from stressed to the first posttonic syllable in Standard Danish is accompanied by two different patterns of larynx height variation: in one the larynx rises, in the other it lowers. The findings are discussed in the light of theories for the production of Fo.

I. INTRODUCTION

The fundamental frequency (Fo) variation in a sentence may be described as the composite result of the superposition of Fo movements associated with units of greater or lesser temporal scope, ranging from the text, through the utterance, sentence and phrase to the prosodic stress group. The segmentally conditioned Fo variation will not be considered here.

The more locally determined Fo variation is commonly agreed to be caused by the activity of the intrinsic laryngeal muscles, primarily the cricothyroid muscle (see e.g. Collier 1975, Atkinson 1978, Honda 1983, and Ohala 1978, who gives an extensive review of the literature), and to some degree assisted by the extrinsic laryngeal muscles, i.e. the inferior extrinsic muscles (the sternohyoid, sternothyroid, and thyrohyoid muscles) for Fo lowering, and the geniohyoid muscle for Fo raising. (See further Ohala 1970, Sawashima, Kakita, and Hiki 1973, Collier 1975, Erickson and Atkinson 1976, Atkinson and Erickson 1977, Erickson, Liberman, and Niimi 1977, Atkinson 1978, Honda 1983.)

The physiological mechanisms that underlie the overall fundamental frequency declination, which seems to be universally associated with terminal declarative utterances, are less agreed upon by researchers, and they have indeed been subject to less experimental research. One explanation was brought forward by Lieberman (1967), who, within his theory that claims subglottal pressure to be the major source of Fo variation, associated this type of Fo declination with the subglottal pressure declination which can be observed during such an utterance. The theory in its general form is hardly tenable (see Ohala 1978 for a review on the "larynx vs. lungs"-controversy), but the results reported in Collier 1975, where the EMG activity of a number of laryngeal muscles were examined together with subglottal pressure in Dutch sentences of varying stress patterns, indicate that the slow Fo declination could only be accounted for by subglottal pressure (whereas the local Fo variation superimposed upon it had to be - and could be - explained in terms of the activity of the laryngeal muscles examined). This explanation is repeated in Cohen, Collier, and 't Hart 1982, but recently (Collier 1983) data have been reported which suggest that it only applies under certain conditions of stress distribution in the sentence.

Another explanation - although also referring to the pulmonic system - has been advanced by Maeda (1974). On the basis of the observation that the vertical position of the larynx - like Fo - shows an overall decline over the utterance, it is hypothesized that the decreasing lung volume causes a continuous lowering of the sternum, which via ligaments and muscular tissue pulls the larynx downwards with a fundamental frequency decline as the result. Other researchers, however, have not found the larynx lowering reported by Maeda; on the contrary, Gandour and Maddieson (1976) and Ewan (1979) report that - if anything - the larynx rises during the utterance.

Maeda (1979) gives a modified version of the hypothesis. In the material presented he also fails to find the Fo decline to be accompanied by a larynx height decline[1], but he finds that the laryngeal ventricle shows an overall tendency to shorten during the utterance. Under the assumption that ventricle length can be used as an estimator of vocal fold length, the following mechanism is hypothesized: The decreasing lung volume exerts a pull upon the trachea. This pull is transferred to the cricoid cartilage and - if the thyroid cartilage is fixed - is thought to tilt the cricoid cartilage in relation to the thyroid cartilage, whereby the vocal folds are shortened and, consequently, Fo lowered.

Now, the explanations outlined above imply that gross fundamental frequency declination is an automatic consequence of inherent properties of the speech production apparatus - or more specifically, of the pulmonic system. The mere fact that Fo declination is so widespread in terminal declaratives among the languages of the world speaks in favour of this view. On the other hand, there is evidence which quite convincingly supports the notion that Fo declination is actively and pur-

posefully controlled by the speaker: Fo declination can convey information about sentence type and function as is the case in Danish (Thorsen 1979, and 1980a) where terminal declarative sentences have the steepest declination, syntactically and lexically unmarked questions have no Fo declination, and non-final periods and interrogative sentences with word-order inversion and/or interrogative particle have intermediate degrees of declination (cp. figure 1 below). The slope of declination tends to be steeper in short than in long sentences (see e.g. Maeda 1974, Sorensen and Cooper 1980, Thorsen 1980b and 1981). This means, if declination were a by-product of the function of the pulmonic system, that the expenditure of air per unit time should be greater in short than in long utterances. This is not very probable. Resetting (partial or complete) of the declination line has been shown to take place without intervening inhalation (Sorensen and Cooper 1980, Thorsen 1980b and 1981).

The aim of the present paper is not to point at answers (let alone decisive ones) to the questions of the nature of the physiological mechanisms responsible for Fo declination, or whether declination is actively controlled or not, but merely to present a small corpus of data on grosser and finer Fo and larynx height variation, which may be taken into account in the further discussion of the questions.

II. METHOD

A. MATERIAL

The point of departure for the design of the material was Thorsen's model for Danish intonation (see e.g. Thorsen 1980a) which is reproduced in figure 1. It is seen that in declarative sentences the prosodic stress groups are superimposed upon a gradually declining line connecting the stressed syllables. The prosodic stress group in Standard Danish is constituted tonally by a relatively low stressed syllable followed by a high-falling tail of unstressed ones.

The test word used was the nonsense word ['fi:fi] in which the local Fo variation would be represented by the rise from the stressed syllable to the first posttonic. The variation through the sentence could be observed by inserting the test word in different places in a carrier sentence. In order to make it easier for the subjects, however, the test word was inserted in three sentences as follows:

1) *i* ['fi:fi] *forkortes vokalen* [i 'fi:fi fʌ'g̊ʰɒ:d̥əs vo'g̊ʰæʔlɳ]

2) *vokalen i* ['fi:fi] *forkortes* [vo'g̊ʰæʔlɳ i 'fi:fi fʌ'g̊ʰɒ:d̥əs]

3) *vokalen forkortes i* ['fi:fi] [vo'g̊ʰæʔlɳ fʌ'g̊ʰɒ:d̥əs i 'fi:fi].

Since the three sentences were identical with respect to stress distribution and vowel length pattern (the stressed syllables, which are underlined, were all long), it was thought justified

to use this procedure and yet be able to treat the data as if they had been extracted from the first, second, and third stress group of one and the same sentence.

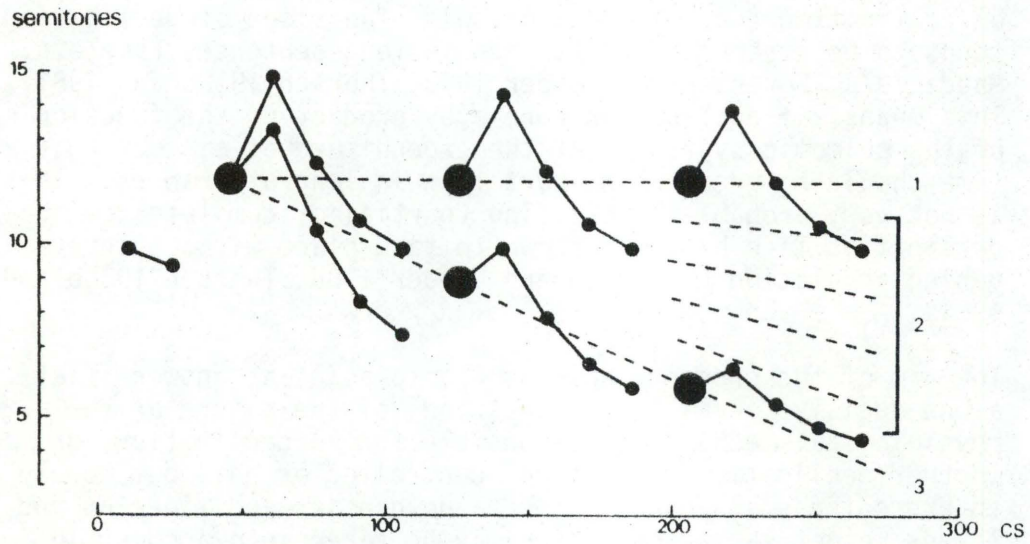


Figure 1

A model for the course of Fo in short sentences in ASC Danish. 1: syntactically unmarked questions, 2: interrogative sentences with word order inversion and/or interrogative particle and non-final periods (variable), 3: declarative sentences. The large dots represent stressed syllables, the small dots unstressed ones. The full lines represent the Fo pattern associated with stress groups, and the broken lines denote the intonation contours. Zero on the logarithmic frequency scale corresponds to 100 Hz. (Reproduced from Thorsen 1980b).

The sentences were arranged in eight random orders in a reading list.

B. RECORDING

The recording equipment consisted of a television camera (Sony AVC-3250 CES) and a video-recorder (Sony U-Matic type 2630). The frame frequency of the equipment was the normal 50 frames per second. The speech signal was recorded on the sound track of the video-tape via a Sennheiser MD 21 microphone placed about 15 cm from the subject's mouth. In order to synchronize speech and video signals a timer signal was recorded on the video-tape using a timing device (FOR-A CO. type VTG 33). On playing back the tape, the timer signal was displayed on the monitor screen in minutes, seconds, and centiseconds and it could, moreover, be registered together with the speech signal

on an ink writer as pulses for seconds and centiseconds. In this manner it was possible to relate each TV-frame to the speech signal.

During recording the subject was seated in a dentist's chair with a fixed head-rest. The camera was placed at the level of the subject's thyroid prominence and at right angles to his mid-sagittal plane at a distance which allowed the area between the subject's chin and sternum to be covered by the field of vision. For calibration purposes the recording of the speech material was immediately preceded by a short recording of a millimeter scale placed in front of the subject's thyroid prominence in his mid-sagittal plane. Each subject read the list once, so that eight repetitions of each test sentence was obtained. Subjects were instructed to use a neutral declarative intonation.

C. SUBJECTS

Recordings were made of three male speakers, PD, PM, and NR (the author). PM and NR are phoneticians, and PD is an engineer and member of the technical staff of the institute. They are all speakers of Standard Danish, although PM, who has grown up in Jutland, has some dialectal influence which, however, did not seem to be reflected to any essential degree in his Fo pattern.

D. REGISTRATION AND MEASUREMENTS

The following acoustic curves were made: duplex oscillogram, two intensity curves, and an Fo curve. The timer signal was also registered on the mingograph. Measurements of Fo and larynx height were made at the midpoint of the stressed and the first posttonic vowels of the test words. The locations on the video-tape of the frames in which larynx height was to be measured were determined from the acoustic curves and the timer curve. Since the interval between frames was 2 cs (the frame frequency being 50 Hz), the temporal inaccuracy of a frame in relation to the corresponding point of measurement as determined from the acoustic curves was $\pm$ 1 cs. The video recorder was equipped with a step function which made it possible during playback to "freeze" the picture and step forward frame by frame and read off larynx height in the frames selected for measurement. Larynx height was determined from a scale drawn on the monitor screen on the basis of the millimeter scale which had been recorded on the tape prior to the reading of the material. The vertical position of the larynx could be measured with an accuracy of $\pm$ 0.5 mm.

Table I

Mean distances in cs of point of measurement from the beginning of the sentence, mean fundamental frequencies in Hz, and mean larynx heights in mm (arbitrary zero), with corresponding standard deviations, in stressed syllables ('V) and first posttonic syllables (˳V) in the three stress groups in the sentence.

			1st stress group		2nd stress group		3rd stress group	
			'V	˳V	'V	˳V	'V	˳V
NR	t	$\bar{X}$	27	47	82	102	154	177
		s	1.246	2.532	1.885	1.669	4.234	4.734
	Fo	$\bar{X}$	112	148	99	125	89	103
		s	5.685	7.285	2.843	4.999	1.769	3.464
	LH	$\bar{X}$	6.9	5.1	6.4	5.1	4.1	4.0
		s	0.835	0.885	1.061	1.246	0.641	0.0
PD	t	$\bar{X}$	21	36	66	82	116	137
		s	1.727	1.727	4.241	3.454	3.871	4.309
	Fo	$\bar{X}$	114	130	110	121	97	114
		s	5.089	5.995	9.376	7.525	2.780	4.790
	LH	$\bar{X}$	32.9	35.9	33.4	35.5	32.1	34.1
		s	1.126	0.991	0.916	1.069	0.641	0.835
PM	t	$\bar{X}$	22	39	65	82	116	138
		s	1.188	1.959	1.069	1.246	1.832	3.012
	Fo	$\bar{X}$	114	137	103	119	94	108
		s	5.731	6.776	1.750	2.840	3.382	5.894
	LH	$\bar{X}$	22.3	24.1	21.6	24.3	19.5	21.8
		s	0.463	0.991	0.916	0.707	0.926	1.581

III. RESULTS

Mean fundamental frequency, larynx height and distance from the beginning of the sentence are given in table I, and fundamental frequency and larynx height means are plotted in figure 2 as a function of time.

A. OVERALL Fo DECLINATION AND LARYNX HEIGHT VARIATION

In order to obtain a quantitative estimate of the overall Fo and larynx height variation, regression lines and correlation coefficients of these variables versus distance from the beginning of the sentence were computed. The computations were made on the basis of the raw data, and stressed and first post-

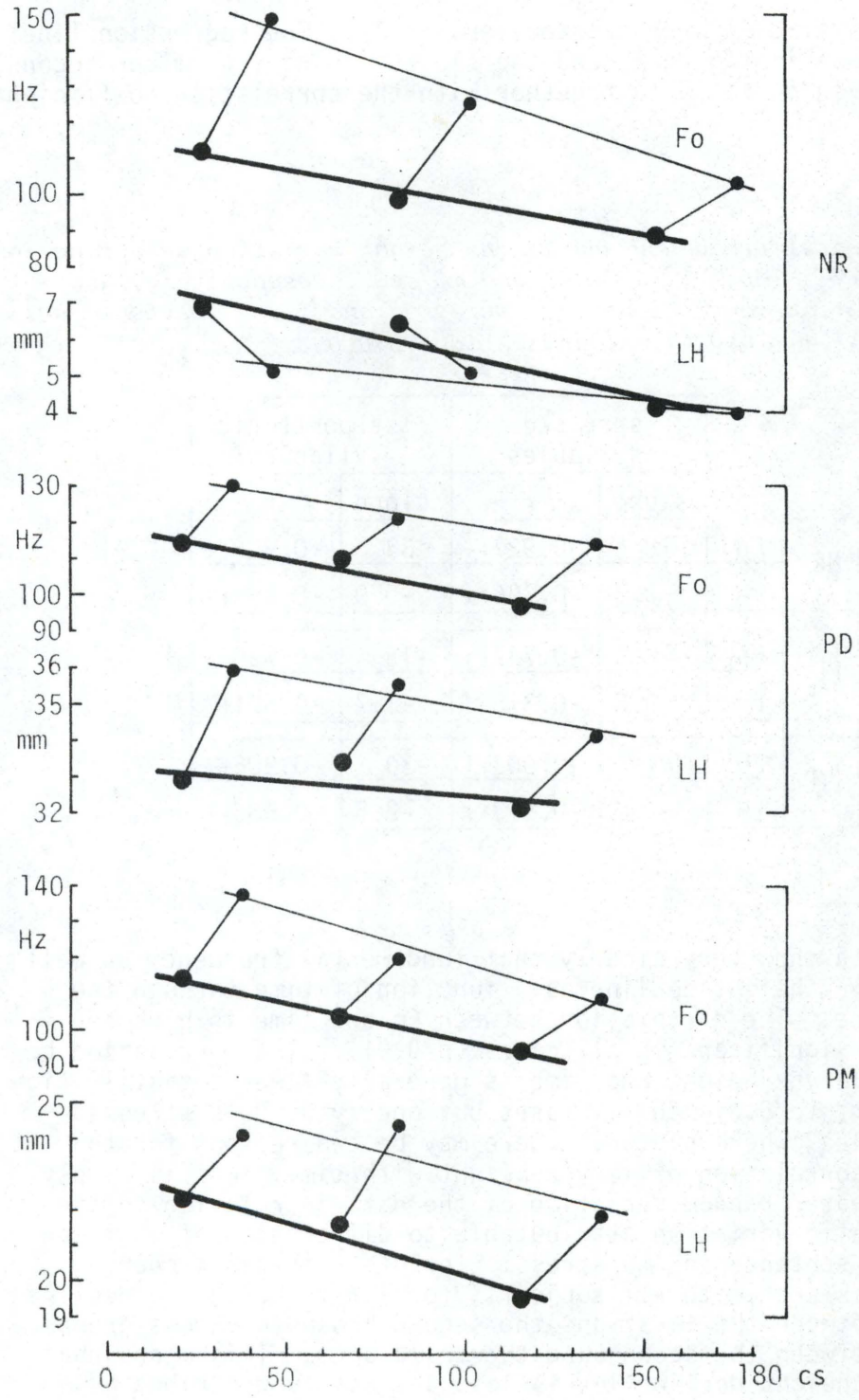


Figure 2

Fo and larynx height (in mm, arbitrary zero) as a function of time through the sentence. Stressed syllables are indicated by large and first posttonic syllables by small dots, and connected by lines within each stress group. The least squares regression lines, computed on the basis of the raw data and for stressed and first posttonic syllables separately are also shown, heavy lines for stressed syllables and light lines for first posttonics.

tonic syllables were treated separately. The regression lines are shown in figure 2, and the slopes (in Hz and mm per second) are given in table II together with the correlation coefficients.

Table II

Fundamental frequency and larynx height variation as a function of time. Slopes in Hz/sec. and mm/sec., respectively, and correlation coefficients for stressed and first posttonic syllables. ++: $p<0.01$, no indication: $p>0.05$.

		stressed syllables		1st posttonic syllables	
		slope	r	slope	r
NR	Fo	-18	-0.927++	-33	-0.955++
	LH	-2.3	-0.796++	-0.9	-0.499++
PD	Fo	-18	-0.737++	-15	-0.735++
	LH	-0.8	-0.318	-1.7	-0.601++
PM	Fo	-21	-0.902++	-30	-0.906++
	LH	-3.0	-0.813++	-2.5	-0.655++

The data show very clearly that fundamental frequency as well as larynx height decline as a function of time through the sentence. The correlation between Fo and time is high and highly significant in all cases ($p<0.01$). The correlation between larynx height and time is generally lower but still significant ($p<0.01$) in all cases but one, viz. PD's stressed syllables, where $p>0.05$. There may be two reasons for the lower correlation of larynx height with time. One is simply the greater random variation of the data in relation to the systematic variation attributable to differences of position in the sentence and to stress. The other is the tendency, which is seen with all subjects, for larynx height to decline less between the first and the second prosodic stress group than between the second and the third ones. This means that larynx height declination is less adequately described by a straight line than is Fo declination. A curvilinear function with an increasing negative slope would presumably give a better fit.

B. LOCAL FO- AND LARYNX HEIGHT VARIATION

Whereas there is in all cases an evident Fo rise from the stressed to the first posttonic syllable, the corresponding larynx height variation follows two different patterns (see figure 2). The two subjects PD and PM have a higher larynx

position in first posttonic syllables than in stressed syllables, i.e. their larynx height pattern closely matches their Fo pattern. A one-tailed t-test showed the differences in all cases to be statistically significant ($p<0.01$) for both Fo and larynx height in stressed vs. posttonic syllable.

For the third subject, NR, the fundamental frequency is also significantly higher in first posttonic syllables than in stressed ones ($p<0.01$), but the larynx height differences go in the opposite direction, the larynx being LOWER in first posttonic than in stressed syllables. NR's pattern of larynx height variation as a function of stress seems to be less consistent than that of PD and PM. The differences are smaller and could only be proved statistically significant in the first two stress groups ($p<0.01$ and $p<0.05$, respectively).

IV. DISCUSSION

A. FO DECLINATION

The data presented above suggest that slow, overall fundamental frequency declination in terminal declaratives is somehow connected with larynx height. But the very fact that the larynx declines over the utterance makes it difficult to interpret the results in relation to the possible explanations for Fo declination outlined in section I above.

In agreement with Maeda's original hypothesis (Maeda 1974) that the decreasing lung volume lowers the sternum which in turn pulls upon the larynx via the muscular and ligamental connections between the two structures, the observed larynx lowering could be seen as the primary cause of Fo declination. Larynx lowering could also, under the subglottal pressure hypothesis, be viewed as a secondary effect of the lung volume decrease without direct influence on Fo. And, finally, if Fo declination is actively controlled, larynx lowering could be the result of activity in the inferior extrinsic laryngeal muscles, which are known to be associated with Fo lowering. Thus, the present larynx height data are compatible with any of these hypotheses.

The finding of a larynx height decline seems to weaken Maeda's (1979) tracheal pull hypothesis, since this hypothesis presupposes that the thyroid cartilage be fixed for the tracheal pull to be effective in shortening the vocal folds, which means that larynx height should remain constant over the sentence.

The possibility that overall Fo declination could be caused by a declining tension of the vocal folds as a result of a gradual decrease of overall activity in either the vocalis muscle or the cricothyroid muscle may also be considered less likely. A decrease of vocalis activity would hardly influence larynx height, and a decreasing cricothyroid activity would either have no effect on observed larynx height (i.e. the position of

the thyroid prominence), if the cricoid cartilage moves as a result of cricothyroid activity, or it would raise the observed larynx position if it is the thyroid cartilage that moves.

Thus, although the observations presented in this study certainly do not make it possible to accept any single one of the hypotheses attempting to explain Fo declination and discard the others, they may to some extent narrow the field of candidates. But, of course, more research on the topic is needed, particularly research focused on the behaviour of physiological parameters under conditions which have been shown to influence Fo declination, such as sentence length and type (cp. section I above).

B. LOCAL FO VARIATION

Whereas there is good agreement among subjects with respect to the relation between overall Fo declination and larynx height, the Fo rise from stressed to first posttonic syllable is accompanied by two different patterns of larynx height variation, two subjects (PD and PM) showing an upward and the third (NR) a downward larynx movement.

Local Fo movements are primarily attributed to the activity of the intrinsic laryngeal muscles, particularly the cricothyroid, and assisted by the activity of the extrinsic laryngeal muscles. In the case of NR, his downward larynx movement may indicate that the Fo rise from stressed to first posttonic syllable is to be explained in terms of cricothyroid activity alone, since contraction of the cricothyroid muscle may cause a lowering of the observed larynx position, under the condition, of course, that cricothyroid activity rotates the thyroid cartilage and not the cricoid cartilage. If the extrinsic muscles were involved, an upward larynx movement should be expected.

The low-to-high larynx movement with subjects PD and PM can hardly be accounted for by reference to the cricothyroid muscle alone. The extrinsic laryngeal muscles will have to be taken into consideration also, and here two possibilities present themselves: either the low larynx - and Fo - of stressed syllables is brought about by contraction of the inferior extrinsic laryngeal muscles which are known to be associated with low or lowering Fo, or the high larynx - and Fo - in first posttonic syllables is due to contraction of the geniohyoid muscle, which can be active in Fo raising (cp. the references in section I above).

It seems difficult to explain the occurrence in the material of two distinctly different larynx height patterns. The conditions of recording were the same for all subjects, and they all used their normal speaking Fo range without displaying extremely large or extremely small Fo excursions. If anything, the greatest activity of the extrinsic laryngeal muscles, and hence the greatest low-to-high larynx displacement, should be

expected for speaker NR, who in fact showed the opposite pattern, since he is the one who has the greatest Fo rises from stressed to first posttonic syllables.

Thus, there seems to be no other explanation than the very general one that speakers within certain limits use different strategies producing the same acoustic pattern.

V. NOTE

1. Maeda gives Fo and larynx height curves for two utterances. In one of them there is certainly no larynx height decline, but in the other there seems - to my eye - to be a larynx lowering of 6-8 mm over an utterance of about 1.5 seconds.

REFERENCES

Atkinson, J.E. 1978: "Correlation analysis of the physiological factors controlling fundamental voice frequency", *J. Acoust. Soc. Am. 63*, p. 211-222

Atkinson, J.E. and Erickson, D. 1977: "The function of strap muscles in speech: Pitch lowering or jaw opening", *Status Rep. Speech Res., Haskins Labs. 49*, p. 97-102

Cohen, A., Collier, R. and 't Hart, J. 1982: "Declination: Construct or intrinsic feature of speech pitch?", *Phonetica 39*, p. 254-273

Collier, R. 1975: "Physiological correlates of intonation patterns", *J. Acoust. Soc. Am. 58*, p. 249-255

Collier, R. 1983: "Physiological explanations for Fo declination", *Proc. 10th Int. Congr. Phon. Sc.*, p. 440

Erickson, D. and Atkinson, J.E. 1976: "The function of strap muscles in speech", *Status Rep. Speech Res., Haskins Labs. 45/46*, p. 205-210

Erickson, D., Liberman, M., and Niimi, S. 1977: "The geniohyoid and the role of the strap muscles", *Status Rep. Speech Res., Haskins Labs. 49*, p. 103-110

Ewan, W.G. 1979: "Laryngeal behavior in speech", *Rep. Phonology Lab. 3, Univ. of Calif., Berkeley*, p. 1-93

Gandour, J. and Maddieson, I. 1976: "Measuring larynx movement in Standard Thai using the cricothyrometer", *Working Papers in Phonetics, UCLA, 33*, p. 160-190

Honda, K. 1983: "Relationship between pitch control and vowel articulation", *Status Rep. Speech Res., Haskins Labs. 73*, p. 269-282

Lieberman, P. 1967: *Intonation, Perception, and Language*, (MIT Press, Cambridge, Mass.)

Maeda, S. 1974: "A characterization of fundamental frequency contours of speech", *MIT Quart. Prog. Rep. 114*, p. 193-211

Maeda, S. 1979: "On the Fo control mechanisms of the larynx", *Séminaire Larynx & Parole, Institut de Phonétique, Grenoble - 8-9 fév. 1979*, p. 245-257

Ohala, J. 1970: "Aspects of the control and production of speech", *Working Papers in Phonetics, UCLA, 15*

Ohala, J. 1978: "The production of tone", *Rep. Phonology Lab. 2, Univ. of Calif., Berkeley*, p. 63-117

Sawashima, M., Kakita, Y., and Hiki, S. 1973: "Activity of the extrinsic laryngeal muscles in relation to Japanese word accent", *Ann. Bull. Res. Inst. Logopedics and Phoniatrics, Tokyo, 7*, p. 19-25

Sorensen, J.M. and Cooper, W.E. 1980: "Syntactic coding of fundamental frequency in speech production", *Perception and Production of Fluent Speech* (ed.: Cole, R.A.) (Erlbaum Associates, Hillsdale, NJ), p. 399-440

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

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

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

Thorsen, N. 1981: "Intonation contours and stress group patterns of declarative sentences of varying length in ASC Danish - Supplementary data", *Ann. Rep. Inst. Phon. Univ. Cph. 15*, p. 13-47

ARIPUC 18, 1984, p. 13-43

VARIABILITY AND INVARIANCE IN DANISH STRESS GROUP PATTERNS*

NINA GRØNNUM THORSEN

Three aspects of the fundamental frequency (Fø) pattern associated with the prosodic stress group in Standard Danish are examined in acoustic analyses of recordings by four speakers. (1) Evidence is presented in favour of previous statements to the effect that short stress groups will have Fø patterns which are truncated rather than compressed editions of those found with longer stress groups. (2) The shape of the Fø pattern can be considered basically invariant (for a given speaker) and independent of the segmental structure of the stress group, but its surface manifestation is modified by intrinsic Fø level differences between vowels of different tongue height, and by the intonational context. (3) Fø patterns in extremely long prosodic stress groups are investigated: With one speaker the falling slope is decomposed into a succession of two shorter ones, with a distinct partial resetting between them, which together describe an overall declination, much reminiscent of the way long sentence intonation contours behave. With three speakers the Fø course through the unstressed syllables is more akin to a mildly undulating wave. The initial fall levels out and falls again, or performs a slow fall-rise-fall.

I. INTRODUCTION

From various previous investigations the picture of the Fø pattern associated with the prosodic stress group in Standard Danish emerges fairly unambiguously. Its most salient features are summarized below.

A prosodic stress group consists of a stressed syllable and all succeeding unstressed syllables, if any. (Unstressed syl-

---

* A revised version appeared in *Phonetica 41*, 1984, pp. 88-102.

lables before the first stressed one in the utterance constitute a unit apart.) Its boundaries are purely prosodically determined, lying always immediately to the left of a stressed syllable. Accordingly, a prosodic stress group may cut across any number of syntactic boundaries, and across syntactic boundaries of any rank below the level of the sentence, cf. Thorsen (1978, 1980a).

More specifically, there is recent evidence to suggest that initial consonants are dissociated prosodically from the stressed vowel, cf. Thorsen (1982a, 1984) and Fischer-Jørgensen (1984), to the effect that prosodically the stress seems to begin with the vowel, and syllable initial consonants join up with preceding material, if any, or constitute a unit apart.

The Fø pattern associated with the prosodic stress group can be seen as an essentially invariant Fø wave, upon which the segments and syllables are superposed. In Standard Danish the stressed syllable hits the wave at the trough before the fairly steep rise to the Fø peak, which generally coincides with the first post-tonic vowel. If stressed plus first post-tonic syllable together are sufficiently short, however, the peak of the Fø wave will only be reached in the second post-tonic syllable, cf. Thorsen (1982a, 1982b). In other varieties of Danish the timing of stressed and unstressed syllables with respect to the trough and peak of the Fø wave will differ, and so may the slope of the rising and falling flanks. Stylized stress group patterns of Standard Danish and two Jutlandish dialects are depicted in figure 1; see further Thorsen and Nielsen (1981) and Fischer-Jørgensen (1984).

The magnitude of the rise from the trough to the peak of the wave is subject to contextual as well as individual variation. (a) The rise is higher early on the intonation contour than late, and it is higher on less steeply falling intonation contours. These facts are reflected in the model in figure 2. The rise is smaller, and under certain circumstances it may even be absent, when the prosodic stress group contains only one unstressed syllable. In Thorsen (1980b) I attempt to account for these phenomena as purely context determined, involuntary variation of an underlying invariant pattern.
(b) Some speakers rise higher than others, and the slope of the rise as well as the fall through the post-tonics is also subject to individual variation. In one investigation (Thorsen 1980a) one speaker demonstrated a syntactically determined variation in the height of the Fø rise from the valley. The rise was highest to a post-tonic syllable belonging to the same word as the stressed syllable, lowest to a post-tonic belonging to the same word as the succeeding stressed syllable, and intermediate when the post-tonic syllable constituted a separate word between the two stressed syllables. In other investigations this same speaker does not, however, mark syntactic boundaries and structures tonally any more than do the other seven speakers I have looked at so far. See further below about distinctness levels.

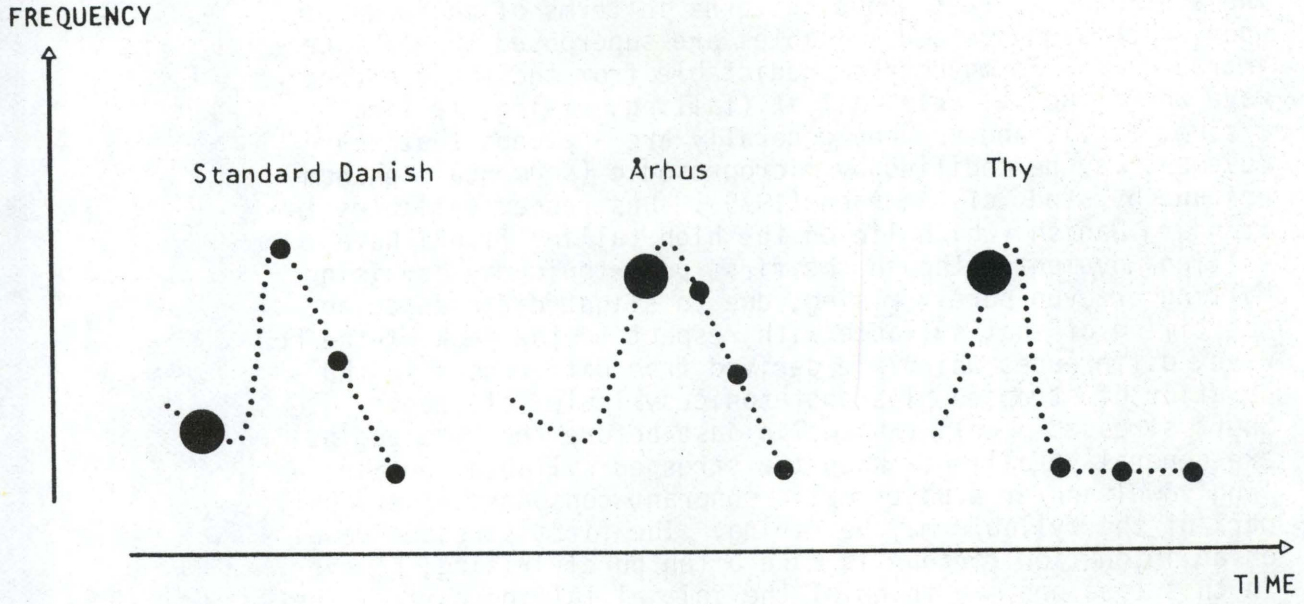


Figure 1

Stylized stress group patterns in three types of Danish: Standard Danish, Århus, and Thy. The big dots represent stressed syllables, small dots unstressed ones.

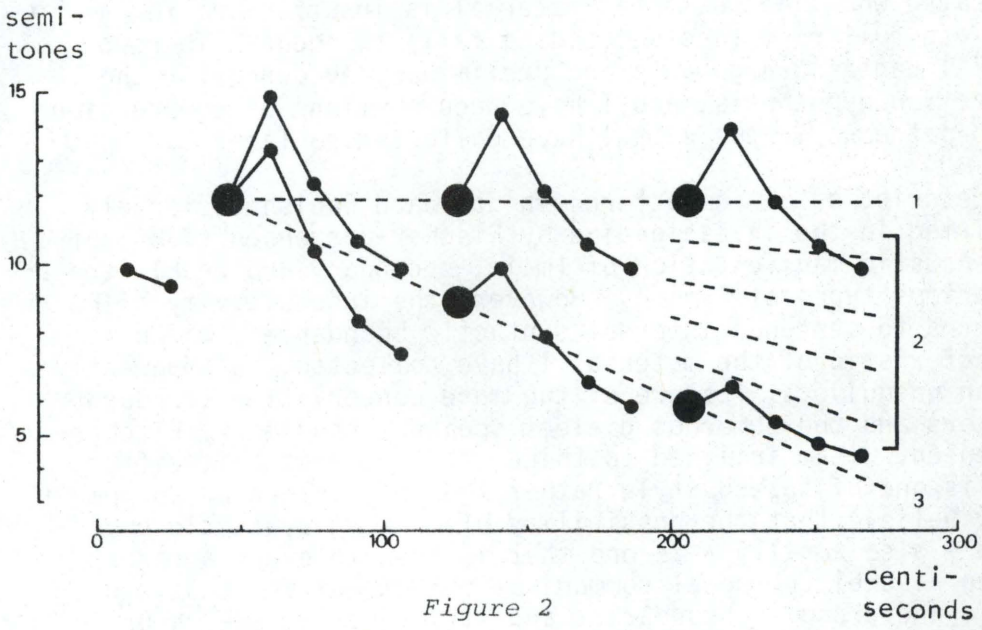


Figure 2

A model for the course of fundamental frequency in short utterances in Standard Copenhagen Danish. (1) Syntactically unmarked questions, (2) interrogative utterances with word order inversion and/or interrogative particle; non-terminal declarative and interrogative sentences (variable), (3) terminal declarative utterances. The large dots represent stressed syllables, the small dots unstressed ones. The solid lines represent the Fø pattern associated with prosodic stress groups, and the dashed lines denote the intonation contours proper.

The account of stress group patterns in terms of an Fø wave upon which segments and syllables are superposed should make intrasyllabic Fø movements predictable from the shape of the wave where the syllables hit it (falling, rising, falling-rising, etc.), and so they generally are - except that any movement may be modified by microprosodic (segmental) phenomena and by stød, cf. Thorsen (1979). Unstressed syllables in Standard Danish (which lie on the high-falling flank) have falling movements, though the first post-tonic may be rising-falling or even purely rising, due to slight differences in the timing of that syllable with respect to the peak of the Fø wave, differences which are derived from differences in the duration of stressed plus post-tonic syllable, cf. above. Short stressed vowels (which lie just before the rise begins) are generally falling. When the stressed syllable contains a long vowel and/or a postvocalic sonorant consonant, the final part of the syllable may be rising. The first stressed vowel on an intonation contour is more often purely rising, however. In this case one may think of the initial falling part of the trough of the wave as being cut off as an adaptation to the high initiation of the intonation contour.

Another consequence of the "passive" alignment of segments with the Fø pattern - the pattern making no timing adjustments to the segments or vice versa - is that Fø patterns should be truncated when the segmental material is insufficient for a complete full rise (and succeeding fall) to occur. Thorsen (1982a) contains arguments and preliminary evidence for the truncation hypothesis, and I have seen no signs of compression of Fø patterns in the data I have collected so far.

The description of Fø patterns in Standard Danish is largely confirmed in the investigation by Fischer-Jørgensen (1984) of the acoustic manifestation of (main, secondary and weak) stress in various types of Danish. However, the insensitivity of Fø patterns to sentence internal syntactic boundaries, which is characteristic of the material I have collected, is apparently not an unequivocal feature of the more conservative standard speakers and the numerous dialect speakers studied by Fischer-Jørgensen. I am inclined to think, though, that this difference is one of speech style rather than of dialect or sociolect, and I believe that the possibility of marking syntactic boundaries - also tonally - is one that is open to every speaker. Jørgen Rischel (personal communication) emphasizes that syntactic and prosodic boundaries and structures in speech production is not an either-or. We may conceptualize a highly articulate (distinct) level of production where the superordinate structure is imposed by the syntax, and at the other extreme a stage with a prosodic structure which works in terms of uniform principles of rhythmization and is independent of any syntactic boundaries (within the sentence). Between these levels are stages where the two structurings intertwine, and a speaker may choose to operate, i.e. produce his actual speech, from levels "above" or "before" the most exclusively prosodically structured one. - This choice would then correspond to what I termed choice of speech style above.

The present paper is an attempt to (1) provide (further) documentation for the truncation of Fø patterns in short prosodic stress groups; (2) provide documentation for the passive alignment of segments and syllables with the Fø pattern, i.e. to establish the (lack of a) relation between the segmental structure of the stressed and post-tonic syllable(s) and the location in time and frequency of the peak of the Fø pattern; (3) provide data on long prosodic stress groups (i.e. stress groups with a large number of unstressed syllables).

II. PROCEDURES

A. MATERIAL

1. Fø PATTERN TRUNCATION

If the truncation hypothesis is correct, a prosodic stress group consisting of one stressed syllable will contain less of an Fø rise than a prosodic stress group with one or more post-tonic syllables, and monosyllabic stress groups should be further differentiated according to the segmental composition of the stressed syllable: long or short vowel, voiced or unvoiced post-vocalic consonant(s).

Eight words with short and long vowel, succeeded by unvoiced obstruent(s) or a sonorant consonant were selected: *kit*, *Keats*, *bit*, *beat*, *guld*, *Kool*, *tin*, *team* [gʰid gʰi·ds bid bi·d gul gʰu·l dˢen dˢi·m]. (Long vowels are indicated here with only one dot after the vowel symbol.) These words were embedded in terminal declarative utterances, as the second of four prosodic stress groups. The carrier sentences are listed in Appendix I,A with translations. The first stress group in these utterances was likewise monosyllabic, so we get a direct succession of three stressed syllables on a declining intonation contour. - The Fø pattern in these monosyllabic prosodic stress groups may be compared with the words with post-tonic syllables in the material to be described in the next section.

2. Fø PEAK LOCATION

If the shape of the Fø pattern is constant (for a given context and speaker, cf. the introduction) and if time compression does not occur, cf. above, the peak of the pattern should be constant in time and frequency with respect to the preceding Fø minimum and independent of the segmental composition of the prosodic stress group (long or short stressed vowel, one or two post-vocalic consonants). A true peak can only be reliably identified in a continuously rising-falling Fø movement, and accordingly the material is made up of words which are voiced from the onset of the stressed vowel through two post-tonic syllables.

Twelve words were selected with short and long, and low and high stressed vowel succeeded by zero, one or two postvocalic

consonants: *kanderne, kinderne, salmerne, skyldnerne, sadlerne, vidnerne, damerne, dynerne, talerne, kuglerne, bagerne, pigerne* ['g̊ʰanʌnə 'g̊ʰenʌnə 'salmʌnə 'sg̊ylnʌnə 'saðlʌnə 'viðnʌnə 'dæ·mʌnə 'dy·nʌnə 'dˢæ·lʌnə 'g̊ʰu·lʌnə 'bæ·ʌnə 'bʰi·ʌnə]. The last two words have no consonant phonetically after the stressed vowel. These words were embedded in terminal declarative utterances, as the second of four prosodic stress groups. The carrier sentences are listed in Appendix I,B with translations.

3. LONG PROSODIC STRESS GROUPS

Seven prosodic stress groups were constructed, containing from three to nine unstressed syllables (the stressed vowels are indicated orthographically here with acute accents; reduced main stress (secondary stress) in the second part of the compounds is indicated with grave accents): *(journa)líst på Poli(tíken), ánsat på Poli(tíken), árbejdsmànd på Poli(tíken), spórtsredaktør på Poli(tíken), móderedaktør på Poli(tíken), údenrigsredaktør på Poli(tíken), údenrigskorrespondènt på Poli(tíken).* These phrases were embedded in short terminal declarative utterances, which were variations on the same syntactic and semantic theme. The utterances are listed in Appendix I,C with translations.

The total of 27 utterances from the three corpora were mixed with 10 other utterances and 11 short texts to be recorded for other purposes. The 48 items were randomized three times, each randomization yielding two pages of reading material, totalling 6 pages, to be read twice by each subject, in order to obtain 6 recordings of each utterance (text).

B. SUBJECTS, RECORDINGS AND TECHNICAL PROCEDURE

Four phoneticians, two males (NRP and JBC) and two females (GB and NT (the author)) read the material, in two sessions on separate days. They all speak a form of Standard Copenhagen Danish. Subjects' style of speech can be characterized as conversational but distinct.

The recordings were made with semi-professional equipment (Revox A-77 tape recorder, Sennheiser MD21 microphone) in a quasi-damped room at the Institute of Phonetics on Agfa PE 39 tape, at 7½ i.p.s.

The tapes were processed by hardware intensity and pitch meters (F-J Electronics) and registered on a mingograph (Elema 800) at a paper speed of 100 mm/s. By adjustment of the zero-line of the pitch-meter to the lower limit of the subject's voice range and full exploitation of the record space of the mingograph galvanometer, a measuring accuracy of 1 Hz for males and 2 Hz for females is attained.

In unidirectional Fø courses with constant slope, only the beginning and end points were measured, according to a procedure outlined in Thorsen (1979, p. 63-66). In more complex Fø courses the turning points (minima and maxima) were measured as well. In the data on long prosodic stress groups the procedure was simplified and the unstressed vowels were represented by a single Fø point, the midpoint (in time as well as frequency), which was an uncontroversial procedure since all such vowels had monotonically falling Fø movements. (The Fø movements are all so short and slight that it does not much matter which point you choose as the one measuring point: the initial, medial or final (or any other) Fø value. Changing the location of the measuring point will transpose all the unstressed vowels by very nearly the same (negligible) amount upwards or downwards, relative to the stressed vowels in figure 6 below.) There are instances where two vowels cannot be reliably segmented from the intervocalic consonant (typically in *-korres(pondènt)* which is pronounced without any intervocalic consonant at all: [g̊ʰɒ(·)ʌs]), and the measuring points are then (somewhat arbitrarily) assigned time coordinates - relative to the onset and offset of the two vowels - in analogy with vowels where a midpoint can be unmistakably located. The distance in time of each point from the first Fø value measured (which is assigned the time coordinate zero) was also determined. Fø and time measurements were averaged over the six recordings by each subject. Average Fø values were converted to semitones (re 100 Hz) and average tracings drawn. No correction was attempted for intrinsic Fø level differences between stressed vowels of different tongue height.

Standard deviations on the average Fø and duration values are generally small, i.e. rarely above 5% of the mean, so production stability across different readings is rather good and the figures must be fairly reliable indications of speakers' behaviour.

III. RESULTS

A. Fø PATTERN TRUNCATION

Average Fø tracings of the eight mono-syllabic test words, together with the preceding and succeeding stressed syllable, are depicted in figure 3. The data will only be evaluated qualitatively.

The second of the three words in the tracings is the one primarily under scrutiny here. First of all, it is very evident that monosyllabic prosodic stress groups have nothing like the Fø rise which characterizes prosodic stress groups with one or more post-tonic syllables, see the words in figure 4. Secondly, the occurrence and magnitude of a rising Fø movement is coupled (though not narrowly correlated) with the duration of the voiced portion of the syllable: A stressed syllable with a long vowel succeeded by a sonorant consonant (*Kool, team*) tends to have a final Fø rise (with NRP the rising part

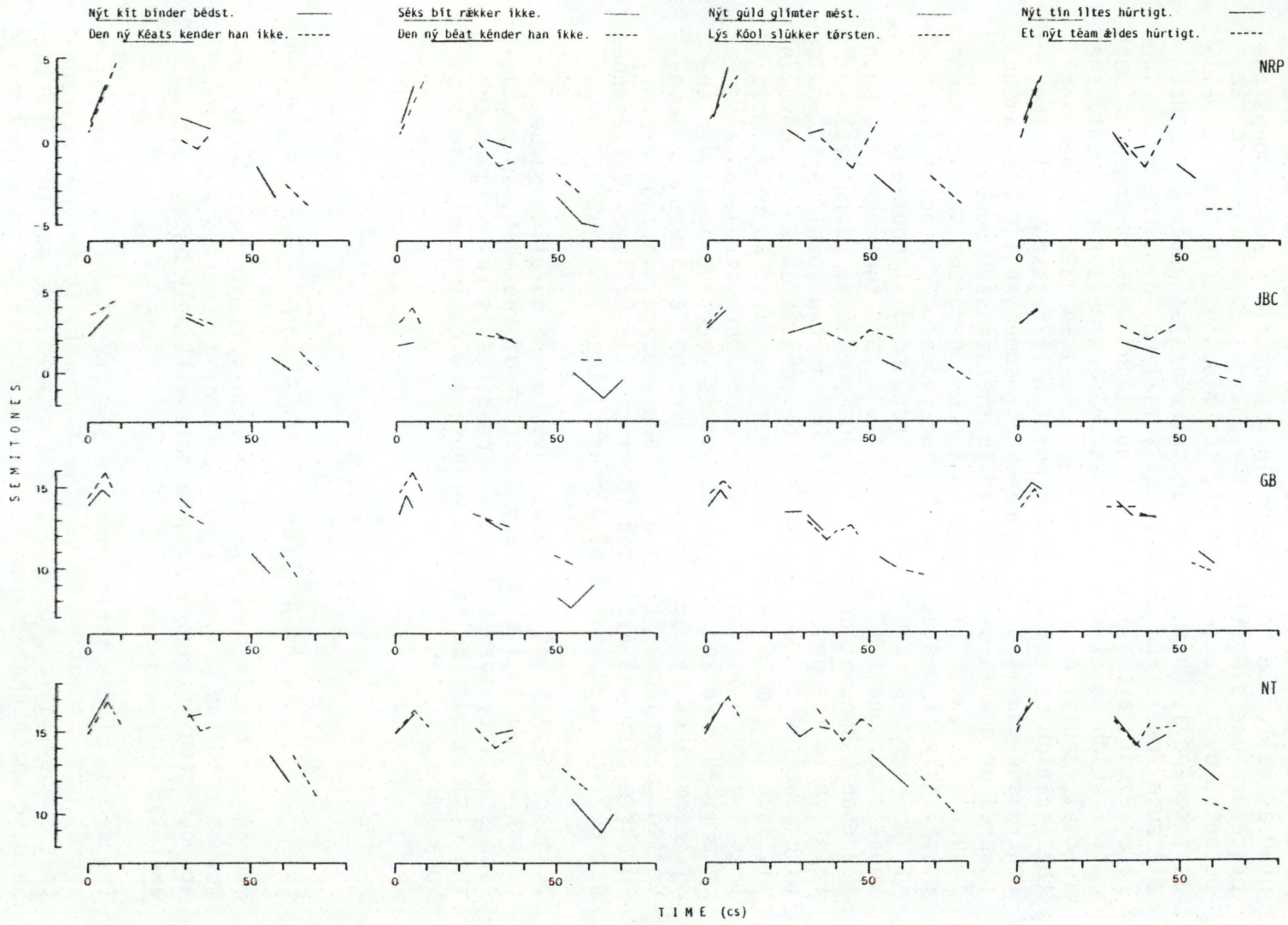


Figure 3

Fundamental frequency tracings (average of 6 recordings) of the underlined sequences identified at the top of the figure. The second word of each sequence contains either a short vowel (full line tracing) or the (nearly) corresponding long vowel (dashed line tracing). Two male subjects (NRP and JBC) and two females (GB and NT). Zero on the logarithmic frequency scale corresponds to 100 Hz.

even exceeds the preceding Fø fall - but in GB's *team* it is actually lacking). A short vowel succeeded by an unvoiced consonant (*kit, bit*) is unidirectional and falling with three subjects, though the fall is slight (extending over less than a semitone); one subject (NT) has a correspondingly very slightly rising movement in this condition. In the intermediate context, with a long stressed vowel succeeded by an unvoiced consonant (*Keats, beat*) or a short vowel succeeded by a sonorant consonant (*guld, tin*), the number of cleanly falling Fø movements equals the number of falling-rising movements (with very slight rises), except that JBC's *guld* is unidirectionally (but very slightly so) rising.

The first word in the tracings in figure 3 is likewise stressed and monosyllabic. Its Fø movement is rising. (Some of the words with long vowel and stød (*ny, lys*) have a rise-fall, where the falling part can most probably be ascribed to influence from the stød, cf. Fischer-Jørgensen, 1984.) The rising movement in these vowels can hardly be seen as a compression of the Fø pattern associated with a stressed plus post-tonic syllable(s), for a number of reasons. (1) A stressed vowel in the first stress group of an utterance will be rising also when it is actually succeeded by one or more unstressed syllables, so rising movements in this position are not peculiar to monosyllabic stress groups. (2) The initial falling part of this supposedly compressed Fø pattern is missing. (3) Lastly, there is no reason to assume that the first stress group in an utterance is subjected to time compression of the low+high-falling Fø pattern, when such is not the case in later prosodic stress groups. - I think a more likely explanation - as mentioned in the introduction - is to see the rising Fø movement in the first stressed syllable of an utterance as an adaptation of the Fø wave pattern to the high initiation of the sentence intonation contour: The first stressed vowel in an utterance is approached from a considerably lower value (or - if it is initial - from zero). The "ideal" slight fall plus rise succeeding this rapid movement or jump is smoothed out into one continuous rise.

The third prosodic stress group in the utterances had post-tonic syllable(s) (which are not depicted in the tracings in figure 3). There is no systematic difference between the stressed vowels of that stress group and the preceding ones in the figure, which is further evidence against compressed Fø patterns in monosyllabic prosodic stress groups. (In *rɛ̀k(ker)* there are three instances of falling-rising vowel Fø movements: this vowel is longer than the other vowels in the same position and therefore contains the beginning of the rise to the post-tonic syllable.)

I think there is every reason to conclude that the Fø pattern associated with a prosodic stress group in Standard Danish is truncated rather than compressed in time when the segmental material is insufficient for a complete (falling-)rising (-falling) gesture to develop. - Likewise, the account of intrasyllabic Fø movements sketched in the introduction can be

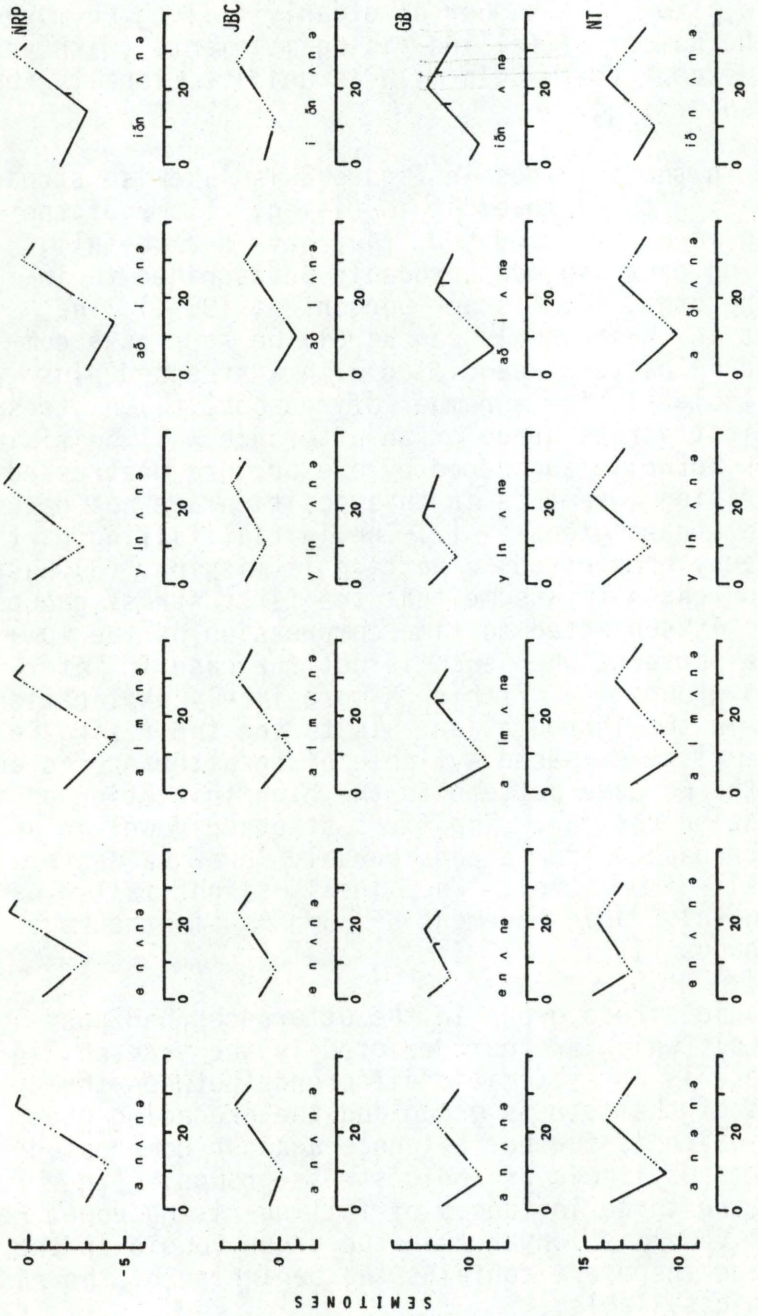


Figure 4

(continued on the next page)

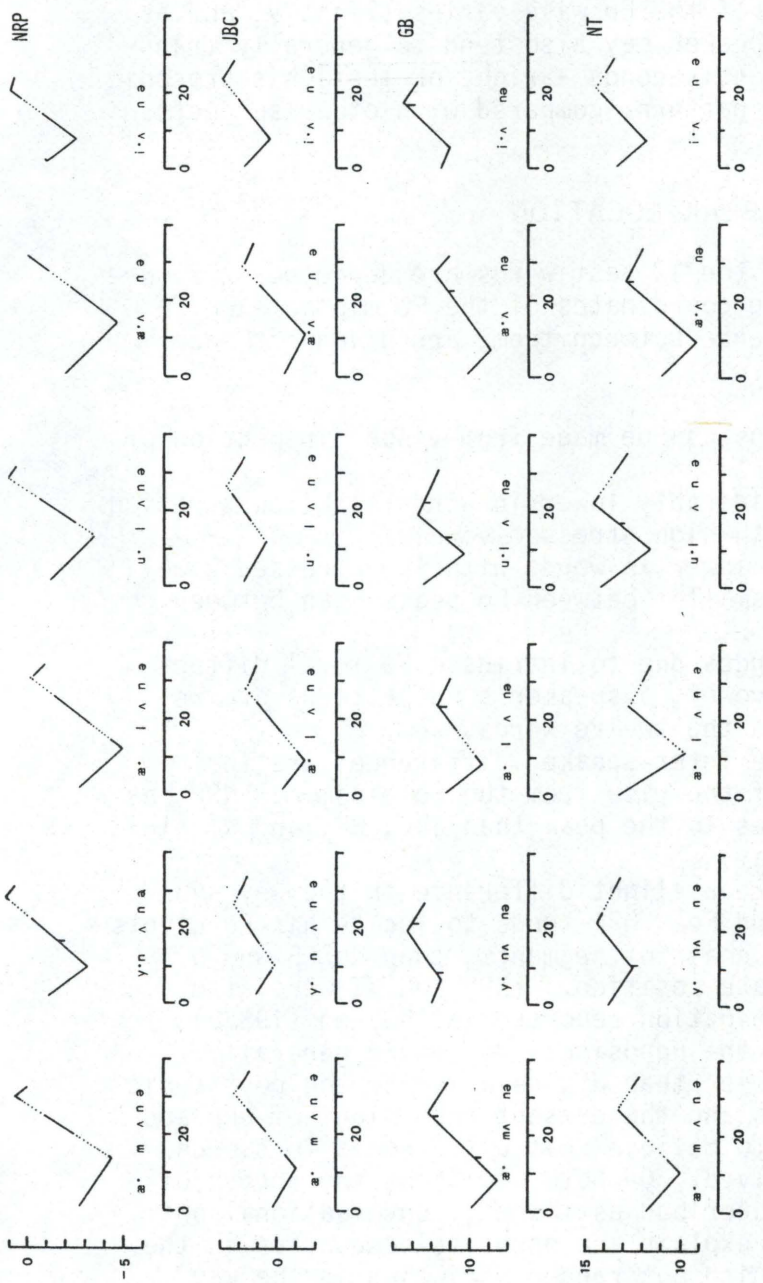


Figure 4

Fundamental frequency tracings (average of 6 recordings) of the words identified at the top of each half of the figure. To the extent that vowels and consonants could be reliably identified and segmented, vowels are traced in full lines, consonants in dotted lines. A vertical stroke indicates a segment boundary in accordance with the transcription beneath each tracing. Two male subjects (NRP and JBC) and two females (GB and NT). Zero on the logarithmic frequency scale corresponds to 100 Hz.

maintained: the first stressed syllable in an utterance is rising, whether it is succeeded by post-tonic syllables or not, in adaptation to the high beginning of the sentence intonation contour. Succeeding stressed syllables are falling, but a final (slight) rise may occur when the vowel is long and/or succeeded by a sonorant consonant. The occurrence or not of this rise is subject to intra- as well as inter-speaker variation. In other words, the timing of the stressed syllable relative to the valley of the Fø wave varies slightly, and as we shall see below a speaker may also tend to generally transpose (by a couple of centiseconds - right or left) his prosodic stress group on the Fø pattern, compared with other subjects.

B. PEAK LOCATION

Average Fø tracings of the 12 test words are depicted in figure 4. Average Fø and time coordinates of the Fø minimum, the Fø maximum and the difference between them, are given for each subject in Appendix II.

A number of observations can be made from visual inspection of figure 4:
1) Fø minima are considerably lower in words with low stressed vowel than in words with high stressed vowel.
2) Fø maxima are also lower in words with low stressed vowel, but the difference is smaller between Fø peaks than between valleys.
3) Apart from differences due to intrinsic Fø level differences in the stressed vowel, a speaker's Fø patterns are remarkably similar across the twelve words.
4) The most noticeable inter-speaker differences are the steepness and extent of the rise from the Fø minimum: NRP has higher and steeper rises to the peak than JBC, GB, and NT (in this logarithmic display).
5) There is furthermore a slight difference in the way subjects align segments and Fø: NRP tends to locate his Fø pivots later, relative to the chain of segments, than do GB and NT; JBC takes an intermediate position. (NRP and NT were also speakers for the investigation reported in Thorsen (1982b), and here the trend was the opposite: NRP would generally locate his Fø peak earlier than NT, relative to the post-tonic segments. The previous and the present recordings of NRP and NT do not give reason to believe that differences in speech style or rate are involved. On both occasions the speech of both subjects can be described as fluent, conversational but distinct, and I cannot explain the opposite tendencies in the two recordings as anything but random variation in the way speakers may choose to coordinate their laryngeal and supralaryngeal gestures, a variation with rather narrow limits or margins which cannot be transgressed, however.)

Some of the qualitative observations above are quantified in Table I.
1) The difference between Fø minima in words with low and high stressed vowel is statistically significant with all subjects, though of different magnitude for males (7-8 Hz,

Table I

Average time and frequency coordinates (mean of means, N=6) to Fø minima and Fø maxima in six words with low stressed vowel and six words with high stressed vowel. The difference between these two points and the slope of the Fø rise between them are also given. The level of statistical significance between the means is indicated, if p<0.05.

		Fø minimum		Fø maximum		Fø max-min		Slope
		Hz	cs	Hz	cs	Hz	cs	Hz/s semitones/s
Subject								
NRP	low	77.7	13.1	102.6	30.0	24.8	16.9	148 = 28.5
	high	84.4	11.0	106.5	26.2	22.1	15.2	147 = 27.6
	p <	0.0005	0.025	0.0005	0.025	0.01	-	-
JBC	low	94.2	10.2	112.0	29.3	17.8	19.2	93 = 15.8
	high	102.5	10.2	115.0	25.6	12.5	15.3	82 = 12.5
	p <	0.0005	-	0.005	0.0005	0.0005	0.0005	0.05
GB	low	167.8	7.6	198.7	24.5	30.8	16.9	184 = 17.3
	high	185.9	6.7	209.1	19.1	23.2	12.4	190 = 16.3
	p <	0.0005	-	0.005	0.0005	0.01	0.005	-
NT	low	177.8	10.4	214.1	25.7	36.3	15.3	240 = 20.8
	high	198.3	8.9	229.3	22.0	31.1	13.0	239 = 19.3
	p <	0.0005	0.025	0.0005	0.005	0.05	0.0005	-

corresponding to approximately 1½ semitones) than for females (about 20 Hz, corresponding to approximately 2 semitones).
2) The Fø maximum is likewise significantly higher in words with high stressed vowel with all speakers, but the difference is smaller (3-4 Hz, or half a semitone, with the males and 10-15 Hz, or one semitone, with the females). Note that all the words ended in the same sequence [-ʌnə], and the difference in the level of the Fø maximum is therefore not due to intrinsic differences in the post-tonic syllable(s). The only way to account for the difference in Fø peak level in the material is as a form of coarticulation or carry-over from the intrinsically different low and high stressed vowels. This is in complete accordance with Reinholt Petersen's (1980) observations and conclusions.
3) Since intrinsic differences between low and high stressed vowels are only partially carried over into the post-tonic vowel, we get a significantly smaller Fø rise from valley to peak after high stressed vowels than after low stressed vowels, though the difference is slight (3-8 Hz).
4) With NRP and NT the Fø minimum has a significantly smaller time coordinate, i.e. it occurs sooner after high stressed vowel onset than after low stressed vowel onset (about 2 cs), and
5) the Fø maximum is significantly earlier with all subjects in words with high stressed vowel (4-5 cs).
6) The effect of (5) is a rise of significantly shorter duration after high vowels with JBC, GB, and NT (2-4 cs) (even though in NT's case it is partially counteracted by the effect of (4)).
7) The slope of the rise differs among speakers, but for a given subject it is constant, irrespective of the nature of the vowel, except that JBC has a slightly but significantly steeper rise after low stressed vowels.

A further dependence upon the segmental structure of the stressed syllable (apart from the nature of the vowel) appears when *kanderne* and *kinderne* are compared with the other five words with low and high vowel, respectively. The stressed syllable of *kanderne* and *kinderne* has a short vowel in a phonetically open syllable, whereas in the other words there is either a postvocalic consonant or a long stressed vowel. The Fø maximum tends to be timed sooner in the two words with short stressed syllable than in the rest of the words (cf. Appendix II). This tendency is most marked with *kanderne*. With JBC and NT it is parallelled by a similar tendency for earlier Fø minima.

Any further systematic differentiation (within or across speakers) in Fø minimum and maximum values and timing, according to the segmental structure of the stressed syllable, is not apparent in the material. In the tracings of figure 5 the time course difference between words with short (*kanderne*, *kinderne*) and long stressed syllable is disregarded, and the grand mean of all six words with low versus high stressed vowel is presented. The line-up point is the Fø minimum.

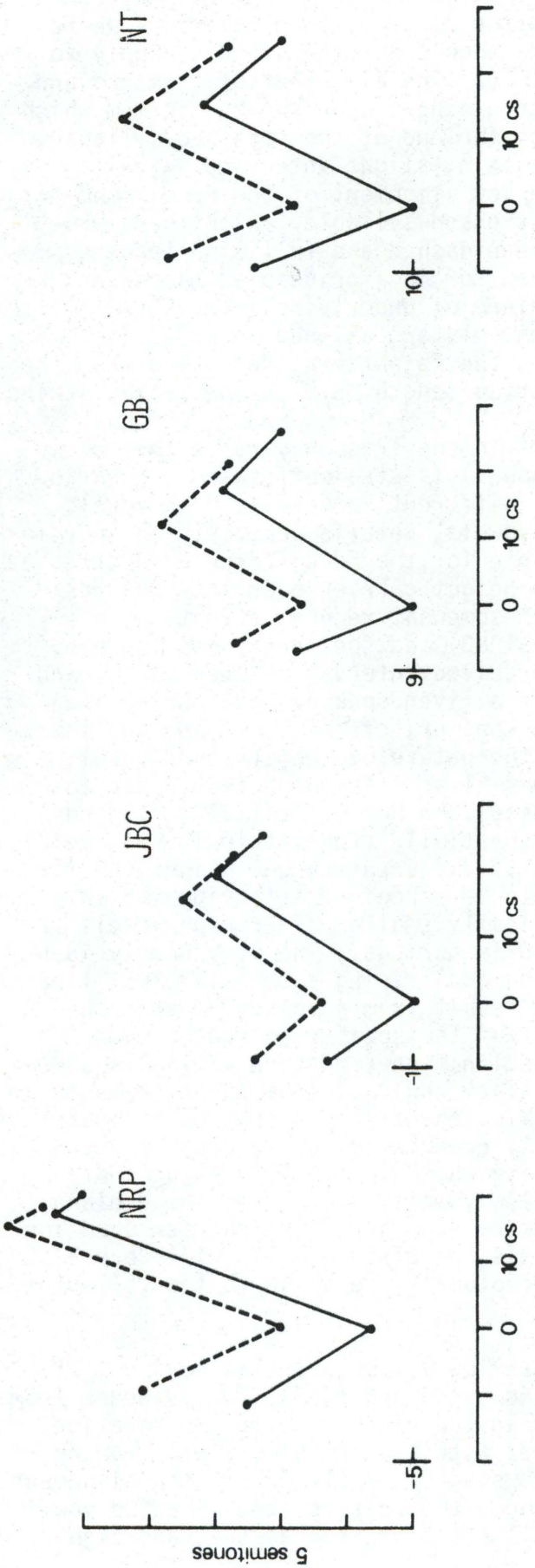


Figure 5

Average fundamental frequency patterns of 6 words with low stressed vowel (full line) and 6 words with high stressed vowel (dashed line) depicted in figure 4 (mean of means).

For the clarity of exposition, let me present first the conclusions I have drawn from the data. (1) Invariance - where Fø patterns are concerned - should be understood to apply to production AND perception. (2) The alignment of segments and syllables with the Fø pattern hinges upon the Fø minimum which is linked up, roughly, with the end of the stressed syllable, but there is room for a certain, slight inter-and intra-speaker variation in the exact alignment of the Fø minimum with the segments of the stressed syllable. With these provisos, the difference between dashed and full line tracings in figure 5 can be accounted for in a principled manner and a hypothesis can be maintained of underlying invariant Fø patterns with which the segments are aligned in a passive fashion, being strung along the Fø pattern, departing from the minimum, like beads of varying length onto an undulating string.

The difference in placement in the frequency range (higher in words with high stressed vowel) is straightforward. High and low vowels have inherently different Fø levels, high vowels having higher Fø than low vowels, ceteris paribus. The physiological mechanism responsible for the Fø difference in stressed vowels of different tongue height carries over into the post-tonic syllable, though with somewhat reduced effect. See further Reinholt Petersen (1980) and the references therein. If the magnitude of the perceived interval between valley and peak is to be the same (for a given speaker, and for a given location on the intonation contour, cf. figure 2 and the introduction), irrespective of the nature of the stressed vowel, then the Fø rise should be smaller after high than after low vowels. This reasoning rests upon Hombert's (1977) findings that listeners compensate (partially) for intrinsic Fø level differences between vowels of different tongue height, so for a low-to-high Fø interval to be perceived as having the same magnitude, it must be physically smaller after high vowels than after low vowels, ceteris paribus. The very nearly identical slopes in the (falling and) rising dashed and full line tracings in figure 5 could result from a voluntary and controlled gesture on the part of the speaker to render such movements with a certain inclination (and then JBC falls somewhat short of the ideal). They could also be due to a muscular or mechanical constraint, i.e. these slopes are the fastest that each speaker can comfortably produce in the recorded speech style and rate (in which case there must be something that speeds up JBC's rise from low vowels). Whatever the explanation, the near constant rising slopes cause a shorter duration of the rise after high vowels, or else the peak will "overshoot" and be too high perceptually, relative to the valley of the pattern (cf. above).

High vowels are shorter than low vowels, ceteris paribus, cf. Fischer-Jørgensen (1964) and Bundgaard (1980) for evidence from Danish. This is reflected in the difference in the duration of the initial fall in words with low and high vowels, being longer in words with low vowels, cf. Table I. If the alignment of segments and Fø hinged upon the onset of the stressed vowel, such a difference in Fø fall had no raison d'être, and high

and low vowels would simply occupy a shorter or longer stretch on an Fø pattern with a time constant pivot. If, on the other hand, we assume that the alignment of segments and Fø pattern hinges upon a coordination of the Fø minimum with some point later in the chain (as presupposed in figure 5) approximately around the end of the stressed syllable, then we would expect the Fø pattern in words with a low stressed vowel to extend further "backwards" from the Fø minimum than in words with a higher, and shorter, stressed vowel. With this account the small difference between *kanderne*, *kinderne* and the rest of the words with respect to the absolute timing of the Fø minimum (earlier versus later) is a natural consequence of the fact that these two words have shorter stressed syllables and therefore shorter initial falling movements, i.e. earlier Fø minima, than the rest of the material.

To conclude this section: Fundamental frequency patterns associated with prosodic stress groups in Standard Danish can be considered underlyingly invariant if we take invariance to apply to both production and perception. Departing from the Fø minimum, located near the end of the stressed syllable, the Fø pattern develops its rise-fall independently of the segmental composition of the stressed and post-tonic syllable (short or long stressed vowel; zero, one or two intervocalic consonants). The one major influence upon the physical manifestation of Fø patterns stems from the (high-low) nature of the stressed vowel. The whole pattern is transposed upwards and the rise is abbreviated in prosodic stress groups with high versus low stressed vowels, in a hypothesized interplay between an intrinsic, physiologically determined constraint and a perceptual demand for equal low-high pitch intervals. The underlying invariance leaves room for a good deal of inter- and intra-speaker variation, however. Different speakers have different magnitudes and/or slopes in the falling and rising part of the Fø pattern. The coordination of Fø minimum with the stressed syllable varies slightly - with the same speaker - from one word to the next, as witnessed by figure 4, but considering the words as a whole there is also an overall trend towards relatively later segment/Fø minimum intersections with some subjects than with others.

As mentioned in the introduction, Fø patterns are subjected to further modification as a function of their intonational context. This contextually determined variation may be automatic and involuntary, or it may be more directly speaker controlled, cf. Thorsen (1980b). Whichever it is, we may still consider the underlying Fø pattern an invariant entity which is subject to certain sandhi modifications.

C. LONG PROSODIC STRESS GROUPS

Average Fø tracings of the seven prosodic stress groups, including the succeeding stressed syllable, are depicted in figure 6. The two stressed vowels (the heavier dots) are designated V_1 and V_2, respectively in the following. (V_1 was not the first stressed vowel in the utterance as a whole, however. It was preceded by one prosodic stress group.) V_2 was an [i] in all seven utterances, whereas V_1 varied. No correction for intrinsic Fø level differences between V_1's of different tongue height is attempted in the figure. - Let me mention first a couple of aspects which tie up with the preceding section, then briefly touch upon a relation which could deserve a more thorough treatment (for which this material is not suited, however), and then I shall proceed to the matter under scrutiny here, namely the tonal behaviour of the unstressed syllables in the seven prosodic stress groups in figure 6.

As in the peak location material above, NRP has the highest rise to the first post-tonic syllable. - The difference in intrinsic Fø level between vowels of different tongue height is clearly reflected in V_1 in figure 6. - We may note in passing that the duration of the prosodic stress group increases progressively with increased number of unstressed syllables.

I have previously stated (1978 and 1980b, for example) that in short utterances (i.e. containing no more than three to four stress groups) the stressed syllables will be distributed along straight lines, whose declination varies with the function of the utterance. These lines constitute the intonation contour proper. (With the rather modest range of variation - well within the octave - exhibited by Standard Danish speakers in the type of utterance and style of speech investigated, intonation contours will be approximately rectilinear, irrespective of the mode of display - logarithmic or linear.) With increased utterance length, in terms of a larger number of (equally long) prosodic stress groups one or more partial resettings of the intonation contour occurs, i.e. the sentence contour is decomposed into a succession of shorter phrase contours, each with its own declination, the ensemble of which exhibit an overall downdrift. The division of the intonation contour into prosodic phrases bears no simple relation to syntactic structure; see further Thorsen (1982c).

The assumption of straight line intonation contours in short utterances implies tacitly that the time and frequency intervals between neighbouring stressed syllables should be directly proportional. The greater the distance in time, the greater the frequency interval. However, till now I have only ever investigated utterances with fairly equal prosodic stress groups (containing two to four post-tonic syllables). With a set of utterances such as the ones treated here we should be able to form a more precise idea of the way the stressed syllables are scaled in relation to each other in terminal declarative utterances. But due to shortcomings in the way the material is structured I shall restrict myself to the formulation of a hypothesis, which is then open to testing in a future experiment.

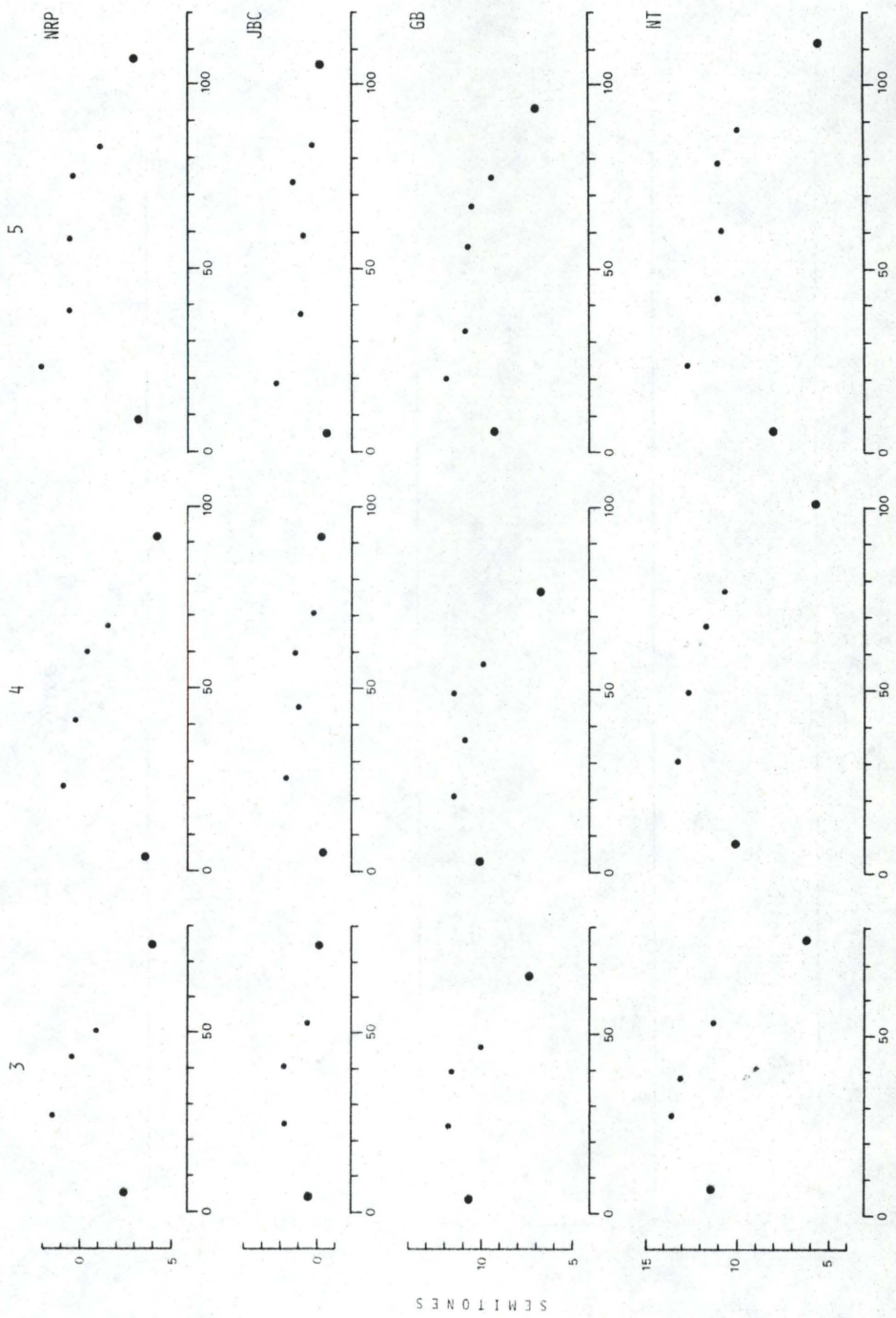


Figure 6

(continued on the next pages)

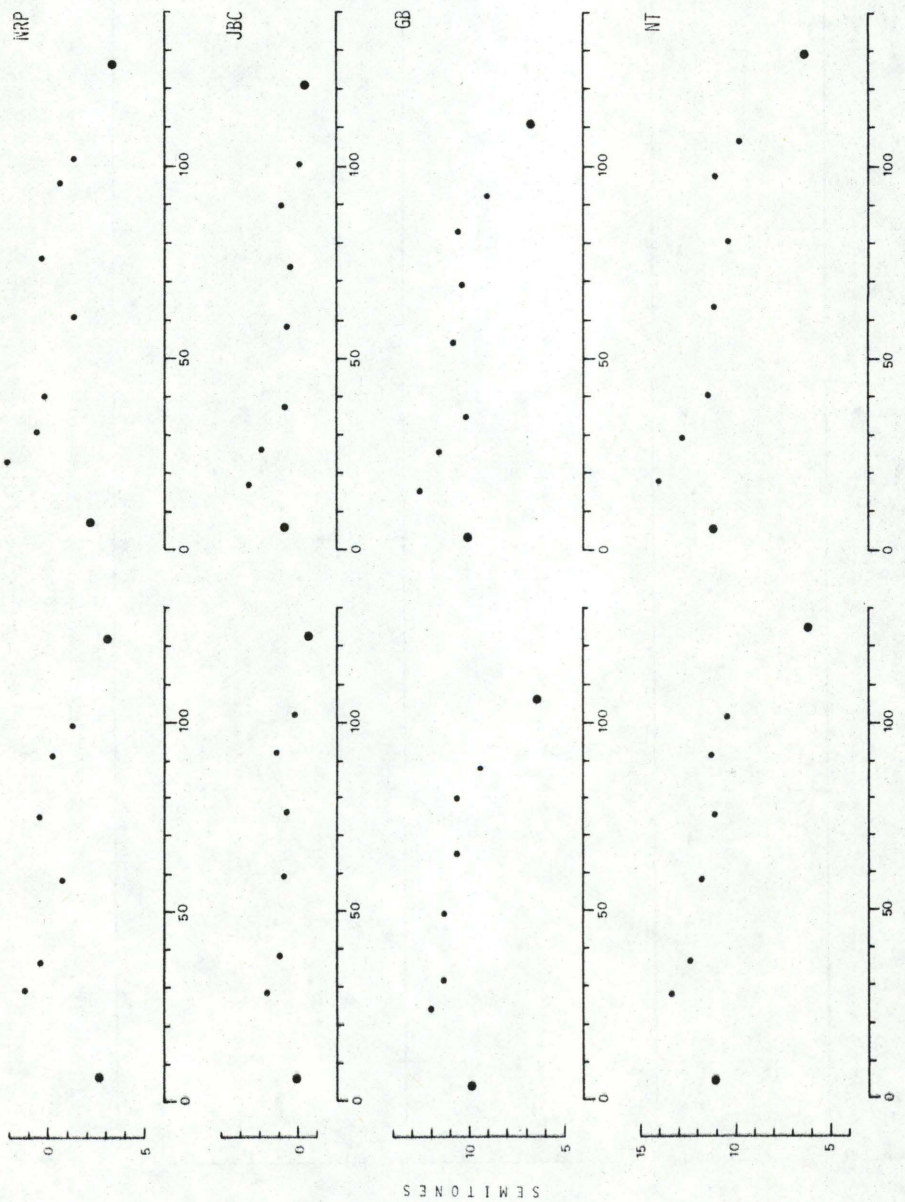


Figure 6

(continued on the next page)

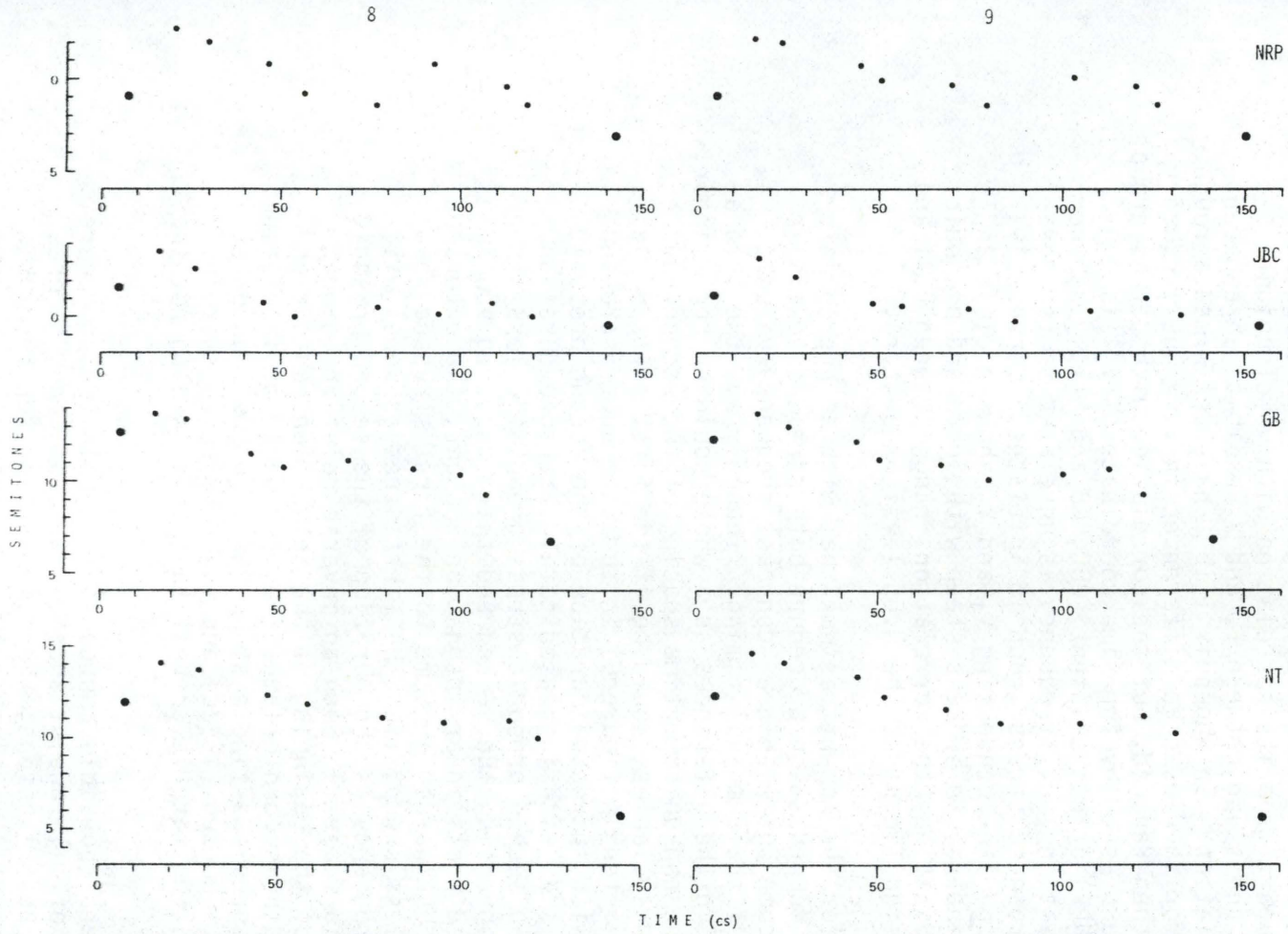


Figure 6

Slightly stylized fundamental frequency tracings (average of 6 recordings) of prosodic stress groups with a varying number (from 3 to 9) of unstressed syllables. The stressed vowel of the succeeding prosodic stress group is included. Stressed and unstressed syllables are indicated with big and small dots, respectively. Two male subjects (NRP and JBC) and two females (GB and NT). Zero on the logarithmic frequency scale corresponds to 100 Hz.

In Appendix IIIa time and frequency coordinates to the two vowels depicted in figure 6 are given, as well as the slope of the straight line that would connect them. It appears that V_2 ([i]), which is the end point of the intonation contour, is very nearly constant, and the very slight variation is not correlated with the duration of the preceding prosodic stress group. The behaviour of V_1 is obscured by the variation in vowel quality, and the relatively large variation shows no clear trend (apart from its correlation with vowel height). Consequently, the slope between V_1 and V_2 exhibits an apparently random variation. In Appendix IIIb I have performed a rough compensation for intrinsic Fø differences in V_1. The frequency variation is then reduced, but the correlation with stress group duration (as expressed by the time coordinate of V_2) is non-significant (NRP) or non-existent (JBC, GB and NT). In other words the variation in V_1 frequency is unrelated to the duration of the prosodic stress group it initiates. Conversely, the slope of the line connecting V_1 and V_2 shows a clear and significant tendency to be less steep with increased prosodic stress group duration (the correlations range between 0.71 and 0.85 and are significant at the 0.05 level or better).

Let me then, for the moment, assume that with an appropriate measuring and compensation procedure both stressed vowels in figure 6 had come out with near constant frequency values, irrespective of their mutual timing. The only way the intonation contours in the utterances as a whole could be rectilinear, in spite of the non-uniform time spacing of the second and third stressed syllables, would be if the first stressed syllable (not depicted in figure 6) in the utterance decreased in frequency with increased duration of the second prosodic stress group, i.e. with increased total utterance duration. (The longer the second prosodic stress group, the less steep is the line connecting the second and third stressed vowels. With an assumed rectilinear intonation contour, the extension of this line "backwards" in time to the first stressed vowel will also be less steep, i.e. the first stressed vowel will be lower in frequency as the duration of the second prosodic stress group increases.) Such an inverse relation between the Fø level of the beginning of the intonation contour and its total length is counterintuitive, and would be in distinct opposition to previous findings, on Danish as well as other languages, cf. Thorsen (1982c) and the references therein. Spot checks in the material reveal that this is also not the case here.

There are several possible sources of error in the procedure. The reduction of all vowels, short as well as long, to one point may not be felicitous; the compensation for intrinsic Fø level differences with equal amounts across all speakers is bound to be unsatisfactory; the influence from different syllable initial consonants is not accounted for at all. Therefore I do not wish to carry the discussion further than to hypothesize that the stressed vowels of an utterance are scaled without regard to their mutual timing. A stressed vowel is lower than the preceding one by a certain amount. However,

this frequency interval is sensitive to utterance length in another way, namely as defined by the total number of stressed syllables. The interval between the only two stressed syllables of a terminal declarative is larger than between a pair of stressed syllables in a declarative with three or four prosodic stress groups, cf. Thorsen (1982c). In other words, we may conclude that (a) variation in prosodic stress group length entails a variation in the slope of the lines connecting neighbouring stressed syllables, and (b) variation in the number of prosodic stress groups (stressed syllables) entails a variation in overall slope.

Two distinct patterns of behaviour emerge where the unstressed syllables in the prosodic stress group are concerned. NRP treats the unstressed syllables in much the same fashion that long sentence intonation contours are produced, whereas JBC, GB and NT employ a different strategy.

When the number of unstressed syllables equals or exceeds six, NRP decomposes the falling slope into two separate, successive slants with a distinct partial resetting between them (of an order of magnitude of 1.5 to 2.0 semitones). ("Partial" because the post-tonic of the second declination does not reach the level of the first post-tonic in the stress group.) The two slopes together describe an overall downdrift. The factor determining the location of the resetting could be one of several: (1) In these utterances it coincides with the first word boundary in the stress group. (2) It occurs after a syllable with secondary stress, i.e. after a syllable which in non-compounded form would have main stress and would accordingly be succeeded by an Fø rise to the succeeding post-tonic syllable. Consequently, this secondary stress and the succeeding unstressed syllables could be said to constitute a pseudo-prosodic stress group. "Pseudo" because the secondary stress does not fall into place on the intonation contour with the truly stressed syllables of the utterance. (3) This same boundary is also the boundary before a prepositional phrase, i.e. a major syntactic boundary. Whether one or the other factor is decisive in NRP's speech can only be found out in a new material.

With JBC, GB and NT the humps and bumps in the declining unstressed syllables hardly deserve the name of resetting, and they are furthermore not consistently located - relative to the chain of syllables. I prefer to describe these Fø courses as very slightly undulating S-shapes or waves, where an initial fall levels out - or rises slightly, depending on the speaker and on the length of the stress group - and then falls again at the end of the pattern.

It is conspicuous that the four speakers apparently attempt not to let the Fø pattern transgress the intonation contour, i.e. the unstressed syllables do not drop below a line connecting the two stressed vowels, except in the two longest stress groups with JBC. In a previous analysis (Thorsen, 1982c), one subject (JR) distinguished himself from other speakers by his

very sharply falling post-tonics which (though their number did not exceed three) often happened to cut through the intonation contour proper. He might, accordingly, have represented yet another distinct pattern of behaviour in this material.

Differences between subjects aside, the long prosodic stress groups here bear witness to a less passive control of the Fø pattern than I have previously assumed, at least when the prosodic stress group is sufficiently long. There are apparently limits (set by each speaker) to the extent of Fø declination tolerated in a series of post-tonic syllables, and when this limit is reached measures are taken to restore the pattern, at least in planned, read speech. Spontaneous speech may be less neatly regulated in this respect.

IV. FINAL COMMENTS

My final comments will serve as a sort of footnote to the title of this paper. A low plus high-falling Fø pattern associated with a stressed plus one or more unstressed syllables, truncated to a greater or lesser degree as the prosodic stress group is cut short of segmental material, is retrievable in every circumstance. The pattern is subjected to considerable "free" (i.e. inter-and intra-speaker) as well as contextual variation. I do not know which of these facts is most significant and amazing from the point of view of speech production and perception, but I do think that we should pay more attention to the variability in speech production, in casu in the production of prosodic phenomena. This is an aspect that models tend to gloss over and ignore: by their very nature linguistic models rather stress the invariance aspect and underrate the plasticity exhibited by speakers in a number of ways and conditions. However, the variation within or across speech production "items" is by force a greater challenge to any model which seeks to quantify its elements, than is the invariance.

ACKNOWLEDGEMENT

Thanks to Jørgen Rischel who made a number of suggestions for clarifications in the manuscript.

REFERENCES

Bundgaard, M. 1980: "An acoustic investigation of intrinsic vowel duration in Danish", *Ann. Rep. Inst. Phon., Univ. Cph. 14*, 99-120

Fischer-Jørgensen, E. 1964: "Sound duration and place of articulation", *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikation 17*, 175-207

Fischer-Jørgensen, E. 1984: "The acoustic manifestation of stress in Danish with particular reference to the reduction of stress in compounds", *Ann. Rep. Inst. Phon., Univ. Cph. 18*, 45-161

Hombert, J.-M. 1977: "Development of tones from vowel height?" *J. Phonetics 5*, 9-16

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

Thorsen, N. 1978: "An acoustical analysis of Danish intonation", *J. Phonetics 6*, 151-175 (A more extensive version was published in *Ann. Rep. Inst. Phon., Univ. Cph. 10*, 1976, 85-147.)

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

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

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

Thorsen, N. 1982a: "On the variability in Fo patterning and the function of Fo timing in languages where pitch cues stress", *Phonetica 39*, 302-316

Thorsen, N. 1982b: "Selected problems in the tonal manifestation of words containing assimilated or elided schwa", *Ann. Rep. Inst. Phon., Univ. Cph. 16*, 37-100

Thorsen, N. 1982c: "Sentence intonation in Danish", *Preprints of Papers for the Working Group on Intonation*, XIIIth Int. Cong. Ling., Tokyo 1982 (summary of two reports in *Ann. Rep. Inst. Phon., Univ. Cph. 14*, 1980, 1-29, and *15*, 1981, 13-47)

Thorsen, N. forthcoming: "Fø timing in Danish word perception", *Phonetica 41*, 1984. (A more extensive version published under the title "Two experiments in the perception of Fø timing in Danish" in *Ann. Rep. Inst. Phon., Univ. Cph. 17*, 1983, 1-20.)

Thorsen, N. and Nielsen, B. Jul 1981: "Fundamental frequency patterning and sentence intonation in two Jutlandish dialects - A first approach", *Ann. Rep. Inst. Phon., Univ. Cph. 15*, 1-12

APPENDIX I

The test words and sequences (underlined) in their carrier sentences with translations. The stressed vowels are indicated orthographically here with acute accents.

A. Fø pattern truncation

Nýt kĭt bĭnder bédst.
New putty holds better.

Den ný Kéats kénder han ĭkke.
He does not know the new (edition of) Keats.

Séks bĭt rækker ĭkke.
Six bits are not enough.

Den ný béat kénder han ĭkke.
He does not know the new beat (music).

Nýt gúld glĭmter mést.
New gold glistens more.

Lýs Kóol slúkker tørsten.
Light Kool quenches your thirst.

Nýt tĭn ĭltes húrtigt.
New pewter is quickly oxidized.

Et nýt téam éldes húrtigt.
A new team grows quickly old.

B. Fø peak location

Glássene og kánderne var méget gámle.
The glasses and the jugs were very old.

Næsen og kĭnderne var méget várme.
The nose and cheeks were very hot.

Téksten til sálmerne var méget gámmel.
The lyrics of the psalms was very old.

De fléste af skýldnerne var méget gámle.
Most of the debtors were very old.

Trénserne og sádlerne var méget gamle.
The bridles and the saddles were very old.

Fórsvareren og vĭdnerne var méget gámle.
The counsel and the witnesses were very old.

Billedet af dámerne var méget gámmelt.
The picture of the ladies was very old.

Óptagelsen af tálerne var méget gammel.
The recording of the speeches was very old.

Púden og dýnerne var méget gámle.
The pillow and the eiderdowns were very old.

Køllen og kúglerne var méget gámle.
The club and the balls were very old.

Slágterne og bágerne var méget gámle.
The butchers and the bakers were very old.

Kárlene og pígerne var méget gámle.
The grooms and the maids were very old.

C. Long prosodic stress groups

Knúdsen er journalíst på Politíken. (3)
Knudsen is a journalist for Politiken.

Péter blev ánsat på Politíken. (4)
Peter was employed by Politiken.

Han stártede som árbejdsmànd på Politíken. (5)
He began as a labourer at Politiken.

Lársen er spórtsredaktør på Politíken. (6)
Larsen is a sports editor with Politiken.

Líssi er móderedaktør på Politíken. (7)
Lissi is a fashion editor with Politiken.

Han éndte som údenrigsredaktør på Politíken. (8)
He ended up as editor of foreign affairs with Politiken.

Han begýndte som údenrigskorrespondènt på Politíken.
He began as a foreign correspondent to Politiken.

The number in parentheses indicates the number of unstressed syllables in the prosodic stress group.

APPENDIX II

Average values of Fø (in Hz) and time (in cs) coordinates of the Fø minimum, the Fø maximum, and the difference between maximum and minimum in 12 words. Standard deviation of Fø means are generally well below 5% of the mean, whereas standard deviations of time values range between 10 and 15%.

	Fø minimum		Fø maximum		Fø max-min	
	Hz	cs	Hz	cs	Hz	cs
Subject NRP						
kanderne	79.3	12.0	103.8	26.7	24.5	14.7
kinderne	84.3	10.7	106.0	23.5	21.7	12.8
salmerne	78.3	13.7	105.3	30.8	27.0	17.1
skyldnerne	85.2	10.2	107.8	26.8	22.6	16.6
sadlerne	77.3	13.7	102.2	30.0	24.9	16.3
vidnerne	83.8	13.8	105.3	29.2	21.5	15.4
damerne	78.2	13.8	103.5	29.7	25.3	15.9
dynerne	83.8	10.0	107.5	28.3	23.7	18.3
talerne	76.0	11.7	100.2	30.5	24.2	18.8
kuglerne	81.8	13.5	106.3	28.5	24.5	15.0
bagerne	77.3	13.7	100.3	32.2	23.0	18.5
pigerne	87.2	7.7	106.3	20.8	19.1	13.1
Subject JBC						
kanderne	94.2	9.5	113.5	26.7	19.3	17.2
kinderne	100.8	8.3	112.5	23.7	11.7	15.4
salmerne	95.7	10.5	112.2	30.2	16.5	19.7
skyldnerne	104.3	11.0	115.0	26.5	10.7	15.5
sadlerne	94.8	9.7	111.3	30.0	16.5	20.3
vidnerne	101.3	11.8	113.3	26.8	12.0	15.0
damerne	95.0	11.3	113.5	31.3	18.5	20.0
dynerne	101.5	10.0	114.7	26.2	13.2	16.2
talerne	92.8	9.2	110.3	28.3	17.5	19.1
kuglerne	103.8	11.8	116.8	26.5	13.0	14.7
bagerne	92.8	11.2	111.0	29.3	18.2	18.1
pigerne	103.0	8.2	117.5	23.8	14.5	15.6

	Fø minimum		Fø maximum		Fø max-min.	
	Hz	cs	Hz	cs	Hz	cs
Subject GB						
kanderne	170.7	7.7	198.3	23.2	27.6	15.5
kinderne	189.0	5.7	204.0	19.2	15.0	13.5
salmerne	169.0	7.3	200.0	25.0	31.0	17.7
skyldnerne	185.3	6.7	204.7	18.3	19.4	11.6
sadlerne	164.7	7.3	196.3	22.0	31.6	14.7
vidnerne	174.3	5.5	202.0	20.7	27.7	15.2
damerne	163.6	7.8	201.3	25.7	37.7	17.9
dynerne	194.0	7.3	216.0	19.5	22.0	12.2
talerne	170.7	8.3	197.0	23.5	26.3	15.2
kuglerne	182.3	9.0	210.0	19.3	27.7	10.3
bagerne	168.3	7.3	199.0	27.5	30.7	20.2
pigerne	190.3	5.8	217.7	17.5	27.4	11.7
Subject NT						
kanderne	185.3	9.2	221.7	23.3	36.4	14.1
kinderne	205.7	7.1	229.3	20.8	23.6	13.7
salmerne	178.7	10.3	215.0	28.7	36.3	18.4
skyldnerne	195.0	9.7	232.0	24.0	37.0	14.3
sadlerne	178.3	11.2	212.7	25.1	34.4	13.9
vidnerne	191.0	9.8	222.3	22.8	31.3	13.0
damerne	178.0	10.4	216.3	27.2	38.3	16.8
dynerne	205.6	9.8	230.0	21.8	24.4	12.0
talerne	176.3	11.7	207.7	24.4	31.4	12.7
kuglerne	194.7	9.8	231.7	21.7	37.0	11.9
bagerne	170.3	9.3	211.0	25.2	40.7	15.9
pigerne	197.7	6.9	230.7	20.7	33.0	13.8

APPENDIX IIIA

Average time (cs) and frequency (semitones re 100 Hz) coordinates to two stressed vowels (V_1 and V_2) surrounding a varying number of unstressed syllables (from three to nine). The slope of a straight line connecting V_1 and V_2 is given in semitones/second. V_2 was an [i] in all utterances, whereas V_1 varied as indicated to the left. Two male subjects (NRP and JBC) and two females (GB and NT).

number of unstressed syllables	V_1	NRP V_1	NRP V_2	NRP Slope	JBC V_1	JBC V_2	JBC Slope
3	[i]	-2.6/5	-4.2/75	-2.3	1.1/5	-0.2/75	-1.9
4	[a]	-4.7/4	-4.3/91	0.5	-0.3/5	-0.3/92	0
5	[ɑ·]	-4.2/9	-4.0/107	0.2	-0.6/6	-0.3/106	0.3
6	[ɒ·]	-3.8/6	-4.1/124	-0.3	0.2/6	-0.5/123	-0.6
7	[o·]	-2.1/7	-4.4/127	-1.9	0.6/5	-0.3/121	-0.8
8	[u·]	-2.0/7	-4.2/142	-1.6	1.7/5	-0.5/141	-1.6
9	[u·]	-1.9/6	-4.1/150	-1.5	1.4/5	-0.6/154	-1.3
		GB			NT		
3	[i]	10.6/4	7.2/56	-6.5	11.4/6	6.2/76	-7.4
4	[a]	10.0/3	6.8/76	-4.4	9.8/8	5.6/101	-4.5
5	[ɑ·]	9.2/6	7.0/93	-2.5	7.8/6	5.4/111	-2.3
6	[ɒ·]	9.8/4	6.4/106	-3.3	11.0/5	6.2/124	-4.0
7	[o·]	10.0/3	6.8/111	-3.0	11.1/6	5.2/130	-4.8
8	[u·]	12.6/5	6.8/125	-4.8	11.8/7	5.8/144	-4.4
9	[u·]	12.4/5	7.0/141	-4.0	12.2/6	5.7/155	-4.4

APPENDIX IIIB

As Appendix IIIa, except that a rough compensation for intrinsic Fø level differences between vowels of different tongue height (V_1) has been performed by adding 0.5 semitones to the values of [i] and [o·], and 2.0 semitones to [a], [ɑ·], and [ɒ·].

number of unstressed syllables	V_1	NRP V_1	NRP V_2	NRP Slope	JBC V_1	JBC V_2	JBC Slope
3	([i])	-2.1/5	-4.2/75	-3.0	1.6/5	-0.2/75	-2.6
4	([a])	-2.7/4	-4.3/91	-1.8	1.7/5	-0.3/92	-2.3
5	([ɑ·])	-2.2/9	-4.0/107	-1.8	1.4/6	-0.3/106	-1.7
6	([ɒ·])	-1.8/6	-4.1/124	-1.9	2.2/6	-0.5/123	-2.3
7	([o·])	-1.6/7	-4.4/127	-2.3	2.6/5	-0.3/121	-2.5
8	([u·])	-2.0/7	-4.2/142	-1.6	1.7/5	-0.5/141	-1.6
9	([u·])	-1.9/6	-4.1/150	-1.5	1.4/5	-0.6/154	-1.3
		GB			NT		
3	([i])	11.1/4	7.2/56	-7.5	11.9/6	6.2/76	-8.1
4	([a])	12.0/3	6.8/76	-7.1	11.8/8	5.6/101	-6.7
5	([ɑ·])	11.2/6	7.0/93	-4.8	9.8/6	5.4/111	-4.2
6	([ɒ·])	11.8/4	6.4/106	-5.3	13.0/5	6.2/124	-5.7
7	([o·])	10.5/3	6.8/111	-3.4	11.6/6	5.2/130	-5.2
8	([u·])	12.6/5	6.8/125	-4.8	11.8/7	5.8/144	-4.4
9	([u·])	12.4/5	7.0/141	-4.0	12.2/6	5.7/155	-4.4

ARIPUC 18, 1984, p. 45-161

THE ACOUSTIC MANIFESTATION OF STRESS IN DANISH WITH PARTICULAR REFERENCE TO THE REDUCTION OF STRESS IN COMPOUNDS*

ELI FISCHER-JØRGENSEN

The present investigation of Danish stress is based on a rather extensive material from Standard Danish and from Jutlandish, Funish and North Zealandish dialects. It is found (1) that unstressed syllables are distinguished from stressed syllables by not being able to have stød or phonological vowel length, by shorter duration, by a different Fo pattern which can be described as a different placement on an undulating Fo wave (but Fo pattern differences are not the same in different dialects), and finally often by lower intensity, at least as far as the endings -e [ə] and -er [ɒ/ɔ] are concerned. (2) Syllables with secondary stress retain stød and phonological vowel length but are distinguished from stressed syllables by shortening (although not as much as unstressed syllables), by Fo, behaving in this respect like unstressed syllables, but hardly by intensity except for the dialects which have low Fo on unstressed syllables. (3) Syllables which are reduced by unit accentuation (in relation to a following stressed syllable) lose stød and (at least for monosyllabics) phonological vowel length; they are somewhat more reduced in duration and partly in intensity than syllables with secondary stress, and they behave, as do syllables with secondary stress, like unstressed syllables in the Fo pattern of the stress groups.

---

*) *This report is a considerably enlarged - and in some details revised, version of Fischer-Jørgensen 1983.*

Content

		p.
I	Introduction	47
II	The Danish prosodic system	51
	A. Vowel length and stød	51
	B. Stress	51
	1. General (traditional) description	51
	2. Number of stress degrees (various phonological descriptions)	53
	3. The problems of phonetic manifestation	56
	4. Stress groups	56
III	Previous descriptions of the acoustic manifestation of stress in Danish	57
IV	Material, informants and measurements	61
V	Results	67
	A. The relation between main stress and weak stress	67
	1. Vowel quality and stød	67
	2. Duration	68
	3. Fo contour	72
	4. Intensity	87
	5. Conclusion	90
	B. Main stress and secondary stress	90
	1. Duration	91
	2. Fo contour	100
	3. Intensity	105
	4. Conclusion	106
	C. Stress reduction in the second member of compounds compared to stress reduction in unit accentuation	106
	1. Duration	107
	2. Fo contour	110
	3. Intensity	112
	4. Conclusion	112
	D. Secondary stress in compounds compared to weak stress	113
VI	General conclusions	114
	References	116
	Appendix I: List of sentence material 1961	123
	Appendix II: " " " " 1982	124
	Explanations of Appendix III-V	126
	Appendix III: Measurements of duration and sketches of Fo and intensity for main vs. weak stress	128
	Appendix IV: The same for main stress and reduced stress (speakers of Standard Danish)	135
	Appendix V: The same for dialect speakers	153

I. INTRODUCTION

It is not my intention to enter into any general discussion of the phonology of stress in this paper, but it may be useful to start by stating what I mean by stress.

I use accent as a more general term covering stress and tone. Stress is used to indicate culminative accent, by which one syllable is made more prominent than other syllables in the given unit, whereas tone or tonal accent characterizes a syllable or part of a syllable or a word without giving special prominence to it. It is inherent in the idea of prominence that something stands out compared to the surroundings. Thus stress is a concept of relations in the speech chain. This makes it problematic to set up a definite number of absolute stress levels for a given language. Phonetically stress is a property of a syllable. Phonologically it may also belong to a syllable. This is the case in simplex words where one syllable is more prominent than the others. But it may also characterize a larger (meaningful) unit, e.g. a member of a compound or a phrase. Word stress may be called potential in the sense that it characterizes that syllable of a word which will be stressed if the word is stressed in the actual sentence. Stresses may be reduced in compounds or in certain phrasal structures. Stress reduction in compounds may be regarded as lexicalized syntactic reduction of stress (see, e.g., Martinet 1954). These reductions are rule governed. Moreover, the speaker is always free to give more or less prominence to words which he finds more or less important in a given sentence, and in the extreme case of explicit or implicit contrast the manifestation is normally conventionalized. This paper only deals with word stress and stress reduction by rule in neutral, declarative sentences. The main problem will be the acoustic manifestation of stress, more specifically the acoustic difference between main stress and reduced stress in compounds.

The search for acoustic differences raises some problems. Some phoneticians want to define stress from the point of view of speech production as increased effort on the part of the speaker. Fónagy (1958 and 1966) emphasizes that stress should be defined physiologically and that the acoustic consequences of the effort may be varied and shifting, and may even be conventionalized so that they are no longer effects of the effort but conventionalized signals, like e.g. the duration of vowels in Russian (compared to the much smaller lengthening in Hungarian which is really caused by more effort), and these conventions may be different in different languages. However, even if it may be true that stress can be defined in a simpler and more consistent way from the point of view of production (more evidence is still needed on this point), information on the physiological effort must be conveyed to the listener through the speech wave. A feature that functions in speech must be found at all stages of the speech communication act.

There is, however, another reservation to be made: Since stress is to a large extent rule governed, it is often predictable. Moreover, a syntactic construction may be sufficiently characterized by grammatical means so that the stress difference is redundant. Chomsky and Halle (1968, p. 25-26) emphasize that stress can be predicted to a large extent from semantic and syntactic facts, so that the listener expects stress to appear at certain points of the utterance and will perceive it even if there are no acoustic cues indicating it. It may be a perceptual fact without any physical basis. They go so far as to maintain that "there is no acoustic evidence to support the view that perceived stress contours correspond to a physically definable property of utterances". - This general formulation is exaggerated. But there is no doubt that it may happen, and it has even been proved by Berinstein (1979). She asked a number of English, K'ekchi and Spanish subjects to indicate the placement of stress in a sequence: *bibibibi* which was manipulated in different ways. In the control case, where all syllables were alike, 80% of the English subjects placed the stress on the initial syllable, and more than 40% of the K'ekchi subjects (who have final stress in their mother tongue) placed the stress on the last syllable. And there was still a certain bias when the syllables were different. But the importance of acoustic cues and the presence of acoustic differences in normal speech has been demonstrated in many experiments. And, as Katwijk (1974) remarks, stresses placed in a wrong position are easily noticed, and they are even very disturbing in normal speech situations. Thus, it may nevertheless be worth while looking for acoustic differences. One must, however, be prepared to find acoustic differences which are not normally used in the perception of stress and, on the other hand, one must also envisage the possibility that cues found in tests with synthetic sounds may not be used normally in the subject's mother tongue. He may know them from a different dialect or language.

In spite of these reservations I think that it is possible through an acoustic analysis to find the properties which may be supposed to be most important in a given language. An acoustic analysis is the first step towards the goal of finding the perceptually relevant cues.

The acoustic properties which have most often been found to correlate with stress are (1) increased duration, (2) increased intensity, and (3) deviations in fundamental frequency (Fo). Another fairly frequent property is precise vowel quality versus schwa or at least a centralized quality; sometimes unstressed syllables are also characterized by lack of aspiration. A specific durational cue mentioned by Bolinger and Gerstman (1957) is "disjuncture", i.e. the distance between syllable centers. This distance may be shortened in compounds compared to a sequence of independent words.

The Fo deviation in stressed syllables need not be a movement upward, though high or rising pitch is the most common feature. It may also be a downward deviation (Bolinger 1958a). The de-

cisive thing often seems to be a relatively abrupt shift of level or direction. Cohen and 't Hart (1967), and 't Hart and Cohen (1973) found that in Dutch a rise-fall in the same syllable was the most prominence lending contour, but a simple rise or a fall (the latter only in the last accented syllable of the phrase) may be sufficient.

The timing may also be important. In Dutch a rise in the beginning of the syllable or a fall at the end have been found to be the most favourable placement for the perception of stress (Katwijk 1969 and 1974). (For a more detailed discussion of timing in Fo contours, see Thorsen 1982a.)

In quite a number of languages it has been found that Fo changes are perceptually most important, e.g. in English (Fry 1955 and 1958), Bolinger (1958a and b), Morton and Jassem (1965)), whereas intensity is the weakest cue. Fo has also been found to be important for the perception of stress in German (Isačenko and Schädlich 1964, and Bleakley 1973) and in Dutch (Katwijk 1974), in Polish (Jassem 1959), and in French (Rigault 1962)). In French there are also obvious differences in duration (Benguerel 1971), but hardly in intensity (Spang-Thomsen 1963).

There is not complete agreement about English. Scholes (1971) found intensity to be the most important cue between syntactically conditioned stress differences as *the good flies quickly past* and *the good flies quickly passed*, at least in natural speech (in synthetic speech the temporal differences turned out to be more important), and Lieberman (1960) found intensity differences to be dominant compared to duration in English pairs like *cónduct - conduct*.

However, both for English and for most other languages the hierarchy which is most frequently set up is (1) change in Fo, (2) increased duration, (3) increased intensity. Hyman (1977) considers this to be a universal hierarchy. But that can hardly be upheld. Berinstein (1979) proposes instead to consider this hierarchy to be the unmarked one, which is valid only for languages without phonemic contrast in tone or vowel length. If one of these properties is phonemic it will be superseded by the other cues in the hierarchy. She finds confirmation of this hypothesis in the fact that in K'ekchi, which has phonemic vowel length, duration is the weakest cue for stress, whereas it is used for this purpose in Cakchiquel, which does not have phonemic vowel length. Similarly, Janota (1967 and 1979) found intensity to be more important than duration in Czech, which has phonemic vowel length, whereas Polish which does not have phonemic vowel length has the unmarked hierarchy (Jassem 1959). Fónagy (1958) also draws attention to the extensive and conventionalized increase of duration of stressed syllables in Russian (which does not have phonemic vowel length) compared to the small differences in Hungarian (which does have phonemic vowel length). But even without interference from the phonemic factor there may be deviations from the unmarked hierarchy. Botinis (1982) thus finds duration and intensity to be more consistent cues than Fo in Greek, and Bertinetto (1980) finds

duration to be the most important cue in Italian. He takes the difference limen for the various cues into account, which has not always been done.

On the whole, there is some uncertainty in comparing the relative importance of properties measured in different units, except in the cases where a cue is less frequently used, or not used at all. Finally, it should be mentioned that some phoneticians prefer to measure the energy integral, not the peak intensity as most have done. In this way duration and intensity are combined into one feature, which for short vowels is justified by the integration time of the ear. Lieberman (1960) finds this to be the best measure for English stress (cp. also Lea 1977). Rossi (1970) and Rossi et al. (1981) do not use intensity but a calculated loudness value, based on compensation for intrinsic vowel intensity and the integration time of the ear. In principle this is, of course, a correct procedure for arriving at the loudness value. But in this way physical duration is reckoned twice (as part of loudness and on its own,as is also the case when the energy integral is used). It is probably safer to avoid these problems by keeping vowel quality, Fo, and duration constant when investigating the perceptual importance of intensity by means of synthetic sounds (which has also been done by various experimentors). In measurements of natural speech one may at least stick to comparisons between vowels of the same quality, or one may compare relations or differences between similar pairs of vowels instead of absolute values.

Most investigators have compared full stress and weak stress. But one cannot be sure that the same cues are used for the distinction between full stress and reduced stress, as found, e.g. in compounds in Germanic languages. It would no doubt have been methodologically preferable to start this study of Danish stress by a study of full stress vs. weak stress (stressed vs. unstressed syllables) and from there go on to the more complicated problems of reduced stress in compounds. But I had a 20 years old material on compounds which had never been utilized. I thought it only needed a few supplements, which then grew into a rather large corpus. I have, however, used this material and some other recordings made for a different purpose to give some necessary information on weak stress (section V.A). The Fo differences between stressed and unstressed syllables in Danish have been thoroughly investigated by Nina Thorsen, but there are hardly any published results for intensity and duration measurements in Danish stressed and unstressed syllables.

As stress in Danish compounds has not been investigated instrumentally before, this is a preliminary pilot study. For such a first study you may (1) choose a well delimited material, perhaps consisting in nonsense words, read by speakers of the same dialect, as a safe starting point for later enlargements of the field, or (2) you may wish to get a wider, but more superficial overview, e.g. using various types of compounds in natural sentences spoken by subjects with different dialectal

background, in order to get an impression of the main problems and find the points which would be worth a more detailed investigation. I have chosen the latter approach. But this, of course, implies that not all complicating factors can be kept apart, and the conclusions must therefore necessarily be preliminary.

II. THE DANISH PROSODIC SYSTEM

A. VOWEL LENGTH AND STØD

Danish has distinctive vowel quantity, which has quite a considerable functional load in words of two or three syllables (e.g. /mi:lə milə/). There is hardly any concomitant difference in vowel quality except for /a/, /o/ and /ɔ/. Postvocalic consonants do not show any consistent difference in duration after long and short vowels (Fischer-Jørgensen 1964, and Petersen 1973).

Almost all monosyllabic words, as well as final stressed syllables of polysyllabic words, with long vowel have "stød". The same is true of the great majority of monosyllabic words and final stressed syllables with short vowel plus sonorant consonant. "Stød" is a specific accent (transcribed [ʔ]), which in its typical form is characterized acoustically by a drop in intensity and Fo, often ending in creaky voice in the latter half of the long vowel or the beginning of the sonorant consonant after the short vowel. But the creaky voice may be practically absent, and there is, on the whole, a rather large variation in the manifestation of the stød. Its presence is to a large extent rule governed (it corresponds historically to accent 1 in Swedish and Norwegian), but there is a number of minimal pairs, mainly disyllabic words with the endings *-en*, *-el*, *-er* where one member of the pair is an old monosyllable, (e.g. *løber* [lø:ʔbɒ] 'runs' vs. *løber* [lø:bɒ] 'runner') and also some monosyllables with short vowel plus sonorant consonant (e.g. *ven* [vɛn] 'friend' vs. *vend* [vɛnʔ] 'turn!'). (For the rules governing the presence of stød see, e.g., Aage Hansen (1943), Hjelmslev (1951 (~~1983~~)), and Basbøll (1972 and 1971-1973); for its manifestation see Smith (1944) and Petersen (1973).

B. STRESS

1. GENERAL (TRADITIONAL) DESCRIPTION

Stress is traditionally described on an auditory basis, and a language may be characterized by its distribution of stresses in simplex words, compounds, phrases, and utterances. (In the following the accent marks ˊ and ˋ are used to indicate main and secondary stress, respectively, in orthographic examples, whereas ˈ and ˌ are used in phonetic transcriptions, weak stress (= unstressed syllable) being left unmarked.)

The Danish stress system is very similar to that of German or Dutch. Simplex native words consist (with very few exceptions) of an initial stressed syllable with full vowel, followed by 0-2 (rarely 3) unstressed syllables with schwa. Foreign words may, however, have stress on any syllable and full vowels in unstressed syllables. Unstressed syllables do not have stød and have only short vowels.

Compounds and derivatives have only one main stress in the standard language with the exception of a very specific type, in which the first member is an intensifying prefix, e.g. *bómbesíkker* 'absolutely sure' compared to *bómbesikker* 'protected against bombs'. I should prefer to regard this as a lexicalized emotional accent. Nina Thorsen has called my attention to the type *júləáften* 'Christmas Eve' and *påskemórgen* 'Easter morning', which I should be inclined to regard as sequences of two words completely parallel to *Søndag mórgen* 'Sunday morning', but often written as one word because of their specific meaning.

In some West Jutlandish dialects a large number of compounds are described as having two main stresses (cp., e.g., Ejskjær 1954). It remains to be investigated whether this type is phonetically different from a syntactic combination of two words and how it is perceived by the dialect speakers. In a number of cases it will be signalled as a compound by means of other features, e.g. lack of stød in the first member, and the identification as a compound will be supported by the relatively large individual variation, some speakers using two main stresses and others stress reduction in the same words.

The majority of compounds and many derivatives in Standard Danish are characterized by having main stress on the first member and secondary stress on the second member, e.g. *en sórtmèjse* 'a coal tit' vs. *en sórt méjse* 'a black tit'. (Only a few derivative suffixes have weak stress, viz. *-ig*, *-lig*, and probably *-ing* and *-isk*, and several, particularly foreign, suffixes have main stress.)

In more complex compounds and derivatives there may be a graduation of secondary stresses depending on the construction, e.g. *(dýrskùe)plàds* 'cattle showground' vs. *krígs(skùeplàds)* 'theatre of war' (where the stronger secondary stress is indicated by "). There is, however, mainly for rhythmical reasons, a tendency to shift the strongest stress to the last member, particularly after monosyllabic first members, thus ˊ ˋ to ˋ ˊ, and this pattern has been lexicalized in several cases.

A limited number of complex compounds and derivatives have main stress on the second member, e.g. *Langfrédag* 'Good Friday', particularly certain types of derived compound adjectives: *barnágtig* 'childish', *sandsýnlig* 'probable' (more often than in German, but less often than in Dutch).

The second member of a compound with main stress on the first member retains stød and vowel length; and in quite a number of cases a second member which has no stød as a simplex word even acquires a stød (e.g. *sende* [sɛnə] 'send', but *opsende* ['ɔbsɛn?ə] 'send up'); and the first member of the compound, when it is monosyllabic, normally loses its vowel length and stød, e.g. *hus* [hu:?s] 'house', but *husmand* ['hus,man?] 'smallholder'. But when the main stress is on the second member of the compound, the first member has not only reduced stress but loses its stød and, at least when it is monosyllabic, also its vowel length, e.g. *løs* [lø:?s] 'loose' *løsagtig* [løs'agdi] 'prostitute'.

Prefixes are often unstressed, e.g. always *be-*, *er-*, whereas *u-* is vacillating. As the root morpheme *for* has stress and the prefix *for-* is unstressed, we get minimal pairs like *fórbènet* 'the foreleg' and *forbénet* 'ossified, pigheaded'.

Danish phrasal stress differs from e.g. German or English, in that in a neutral Danish utterance all main stresses are of the same weight. Danish has no obligatory "sentence accent" or "focus", and there is no tendency to have a heavier stress in the last stress group. Nor is there any general tendency to stress the second member of a noun or verb group more than the first. On the other hand, a number of special close-knit constructions have reduction of stress on the first member and are clearly distinguished from the above mentioned sequences of equal stresses. These two factors are obviously connected. Examples of stress reduction, "unitary stress" (or "unit accentuation", to use Rischel's terminology (1982)) are *et glas øl* 'a glass of beer', *fru Hánsen* 'Mrs. Hansen', *gå i vándet* 'go swimming' in contradistinction to *gå i vándet* 'walk in the water', *læse románer* 'read novels' in contradistinction to *læse románen* 'read the novel', *stå óp* 'get up' but *stå óp* 'stand'. These latter distinctions are characteristic of Danish, and a large number of minimal distinctions could be quoted. Some rules are given by Jespersen (1922) and a thorough discussion with formulation of new rules is found in Rischel 1980 and 1982. By this reduction the reduced member loses stød and, at least when it is monosyllabic, also vowel length, i.e. it is treated in the same way as the first part of compounds with main stress on the second member. Rischel uses the term "unit accentuation" in both cases (for compounds: "intra-word unit accentuation").

2. NUMBER OF STRESS DEGREES (VARIOUS PHONOLOGICAL DESCRIPTIONS)

It has been discussed how many degrees of stress should be set up on the basis of the above mentioned facts. Jespersen (1897-99 and 1922) sets up four degrees (apart from emphasis), but at the same time he remarks that this is arbitrary. And if it is done on a purely auditory basis, he is right that it is arbitrary.

Andersen (1954) sets up four degrees of what he calls "weight" (he uses "stress" in a somewhat different sense): "levissimus", "levis", "semifortis" and "fortis". They are mainly defined by their segmental structure and their possibility of having distinctive vowel length and stød. Levissimus syllables only contain the vowel [ə]; levis syllables may contain all vowel qualities, but do not distinguish long and short vowels and cannot have stød. This covers weak syllables in foreign words (e.g. *veránda*), weak suffixes (*-ig*, *-lig*, *-ing*), syllables reduced by unit accentuation, and a few particularly weak second members of compounds (e.g. *móndag* [man?da] 'Monday'). Both semifortis and fortis may distinguish long and short vowels and may have stød, but the fortis syllable is perceived as dominating the semifortis syllable. Fortis thus covers what was called main stress in compounds above, and semifortis covers reduced stress on second members of compounds. Other Danish dialectologists have taken over these four degrees and their definitions (although they generally allow levissimus syllables to contain the vowel *i* (e.g. Ella Jensen 1944 and Inger Ejskjær 1954), but they call them degrees of stress, not of weight.

A surface phonological description may combine the criterion of syllable structure with the commutation criterion.

"Fortis" and "semifortis" are evidently commutable, cp. the example *en sórt méjse* 'a black tit' and *en sórtmèjse* 'a coal tit' above. There are many examples of this type. It may also be possible, at least for a very distinct style of speech, to set up two degrees of secondary stress based on examples of the type *dýrskùepl̀ds* vs. *krĭgsskùepl̀ds*, mentioned above.

The distinction between levis and levissimus is more problematic. If [ə] is described as a variant of /e/ and /ɛ/ as done by Hjelmslev (1951), there must be a separate very weak stress level conditioning this reduction of vowel quality. I would, however, find it simpler to regard [ə] as a separate phoneme which may sometimes be in free variation with other phonemes. In this case the syllables with [ə] need not constitute a separate stress level.

If, for the time being we consider syllables with weak suffixes like *-ig*, *-lig* and weak syllables in foreign words, like *veránda melodĭ*, etc. as constituting phonologically unstressed syllables, it can be stated that there is commutation between this level and the level of main stress, although there are very few minimal pairs, e.g. *plástik* 'PCV' vs. *plastík* 'plastic gymnastics', and *bílligst* ['bilisd] 'cheapest' vs. *bilíst* [bi'lisd] 'motorist'. Rischel (1970) succeeded in giving rules for the stress placement in Danish morphemes, including most foreign words, but there were some exceptions, e.g. *Áugust* (name) vs. *augúst* 'August', so that stress is not completely predictable.

Syllables with reduced stress due to unit accentuation create difficulties. There is no perceptible difference between pre-

tonic syllables in foreign words and the first part of compounds or phrases with unit accentuation, e.g. *sandálen* 'the sandal' vs. *sandsýnlig* 'probable', or *en banderóle* 'a revenue label' and *en bande røvere* 'a gang of robbers' (and in this paper they are not provided with any stress marks). On the other hand, the distinction between this type of stress reduction and secondary stress is not clear. There are no commutation examples since the syntactic surroundings are always different. The distinction must therefore be based on the possibilities of distinctive vowel length and stød. This criterion is, however, not always quite reliable. In complex compounds there may be loss of stød in syllables with second degree reduction in medial position, e.g. in *vánd* [van?] 'water' in *úndervàndsbåd* 'submarine'; on the other hand, it happens in very distinct speech that a disyllabic word reduced by unit accentuation retains a weak stød, e.g. *han springer óp* 'he jumps up', and a long vowel in a disyllabic word is rarely so strongly reduced in duration that it merges with a phonologically short vowel, e.g. there is a small difference of duration between *læse óp* 'to read aloud' and *læsse óp* 'to load'. Finally, in syllables with short vowel plus obstruent where neither vowel length nor stød are involved, it is a real problem whether there is any difference of prominence, and the syllables must then be assigned to different degrees of stress by generalization from the clear cases. But this may be considered a rather dubious procedure.

In normal conversation - and particularly in allegro speech - there may be what Rischel (1982) calls "loss of ranking difference" or "shrinkage of structure", so that all stresses below main stress are reduced to weak stress. In that case secondary stress in compounds may coincide with weak stress, e.g. with the stress of weak suffixes, as *-ig* and *-lig*, or posttonic syllables in foreign words. Rischel assumes that this happens. In an oral communication Rischel has given the example *útŷske* 'monster' vs. *emfátìske* 'emphatic'. In this case the position after the main stress is the same, so that one might in principle apply the commutation test, but it is hardly possible to find minimal pairs. Personally I am inclined to believe that there is a small difference in syllables immediately after the main stress, e.g. in *býbùs* 'town bus' vs. *rébus* 'rebus', or in *natúrgàs* 'natural gas' vs. *Dálgas* (name), or *úlìge* 'unequal' vs. *mílige* 'possible', but not if there is an intervening weak syllable, in which case the weak derivative ending receives a certain rhythmically conditioned prominence, e.g. *péngegrìsk* 'avaricious' = *málerisk* 'picturesque', or *fláskegàs* 'bottled gas' = *ánanas* 'pineapple' (foreign words may even get stød in this case, e.g. *Páradis* ['pʰɑɑˌdi:ʔs] 'Paradise').

In a more abstract phonological analysis it is evidently possible to reduce the number of stress degrees considerably, deriving most of the differences from the syntactic structure including the relations within compounds. I have argued briefly for this point of view in an old paper (1961 (1948)), maintaining that it is completely arbitrary to set up 3 or 4 degrees

of reduced stress in a language. It is only a way of stating its syntactical possibilities. The important thing is always a comparison between two members and two members only, but this comparison may take place on different levels. It is not possible to identify the degree of stress from one syntactic group to another. Hjelmslev (1951) sets up two degrees of stress in Danish, and Rischel (1972) has discussed these problems in much more detail. He states that it is possible to describe the facts by means of cyclical rules, but that it is simpler to derive the prominences in one operation directly from the syntactical tree structures. Basbøll (1978) finds it still simpler to base the rules on boundaries of different degrees. One thus comes down to a distinction of only two phonological stresses (cf. also Liberman and Prince (1977)).

3. THE PROBLEMS OF PHONETIC MANIFESTATION

Reducing the stresses to two on an abstract phonological level does not, of course, make it superfluous to look for acoustic and physiological manifestations not only of strong and weak stress, but also of stress reductions. They still have a function in distinguishing different constructions.

Some of the questions that may be raised are: What are the acoustic properties (if any) distinguishing different degrees of stress and different types of stress reduction (apart from stød, vowel length and specific vowel quality ([ə]))? Is it duration, Fo, or intensity, or a combination of two or all of them, and are these combinations different in different types of reduction? In the first place one might compare main stress and weak stress. The material used in the present investigation was, however, not intended specifically for this purpose, and it is rather limited what can be said about duration and intensity, whereas the Fo contours were so regular in spite of segmental differences that they have been treated in more detail. The main purpose was a comparison between main stress and secondary stress (of the first degree), e.g. *mále bøtten* and *málebøtten*, and between secondary stress and stress reduced by unit accentuation. Although they do not occur in the same syntactic surroundings, they may contain the same segments and occur in the same prosodic surroundings, e.g. *Ótto tog málebøtten frém* 'Otto took out the painting box' and *Ótto vil mále bøtte fém* 'Otto will paint box number five'. As for the difference between secondary stress and weak posttonic stress, I have only (at the last moment) analysed one pair: *útỳske* vs. *Herrenhútiske* (or *polítiske*).

4. STRESS GROUPS

On the preceding pages the Danish prosodic system has been seen in relation to the grammatical structure, to simplex and complex words and phrases. This is useful for a comparison with other languages, and as a background for the description of the acoustic manifestation. But when describing this mani-

festation it may be more useful to start from actual utterances with various stress reductions, and to divide the utterances into stress groups each containing one and only one main stress. At the abstract level where the reduction rules operate, the boundaries between such groups must be syntactically determined, e.g. *óverlægen / drák / et glas øl* 'the chief physician drank a glass of beer', where *lægen* is reduced in relation to *óver*, and *glás* in relation to *øl*. This is the type of boundaries used by Rischel (1982) and Basbøll (1972) for the purpose of the phonological analysis of reduction rules, and also by Andersen (1954). However, for a phonetic description it may be different. That will depend on the degree to which syntactic boundaries are reflected in the prosodic structure. There may be expected to be various interferences in different norms and styles of speech. For Advanced Standard Copenhagen (ASC) Thorsen considers each prosodic stress group to start with the stressed syllable, as it will be described in more detail in the next section.

III. PREVIOUS DESCRIPTIONS OF THE ACOUSTIC MANIFESTATION OF STRESS IN DANISH

Jespersen (1897-99, 1914, 1922) defines stress physiologically. A stressed syllable is pronounced with a higher amount of total energy, including expiration, vocal cord movements and supraglottal movements than an unstressed syllable, and the listener is able to judge the energy used by the speaker by comparison with the way he would pronounce the sequence himself. Stressed syllables may be higher or lower in pitch than unstressed syllables. They are normally higher, and this is the case in Danish. The higher pitch of stressed syllables in Danish is also assumed by Bo (1933) and Jerndorff (1896, quoted by Bo). But Jespersen and Bo state that there may be deviations in particular styles. Bo finds, e.g., a tendency for small children to reverse the pattern. But in order to find examples of a normal lowering of stressed syllables these authors refer to foreign languages. Jespersen mentions Swedish and Norwegian as examples of languages where the stressed syllables may have lower pitch. Forchhammer (1954) quotes Sievers for the observation that this type of pitch movement is used in Southern German dialects, whereas in Danish it may be a sign of irritation, or it may be an individual peculiarity which will be very unfortunate for one who wants to become an actor. None of these authors refer to the occurrence of this peculiarity in Copenhagen speech. Arnholtz (1939-40) mentions the low pitch level of stressed syllables as a recent development in vulgar Copenhagen speech which is now spreading to other classes. He considers this to be a very dangerous development which will render the whole of our classical poetry inaccessible (!), and which must therefore be fought down. (In the following issue of the same periodical (1941) Bergsveinsson publishes an Fo curve based on a gramophone record spoken by Arnholtz, and which shows low falling Fo on stressed syllables.)

Kroman (1947) has a different attitude. He states that there is no standard pitch movement in Danish, but different norms in different parts of the country according to the dialect of the area in question. Most Danish dialects have higher pitch on stressed syllables than on unstressed syllables. But in Northern Zealandish it is different. The main part of North Zealand has a low falling-rising pitch on stressed syllables with a higher pitch on following unstressed syllables in words without stød, but a rising-falling contour with lower pitch on the following unstressed syllable in words with stød, whereas a few dialects on the East coast of Zealand and on Amager have low rising pitch on the stressed syllable in all word types (and the same is true of Bornholm). The Copenhagen pronunciation is said to be influenced both by the coastal dialect and by other North Zealandish dialects. Poul Andersen (1949) is very sceptical as to the existence of a difference between words with and without stød in North Zealand, but agrees that there is low pitch on the stressed syllable and high pitch on a following unstressed syllable. He also mentions this in his textbook (1954) and adds that the same is often heard in Copenhagen speech; but in Standard Danish the stressed syllable has higher pitch, and in disyllables the contour is rising-falling. Andersen also describes the stress of compounds with main stress on the first member. This first member is dynamically stronger, and is said on a higher pitch than the second member (except at the end of questions). Moreover, the two members are temporally closer together than separate words, and the first member is characterized by a dynamic and tonal delay or gliding on, announcing that more is coming. This is no doubt a very good description of Andersen's own pronunciation which (although he is from Copenhagen) is not in its prosodic aspects influenced by the younger Copenhagen norm but rather by a lifelong occupation with Funish dialects, and I think it also covers the conservative standard (described by Jespersen and Bo) which I have spoken myself. Smith and Thyme (1978) consider the tempo to be the most important feature in compounds.

In a long series of studies (1976-1983) Nina Thorsen has demonstrated convincingly that low pitch in stressed syllables is now the general norm in Standard Copenhagen. Her informants are relatively young (in their mid twenties to mid forties), and in her first papers she cautiously talks about "Advanced Standard Copenhagen", but in her latest papers she describes it as simply Standard Copenhagen, and this is certainly justified. The quotations above from older phoneticians seem to indicate that it is a relatively recent phenomenon in the higher sociolect, but the intrusion into the higher norm may well have taken place earlier than these quotations seem to indicate. From my childhood, about 60 years ago, I remember very well the low pitch on stressed syllables in the speech of my Copenhagen cousins, and also in the speech of an uncle and aunts from Copenhagen, born in the eighties of the last century, and that was the higher sociolect. My attitude to this norm was very negative at that time (that is why I remember it clearly), but I have adopted it since then (though

it took me a long time to realize that I had changed my way of speaking).

Thorsen (e.g. 1979 and 1980c) gives the following description of the relation between stress and Fo in the Copenhagen Standard: A Danish utterance may be divided into prosodic stress groups, each beginning with a syllable with main stress. The stressed syllable is said on a relatively low pitch, and from there the pitch glides or jumps (about 2-3 semitones) up to the first posttonic syllable. The following posttonic syllables have gradually decreasing pitch. If there are less than four unstressed syllables they generally remain above the level of the next stressed syllables, and if there is only one posttonic syllable, it may remain at the level of the preceding stressed syllable. Unstressed syllables preceding the first stress group are normally said on a lower pitch, at least after a pause. The extent of the fall depends on the number of posttonic syllables and on the individual. Some speakers have steeper falling slopes than others. She places the boundary between the prosodic stress groups at the start of the stressed syllable and not in accordance with syntactic boundaries, because she has found that in Advanced Standard Copenhagen syntactic boundaries are not reflected in the Fo contour (e.g. 1980c), and the boundary is placed before and not after the stressed syllable because the relation between the stressed syllable and the posttonic syllables is more stable and can be described in a much simpler way than the relation between the stressed syllable and the preceding unstressed syllables. More specifically, the prosodic stress group boundary seems to be located immediately before the stressed vowel, excluding any initial consonants which associate tonally with the preceding stress group, if any, or constitute a prosodic unit apart (Thorsen 1983a, p. 189-190, and 1983b).

The Fo patterns of the stress groups are superposed on the general intonation contour, which is decreasing in statements, level in questions not otherwise characterized as questions, and with intermediate slopes in questions characterized as such by other means (inversed word order or question words) and in non-final phrases. The intonation contour is carried by the stressed syllables,which thus have successively decreasing pitch in statements and the same pitch in questions. This description is summarized graphically in figure 1 (from Thorsen 1980b, p. 122).

The intrasyllabic Fo movements are determined by the Fo pattern and by segmental factors. Long stressed vowels are generally falling-rising with predominant fall, short stressed vowels are falling (there is thus truncation, not compression). Both may, however, be rising in the first stress of the utterance and in cases where the preceding unstressed syllable(s) go(es) down below the next stressed syllable (1980b). The first posttonic syllable (which is placed at the peak of the pattern) may be falling, rising-falling, or (more rarely) rising. All following posttonic syllables have falling pitch.

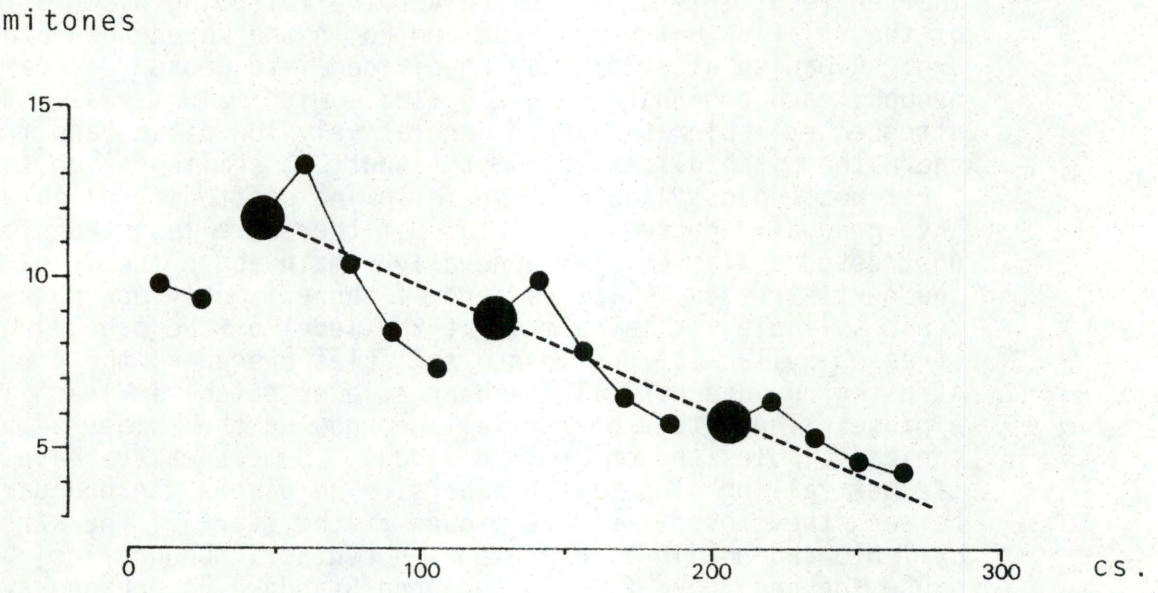


Figure 1

Thorsen's model for the course of fundamental frequency in a short terminal declarative sentence in Copenhagen Standard Danish. The big dots represent stressed syllables, the small dots unstressed syllables.

Secondary stress in compounds is only mentioned briefly (1980b, p. 125 footnote). Syllables with secondary stress are said to behave tonally like unstressed syllables (which is also in accordance with my own earlier observations), whereas they resemble stressed syllables in all other respects (vowel quality, vowel quantity and stød).

Thorsen recognizes three degrees of stress: main stress, secondary stress, and weak stress. Danish has no obligatory sentence accent, but there may be emphasis for contrast, which is characterized by a high rising pitch on the syllable in question, and (more important) a deletion of the Fo deflections (the rises from the stressed syllables) of the neighbouring stress groups to the effect that the immediate surroundings except the first posttonic syllable fall away sharply from the stressed syllable of the emphasized word (1979 and 1980b).

In a paper with Jul Nielsen (1981) Thorsen gives brief descriptions of Fo patterns in Jutlandish. The material is limited, a small number of sentences read by two dialect speakers from Thy in Northern Jutland and two speakers of Regional Standard Danish from the Århus area, but the differences from the Copenhagen pattern are obvious. All four informants have high

rising pitch on the stressed syllable and lower falling pitch on the posttonic syllables with the difference that the first posttonic syllable starts at a rather high level in the speech of the Århus informants, but very low in the sentences of the Thy dialect speakers.

Schematically the three patterns are given in figure 2 (Thorsen 1982a):

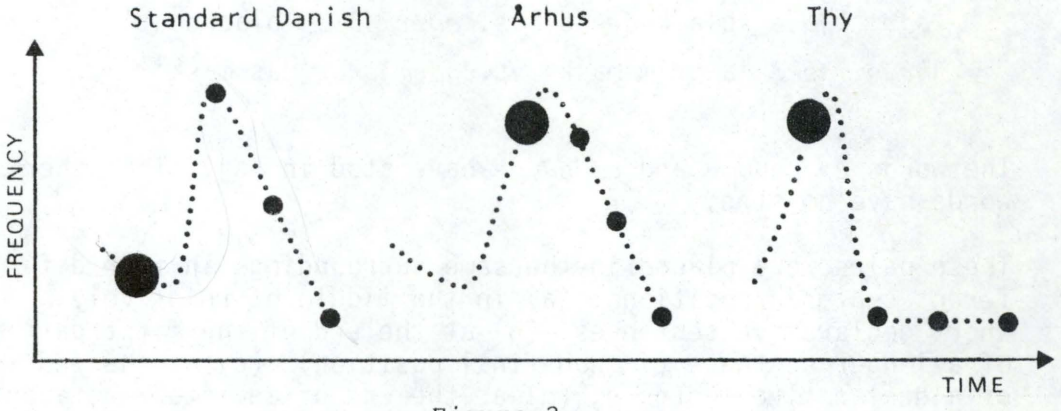


Figure 2

Stylized stress group patterns according to Thorsen (1982a) in three types of Danish: Standard Danish, Århus regional speech, and Thy dialect in West Jutland. The big dots represent stressed syllables, the small dots unstressed syllables.

The patterns in Copenhagen and Jutland can be described as inversed, but it is also, according to Thorsen, possible to describe them as the result of different timings in relation to the same undulating Fo wave. This has the advantage that the direction of the intra-syllabic movements can be more or less predicted from their position on the wave.

Thorsen has not published any measurements of duration and intensity, but she has stated that it is her general impression that the hierarchy of acoustic cues for stress in Danish is (1) Fo, (2) duration and quality, and (3) intensity (1980b), in spite of the fact that vowel duration is phonological in Danish and should thus, according to Berinstein, be the least important cue.

IV. MATERIAL, INFORMANTS AND MEASUREMENTS

The present paper is based on two recordings, one from 1961 and one from 1982. The 1961 material was only partly processed at that time and put aside due to lack of time. It is here considered as a pilot investigation and is only partly utilized.

The 1961 material comprised a number of word groups with two main stresses and comparable compounds with main stress plus secondary stress, consisting of the same words. Some had change of stød, and some were rather complicated. For the present investigation therefore only the following were utilized:

sød súppe 'sweet soup' vs. *sødsùppe* 'a kind of fruit soup'

sórt méjse 'black tit' vs. *sórtmèjse* 'coal tit'

sórt kjóle 'black dress' vs. *sórtkjòle* 'blackcoat'

hvás máske 'a grim mask' vs. *gásmàske* 'gas mask'

The words *sød súppe* and *sødsùppe* have stød in [ð]. The other words have no stød.

These pairs were placed in the same surroundings in four different sentence positions: (a) in the middle of relatively short declarative sentences, (b) at the end of the first part of a longer sentence (in non-final position), (c) at the end of a declarative sentence, (d) at the end of an interrogative sentence. (a) or (c) were used as answers to (d), e.g. (d) *Har du nógensìnde sét en sórtmèjse?* (a) *Já, jeg sǻ en sórtmèjse i skóven i går.* 'Have you ever seen a coal tit?' 'Yes, I saw a coal tit in the forest yesterday'.

Moreover, there were a number of word groups with stress reduction due to unit accentuation compared to groups with two main stresses. The following have been measured:

kássen blev sát óppe på lóftet 'the box was stored in the loft'
vs. *kássen blev sat óp på lóftet* 'the box was moved to the loft'
and *han sátte kássen óppe på lóftet* 'he placed the box in the loft'
vs. *han satte kássen óp på lóftet* 'he moved the box to the loft'.

Further measurements were made of the words *en kánde* 'a jug', *en vándkànde* 'a water jug', *en kande vánd* 'a jug of water' and *en tóm kánde* 'an empty jug', placed in the frame *han fandt ... i køkkenet* 'he found ... in the kitchen'. All the sentences used are listed in Appendix I.

This material was read once by 10 informants. Three were left out here because their type of speech was considered to be sufficiently represented by the others. The seven informants whose curves were measured were: (a) BS, born 1922, and HP, born 1938 , both representing "Advanced Standard Copenhagen speech" (ASC); (b) PD, born 1905, EF born 1911, and JR, born 1934, all representing Standard Danish of a somewhat more conservative type (CSC). They have grown up outside of Copenhagen in civil servant families speaking Standard Danish, but they had all lived for many years in Copenhagen; (c) BF, born 1935,

and SR, born 1940, both in Southern Jutland; they speak Standard Danish on a (Jutlandish) dialectal background (DSC). (The sentences with *sórt méjse* and *sórtmèjse* were only spoken by four informants.)

The 1982 material likewise comprised word groups with two main stresses compared to compounds with main stress on the first member, but moreover in various cases comparable word groups with unit accentuation. The words were varied in syllable type and syllable number. The following triplets (and pairs) were recorded:

1 *sød súppe* 'sweet soup', *sødsùppe* 'a kind of fruit soup', *en sød suppedás* 'a pretty kettle of fish' (more idiomatic: *en køn suppedas*)

2 *mát lámpe* 'dim lamp', *nátlàmpe* 'bedside lamp', *mát lampét* 'dim bracket lamp'

3 *Han lod Spíes réjse ságen* 'he let Spies bring the matter up'
Han tog Spíesrèjsen méd 'he included the Spies-travel' (Spies is the name of a well-known travel agent),
Han lod Spíes rejse méd 'he let Spies come along'

4 *Péters vén skáber sig* 'Peter's friend puts on an act'
Nú giver vénskàbet problémer 'Now the friendship gives problems'
Péters vén skaber problémer, 'Peter's friend gives problems'

5 *gás brænder* 'gas burns', *gásbrènder* 'gas burner', *gás brænder néd* 'gas burns down'

6 *måle kássen* 'paint the box', *málekàssen* 'painting box', *måle Casanóva* 'paint Casanova'

7 *måle bøtten* 'paint the pot', *málebøtten* 'the painting pot',
måle bøtte fém 'to paint pot number five', *male bøtten grøn* 'to paint the pot green'

8 *lámme lår* 'lame thighs', *lámmelår* 'lamb's thigh'

9 *det várme apparát* 'the hot instrument', *vármeapparàtet* 'the heater', *varme apparátet óp* 'to heat the instrument'

10 *mís forstår* 'Pussy understands..', *mísforstår* 'misunderstands..', *mís forstår spøg* 'Pussy understands jokes'.

For numbers 3 and 4 the whole sentences are quoted in order to facilitate the translation, but the other words and word groups were also placed in sentences, which were as similar as possible, e.g. *Otto vil måle bøtten straks i mórgen* 'Otto will paint the box immediately tomorrow morning', *Otto tager málebøtten frém i mórgen* 'Otto will take out the painting pot

tomorrow', and *Ótto vil mále bøtte fém i mórgen* 'Otto will paint pot number five tomorrow'. - The words and word groups were always placed in the middle or beginning of declarative sentences since the 1961 material did not show any basic differences between the different positions, and since measurements turned out to be difficult for words placed finally in the sentences. There was also always a stressed syllable before the test word except for the examples *gásbrènder*, *lámmelǎr* and *málekàsse* and their counterparts. The sentences are listed in Appendix II.

As mentioned in section II disyllabics reduced by unit accentuation retain some vowel length and are thus not quite as weak as monosyllabics immediately before a strong stress. Thus, of the above examples only *lampét* is a reliable example of an unstressed syllable with full vowel. A few more examples of this type were recorded in order to be able to compare weak stress and main stress, e.g. *sandálen* 'the sandal' vs.*Sánddàlen* 'Sand Valley' (a place name), *barnágtig* 'childish' vs. *gárnàgtig* 'thread-like'. Moreover, one informant read the words *bílligst* 'cheapest' and *bílist* 'motorist'. A recording of the same two words spoken by five subjects for a different purpose was also utilized.

Finally, secondary stress in a compound was compared to a weak derivative suffix in the pair *útỳske* 'monster' vs. *polítiske* 'political' and *Herrenhútiske* 'Herrenhutish'. These words were also placed in sentences, which are listed in Appendix II.

All sentences were mixed in four different randomizations, and the list was read twice by all informants. There are thus eight examples of each sentence per informant. Most sentences were read by 6 informants, but numbers 5 and 8 were read by only four informants, the pair *barnágtig* and *gárnàgtig* by three, and numbers 3 and 4 only by one speaker.

The lists were mixed with sentences used for measurement of segment duration (Fischer-Jørgensen 1982). Those sentences contained some compounds, which can be compared to the first member used as a simplex, i.e. *náttefròst*, *Dánnevàng*, *Dánfòss*, *mándfòlk*, *mándefàld*. They also contained a number of words with weak syllables in *-e* [ə] and *-er* [ɒ] or [ɔ], which together with some words from the main corpus were used in a more cursory analysis of the difference between such syllables and syllables with main stress. The words used were (1) in sentence medial and final position: *dánne*, *nátte*, *báste*, *lákke*, *mĭsse*, *mánse*, *sønner*, *láser*, *kaláser*, *pláner*, *spĭser*, *mulátter*, *básser*, *polákker*; (2) in sentence medial position only: *dánse*, *básse*, *tále*, *Dánnevàng*, *mĭssekàt*, *mǎleskĭve*, *tálestrøm*, *náttefròst*, *sønnerne*, *máser*, *plánerne*, *básserne*; (3) in sentence final position only: *máse*, *mále*, *ládte*.

The main informants used in this investigation were the same as those used in the investigation of segment duration (1982). There were 6 main informants: NR, born 1942, representing Advanced Standard Copenhagen (ASC), EF, born 1911, represent-

ing a somewhat more conservative norm (CSC), PH, born 1947, speaking Standard Danish with a very slight tint of Lolland regional language. His deviations from the ASC norm are, however, probably due rather to the more conservative norm of his parents than to Lolland dialect. He is therefore regarded as a CSC speaker. Moreover, OT, born 1928 in Himmerland in East Jutland, SR, born 1940 in Hjerting in South Jutland, IE, born 1926 in Vinderup in West Jutland, all speaking Standard Danish on a perceptible Jutlandish basis, which is relatively weak for OT, but very clear in the case of IE. In his treatment of weak syllables OT belongs rather to the CSC group, but in the contour of compounds he is in some respects closer to the other speakers with Jutlandish background, who will here be called DSC speakers. All main informants have lived in Copenhagen for many years (between 16 and 52 years). The triplet *mále kássen* etc. was not read by SR, but by two further informants, NK, born 1915, speaking Standard Danish with a perceptible Funish accent, and by BJ, born 1946, speaking Standard Danish with a perceptible Jutlandish accent. (Somewhat more detailed information is given in Fischer-Jørgensen 1982.)

Some of the sentences were also read by a number of dialect speakers. The pair *mále bøtten* and *málebøtten* was read by four Funish dialect speakers (LA, MA, HV and IP) and two speaking the Regional Standard Danish (RSD) in Odense on Funen (EK and EH). IP also read the pair *mát lámpe / nátlàmpe*, and this pair was also read by a speaker of the lower Copenhagen sociolect (ST). The same two pairs and, moreover, the pairs *Sánddàlen / sandálen* and *sød súppe / sødsùppe* were read by six bilingual dialect speakers from Jutland, who read the sentences both in their dialect and in their Regional Standard Danish, viz. TA from Thy, EA from Mors, JD from Fjends Herred, BT from Uldum (all speaking Western Jutlandish); PN from Vendsyssel (Northern Jutlandish), and PM from Himmerland (Eastern Jutlandish). The same pairs were finally read by LH, who speaks Århus Regional Standard Danish. (More detailed information on the informants is given in Fischer-Jørgensen 1982.)

The dialect speakers also read a number of sentences containing simplex words used for the measurement of segment duration. These sentences are used as basis for a cursory analysis of unstressed syllables, and particularly the general Fo-contour. Recordings of six speakers of East Jutlandish dialect, and seven speakers of North Zealandish were used for the same purpose.

The recordings of the main informants were made on a semi-professional tape recorder in a sound-treated room at the Institute of Phonetics, University of Copenhagen.

The Jutlandish speaker PM, the Funish speaker JP, and the Copenhagen speaker ST were recorded in Copenhagen. The recordings of the other Jutlandish speakers were made on a Nagra tape recorder at the Institute for Jutlandish Language and Culture at the University of Århus. Two recordings of Funish

speakers were made at the Institute of Linguistics in Odense. The recordings of the other Funish speakers and those of the Zealandish speakers were made in their private homes.

All tape recordings were processed at the Institute of Phonetics in Copenhagen; they were registered on mingograms, comprising a duplex oscillogram, an Fo curve, and two intensity curves (one unfiltered, one highpass filtered at 500 Hz).

The duration of all segments and of the whole word (or member of a compound) was measured, generally with an accuracy of ± 2.5 ms. However, [ø] and [ðʔ] in *sød* could not be delimited, and the same was often true of [l] and [ə] in [ma:lə].

The words *vârmeapparàtet* and *vârme apparât(et)* made difficulties and could only be partly segmented (these examples were intended specifically for a comparison of the intensity of the vowels). In some cases the initial or final consonant could not be delimited from the surrounding words (e.g. *en mât lâmpe*). The dialect recordings presented considerably more problems.

The Fo contours were traced for three examples of each word and sketched for all words. They turned out to be extremely stable, so that it was considered superfluous to measure all curves and to construct average curves. In unstressed syllables the duration is so short that no glide can be heard, and one might choose to indicate only the point in time at a distance of 2/3 from the start of the vowel, which according to Rossi (1971a) and Rossi et al. (1981) corresponds to the perceived pitch level, but for stressed vowels some glides may be heard, and it is possible that the movement of a following sonorant should be added, so for the time being I have preferred to simply trace the Fo movement and to present some stylized curves describing the main features. According to various studies in other languages (e.g. Fry 1958), the absolute distance in Fo is not important for the perception of stress. It is generally found to be an all-or-none phenomenon. However, this may not always be the case. In a perceptual experiment with synthesized isolated two-syllabic nonsense words and higher Fo on the second syllable Rosenvold (1981) found that Danish ASC listeners heard the first syllable as stressed when the second was much higher, and the second syllable as stressed when it was only slightly higher (but there was a fairly clear cross-over point). Thorsen (1983b) found an opposite tendency, and assumes that the second syllable in Rosenvold's test material may have been too high to be a likely stressed syllable in a one-word non-emphatic terminal declarative utterance.

As for the intensity I have measured peak intensity in the vowels, but because of difficulties with the calibration the measurements were in mm. On the whole, the measurement of intensity gives many problems because the relations to perception and to Fo are not sufficiently investigated. Very often a secondary stress has a higher intensity than the main stress in Danish, probably because Fo is higher. But what do we perceive? I will return to these problems in section V.A.1.

V. RESULTS

In the following sections different speakers will be compared. It may therefore be practical to list the informants for Standard Danish again, divided into groups.

		1961 material	1982 material
ASC speakers:	Speakers of Advanced Standard Copenhagen	HP, BS	NR
CSC speakers:	Speakers of a somewhat more conservative norm	PD, JR, EF	EF, PH, OT
DSC speakers:	Speakers with a clearly audible Jutlandish background	BF, SR	IE, SR
	Speaker with a clearly audible Funish background		NK

A. THE RELATION BETWEEN MAIN STRESS AND WEAK STRESS

As the material was not directly intended for the purpose of comparing main stress and weak stress, it is rather heterogeneous in this respect, but it contains a large number of both categories, and a few examples were constructed specifically with this comparison in mind.

1. VOWEL QUALITY AND STØD

Most unstressed syllables in native Danish words have the vowels [ə] or [ɒ], or a syllabic sonorant. *-er* is pronounced [ɒ] in a more conservative norm, [ɔ] in a more advanced norm. It is here transcribed [ɒ] in accordance with my 1982-report on segment duration. With the exception of [ɒ] none of these sounds can appear in stressed syllables, and they are thus a reliable signal for lack of stress (= weak stress). But many unstressed syllables in foreign words and in Danish words that usually or occasionally lose their stress in the sentence have full vowels, so that a full vowel does not give any information about stress. On the other hand, only syllables with main or secondary stress can have stød, so that a stød is a reliable signal of stress. Just to give an idea of the extent to which these cues may be used, I counted their number in a 1000 syllable corpus from four not very technical book reviews in a newspaper. There were 142 occurrences of stød and 189 occurrences of schwa or syllabic consonants. This means that 331 or 33% of the syllables were characterized as stressed or unstressed by these means.

Stød and schwa (or sonorant consonant) are thus important cues for stress in Danish, but not sufficient cues.

2. DURATION

Table I gives a survey of the duration of unstressed vowels in the main material. There were 92 averages of the vowel [ə] in utterance medial position, distributed over seven informants. [ə] is always very short in this position, the grand mean for all speakers being 4.4 cs. Since the material was not intended for the measurement of unstressed vowels, it is rather unsystematic as far as the quality of the preceding stressed vowel is concerned. In most cases the preceding stressed vowel was a short [a] which is, of course, much longer than [ə], the average difference being 8.3 cs. But [ə] was also significantly shorter (at the 1% level) than preceding short vowels with smaller intrinsic duration than [a]. In utterance final position the delimitation was rather uncertain.

There were 84 individual averages of the weak ending *-er*, pronounced [ɔ] or [ɒ]. It is considerably longer than [ə], the grand mean in utterance medial position being 7.2 cs. It is always significantly shorter than a preceding stressed short vowel of the same intrinsic duration ([a]). In proclitic position (in the prefix *for-*) it is shorter than in enclitic position. It is even shorter than a preceding stressed [i].

In four averages the quality of the stressed and the unstressed vowel were identical ([a]). Here, too, there was a significant difference in duration (1.9 cs) between the two vowels.

Some examples were constructed with the specific aim of comparing stressed and unstressed syllables with the same (or almost the same) segmental structure (see section IV), i.e.: *Sánddàlen / sandálen*, *gárnàgtig / barnágtig*, *en mát lámpe / en mát lampét*, *billigst / bilíst*, and the nonsense words *mámam / mamám*. In every pair it is possible to compare a stressed syllable in one member with an unstressed syllable in the other member of the pair, and in the two last pairs both syllables can be compared. Moreover, the stressed vowel can be compared with an unstressed vowel of the same quality in the same word or phrase in *barnágtig*, *en mát lampét*, and in both members of the *bilíst* and *mamam* pairs. As for *sandálen* this is not possible, because *a* and *l* could not always be delimited and because of the stød in the *a* of *dalen*.

The main durational differences are given in table II, and more detailed measurements for the individual speakers together with information on Fo and intensity are found in Appendix III, 1-5. For two of the pairs there is more than one reading by NR. They have been averaged before the grand mean of all speakers was calculated.

The first three pairs were placed in the middle of sentences with four stress groups. The pair *billigst / bilíst* was placed in a frame *det er ... de siger*, and *mámam / mamám* were said in the frame *de sagde ...* The pronunciation of this latter pair did not always sound quite natural (although the second member has a meaning as a nursery word for food), and for

Table I

Duration of unstressed vowels in Danish (in cs). N = number of individual averages, each comprising 8 tokens. Measure (a) in final position includes final weak vibrations, (b) stops approximately when the formants stop. ɒ *(1) is posttonic,* ɒ *(2) is pretonic.*

A. utterance medial position

	NR	PH	EF	OT	BJ	SR	IE	NK	av.
	(N)	(N)	(N)	(N)	(N)	(N)	(N)	(N)	
ə	(16) 4.3	(16) 3.3	(26) 4.4	(14) 4.3	(7) 6.1	(6) 4.3		(7) 4.4	4.4
(1) ɒ	(15) 6.8	(15) 7.4	(24) 7.1	(15) 7.4	(5) 8.9		(5) 6.8	(5) 6.3	7.2
(2) ɒ	(3) 5.2	(3) 3.8	(5) 5.3	(3) 6.4		(3) 5.3	(3) 6.9		5.5
i	5.3	(1) 6.1	(1) 5.8	(1) 5.8	(1) 6.5		(1) 8.4		6.3
a	(2) 6.0	(1) 6.5	(1) 5.5						6.0

B. utterance final position

	NR		PH		EF		OT			
	(N) a	b	(N) a	b	(N) a	b	(N) a	b	a	b
ə	(4) 9.2	5.3	(4) 6.6	4.6	(12) 8.7	5.7	(5) 6.6	5.1	7.8	5.2
ɒ	(3)13.9	8.0	(3) 8.5	5.7	(15)10.0	7.0	(3) 9.9	6.7	10.6	6.9

Table II

Differences in cs between stressed and unstressed syllables and vowels in comparable word pairs (A), and between vowels in consecutive syllables of the same word (B). Numbers in parentheses indicate the number of individual averages, each comprising 8 tokens.[1]

	diff.	'mad[2] 'lambə 'mad lam'pʰɛd	'san,da:ʔlɳ san'da:ʔlɳ	'ga:n,agdi ba(:)n'agdi	'bilisd[3] bi'lisd	'mamam ma'mam	bi'lisd 'bilisd	ma'mam[4] 'mamam	av.
	N	(6)	(6)	(3)	(5)	(6)	(5)	(6)	
A	vowel	4.3	2.4	6.9	2.7	2.9	1.7	0.8	3.2
	syll.	7.4	6.8	10.2	6.7	4.8	5.2	3.1	6.3
B	vowel	'mad-lam 3.4	'bi-li 1.4	'ma-ma 1.7					2.2
	vowel	ban-'agd -3.6	bi-'li -3.0	ma-'ma -2.0					-2.9

[1]) NR made three readings of *billigst/bilist* (there are thus 24 tokens) and two readings of *Sánddàlen/sandálen* (16 tokens)

[2]) *bdg* are generally voiceless.

[3]) *-st* was left out in *billigst/bilist* (*t* could not be delimited, and *s* was of approximately the same duration; the *g* in *billigst* is not pronounced).

[4]) The last *m* in *mámam* and *mamám* was not included, because it could not be delimited.

speaker IE they sounded so unnatural that they have been left out in the grand mean.

It appears from table II A that the stressed syllable is always considerably longer than the corresponding unstressed syllable in the other member of the pair (6.3 cs on the average). The differences are significant for all individual speakers except for the difference between the second syllables of *mámam* and *mamám* for two speakers; the magnitude of the difference is such that it will in almost all cases be clearly perceptible, the difference limen for vowel duration being around 2.5 - 3.5 cs (see the references in Eli Fischer-Jørgensen 1982, p. 159).

The differences in duration between stressed and unstressed syllables are distributed over the segments, and there are individual differences on this point (see Appendix III), but the vowel is significantly shorter in unstressed than in stressed position in almost all individual averages. In the second syllable of *bílligst* and *bilíst* the main difference is (for all speakers) in the *l*, which is 3.5 cs longer in *bilíst* than in *bílligst*.

A comparison between the vowels within the same word (or phrase) also shows that the vowel in unstressed syllable is significantly shorter than the vowel in a preceding or following stressed syllable in *mát lampét*, *barnágtig* and *bilíst* but rarely in *bílligst*, where the main difference is in the initial consonant, and not always in *mámam* and *mamám*.

It is conspicuous that the first [ɑ] is 3.6 cs shorter than the second [ɑ] in *barnágtig*, although the word *barn* [bɑ:ʔn] has a phonologically long vowel with stød in the norm of all the speakers, i.e. both stød and vowel length are lost in *barnágtig*.

Thus, although vowel length is phonological in Danish, it plays an important role in the manifestation of stress, contrary to Berinstein's (1979) assumption. Moreover, the reduction of duration in unstressed syllables also contributes to the reduction of loudness. The stressed vowels of the syllables *lám*, *sán*, *gárn*, *bí*, and *má* (first syllable) have an average duration of 11.3 cs and the corresponding unstressed syllables of the other members of the pairs an average duration of 7.3 cs. According to the Munson curve (see, e.g., Rossi et al. (1981)), a difference between 7.3 and 11.3 cs should give a reduction of about 4.5 phones.

Duration seems to play the same role in the dialects as in the standard language. Since all Jutlandish dialects and some of the Funish and Zealandish dialects have apocope, there are not many examples of [ə] in the dialect recordings. In the Zealandish dialects without apocope the duration of [ə] is 5.3 cs on the average, and the duration of *-er* is 6.1 cs. In the Jutlandish dialects the duration of [ə] in *Kalle* is 5.9 cs, and the duration of *-er* is 5.5 cs, i.e. [ə] is slightly longer

and *-er* slightly shorter than in the standard language, and in Jutland they are not distinguished in duration, and they may not be distinguished in quality either, *-er* being [ʌ] or [ə]. But on the whole, the durations are comparable to those in the standard language.

The name *Nánna* was read by seven East Jutlandish speakers and by six North Zealandish speakers in initial position of the sentence. The second [a] was shorter by, on the average, 5.4 cs with the East Jutlandish speakers and 3.9 cs with the North Zealandish speakers, and this difference was significant for all individual informants.

The pair *Sánddàlen / sandálen* was read by some of the Jutlandish informants. Some of them had, however, a long [ɔ:] in the first syllable of *Sánddàlen*, so that their recording could not be used (see Appendix III, 6-7). For the West Jutlandish speakers and the North Jutlandish speaker, the syllable *san-* was 6.4 cs shorter than *Sán-* and the vowel 2.7 cs shorter. The East Jutlandish speaker PM and the speaker of the Århus RSD had a difference of 6.4 cs in the syllable and 2.3 cs in the vowel. This is in good agreement with the standard speakers.

3. THE Fo CONTOUR

All informants have a difference in Fo between stressed and unstressed syllables, but not quite the same. It is therefore practical, for the purpose of this description, to distinguish between the groups set up in the beginning of section V: Advanced Standard Copenhagen (ASC), a somewhat more conservative norm (CSC), and Standard Copenhagen on a dialectal (i.e. Jutlandish or Funish) background (DSC). (As far as the Fo contour of stressed and unstressed syllables are concerned, OT belongs to the CSC-group.) It is also practical to treat the dialect speakers together with the speakers of the standard language because the DSC-speakers are influenced by their dialectal background. The growing influence can be seen by comparing SR's two recordings from 1961 and 1982. In 1961 he had lived in Copenhagen for only one year, and at that time he had a very small rise (about 1½ semitone) on the first posttonic syllable and sometimes none at all. In the 1982 recording his rise is higher and consistent (at least after single consonants; there were no examples of clusters).

a. Different types of contours seen as different timing

It is possible to describe some of the deviations in the way suggested by Thorsen, i.e. as a different timing in relation to the same or a similar Fo wave. Thus both the relative height and the direction of the Fo movement in stressed and unstressed syllables can be seen as a consequence of their placement on the wave. In figure 3 I have depicted some typical cases. These sketches should, of course, only be seen as rather crude indications of the main trends.

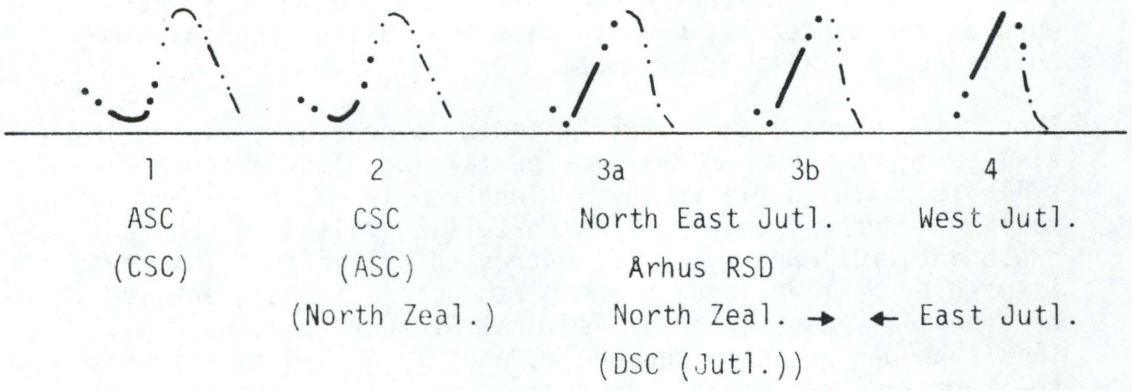


Figure 3

Stylized stress group patterns for various types of Danish. Thick lines represent stressed vowels, thick dots surrounding sonorant consonants, and thin lines unstressed syllables.

Type 1 is the typical ASC contour as described by Thorsen. The stressed syllable is low, often falling (i.e. starting at the end of the falling slope of the previous wave) with a jump or glide up to the first posttonic syllable, followed by a fall through the following unstressed syllables. This type is used by all the ASC speakers of the present investigation (NR, HP, BS, two young Copenhagen speakers JBC and ODL, and the speaker of the lower sociolect ST). The second unstressed syllable may be higher than the first posttonic if the stressed plus first posttonic syllables are too short to reach the peak of the pattern (Thorsen 1982b). Examples from the present investigation are, e.g., *mon sønnerne kómmer*, and *Nánna tákker for gáven*.

Type 2 differs from type 1 by having a more rising stressed syllable. It is often used by the CSC speakers, who may, however, also use type 1. Figures 4 and 5 give some examples of the difference between type 1 and 2. The sentences in figure 4 are read by the ASC speaker HP and the CSC speaker PD from the 1961 material, the sentences in figure 5 by NR (ASC), PH (CSC) and EF (CSC) from the 1982 material. The CSC speakers also often seem to have a somewhat steeper rise and fall than the ASC speakers, i.e. larger intervals; but there are, of course, individual differences in this respect (Thorsen also found a tendency towards a steeper fall for the CSC speaker JR (1980a)). Anyhow, the most conspicuous characteristics of the ASC pitch contour from a perceptual point of view, at least in the subtype used by speakers of the lower sociolect, is not the height of the unstressed syllables but the slow falling-rising movement of the stressed syllable. - In both types an

initial voiced consonant normally has falling Fo (i.e. it belongs to the end of the preceding wave, also when the stressed vowel is rising), except when the word is the first stressed word of the sentence, in which case the initial consonant may be rising, or falling, or even.

Type 4 in figure 3 is characterized by a generally rising initial sonorant consonant (it may be falling after a stressed syllable), a high rising vowel (long vowels may be rising-falling), and a falling or rising-falling following sonorant consonant, followed by lower unstressed syllables. This type is used by West Jutlandish and North Jutlandish dialect speakers, normally also in their Regional Standard Danish (RSD). This type was also, on the whole, used by the East Jutlandish speakers analysed in this investigation, with the exception of PM and LH. (The location of these dialects can be seen on the map in my report on vowel duration (Fischer-Jørgensen 1983b).)

Type 3 deserves a somewhat more detailed analysis. It is in between type 2 and type 4 and is characterized by a rising Fo in the stressed vowel, having a low start and a strong rise after a sonorant consonant (which is generally falling) and a somewhat higher start after voiceless consonants. A following sonorant consonant continues the rise, but from there the curve decreases rather abruptly. In figure 3 two subtypes (3A and 3B) are distinguished. The difference between them depends on the duration of the medial consonant(s). After a single consonant (except when it is particularly long, which may be the case for *s*) and after a few very short clusters (e.g. sometimes *mb*) the first posttonic vowel will be higher than the preceding stressed vowel, whereas after almost all clusters (and sometimes *s*) it will be lower than the preceding stressed vowel. The contour is in fact the same in the two cases, but it will have reached a lower level at the vowel start after a longer consonant (cluster). It looks as if the Fo contour is planned independently of the segmental structure (cf. also Rossi et al. (1981, p. 31-32) and Thorsen (1982b)). This independent planning is, of course, not restricted to type 3, but it is more conspicuous in type 3 because it has consequences for the relative height of stressed and unstressed syllables. In types 1 and 2, where the rise continues for a time after the stressed syllable, the posttonic will also be high after clusters (but as mentioned above the rise may continue in the second posttonic after a short intervening consonant in type 1). In type 4 the first posttonic will always be lower than the stressed vowel (except sometimes in an initial word with single short intervening consonant). Figure 6 shows some superposed Fo contours with the start of the stressed vowel as line-up point, demonstrating different locations of unstressed vowels on the same contour after single consonants and clusters.

Type 3 is used by the North-East Jutlandish dialect speaker PM (Molbæk Hansen, who first drew my attention to the two subtypes), by the speaker of the Århus Regional Standard, and by most North Zealandish dialect speakers. However, the latter

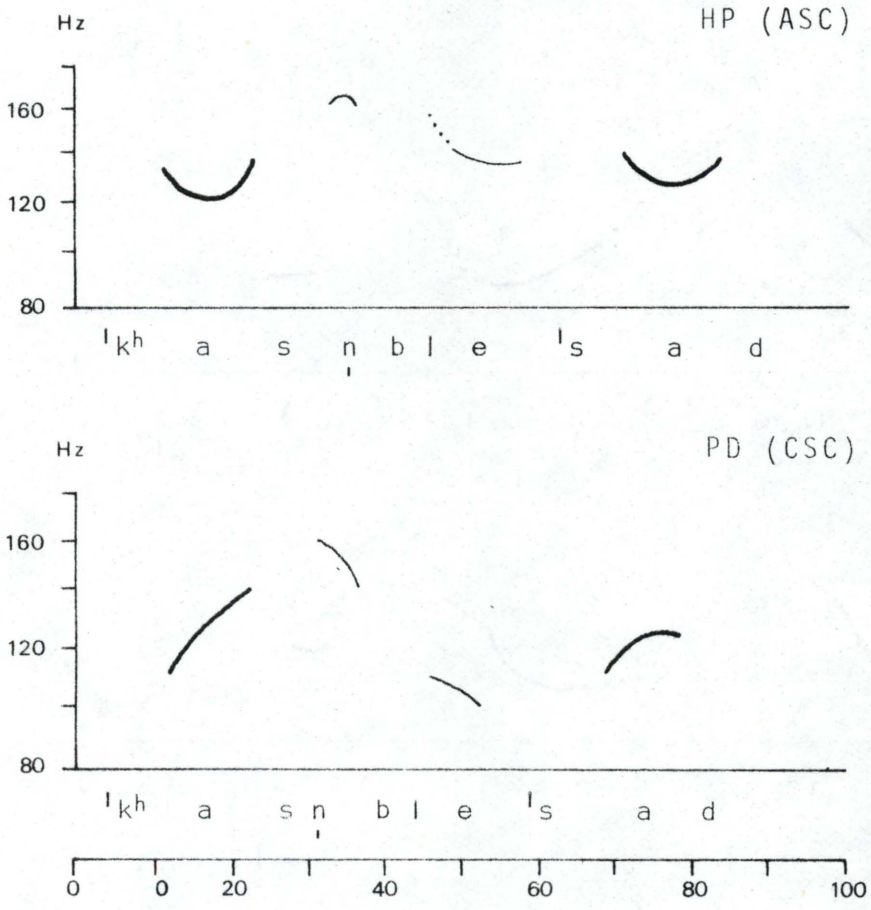


Figure 4

Curves showing the difference between ASC and CSC speakers in the direction of the Fo movement in stressed syllables (1961 material), ——— vowels and syllabic consonants, ... non-syllabic sonorant consonants.

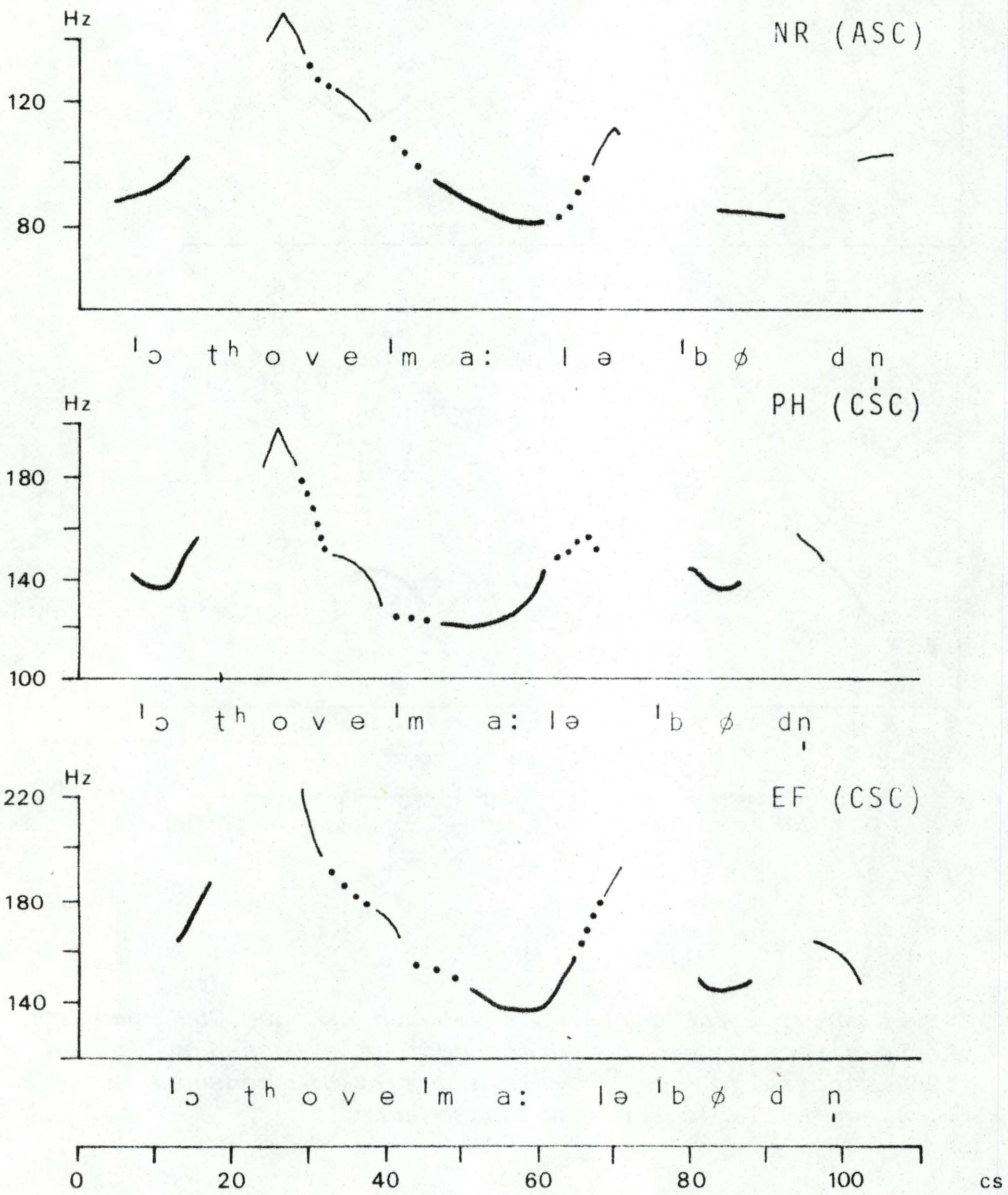


Figure 5

Curves showing the difference between ASC and CSC speakers in the direction of the Fo movement in the non-initial stressed syllable [ma:] *(1982 material),* —— *vowels,* .... *sonorant consonants.*

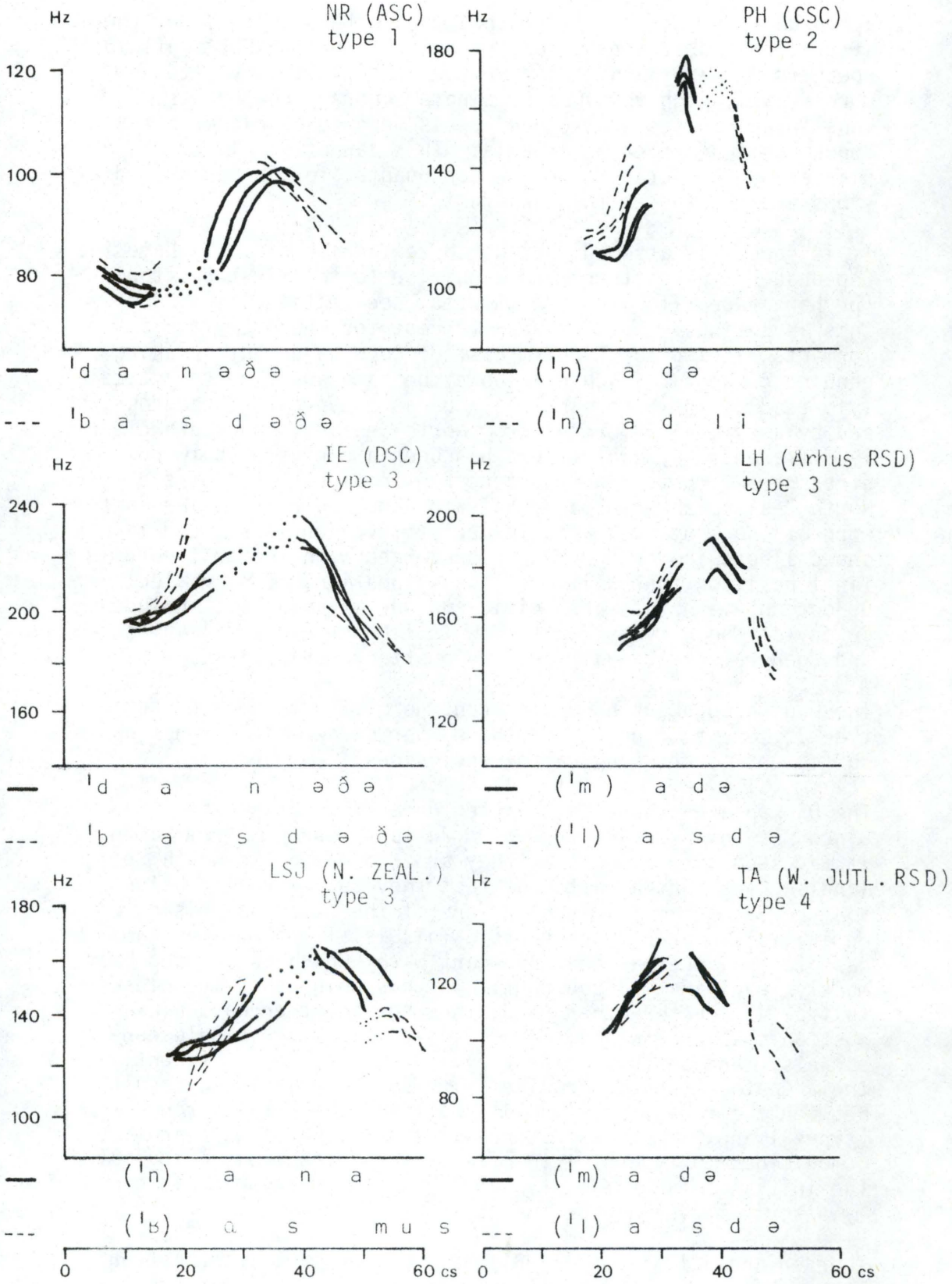


Figure 6

Curves from different types of speakers showing the location of unstressed syllables after single consonants and clusters on the same Fo wave, —— vowels in words with single medial consonant, ---- vowels in words with medial clusters, ... medial sonorant consonants.

approach type 4 in the sense that they often have a lower vowel after single obstruents. On the other hand, the East Jutlandish speakers may approach type 3 in the sense that they often have a relatively high vowel after single sonorant consonants. Thus in both cases unstressed vowels tend to be higher after sonorants than after obstruents. This cannot simply be explained by the duration of the consonants, for in Danish medial stops are not longer than sonorants.

It is generally assumed that North Zealandish dialects have the Copenhagen type pattern with high rise to the first posttonic (perhaps except for words with stød, see section III). But this is obviously not the case (except for one of my six informants, living about 20 km east of Copenhagen but speaking genuine dialect). Figure 7 shows the same sentence (practically), read by the informant ST, who has a typical ASC contour, and by a typical informant from North Zealand. They are clearly different. Probably there has been a confusion between stress group contour and sentence intonation. Most of the North Zealandish informants had a pronounced rise in the last word of the sentence, also in declarative sentences, and if one collects isolated words or sentences with the test word in final position, one will get this intonation. (The dialect in Dragør on Amager has a specific contour which is not included in any of the 4 types - with low falling stressed syllables and (generally) gradually rising posttonic syllables.)

The Funish speakers have different norms. Some come close to type 3, some to type 4, but the number of examples is not quite sufficient to draw any valid conclusions.

The DSC speakers have been listed under type 3 in figure 3. Since most of them are from West Jutland, and one from Funen, it may seem surprising that they are closer to the North Zealandish type. As a matter of fact they do not really belong to type 3. In the first place one of them (IE) has rising Fo on initial sonorant consonants. This is also sometimes the case for BJ, whereas the informant with Funish background has mostly level initial sonorant. On this point they are closer to type 4. On the other hand, they now and then have a high posttonic also after consonant clusters as after single consonants. And on this point they approach type 2. Evidently their contour is a compromise (with somewhat variable results) between their dialect (type 4) and their present environment (types 1 and 2). It is not really type 3, but the compromise comes close to type 3. (In Fischer-Jørgensen 1983a I indicated the initial consonant of type 3 as rising, because I had based my description mainly on the DSC speakers.)

Thorsen and Jul Nielsen (1981) have placed the first posttonic lower than the stressed syllable in the schematic pattern for their Århus speakers (Figure 2); but in their tracings it may be slightly higher or at the same level for speaker BBA, so that there does not seem to be any real deviation from my description. As for their Thy speaker, they have placed the first posttonic quite low in the schematic pattern (Figure 2),

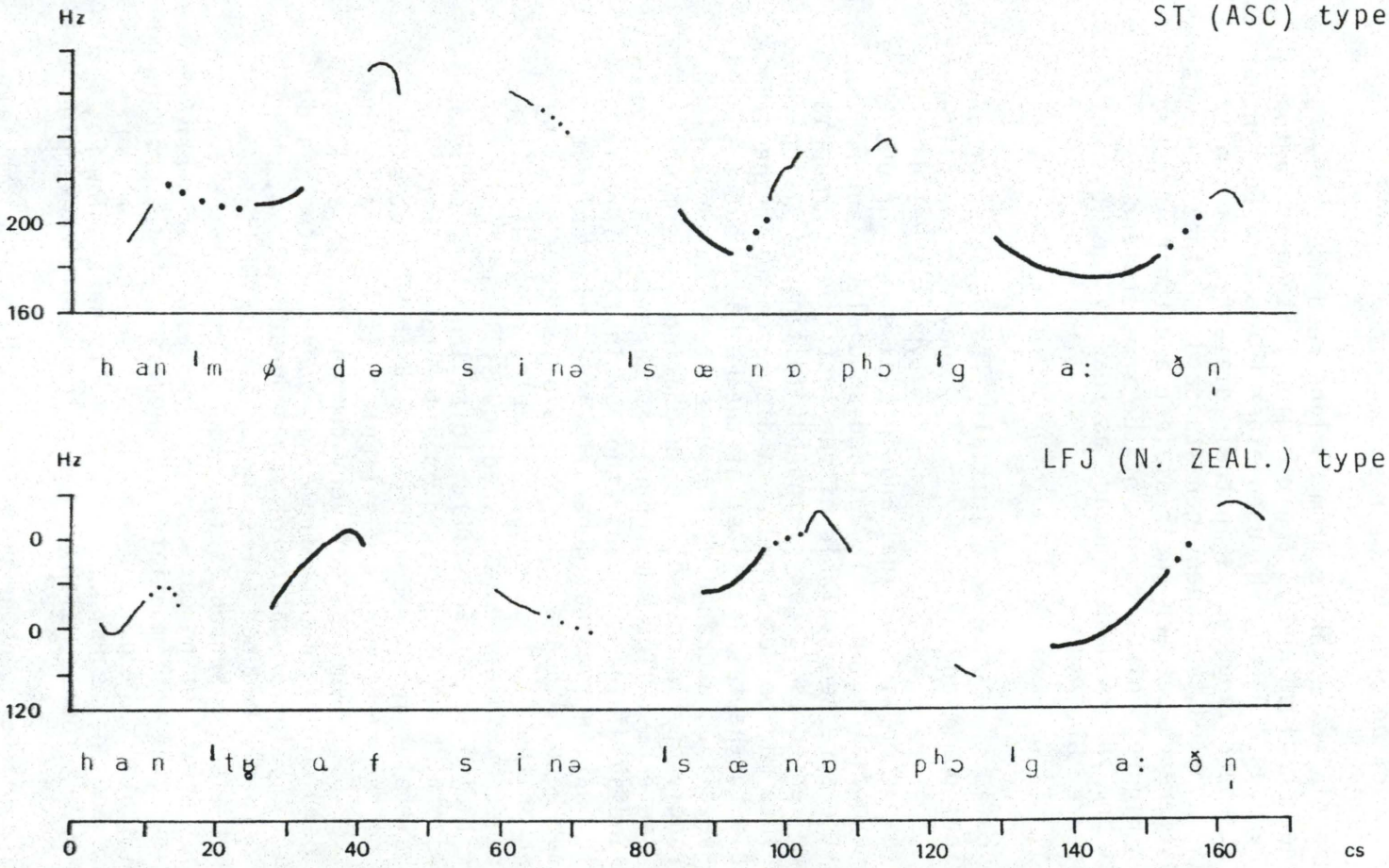


Figure 7

Comparison of Fo curves by ST, who uses the ASC Fo pattern, and LFJ from North Zealand (near Frederiksværk), who is typical for the North Zealandish pattern.

but according to their tracings of individual sentences he does not really go all the way down, so on this point there is no contradiction between our results either.

Some of the dialect speakers from Jutland read the sentences both in their dialect and in their Regional Standard language. Most of them (TA, EA, JD, PN) used exactly the same Fo contours in their standard language as in their dialect. But two (PM and BT) had adopted the Copenhagen standard, so that there is a clear difference between their dialect and their standard language. Figure 8 gives curves of the same sentence in dialect and in standard language as read by the West Jutlandish speaker BT.

b. Modifications in utterance initial and final position

The description given above of the different types is valid for medial stress groups. In final and initial stress groups there may be some modifications. In Jutlandish dialects and regional standards the stressed syllable of the last stress group in declarative sentences is not rising but falling or rising-falling. This may be described as a displacement of the stressed syllable to the falling slope of the Fo wave, but I should prefer to see it as a change of the wave itself under influence of the strongly declining sentence intonation in terminal declarative sentences. The same change is rare in Funish or North Zealandish dialects because the terminal declarative sentence intonation does not generally have this declination. On the other hand, ASC and CSC speakers often have a rising Fo movement on the first stressed syllable, particularly if it is preceded by initial unstressed low syllables (cf. Thorsen's description in section III). Initial unstressed syllables need not, however, be said on a low tone. There are very many examples in my recordings of a higher start with fall down to the first stressed syllable. It is rare in the speech of the DSC speakers, but very common for ASC and CSC speakers, particularly in the frame *de sagde* ... 'they said ...' (used in my investigation of vowel duration 1982, where *sagde* has strongly reduced stress which may count as weak stress), but also in sentences like *Han tog málebøtten frém* 'he took out the painting box', or *Han vil måle bøtten stråks* 'he will paint the box immediately', whereas a single pronoun starting a sentence, e.g. *Jeg købte* ... 'I bought ...', or *Hans sønner* ... 'his sons ...' will normally have a low Fo, and this may, of course, also be found in the other examples.

c. The relation between the prosodic structure and grammatical boundaries

Unstressed medial syllables always belong to the falling slope of the preceding stress group, also where they belong to the following stressed syllable syntactically, or even as part of the same word. This is understandable in types 1 and 2, where the stressed vowel starts at the bottom. But in types 3 and 4, where a stressed vowel after a voiceless con-

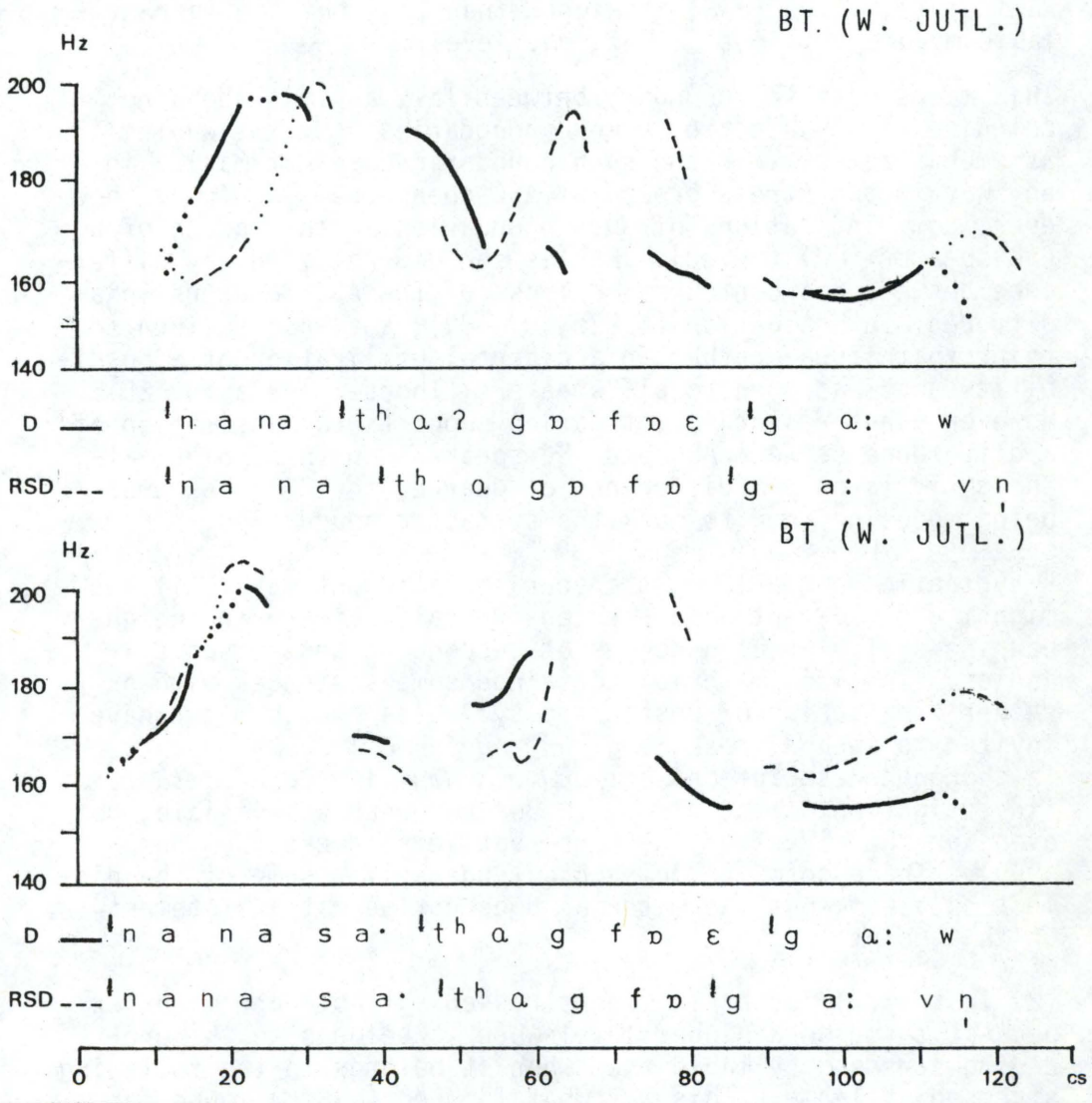


Figure 8

Sentences read by BT in her West Jutlandish dialect (——) and in her RSD (----) which, in contradistinction to the RSD of other West Jutlandish speakers, is very close to the Copenhagen CSC norm (.... = voiced consonants).

sonant may start at a rather high level, one might envisage the possibility of a weak prefix starting the rise. This is, however, not the case. The unstressed syllable is falling or low level. There is only one sentence where three of the West Jutlandish speakers pronounce the last unstressed syllable on a higher level than the preceding one, i.e. [han kʰa'se:ɒð ɛ san'da:ʔl], where [san] is higher than [ɛ], but the intrasyllabic movement is not rising, but level.

This means that the boundary between fall and rise does not coincide with syntactic or word boundaries. On the whole, as emphasized by Thorsen, such boundaries are not marked in any way in the stress groups of ASC speakers. She found, however, some indications of such boundaries in the speech of a CSC speaker (JR) (1980c). But as she did not find any difference between his contours and those of the ASC speakers in a different investigation (1982b, p. 189), she was inclined to think that it was rather an accidental utilization of a possibility that was open to all speakers (Thorsen 1980a, p. 25). However, the present investigation supports the assumption of a difference between ASC and CSC speakers on this point, although it is only a difference of degree, the CSC speakers being more inclined to mark the syntactic boundaries.

A syntactic boundary might appear in different ways: (1) One might, e.g., expect a more extensive fall between two neighbouring syllables of a series of descending unstressed syllables. The 1961 material contained some sentences with rather long series of unstressed syllables, which might have invited to such a break, e.g. *en índgående | beskrívelse* 'a thorough description' or: *úndersøgelse | af den gámle ..* 'investigation of the old ..'. But no break was visible, not even for the oldest of the conservative speakers (PD, born 1905). There seems to be such a tendency for some of the dialect speakers, but the material does not permit a statement on this point.

(2) In the case of a single unstressed syllable one might expect it to be on a higher level when it belongs to the preceding stressed syllable than when it belongs to the following stressed syllable. This was what Thorsen (1980c) found for JR, and a similar tendency can be found in the present data. The CSC speakers of the 1961 material (particularly PD and JR) often pronounce the proclitic indefinite article *en* on a lower Fo than the preceding stressed syllable, e.g. *fándt en kánde* 'found a jug', whereas the enclitic definite article is always said on a higher Fo, e.g. *etnográfen* 'the ethnographer'. The lower Fo of the proclitic article is also used by the DSC speakers but hardly ever by the ASC speaker HP. In the sentence *Hánne mísforstår spøg* (of the 1982 material), in which there is no word boundary between *mís* and *for*, all speakers (except SR) have a higher Fo on *for* than on both *mís* and *står*. In *Hánnes mís forstår spøg* 'Hanne's pussy understands jokes' only the ASC speaker NR and two of the CSC speakers (PH and OT) have a higher Fo on *for*, whereas the CSC speaker EF and the DSC speaker SR and (partly) IE have *for* slightly lower than *mís*. Finally, in *Hánnes mís forstår spøgen* 'Hanne's pussy

understands the joke', where *for* belongs to a following stressed syllable, only the ASC speaker NR has a consistently higher Fo on *for*, whereas EF, OT and PH have an Fo which may be sometimes higher, sometimes lower than *mís*, but at least not as high as in the first sentences, and SR and IE have *for* lower than both *mís* and *stár*. In the sentence *hans plán forkástes* 'his project is rejected' (in contradistinction to *hans pláner frémmes* 'his projects are furthered'), one of the three CSC speakers has a lower Fo on *for* than on *plán*. Finally, in the sentence ...*en mát lampét*... 'a dim bracket lamp' the ASC speaker NR always, and OT mostly, have a higher Fo on *lam* than on both *mát* and *pét*. EF has this pattern in a few cases, but otherwise she and PH have *lam* lower than *pét*, and SR and IE have *lam* lower than both *mát* and *pét*.

The 1961 material also contains the examples *en stór bedráger* 'a great swindler' and *en stórbedràger* 'a great (professional) swindler'. All three CSC speakers, and one DSC speaker had a difference, the syllable *be* being always higher than *stor* in the compound, but generally lower in the separate words (or sometimes slightly higher, but not as high as in the compound), whereas the ASC speaker HP did not have any difference, *be* being always higher than *stór*.

Nina Thorsen (oral communication) suggests that in the examples *mís forstár*, *mát lampét* and *stór bedráger* the relatively low unstressed syllable may be due to the fact that it is the only unstressed syllable of the contour (cp. Thorsen 1982b, p. 125), which is not the case in the compound, where the member with secondary stress behaves like an unstressed syllable as far as the Fo contour is concerned (see below, V.B.2). That may be of influence, but at any rate there is a difference between the ASC speakers, who have high Fo in all these examples and the others, who often have a lower Fo, and who have this lower Fo particularly when the syllable belongs syntactically to the following stressed word, cf. that the CSC speakers in the 1961 material always had a high Fo in the last unstressed syllable of, e.g., *etnográfen*, although this was also the only unstressed syllable.

(3) Closely related to the behaviour of a single unstressed syllable is the tendency for the last of a series of unstressed syllables to go down below the next stressed syllable, particularly in the cases where it is part of the next word or when it belongs to the following word as part of a unit accentuation group.

In the 1961 material the CSC speakers PD and EF have more examples of this kind than the ASC speaker HP, and it is not because they start the posttonic syllables at a lower level. In the 1982 material this phenomenon is hardly ever found in the recordings of the ASC speaker NR, but it is not infrequent for the other speakers. Three examples can be quoted where a syllable is part of the following word. In the sentence *Søren kóm med nogle barnágtige údtàlelser* 'Søren made some childish remarks' EF has a lower Fo on *barn* than on *ágt* in

six out of eight readings, whereas the other two informants who read this sentence (NR and PH) have a higher Fo on *barn*. In the sentence *han kassérede sandálen sídste sómmer* 'he discarded the sandal last summer' the CSC speaker PH and the DSC speakers IE and SR have a lower Fo on *san* than on *dálen*, whereas NR, EF, and in almost all cases OT have a higher Fo on *san*. (All the Jutlandish speakers (except for PM) have, in accordance with their whole pattern, low Fo on *san*.) In both the above quoted sentences the low Fo is favoured by the number of preceding unstressed syllables. In the sentence *han vil mále Casanóva* 'he will paint Casanova' (where *Casanóva* is the last word of the sentence) only the DSC speakers IE, BJ and NK have a lower Fo on *Casa* than on *nóva*.

In the examples with phrasal unit accentuation the distinction between the ASC speaker NR and all the others (the CSC speakers PD, EF and OT, and the DSC speakers IE and SR) is very clear. In the sentence *Ótto vil male bøtten grøn imórgen* 'Otto will paint the box green tomorrow' all except NR have a lower Fo on *male* than on *bøtten*, and in *Ótto vil mále bøtte fém imórgen* 'Otto will paint box number five tomorrow' all except NR have a lower Fo on *bøtte* than on *fém* (OT has one and EF two counterexamples out of eight). In the sentence *Óle spiser sødsùppe til míddag* (and the same with *sød súppe*) *spiser* is lower than *sød* for all informants except NR, who, however, in this case has two examples with lower Fo on *spiser* (for the CSC speakers *spiser* is also low in relation to the preceding stressed word *Óle*). The relation between the unstressed syllable and the following stressed syllable in these cases may be due to a particularly high Fo on the syllable carrying the unit accentuation more than to a particularly low Fo on the unstressed syllable (see section V.C.2). These differences between ASC and CSC speakers can only with difficulty be interpreted as a displacement of the syllable on the Fo wave. The problem of the influence of grammatical boundaries is complicated because various factors may interfere: the number of unstressed syllables, the steepness of the fall in various norms, the distinctness of the speech style, etc. However, as far as I can see on the basis of the present material, CSC speakers differ from ASC speakers particularly in cases where there are grammatical boundaries. These preliminary results should, of course, be controlled by further investigations with this specific purpose.

d. Syllables with stød.

Syllables with stød, which are not treated by Thorsen, raise some problems. They often have a decrease of Fo in the latter half of the syllable, and this favours a lower Fo on the following unstressed syllable. For the ASC speakers it is, however, very rare that it gets lower than that of the preceding syllable. There is not a single example for the speaker NR and only a few for HP and BS, although both texts contain quite a number of words with stød. But the CSC speakers often have a lower unstressed syllable after a stød. The degree of tongue height of the stressed vowel plays a role. High vowels, which

have a higher intrinsic Fo, are more often higher than a following unstressed vowel than low vowels. In the 1961 material the CSC speakers have, e.g., a higher Fo on the unstressed [ɲ] in ['ɛdno'gra:ʔfɲ] but often a lower Fo on the unstressed syllables in [be'sgri:ʔvɻsə] and [be'sgre:ʔvəð], and the CSC speakers of the 1982 material generally have a higher Fo on the final unstressed syllables of [ɑbɑ'ʁɑ:ʔdð̩], [san'da:ʔlɲ], ['blɔmʔsdɲ] and ['pla:ʔnɒ], but often a lower Fo on the unstressed vowel of ['sbi:ʔsɒ], sometimes [pʰe:ʔdɒ], and all (except the ASC speaker NR) have consistently lower Fo on the posttonic syllables in the words [kʰa'se:ʔɒðə] and [pʰa'se:ʔ-ɒðə]. These differences between words with higher and lower vowels cannot be explained by the number of following unstressed syllables. The DSC speakers have almost regularly lower Fo after stød; the same is true of PM's Himmerland dialect and mostly for the North Zealandish speakers. Kroman (1947) was thus right that there tends to be a difference between words with and without stød in the dialects of North Zealand, and Poul Andersen's criticism (1949) was not justified on this point. Figure 9 shows the sentence *han passérede Sánddàlen*, read by

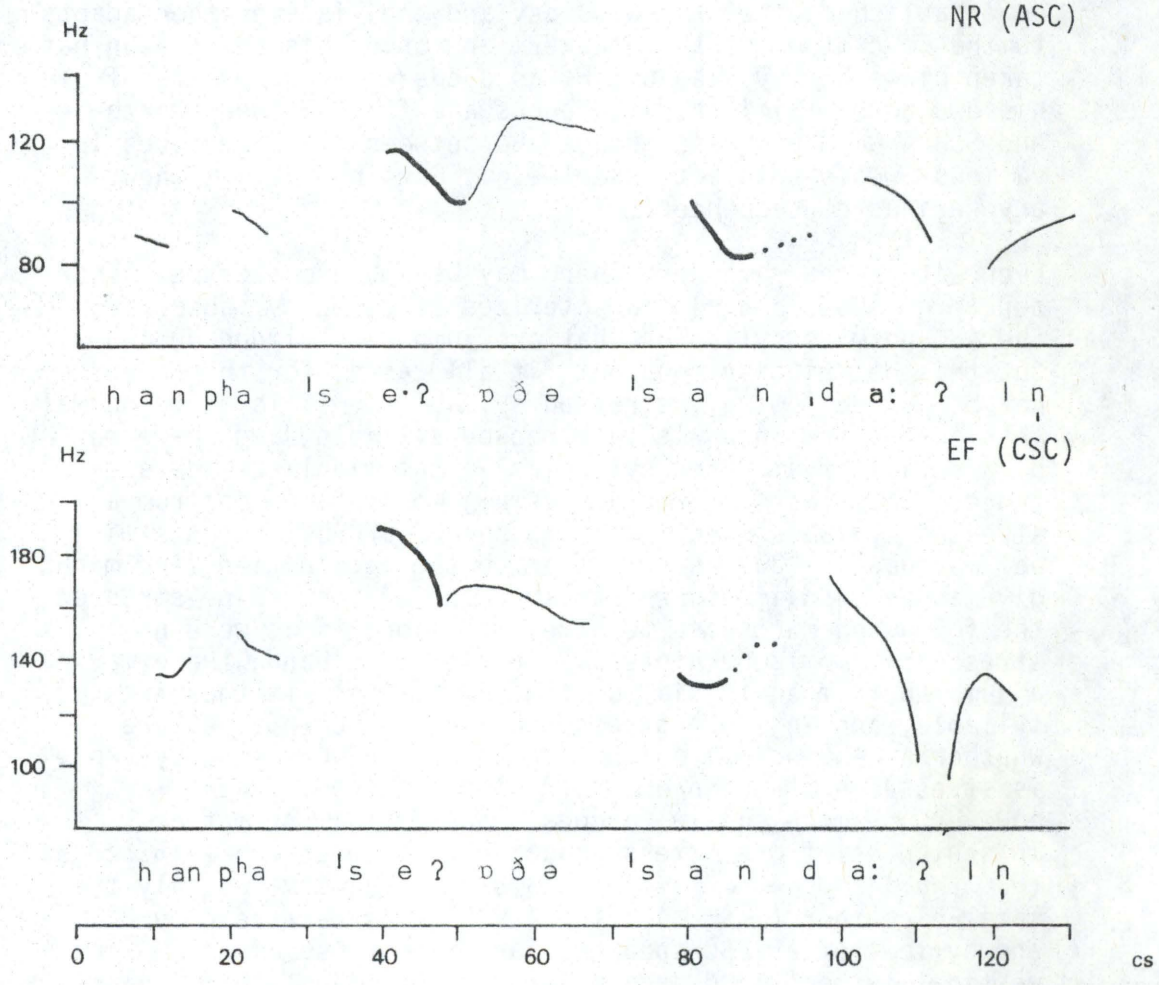


Figure 9

Fo tracings of two sentences with stød read by an ASC (NR) and a CSC speaker (EF).

the ASC speaker NR and the CSC speaker EF. It is difficult to fit the description of the latter curve into the pattern of an undulating wave with possible displacements of the syllables. There is simply no peak (nor any valley) between the two stressed syllables, although there are two unstressed syllables between them.

e. The perceptual relevance of the Fo contours.

It has appeared from the preceding pages that, within the same norm, the same dialect or sociolect, the Fo relations between stressed and unstressed syllables are very stable. It must thus be possible for the listener to use Fo as a perceptual cue for stress. But the differences between different norms are conspicuous. In type 1 and type 4 the relations are almost reversed. And types 3 and 4 are not only used in dialects but in the regional standards of dialect speakers. Almost all Danes are thus exposed to all these norms, e.g. in radio and television, and this does not seem to give any difficulties (with the exception of some compounds, see section V.B.2). Probably the other cues (vowel quality, stød, duration) will be established after a few words, and the listener then adapts to the Fo contour of the speaker, and once this adaptation has taken place he may also use Fo as a cue when it deviates from his own norm. Similar differences are found between North- and South German stress groups and between the word tones in various Swedish dialects (see, e.g., Gårding 1977) without any serious consequences.

Even within the same norm there may be ambiguous cases. Thorsen (e.g. 1980a and b) characterizes stressed syllables in the ASC norm as syllables that are jumped or glided up from. But this description does not fit all cases, for there need not be any following unstressed syllables. It is quite normal that a sentence ends with a stressed syllable, and there may be a series of stressed syllables in the middle of the sentence. Thus there is not always a jump or glide up from a stressed syllable, and a falling series of Fo movements will be ambiguous. (But the relation to the declination line might give a cue, and if other cues show one of them to be stressed, the following must also be stressed, since if it were unstressed it would be higher.) On the other hand, there may be a jump up from an initial unstressed syllable to the stressed syllable, and in a one-stress sentence one cannot be sure whether a sequence of a low level plus a high rising syllable is stressed-unstressed or unstressed-stressed. There may, however, be more subtle Fo cues which decide the matter. In sentences of one stress group the sequences *det er billigst* (*de siger*) and *det er billist* (*de siger*) may have roughly the same Fo contour (- - ´). These sentences were read by NR and by four other ASC speakers for the purpose of an EMG-investigation, and by NR again for the purpose of this investigation. There were clear differences of duration, but the Fo contours were similar (see Appendix III,4). However, NR had a much higher rise in the unstressed than in the stressed [lisd] (but this difference was not found for all the other

speakers). Some had a slight decrease at the end of the unstressed [lisd], but the most obvious and really consistent difference was in the timing of the rise: in *billigst* the rise starts in the *l* or sometimes already in the first *i*, whereas in *bilíst* it never starts until the boundary between *l* and the following *i*. Thorsen (1982a and 1983b) has shown that the Fo level of the *l* alone may be a sufficient perceptual cue for the stress difference, in synthetic speech with no other cues present. The same timing of the rise as in *bilíst* is found in some other words of the same type, e.g. *kalás*, *muláttər*. The frame *det er ... de siger* will give further Fo cues. *det er* [de:] will be slightly lower before a stressed syllable and (more important) the following *de* will be higher after a stressed syllable than after an unstressed syllable.

4. INTENSITY

The intensity of weak syllables has not been measured, but visual inspection of the curves showed that in by far the most cases the unstressed vowels [ə] and [ɒ] of the Copenhagen standard have lower intensity than the preceding stressed vowel, although they generally have a higher Fo. But the speaker of the Copenhagen lower sociolect ST, who has a high rise in Fo, generally also has higher intensity. The West Jutlandish dialect speakers have both lower Fo and lower intensity in unstressed syllables. The same is true in most cases for East Jutlandish and North Zealandish speakers. In the cases where they have a higher Fo on the unstressed syllable, i.e. often after sonorant consonants, they also mostly have lower intensity, but some of the East Jutlandish speakers and one of the North Zealandish speakers often have higher intensity in this case. The North Zealandish speaker who follows the Copenhagen norm and PM, who also in his Standard Danish follows the Copenhagen norm, have higher Fo but lower intensity on unstressed syllables, like the Standard Copenhagen speakers. The intensity contour of the individual speaker is, however, more variable than duration and Fo.

In some cases it is possible to compare syllables with the same full vowel quality. The examples with *Sánna* as first word of the sentence all have a rising Fo contour in the standard language, but they have lower intensity in the second unstressed syllable. In *Nánna*, however, NR has generally a rising intensity contour, and the same is true of the North Zealandish dialect speaker who follows type 2, as well as of PM and often of some of the other East Jutlandish and North Zealandish speakers.

Of the examples which can be compared pairwise (see Appendix III), the pair *Sánddàlen / sandálen*, which has been read by 14 informants, partly with different Fo contours, is the most informative. All the Jutlandish dialect speakers (Appendix III,6-7) have lower intensity on the weak syllable *san-* than on the following stressed syllable, and almost all have higher

intensity on the stressed *Sán-* than on the following syllable with secondary stress (JD has *Sán-* slightly lower, but the difference is much smaller than in *sandálen*). This is in accordance with their Fo contour (except for PM's RSD). The same is true of the standard speakers with Jutlandish background. The only Jutlandish speaker who has a more rising intensity contour (in her RSD) in *Sánddàlen* than in *sandálen* is the one (BT) who has taken over the Copenhagen Fo contour in her RSD. Thus the difference between the intensity contours of *Sánddàlen* in her dialect (¯-) and in her RSD (-¯) follows the difference in her Fo contours. Of the non-Jutlandish standard speakers, who all have a high Fo fall in the second syllable of *Sánddàlen*, two have approximately the same intensity contour in the two words and one has intensity contours that follow his Fo contour very clearly, so that both Fo and intensity are reversed compared to the stress relation, i.e. lower Fo and lower intensity on the stressed syllable. One, however, (NR) has an Fo-independent intensity contour which parallels the stress relation. He has the same Fo-independent intensity contour in the pair *billigst / bilíst*. For the other three pairs NR and the other standard speakers (except those with a Jutlandish background) have an intensity contour which either follows the Fo contour against the stress relation, or which is the same for the two words.

Thus, if a weak syllable has low Fo, it also has low intensity. If it has high Fo there are individual differences, and the same individual may behave differently from one word to the next (NR).

The reason why I have only looked at relations and not made any measurements (except in mm for the word pairs) is that measurements would not tell us much more; and as long as we do not know more about the perceptual value of intensity and its relation to Fo, quantitative evaluations do not seem worth while, except for a limited and specifically constructed material. Rossi et al. (1981) consider Fo and intensity to be mutually independent, because Fo is produced by the laryngeal muscles and intensity by the subglottal pressure. But that is not certain. He gives an example of an isolated [a:] with rising Fo but falling intensity in the latter half. It is possible that this decrease of intensity in the one-sound-sentence is due to the decrease of the subglottal pressure at the end of a sentence (cp. Lieberman 1967), but in the interior of a sentence the subglottal pressure seems to be rather constant, and variations in connection with stress, except for emphatic stress, have been found to be rather small (see e.g. Katwijk 1974). Moreover, a tension of the vocal folds may very well result in a longer closure phase and a larger amplitude and thus contribute to both Fo and intensity rise (see Hirano and Ohala 1969, and Ohala 1973 and 1977). Further, acoustically, higher frequency of vibration will automatically give higher intensity. This may be counteracted by compensatory adjustments in the production, but it is at least dubious whether such compensations take place for full vowels within a limited range of the chest voice. One may therefore expect Fo and intensity to follow the same trend, and the contours are in fact often similar. The traditional view

is that the relation of cause and effect is the opposite, i.e. that higher subglottal pressure in stressed syllables causes Fo to increase. But in the first place the subglottal pressure differences observed in speech can only cause a relatively small rise in Fo (see, e.g., Ohala and Ladefoged 1970 and Ohala 1970 and 1973), and in the second place nobody will probably assume that a high Fo in unstressed syllables, as it is the norm in the Copenhagen standard, is caused by a heightened subglottal pressure. But the similar trend of Fo and intensity is not a necessity; there may be compensations and individual production habits.

The problem is now how perception works.

Stress may be signalled conventionally by either high or low Fo, but low intensity in stressed syllables can hardly ever be a signal for stress. If both intensity and Fo decrease in unstressed syllables - as they do in West Jutlandish dialects - they may support each other perceptually, and if Fo increases and the intensity decreases in unstressed syllables, as was the case in several examples with [ə] and in some examples with full vowels in Copenhagen speech, intensity may perhaps act as a separate cue. But if both rise, as it is often the case, one may assume either that the intensity contour counteracts the perception of stress, or that it passes unnoticed because Fo and intensity are expected to move in the same direction. One cannot know for sure which expectation is the stronger: the expectation to have the same contour in intensity and Fo, or the expectation to have lower intensity in unstressed syllables, and this may determine possible perceptual compensations. More research is needed here.

In the case of weak [ə] and [ɒ] following a vowel of different quality the problem is complicated by the phenomenon of intrinsic intensity. The open back vowel [ɒ] is known to have high intrinsic intensity, so that when it is of lower intensity than the preceding vowel there seems to be an intended intensity reduction in the unstressed syllable. As for [ə], its intensity does not seem to have been investigated. It can only be compared with other vowels in unstressed syllables. One of the informants has read the words [mána] and [mánə] in the same environment. Here [ə] had a definitely higher intensity than unstressed [a] (Fo was about 1 semitone higher). This points to a rather high intrinsic intensity of [ə], but further investigations are required (and a complication is that the degree of openness of [ə] is rather variable). In perception the loudness relations may be changed relative to the intensity relations, because the listener compensates for the intrinsic intensity although we do not know whether this compensation is due to a comparison with his own effort (cp. Ladefoged 1967), or to social experience (which is probably the case with intrinsic pitch and duration), or dependent on the auditory mechanism because the vowels with low intrinsic intensity seem to be just the ones which have their formants distributed over more critical bands (see, e.g. Rossi et al. 1981).

5. CONCLUSION

The difference between stressed and unstressed syllables seems to be manifested acoustically by a combination of duration, Fo, and intensity.

The differences in duration are clear and significant and common to all speakers and dialects. Moreover, since the duration of unstressed vowels is so short that the integration time of the ear plays an important role, duration also influences loudness.

There are also clear and stable differences in the Fo contour, but they are not the same for all speakers and dialects. The intensity differences are more problematic, but intensity seems to play an independent role at least for the weak vowels [ə] and [ɒ], and also in some other cases, though very irregularly. In Appendix III indications of duration, Fo, and intensity are combined for the word pairs with different stress.

B. MAIN STRESS AND SECONDARY STRESS

As mentioned in section II Danish compounds with main stress on the first member are considered to have secondary stress on the second member. The problem is now how this secondary stress differs from main stress, or more precisely: What is the difference between a combination of two main stresses and a combination of main stress and secondary stress?

Both the 1961 and the 1982 material contain a number of sentences with compounds that can be compared with a combination of separate words with the same or nearly the same segmental structure, e.g. *en mát lámpe* 'a dim lamp' and *en nátlàmpe* 'a bedside lamp'. In the 1961 material each sentence was spoken only once in four different sentence positions, which makes it impossible to distinguish accidental and systematic differences for the individual speakers. Therefore this material will only be treated briefly. A more detailed analysis will be given of the 1982 material, where each sentence was read eight times. In the 1961 material one might expect to find some differences according to sentence position, for example more shortening of the second member medially in the sentence than finally. But no such tendency could be seen. The only recurrent feature is that the Fo differences are less clear at the end of a statement, and sometimes they could not be measured in this position.

As mentioned in section II syllables with secondary stress retain their vowel quality, their phonological vowel length, and their stød (they may even acquire one), so that fewer cues are at the disposal of the listener than for the distinction between main stress and weak stress. The differences must thus be looked for in duration, Fo, or intensity.

1. DURATION

Compounds might be expected to involve some shortenings, particularly of the second member, but perhaps also of the first member because of the increased number of syllables in the stress group, and perhaps a shortening of the distance between the stressed vowels giving an impression of a more intimate connection. The investigation has therefore been concentrated on these possible differences between compounds and separate words.

The main material contained 10 pairs for comparison, read by one to seven ASC, CSC and DSC informants. The pairs are listed (with translations) in section IV and the full sentences in Appendix II. The compound members of the pairs are also listed in Tables III-VII. The most important data for each pair are given in Appendices IV-V together with data for some comparable examples of unit accentuation, which will be treated in a later section. For each word the appendix gives the duration of the first and second member as well as the duration of their vowels if they could be delimited; moreover, the differences between the different types are given. The sequence with two main stresses is always listed as number one, the sequence with main stress plus secondary stress (compound) as number 2, and the sequence with unit accentuation and weak first member (within one word or a syntactic group) as number 3, e.g. (1) *mát lámpe*, (2) *nátlàmpe*, (3) *mát lampét*, or (1) *mále bøtten*, (2) *málebøtten*, (3) *mále bøtte fém*, and the line "1-2" gives the difference between the sequence of two words with main stress and the corresponding compound. For the present, only this difference will be treated.

a. Reduction of the first member of a compound

In recordings of the same informants made for a different purpose (see Fischer-Jørgensen 1982), there were sentences of the type *han sagde mále tó gánge*, *han sagde málekàsse tó gánge*. In these examples there was almost always a reduction in the duration of the first member of the compound compared to the simplex word in the same frame. In the examples *mále* - *málekàsse*, spoken by 4 informants, the shortening of the first member was 4.2 cs on the average, and the difference was significant for all four subjects (the vowel [a:] was shortened by 2.7 cs on the average). In the pair *dánne* - *Dánnevàng* (5 speakers), the shortening of the first member was 3.5 cs, in *nátte* - *náttefròst* (7 speakers), the shortening of the first member was 5.3 cs, in *mánde* - *mándefàld* (1 speaker) 4.6 cs, *músse* - *míssekàt* (1 speaker) 4.1 cs, *dan* - *Dánfòss*(6 speakers) 3.1 cs, *man* - *mándfòlk* (1 speaker) 2.5 cs, *mis* - *míslyd* (1 speaker) 4.0 cs, *nat* - *nátpòtte* (1 speaker) 3.3 cs, but *nat* - *nátlàmpe* (6 speakers) no shortening. The shortening was distributed over the segments, the vowel being responsible for between one third and half of the shortening. In *nátlàmpe* the *t* was lengthened, the vowel shortened by 0.9 cs. Thus, with the exception of the last pair, there was always an appreciable shortening of the first member, and it was significant for all speakers.

In the present material the frame was different. The words were spoken in normal sentences and there was always the same number of syllables in the sentences, thus: *Óle spiser sød súppe til míddag* vs. *Óle spiser sødsùppe til míddag*, or *Ótto vil mále bøtten stráks i mórgen* vs. *Otto tager* [tɑ:] *málebȍtten frém i mórgen*. This gave a different result. Table III gives a survey of the data. More detailed measurements are given in Appendix IV, 1-10.

I tried to find examples of different types: mono- or disyllabic first and second members, long and short vowels, second member with prefix, etc., but the choice was restricted by the possibilities of finding sequences of the same (or almost the same) segmental structure that could be used in natural sentences.

Table III

*Reduction in duration (in cs) of the first member of compounds compared to the first word of a sequence of two separate words with main stress and (practically) the same segmental structure. * indicates significance at the 1% level. There were 8 readings of each word. The numbers 1-10 correspond to the numbers in Appendix IV.*

		NR	PH	EF	OT	SR	IE	BJ	NK	average
1	ˈsøðʔˌsobə	3.0*	0.5	-0.2	1.6*	3.6*	4.9*			1.2
2	ˈnadˌlɑmbə[1]	0.4	0.3	-1.3	0.2	0.2	2.4*			0.4
3	ˈsbi:ʔs ˌʁɑisɳ			1.3						1.3
4	ˈvɛn ˌsgɑ:ʔbǒ			1.6*						1.6
5	ˈgasˌbʁænɒ[3]	1.2	1.5	0.3	0.5					0.9
6	ˈma:lə ˌkʰasɳ	1.4	-0.3	3.1*	1.0		-0.6	-0.3	4.3*	1.2
7	ˈma:ləˌbødɳ	3.0*	-0.4	2.6*	1.8	-0.5	-0.9			0.9
8	ˈlɑməˌlɒ:ʔ	-0.1	2.1	1.8	3.1*					1.7
9	ˈvɑ:məɑbɑ ˌʁɑ:ʔdǒ[2]	2.0*	1.5	-0.5	1.8	-0.9	1.5			0.9
10	ˈmisfɒ ˌsdɒ:ʔ	-0.5	0.2	1.4*	4.0*	2.3*	0			1.2
average		1.3	0.7	1.0	1.8	0.9	1.2	-0.3	4.3	1.4 \ 1.2

[1]) In *nátlàmpe* and *mísforstȁr* the initial consonant has not been included because it could not be delimited from the preceding word.

[2]) In *várme-* the [ə] has not been included because it could not be delimited from the following *a*.

[3]) EF had two readings of *gásbrènder/gás brænder*. They have been combined in this table and in tables IV-V and VIII-X but not in the counting of individual averages in the text.

(All individual averages have been included in the grand mean in tables III-X. Except for tables IV and X it does not make much difference whether the two speakers BJ and NK, who read only one pair, or the two pairs read by only one informant are included or not.)

Table III shows that there is a tendency to shortening of the first member (73% of the averages have a shortening), but it is slight (1.4 cs averaged over speakers) and inconsistent. It does not exceed 2.5 cs (which may be considered the lowest limit for an audible difference) except in 9 cases out of 48, and it is only significant in 15 (31%) of the averages.

There does not seem to be any clear difference due to the structure of the sequences (i.e. mono- or disyllabic first member, long or short vowel, etc.). The only consistent difference is that all speakers have more shortening in *sǿdsùppe* than in *nátlàmpe*. This is hardly accidental, for the dialect speakers make the same difference, and moreover, *sǿdsùppe* was the word with most shortening in the 1961 material, and *nátlàmpe* the word with least shortening in the sentences used for the duration analysis mentioned above. One reason may be that *sǿd súppe* (in two words) is not very commonly used, whereas *sǿdsùppe* is a current word, and some speakers seemed to take pains to pronounce *sǿd súppe* in two words. This was also the only pair which, in the given context, would be ambiguous disregarding the stress difference. There is, as might be expected, a certain similarity between the treatment of *málekàsse* and *málebǿtte*. The same three speakers have shortening in both words, and the same two have lengthening. But on the whole, many differences seem to be accidental. Nor is there any clear grouping of speakers. The differences between BJ and NK may be accidental, since they have only spoken one pair each, and in the examples in frame sentences NK does not have more shortening than the other speakers.

A shortening that is so inconsistent and of such small extent can hardly have any function for the perceptual distinction of compounds and sequences of two words.

It is a problem why the shortening in the normal sentences is much less pronounced than in the frame *de sagde ... to gange* mentioned above. It cannot be due to speakers or words. The speakers were the same, and the words were of the same type. One pair was identical: *måle* vs. *målekàsse*. For the four speakers who read this pair both in the frame and in the normal sentences, the shortening of the word was 4.2 cs in the frame and 1.3 cs in the sentence, and the shortening of the [a:] 2.7 and 0.5 cs, respectively. In both cases the relevant stress group was lengthened by two syllables. One might think of an influence from the total sentence length.(Lehiste (1980) has shown that the length of the frame may be of influence). The sentence with *målekàsse* in a frame was two syllables longer but had the same number of stress groups as the sentence with *måle*. The normal sentence with *målekàsse* had the same number of syllables but one stress group less (3 versus 4) than the

sentence with *mále*. One might then conclude either that an increase of the total number of syllables causes shortening, or that reduction of the number of stress groups prevents the shortening. I am inclined, however, to think that the explanation should be sought in the sentence type. There is a stronger boundary between a test word and its frame than between verb and object in a normal sentence, and this boundary may cause lengthenings (cf. Fischer-Jørgensen 1982, p. 179). The word *mále* is longer in the frame than in the normal sentence for the four speakers common to both recordings (5.8 cs), and the same is true of the vowel [a:] (3.7 cs). In the normal sentence the [a:] of *mále* is only 14.9 cs, i.e. it is rather close to a short [a], and therefore it does not tolerate as much shortening as the longer [a:] of the frame sentence.

b. Reduction of the second member of a compound

Table IV gives a survey of the reduction of the second member (with secondary stress) (for more details, see Appendix IV). The shortening of the second member is much more consistent than the shortening of the first member. It is found in 43 of the 48 averages (90%); it is significant in 31 cases (66%), and the grand mean is 2.9 cs averaged over speakers, 3.2 cs averaged over words. But it is not quite consistent. There is a conspicuous lack of shortening in *gásbrènder*. It may be partly due to the fact that in this case the words were not completely identical. There is stød in the *n* in *gás bréndər*, but not in *gásbrènder*, and a stød may have a shortening effect. On the other hand, the conspicuous shortening in *vénskàbet* may be due to the fact that *-skàbet* has stød in contradistinction to the separate word *skáber*. Moreover, in this example the full word *skáber* was the last stressed word of the sentence (in the other cases a stressed syllable followed), so that it may be somewhat lengthened. But there is more to it, for also the initial consonant (*b*) is shortened less in *gásbrènder* (0.3 cs) than other initial consonants (*-sùppe* 2.0 cs, *-làmpe* 1.0 cs, *-kàsse* 1.4 cs, *-bǿtte* 1.7 cs, and *-lǎr* 1.5 cs). If the two words *vénskàbet* and *gásbrènder* are left out, the averages will be somewhat higher for NR, PH, OT and EF, and the grand mean will be 3.1 cs. Leaving out NK and BJ, who read only one pair, will make a difference here, increasing the average by 0.5 cs.

Again all informants have more shortening in *sǿdsùppe* than in *nátlàmpe*. But the second member of *sǿdsùppe* is not much more shortened than the second member of e.g. *málebǿtten* (for four of six speakers it is shortened less), so one should perhaps not only look for special reasons for the large difference between *sǿdsùppe* and *sǿd súppe*, but also for the small difference in *nátlàmpe* vs. *mát lámpe*.

One might think of an influence from syntactic differences, but this is not obvious. Both *sǿd súppe* and *mát lámpe* represent attributive adjective plus substantive, and these sequences as well as the corresponding compounds are all objects

Table IV

*Reduction in duration (in cs) of the second member of compounds compared to a sequence of separate words with main stress and (practically) the same segmental structure. The reduction of the vowel alone is given in parentheses. * indicates significance at the 1% level.*

	NR	PH	EF	OT	SR	IE	BJ	NK	av.
1 'søð?ˌsobə[1]	6.2* (2.5*)	4.3* (2.4*)	5.2* (2.3*)	5.3* (3.1*)	4.3* (2.2*)	6.8* (1.7)			5.4 (2.4)
2 'nadˌlɑmbə	2.0 (1.1)	1.7 (0.8)	4.9* (0)	0.4 (-0.4)	3.1* (-0.2)	5.2* (1.5)			2.9 (0.5)
3 'sbi:?sˌʁɑisn̩			1.0 (0.8)						1.0
4 'vɛnˌsga:?bð̩			5.4* (3.1*)						5.4 (3.1)
5 'gasˌbʁænɒ[2]	-0.3	1.4*	0.3	0					0.5
6 'ma:ləˌkʰasn̩	4.4* (2.0)	3.4* (0.8)	3.6* (0.5)	4.6* (1.4)		2.2 (1.2)	1.3 (-0.6)	2.0 (0.4)	3.1 (0.8)
7 'ma:ləˌbødn̩	5.0* (1.8*)	4.6* (1.1)	5.8* (2.1*)	5.7* (2.6*)	0.7 (-1.0)	7.6* (3.0*)			4.9 (1.6)
8 'lɑməˌlɒ:?	4.5* (2.8*)	3.4* (2.3*)	4.3* (1.7*)	3.9* (1.8*)					4.0 (2.2)
9 'vɑ:məɑbɑˌʁɑ:?dð̩	3.1* (2.5*)	0.9 (0.4)	1.7 (1.5)	3.9* (5.0*)	0.8 (0.2)	7.6* (4.6*)			3.0 (2.4)
10 'misfɒˌsdɒ:?	2.9* (2.2*)	4.8* (2.7*)	3.3* (2.6*)	4.2* (4.0*)	-1.7 (-1.7*)	-0.3 (-0.4)			2.2 (1.6)
average	3.5 (2.1)	3.1 (1.5)	3.6 (1.6)	3.5 (2.5)	1.4 (-0.1)	4.9 (1.9)	1.3 (0.6)	2.0 (0.4)	2.9 3.2[3] 1.3 1.7

[1]) The final unstressed vowel or syllabic consonant was not included except for *súppe* because of difficulties of delimitation and because it was not always the same in the two members of the pair.

[2]) The second vowel in *gásbrènder* could not be delimited from the [ʁ]

[3]) The difference between the average over speakers and over words is diminished if the speakers who read only one pair (and the pairs read by only one speaker) are left out: 3.4/3.3.

in the sentences: *Óle spiser sødsùppe (sød súppe) til míddag* 'Ole has (a special type of) fruit soup (sweet soup) for dinner' and *jeg købte en mát lámpe (en nátlàmpe) til sóvekàmmeret* 'I bought a dim lamp (a bedside lamp) for the bedroom'. And in the example *gás brænder / gásbrènder* the words *gás brænder* are subject plus predicate, and here one should expect a less intimate connection so that one might expect a clearer shortening in the compound *gásbrènder*. It might also be possible that the consonant cluster [sbʁ] and in *nátlàmpe* the cluster [dl] invite to a slower pronunciation.

It is not possible either to find any reason for the differences between speakers. SR has the slightest shortening, IE the strongest, and they are both from Jutland. SR has the lowest average of all subjects, but he has a very unequal distribution: some words are shortened considerably, some not at all.

For the longer words (*vármeapparàtet* and *mísforstår*), there is great divergency among the speakers, but not in the same way. IE has more shortening than everybody else in *vármeapparàtet*, less than everybody else (except SR) in *mísforstår*. No speaker is quite consistent. This seems to show that shortening of the second member of the compound is very common, but that it is not a necessary feature of compounding. In this connection it should be mentioned that OT's pronunciation of *sødsùppe* generally sounds to me as two words, although he has just as much difference in duration as the other speakers. And SR's *sødsùppe* and *nátlàmpe* sound a few times like two words, but they are just as much shortened as in other cases (see below, section V.B.2).

c. Reduction of the distance between the sounds of the first and second member in compounds

It appears from table IV that the vowel is responsible for only about half of the shortening of the second member. Particularly the preceding consonant seems to be involved. That may have something to do with a disjuncture in separate words compared to compounds.

Table V gives the reduction in the distance between the vowels in the first and second member of a compound (the vowels with main and the vowels with secondary stress) compared to the distance between the two vowels with main stress in a corresponding sequence of two words. The intermediate consonant cluster has also been measured, but vowel distance (i.e. inclusion of the first vowel) gives larger differences and may be more interesting because it seems to be a plausible hypothesis that one perceives the syllable as starting approximately with the vowel. Experiments have shown that it is not that simple (see, e.g., Marcus (1979) with references). The whole structure of the syllable and particularly the type and number of initial consonants play a role. But since the pairs compared have the same structure, this is not very important, and vowel start seems to be a reasonable choice.

Table V

*Reduction of the distance (in cs) between the start of the first vowel and the vowel with secondary stress in compounds compared to the distance between the vowels in a sequence of two words with main stress and (practically) the same segmental structure. * indicates significance at the 1% level.*

		NR	PH	EF	OT	SR	IE	BJ	NK	av.
1	'sø̲ð?ˌso̲bə	4.9*	1.3	3.2*	3.1*	4.4*	5.2*			3.7
2	'na̲dˌla̲mbə	0.7	1.1	0.2	0.8	1.5	3.6*			1.3
3	'sbi̲:?sˌʁa̲isɳ			1.7						1.7
4	'vɛ̲nˌsga̲:?bð̲			3.0*						3.0
5	'ga̲sˌbʁæ̲nɒ	0.9	0.6	0.9	1.3					0.9
6	'ma̲:ləˌkʰa̲sɳ	2.7*	1.7	5.2*	2.1*		-0.3	0.8	5.0*	2.5
7	'ma̲:ləˌbø̲dɳ	3.1*	3.2*	3.4*	3.8*	1.0	0.7			2.5
8	'la̲məˌlɒ̲:?	2.8*	2.3*	1.7*	2.8*					2.4
9	'va:məabaˌʁa̲:dð̲	5.5*	1.2	-2.5	4.3*	0.3	5.3*			2.4
10	'misfɒˌsdɒ̲:?	0.7	3.2*	3.0*	4.5*	4.3*	1.5			2.9
average		2.7	1.8	2.0	2.8	2.3	2.7	0.8	5.0	2.5\2.3

The distance is reduced in compounds in 47 of 48 averages (98% of the cases), but it is significant in only 25 averages (52%); this reduction has a somewhat more even distribution across words and speakers, but it is of slightly less magnitude than the reduction of the second member (2.5 cs as an average over speakers vs. 2.9 cs). Of course this measure is a combination of shortenings in the first and second member and not an independent feature, but it might have been expected to be the best measure. Probably the best measure is the combined shortening of first and second member. This is shown in table VI. The combination of shortenings in first and second member gives an average of 4.3 cs. This means that in many cases it will be audible but not always for all speakers (cf. the average 2.3 cs for SR).

It can be concluded that there is very often a reduction of duration in compounds, and in many cases it will be audible, but it cannot be a sufficient cue.

d. Supplementary evidence from the 1961 material and the recordings of dialect speakers

Four pairs from the 1961 material have been analyzed, i.e. *sórt méjse* vs. *sórtmèjse*, *sórt kjóle* vs. *sórtkjòle*, *sød súppe* vs. *sødsùppe*, and *hvás máske* vs. *gásmàske*. The sentences are listed in Appendix I.

The first member of the compound was shortened in 58% of the cases but not very much, the grand mean being 1.3 cs, and there were differences according to words and to speakers, but no clear distribution according to speaker

Table VI

Standard Danish

Reduction (in cs) of first and second member and total reduction of compounds compared to a sequence of two separate words with main stress and (practically) the same segmental structure. The difference in distance between first and second vowel is also given (10 word pairs).

	NR	PH	EF	OT	SR	IE	BJ	NK	av.
first member	1.3	0.7	1.0	1.8	0.9	1.2	-0.3	4.3	1.4
second member	3.5	3.1	3.6	3.5	1.4	4.9	1.3	2.0	2.9
total reduction	4.8	3.8	4.6	5.3	2.3	6.1	1.0	6.3	4.3
distance	2.7	1.8	2.0	2.8	2.3	2.7	0.8	5.0	2.5

categories. The speaker who had the most consistent shortening (14 out of 16 cases, or 88%) was the CSC speaker PD.

There was a much more pronounced tendency to shorten the second member of the compound. This shortening was found in 83% of the cases with relatively small differences between speakers and words. The grand mean was 3.4 cs.

There was also a tendency to shorten the distance from the start of the first vowel to the start of the second vowel, the grand mean being 2.6 cs and the shortening taking place in 73% of the cases. There is thus complete agreement with the results of the main material.

The pairs *sød súppe / sødsùppe*, *måle bøtten / målebøtten* and *mát lámpe / nátlàmpe* were also read by a number of dialect speakers, four West Jutlandish speakers (BT, TA, EA and JD), one from Vendsyssel (North Jutland) (PN), and one from North-East Jutland (PM) and moreover by a speaker of the Århus Regional Standard. The pair *måle bøtten / målebøtten* was also read by six Funish dialect speakers. The Jutlandish dialect speakers read the sentence both in their dialect and in their Regional Standard Danish (RSD). A detailed account of the measurements is given in Appendix V followed by surveys of the differences corresponding to tables III-VI for the Standard Danish speakers. A brief summary corresponding to table VI for the standard language is given in table VII. It appears from this table that the Jutlandish speakers have a slight shortening in the first member of a compound, the average being 1.1 cs. This shortening is found in 68% of the individual averages, but it is only significant in 26%.

Table VII

Reduction (in cs) of first and second member and total reduction of compounds compared to a sequence of two separate words with main stress and (practically) the same segmental structure. The difference in distance between first and second vowel is also given.

A. Jutlandish (3 word pairs)

	BT	TA	EA	JD	PN	PM	LH	av.
first member	1.7	0.9	0.8	3.2	0.8	0.3	-0.2	1.1
second member	2.4	2.0	1.6	3.5	3.1	2.3	2.1	2.4
total reduction	4.1	2.9	2.4	6.7	3.9	2.6	1.9	3.5
distance	2.0	2.0	1.8	4.3	3.5	1.4	1.6	2.4

B. Funish (1 word pair)

	MA	LA	HV	EK	EH	av.
first member	-1.1	0.5	2.7	3.8	1.7	1.5
second member	0.2	0.5	0.6	3.2	1.7	1.2
total reduction	-0.9	1.0	3.3	7.0	3.4	2.8
distance	-3.4	0.5	5.9	2.3	1.1	1.3

In the second member the shortening is more pronounced (2.4 cs); it is found in 82% of the averages and is significant in 50%. Finally, the shortening of the distance between the first and the second vowel is 2.4 cs. It is found in 88% of the averages and is significant in 56%. This corresponds to the shortenings in the standard language. The Funish speakers have a shortening of 1.5 cs in the first member of *málebøtten*, 1.2 cs in the second member, and 1.3 cs for the distance, that is less shortening than for the other speakers but with a good deal of variation. (IP, who had a large reduction in the second member, has been left out, because her pronunciation of this word was unnatural).

In some cases compounds read by the Jutlandish speakers sounded as two words, also when the durational relations (reductions) were as expected (see the next section).

2. Fo-CONTOUR

a. Different types of compounds and speakers

For the analysis of the Fo contour it is practical to distinguish between different types of compounds according to the number and placement of unstressed syllables. The examples used in the main material belong to five different types (˘ indicates unstressed syllables here, — main and secondary stress and | the grammatical boundary; only the compound is cited as example, and the number of speakers is indicated in parentheses):

(a)	— | — ˘	*sødsùppe* (6), *nátlàmpe* (6), *gásbrènder* (4), *Spíesrèjsen* (1), *vénskàbet* (1)
(b)	— ˘ | — ˘	*málekàssen* (7), *málebøtten* (6)
(c)	— ˘ | —	*lámmelår* (4)
(d)	— | ˘ —	*mísforstår*
(e)	— ˘ | ˘ ˘ — (˘)	*vármeapparàtet*

(*gásbrènder* was read in two different sessions by EF, so that there is a total of 5 averages of this word.)

The four pairs used from the 1961 material all belonged to type (a). Most of the examples thus belong to types (a) and (b). These types are used here as examples of very clear differences between the Fo contours of sequences of two words with main stress and compounds with main plus secondary stress, and as examples of differences between various categories of speakers. The three remaining types are used to show that the location of unstressed syllables and the presence of stød may sometimes blur the distinction, so that there may be cases where Fo is not a sufficient cue for compounds. Sketches of the different contours are given in figure 10. They are only meant as rough indications of the main trends. (There are, of course, individual differences.) (See Appendix IV.)

In a sequence of two main stresses the second will start a new stress pattern but, due to the general declination in declarative sentences, on a slightly lower level. As described in section V.A.2, the stressed syllable will be placed at the bottom of the Fo wave for the ASC and CSC speakers and often be falling-rising, with a tendency to be more falling for the ASC, and more rising for CSC speakers. For the DSC speakers (who, in practically all the examples are represented by IE and SR, who have a West- and South-West Jutlandish background), and for the West Jutlandish dialect speakers the start is higher and the movement is rising, particularly for the dialect speakers (in syllables with stød most speakers have falling Fo, see, e.g., the Fo contour of the syllables *sød* in Appendix IV.1).

In compounds, however, the syllable with secondary stress behaves like an unstressed syllable in the Fo pattern. This also means that there will be differences between speaker cate-

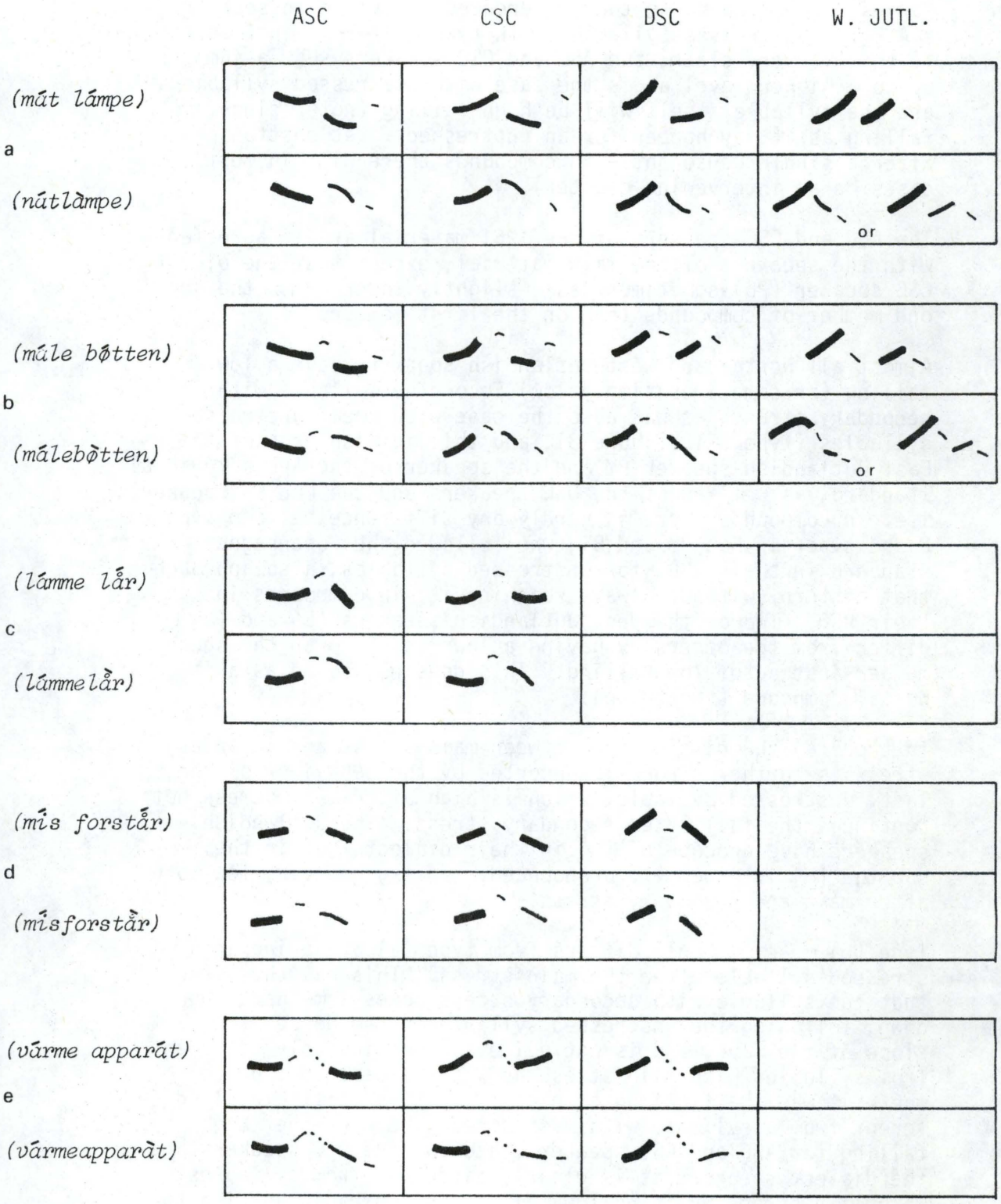


Figure 10

Rough sketches of Fo patterns for various types of compounds and various categories of speakers. Only the vowels are indicated. Thick lines = vowels with main stress; thinner lines = secondary stress, thin lines = weak stress.

gories corresponding to the differences described in section V.A.2 for unstressed syllables. In type (a) —ˈ— ˘ in figure 10 this is very clear; the ASC and CSC speakers show a jump up to a higher level as is the case with unstressed syllables, and the syllable itself will be high-falling (not rising-falling as it may happen for an unstressed first posttonic after a single consonant - in compounds there will in most cases be an intervening cluster).

The ASC and CSC speakers of the 1961 material are in agreement with the speakers of the main material, except that the oldest CSC speaker (PD) sometimes has a slightly lower Fo on the second member of compounds than on the first member.

Almost all North- and West-Jutlandish speakers have a low falling (or sometimes simply low) Fo on the syllable with secondary stress, as was also the case with their unstressed syllables (type 4 in figure 3), and this is also true of the East Jutlandish speaker PM and the speaker of the Århus Regional Standard, as well as of the DSC speakers and the Funish speakers; i.e. in compounds there is hardly any difference between type 3 and type 4. But PM and BT, who followed the Copenhagen standard in their RSD, for unstressed syllables, also approach that pattern (without always reaching it) in compounds in their RSD. Two of the West Jutlandish speakers (EA and TA) differ from the others by having a low rising Fo on the second member instead of low falling. This does not sound as a normal compound (see below).

In type (a) the distinction between main stress and secondary stress is further strongly supported by the behaviour of the last, unstressed syllable, which is high after main stress but continues the fall after secondary stress. The Jutlandish speakers have apocope of [ə] in their dialect, but in their RSD the [ə] is generally pronounced. It is, however, low both after main and secondary stress.

Type (b) — ˘ ˈ — ˘ only differs from type (a) by having an unstressed syllable after the main stress. This has the effect that the syllable with secondary stress comes somewhat lower down, following the unstressed syllable on the descending slope of the wave, and is not quite as conspicuously different from syllables with main stress as was the case in type (a) where it was the first posttonic. But it is still clearly different from a syllable with main stress, being rather high falling for ASC and CSC speakers, and for the DSC speakers and the dialect speakers it is clearly different from syllables with main stress by being low falling in contradistinction to the high rising syllable with main stress. Moreover, the last unstressed syllable which for the standard speakers is high after main stress and low after secondary stress makes the difference quite clear. The DSC speakers often have a higher Fo on the syllable *le* in *male* in the compound.

Types (c) — ∪ | — and (d) — | ∪ — differ from types (a) and (b) by not having a final weak syllable to indicate whether the preceding syllable has main stress or not. Moreover, in the examples chosen the syllables *lår* and *står* have stød and therefore a falling Fo movement. There will thus only be a difference in level between secondary and main stress, and this is not always quite clear. There are some cases of overlapping, where it is not possible to decide from the Fo contour alone whether the syllable has main or secondary stress. In (c) there is, however, a cue in the following word, which has reduced stress and is thus higher after the stressed *lår* than after the *lår* with secondary stress.

Whereas in (c) the weak syllable following the first stressed syllable belongs syntactically to this preceding syllable (*lámmelår*), it belongs to the following syllable in (d) (*mís-forstår*). In the sequence of separate words (*mís forstår*) there is, moreover, a word boundary between *mis* and *for* in contradistinction to what is the case in the compound. Now, as mentioned in section V.A.2, the CSC and DSC speakers have a tendency to place an unstressed syllable at a lower level when it belongs grammatically to a following stressed syllable, and thus they make a difference between *mís forstår* and *mís-forstår*, which is not quite consistent for the CSC speakers but which is consistent for the DSC speakers. Only the ASC speaker NR has complete overlapping (cf. the discussion in section V.A.3, p. 82).

In type (e) - ∪ ∪∪ — (∪) there is the further complication that there are three intermediate unstressed syllables, so that the last syllable comes rather low down. The CSC and DSC speakers have more often rising Fo on the syllable *rát* with main stress (although it has stød) than on the *ràt* with secondary stress, but all speakers have overlapping cases.

The texts contained a number of other compounds which did not have any counterparts consisting of separate words. These words were also read by the DSC speakers NK and BJ. They belonged to the types a, b and c, and their Fo contour was in complete accordance with the description given above of the compounds in the test words.

b. "Equal stress" in some Jutlandish compounds. As mentioned in section II, some compounds in Jutlandish dialects are described as having equal stress on both members. In the present material the words *nátlàmpe*, *sødsùppe* and *málebøtten*, read by the West Jutlandish speaker EA, sounded to me as two separate words, and in TA's pronunciation I could not decide whether they were two separate words or not, sometimes I tended to hear one word, sometimes two. The word *sødsùppe* has been left out in the measurements for these two speakers because there was nothing in the environment that could indicate what word was meant,

and so it could not be decided whether they had used the wrong word or had intended the right word; moreover, this word was not natural for them. There was no clear difference in these pairs in their RSD either; some other compounds in the texts which were not members of pairs (*Dánfòss*, *sóvekàmmeret*) sounded as normal compounds. In the pronunciation of the RSD speaker from Århus the words *sødsùppe* and *nátlàmpe*, but not *málebøtten*, also sounded as two words. Some of the compounds read by the DSC speakers made the same impression of being two separate words. In the 1961 material this was true of *sødsùppe* and *sórtkjòle* read by SR, and of these words plus *sórtmèjse* read by BF. A small informal listening test with two ASC listeners and one CSC listener showed much uncertainty in the identification of these words. In the 1982 material I heard two of SR's examples of *sødsùppe* and all OT's examples of the same word as two separate words, and I was in doubt about the stress in some examples of *nátlàmpe*.

Asked directly both OT and EA declared that there was a difference; they had meant them to be compounds. And of course these words are grammatically compounds and not simply separate words. In the case where they consist of attributive adjective + substantive (e.g. *sødsùppe*, *sórtkjòle*) they might be separate words as long as they have no definite article. But both OT and the DSC and Århus speakers who used these words would place the definite article at the end (*sødsùppen*, *sórtkjòlen*), whereby they are characterized as compounds. So did the West Jutlandish speakers in their RSD in the case of *málebøtten*. And words like *málebøtte* and *nátlàmpe* do not give any grammatical sense except as compounds. But they may be described as compounds with equal stress, i.e. they may constitute two stress groups.

The durational relations in the words heard as separate words do not generally differ from those found for other speakers whose compounds were heard as quite normal. OT even has a very clear durational difference between *sødsùppe* and *sød súppe* But in most cases there were deviations from normal compounds in the Fo contour (see Appendix IV and V).

Both EA and TA (particularly EA) had rising Fo in the syllable which should have secondary stress, i.e. in this respect it behaves as a syllable with main stress starting a new stress contour. For OT and SR the difference seems to lie in the last unstressed syllable of *sødsùppe* (and *nátlàmpe*). It was slightly higher than the preceding syllable which was rather level, and for SR this only happened in the two examples which I heard as having two main stresses. LH had rising Fo in the sonorant consonant following the vowel in *-làmpe* and no Fo difference between *sød súppe* and *sødsùppe*. This seems to indicate that the Fo movement is a more important cue than duration, at least for ASC and CSC listeners. In the 1961 material, however, it was difficult to find any explanation of the differences. It would be interesting to undertake some well planned listening tests both with Copenhagen and dialect listeners. They might

show more clearly what are the important perceptual cues. The problem is not quite simple. The Copenhagener does not simply use his own norm in listening. He is not confused by the fact that most dialect speakers have low falling Fo in unstressed syllables and in syllables with secondary stress, whereas he himself has high falling Fo. The listener probably quickly adapts to this difference. But when the dialect speaker has rising Fo he hears a stressed syllable because in the dialect this normally signals stress, and he expects syllables with secondary stress to be treated like unstressed syllables as far as the Fo contour is concerned.

3. INTENSITY

Peak intensity has been measured in mm, and the sequence of two main stresses was compared to the sequence of main stress plus secondary stress for each pair. If the compound has a more extensive decrease or a smaller increase in the second part than the sequence of two words (i.e. ˉ _ vs. ˉ – , or – – vs. _ ˉ), the contribution of the intensity to the reduction of the second part of the compound was considered positive (see Appendix IV). This was the case in 80% of the pairs for IE, SR and EF; and for IE 67% of the individual averages were significantly different, but for EF and SR only 9 and 17%, respectively. For PH there were 50% positive relations, for NR 25%, and for OT 13%. There are thus great individual differences, and the positive cases are very rarely significant except for IE. It is therefore not probable that these differences have any value as cues for stress differences. The consistent difference for IE can probably be explained by the fact that she has a very low Fo on syllables with secondary stress. SR also has a relatively low Fo combined with secondary stress, but EF has a high Fo.

In the 1961 material the intensity relations are also irregular, but here too, EF often combines high Fo and low intensity.

For the Funish and Jutlandish speakers the result is quite different (see Appendix V). There were 41 individual pairs, 76% of these were positive, and for 49% the difference was significant. Six of the ten exceptions were due to the recordings in standard language of BT and PM, and they do not really use a regional standard but the Copenhagen standard. If they are excluded, there will be 89% positive pairs, of which 57% have a significant difference.

This means that almost all the speakers who have low falling Fo on syllables with secondary stress have a relative decrease in intensity, whereas those who have a high falling Fo, do not have this decrease in intensity. Thus the intensity difference seems to depend on the Fo difference and may not be an independent cue. But EF has high Fo combined with low intensity. She thus seems to use intensity as a separate factor, but even in this case it is only a tendency, since only one of her nine positive averages shows a statistically significant

difference. This generally negative result was one of the reasons why I did not find it worth while for the time being to re-record the whole material in order to get dB-measures instead of mm.

The Jutlandish examples read by TA, EA, and LH, which sound to me as having two main stresses, have a clear intensity difference like the other compounds in Jutlandish. This does not seem to have influenced my identification. It is possible that the Jutlandish speakers themselves use different perceptual criteria. However, dialectologists with a Jutlandish background also describe a number of compounds as having two main stresses (e.g. Ejskjær 1954).

4. CONCLUSION

The analysis of compounds with secondary stress on the second member shows that they differ from a sequence of two words with main stress both in duration and in Fo. The shortening is small and inconsistent in the first member of the compound, whereas the shortening of the second member and of the distance between the start of the two vowels is more pronounced and more regular. The total shortening is quite considerable, 4.3 cs for the standard speakers and 4.2 cs for the Jutlandish speakers on the average.

The difference in Fo contour is in almost all cases quite clear, but it is not the same for different categories of speakers, i.e. for different dialects and sociolects. This does not seem to hamper mutual understanding except in the cases where some Jutlandish speakers use what listeners from other dialects hear as equal stress on the two members of the compound.

The differences in intensity seem to follow the differences in the Fo contour: Speakers who have high falling Fo on the second member of compounds tend to have higher intensity too, whereas those who have low falling Fo also have low intensity. There are, however, exceptions; higher Fo may be combined with lower intensity, particularly for informant EF, but it is only a tendency. There seems to be individual differences in the degree to which Fo and intensity vary in parallel.

Fo seems to be the decisive factor in the cases where CSC listeners hear two equal stresses in Jutlandish dialects or Standard Danish influenced by Jutlandish.

C. STRESS REDUCTION IN THE SECOND MEMBER OF COMPOUNDS COMPARED TO STRESS REDUCTION IN UNIT ACCENTUATION

As mentioned in section II there is a clear difference between the reduction of stress in second members of compounds and in first members of a unit accentuation group, since the reduced second members of compounds retain their stød and their phono-

logical vowel length difference, whereas syllables reduced by unitary stress loose their stød and - at least for monosyllabics - also the vowel length difference. But the question is whether there is any difference between the two types in the cases where there is short vowel and no stød, i.e. whether there are also differences in other features which might be relevant as cues for stress perception.

The main material contains a number of examples of stress reduction by unit accentuation which may be compared both to words with main stress and to words with reduced stress in second members of compounds. The pairs are listed in section IV with translation and the sentences are listed in Appendix II, but it may be practical to repeat the triplets here (with number of speakers in parentheses): *sød súppe / sødsùppe / sød suppedás* (6), *mát lámpe / nátlàmpe / mát lampét* (6), *mále kássen / málekàssen / mále Casanóva* (7), *mále bøtten / málebøtten / mále bøtte fém* (6), *gás bréndər / gásbrèndər / gás brænder néd* (4), *mís forstår spøgen / mísforstår spøg / mis forstår spøg* (6), *Spíes réjse / Spíesrèjsen / Spíes rejse méd* (1), *vén skáber sig / vénskàbet / vén skaber problémer* (1). (The sentence *Sánna mísforstår spøg* may also be pronounced with unit accentuation, but a complex compound need not be weakened as first member of a unit accentuation group.) There were further two examples where only main stress and unit accentuation reduction could be compared: *det várme apparát / varme apparátet óp* (6), and *mále bøtten / male bøtten grøn*. Three examples from the 1961 material were also analyzed: *vándkànde / kande vánd*, and (only for comparison with main stress): *sát óppe / sat óp* and *sátte kássen óppe / satte kássen óp*.

1. DURATION

As mentioned in section V.B.1, there is only a very weak tendency to shorten the first member of compounds (1.4 cs on the average). In unit accentuation there is no shortening at all of the preceding stressed syllable (0.2 cs on the average). This means that the stressed syllable of the preceding prosodic stress group which is expanded by the reduced word, e.g., *mále bøtte | fém* (vs. *mále | bøtten*) is not shortened.

The reduced syllable is, however, shortened, and more so than in the second member of compounds. Table VIII gives the differences in duration between syllables with reduced stress and syllables with main stress. The reduction in duration is considerable: 5.3 cs on the average. There is shortening in 46 (92%) of the cases. There is no clear grouping of speakers: the ASC speaker and the DSC speaker IE have the strongest reductions, and the DSC speaker SR the smallest reduction, but still 3.4 cs, which is probably audible. There are more conspicuous differences between words. The shortening is of considerably larger magnitude in intraword unit accentuation (7.4 cs) than in phrasal unit accentuation (4.0 cs), but the example *mále bøtte fém* comes very close to the reduction in intraword unit accentuation (6.8 cs). The pair with the least

Table VIII

Standard Danish

*Reduction in duration (in cs) of a word or syllable weakened by unit accentuation compared to a word or syllable with main stress and the same segmental structure, following the same stressed word (A), or preceding the same stressed word (B). * indicates significance at the 1% level.*

A.		NR	PH	EF	OT	SR	IE	BJ	NK	av.
1	'søð? sobə 'da:?s	6.2*	5.6*	5.3*	5.3*	8.5*	10.4*			6.9
2	'mad lam'pʰɛd	9.1*	6.0*	6.7*	7.8*	5.9*	9.0*			7.4
3	'sbi:?s ʁaisə 'mɛð			5.3*						5.3
4	'vɛn sga:bɒ pʁo'ble:?mɒ			4.0*						4.0
5	'gas bʁænɒ 'neð?	1.6	2.0*	0.7	-0.2					1.0
6	'ma:lə kʰasa 'no:va	12.7*	7.2*	5.3*	9.3*		10.9*	7.3*	3.5*	8.0
7c	'ma:lə bødə 'fɛm?	8.9*	2.8*	7.2*	6.2*	5.3*	10.6*			6.8
10	'mis fɒsdɒ 'sbɔi?	10.0*	4.2*	6.2*	6.5*	-0.9	4.9*			5.2
B.										
7d	ma:lə 'bødn̩ 'gʁœn?	6.1*	2.8*	3.9*	5.1*	2.4*	3.2*			3.9
9	va:mə aba 'ʁa:?dð̩ 'ɔb	1.4	2.2*	3.0	1.2	-0.8	2.9*			1.7
	average	7.0	4.1	4.8	5.2	3.4	7.4	7.3	3.5	5.3 5.1

shortening is *gás brænder néd* vs. *gás bránder* (1.0 cs). ['bʁænɒ] was also hardly shortened as second member of a compound (see above V.B.1, where it was supposed to have something to do with the loss of stød). In *varme apparátet óp* the relatively slight shortening is probably due to the long distance between the reduced syllable and the following main stress. The long vowel [a:] of *varme* [va:mə] is only reduced by 1.6 cs and certainly not reduced to be as short as a phonologically short vowel. The other example with a long vowel is *male* in *male bøtten grøn*. The vowel is here shortened by 2.5 cs.

The shortening in these words is somewhat more pronounced than the shortening in second members of compounds. In table IX the two types are compared. The syllable with unit accentuation has more reduction than the compound in 31 (79%) of the 39 comparable pairs, and the difference is significant in 24 pairs (62%). IE has

Table IX

*Reduction in duration (in cs) of a word or syllable weakened by unit accentuation compared to a sequence with secondary stress and the same segmental structure following the same stressed word (reduction in unit accentuation - reduction in the compound). * indicates significance at the 1% level.*

		NR	PH	EF	OT	SR	IE	BJ	NK		av.
1	'søð? sobə 'da:?s	0	1.3	0.1	0	4.2*	3.6*				1.5
2	'mad lam [1] 'pʰɛd	7.7*	4.8*	3.5*	6.3*	4.2*	6.5*				5.5
3	'sbi:?s ʁaisə 'mɛð			4.3*							4.3
4	'vɛn sga:bɒ pʁo'ble:?mɒ			-1.4							-1.4
5	'gas bʁænɒ 'neð?	1.9*	0.6	0.4	-0.2						0.7
6	'ma:lə kʰasa 'no:va	8.3*	3.8*	1.7*	4.7*		8.7*	6.0*	1.5		5.0
7	'ma:lə bødə 'fɛm?	3.9*	-1.8*	1.4	0.5	4.6*	3.0*				2.5
10	'mis fɒsdɒ [2] 'sbɔi?	7.1*	-0.6	2.6*	2.3*	0.8	5.2*				2.9
	average	4.8	2.0	1.6	2.3	3.5	5.4	6.0	1.5	3.4	2.6 [3]

[1]) In no. 2, *lam* in the compound is compared to *lam* in unit accentuation, whereas in table IV the [b] of *lampe* was included in the comparison with main stress.

[2]) EF had two readings of *mísforstår* vs. *mís forstår spøg*.

[3]) The difference between the average over speakers and over words is diminished if the two pairs read by only one speaker (and the two speakers reading only one pair) are left out: 3.2 3.0 cs.

a significant difference in all five pairs, and the smallest difference is 3.0 cs; thus she seems really to use this distinction consistently. For the other subjects there are one or two word pairs with very little difference. The types are thus not kept consistently apart by means of duration. *lampét* compared to *lámpe* is the only word pair with a consistent and clear difference for all speakers (average 5.5 cs, minimum 3.5 cs). In other words, some speakers have very small differences, but the tendency is very clear. The only clear counterexample is *vén skaber problémer* vs. *vénskàbet*. This may perhaps be due to the lack of stød in *skaber* vs. *skàbet* (cf. the example *gásbrænder* above).

One might also expect to find a difference in the distance from the preceding word (see table X), because the second member of the compound belongs to the preceding part (and there was a difference in distance between the vowels of the compound compared to two words with main stress), whereas the reduced syllable with unit stress belongs to the following word syntactically, and one should therefore not expect a shortening of the distance from the preceding word. However, the distance is in fact shortened in this case too. The first 8 cases in table X can be compared to the shortening in compounds. The shortening is here seen in relation to an identical preceding stressed word. The average shortening of the distance is here 1.1 cs. It is found in 71% of the averages but is only significant in 21% of the pairs. It is thus less than in compounds where the shortening was 2.3 cs with 98% positive and 54% significant cases. But that there is a certain shortening shows that the word boundary is not marked very clearly (the difference from compounds is only significant in 15% of the averages). The two last examples show the reduction in distance to a following stressed syllable to which the reduced word belongs syntactically. This shortening seems to be somewhat more pronounced in *male bøtten grøn*. The relatively small reduction in *varme* may be due to the many intervening weak syllables. The syntactic connection thus seems to play a certain role, but more examples would be needed to prove it.

In the 1961 examples the syllable *sat* in *sat óp* was shortened by 7.5 cs compared to *sát óppe*, and the vowel by 3.7 cs, and *sat* in *satte kássen óp* was shortened by 8.3 cs and the vowel by 3.5 cs compared to *sátte kássen óppe* (seven speakers). For all speakers the *a* in *kande* was shortened more in *kande vánd* than in *vándkànde*, and the same was true of the whole word, but the differences were small.

2. THE Fo CONTOUR

Like syllables with secondary stress syllables with reduced stress due to unit accentuation are also treated as unstressed syllables as far as the Fo contour is concerned, i.e. they form part of the descending slope of the preceding stress group, and if the preceding word is monosyllabic the reduced syllable will therefore be high-falling in the pronunciation of ASC and CSC speakers, whereas it will be somewhat lower if it is preceded by weak syllables. In all cases the difference from main stress is clear, except for *mís forstår* and *mís forstår spøg* because of the fall of stressed *står*, which has stød (as was the case for the difference between a compound and two words).

In principle the Fo contour will thus be the same in the second member of compounds and in syllables reduced by unit accentuation, e.g. *málebøtten* and *måle bøtte fém*. The individual speakers often do have small differences between the two types, and there is a consistent difference (i.e. the same in all eight tokens) in 15 out of 39 pairs (thus in 38%), but the difference,

Table X

Reduction of the distance (in cs) between the start of the vowel with reduction due to unit accentuation and the vowel of the preceding stressed syllable (1-10), or the start of the following stressed vowel (9 and 7d) compared to the start of vowels in two consecutive stressed syllables.

		NR	PH	EF	OT	SR	IE	BJ	NK	average	
1	'sø̲ð? so̲bə 'da̲:?s	2.3	1.4	2.5*	3.2*	4.0*	5.0*			3.1	
2	'mad la̲m 'pʰɛ̲d	1.0	1.1	1.5	1.2	2.2	3.9*			1.8	
3	'sbi̲:?s ʁa̲isə 'mɛð			0.1							
4	'vɛ̲n sga̲:bɒ pʁo̲'ble̲:?mɒ			1.0							
5	'ga̲s bʁæ̲nɒ 'ne̲ð?	-0.2	-0.4	-0.4	-0.1					-0.3	
6	'ma̲:lə kʰa̲sa 'no̲:va	1.8	1.1	3.3*	2.3		1.2	0.3	0.3	1.5	
7c	'ma̲:lə bø̲də 'fɛm?	2.4*	1.0	1.8	1.6	1.8*	-0.1			1.4	
10	'mis fɒsdɒ 'sbɔi?	-0.5	0.7	-0.1	-0.2	-1.1	-0.4			-0.3	
	average	1.1	0.8	1.1	1.3	2.1	1.7	1.5	0.3	1.1	1.0
7d	'ma̲:lə 'bø̲dn̩ 'ʏʁœn?	4.4	2.2	2.5*	5.0*	2.5*	3.8*			3.4	
9	vɑ̲:mə ɑba 'ʁɑ̲:?dð̩ 'ɔb	2.4	-0.5	-1.1	3.8*	-0.5	2.7*			1.1	
	average	3.4	0.8	0.7	4.4	1.0	3.3			2.3	2.3
total average		1.7	0.8	1.1	2.1	1.5	2.3	0.3	0.3	1.3	1.4

which mainly consists in a different height of the weakened syllable, is not the same for all speakers, nor is it the same for the different words read by the same speaker, and therefore these differences hardly have any value as perceptual cues.

There is only one more general tendency: The CSC and, still more, the DSC speakers, in contradistinction to the ASC speakers, show a certain tendency to have a relatively lower Fo on a weak syllable which belongs syntactically to the following word (see section V.A.2). Seen in relation to the preceding stressed syllable this tendency is not very obvious for the CSC speakers in the 1982 material except for two of the three speakers in

mísforstår (see section V.a.2), and in *mále bøtte fém* vs. *málebøtten*. It is true that *suppe* is lower in *en sød suppedás* than in *sødsùppe* for all three CSC speakers, but that may be because *suppedás* is the last word of the sentence. A relatively low level in relation to the following stressed word is counteracted in most of the examples by the final position of this latter word. But it is clear in the two examples where the following stressed syllable is not the last stressed syllable of the sentence, i.e. *Ótto vil mále bøtte fém i mórgen* and *Jeg købte en (gámmel) mát lampét til sóvekàmmeret (igår)*. This is, however, mainly due to a particularly high level of the word which carries the unit accent. The syllables *fém* and *-pét* do not lie on the general declination line of the sentence, but are either on the same level or higher than the preceding stressed syllables for all speakers. This cannot be due to a difference in intrinsic Fo, for the intrinsic difference between [a] and [ɛ] is less than three Hz (according to Petersen 1976), and for EF and PH the difference is considerably larger. It cannot, either, be a simple resetting in a longer sentence. It is true that PH has a rather irregular declination line, but EF does not have resetting in other sentences with four to five stresses. The relatively high Fo is mostly accompanied by higher intensity. In the sentences *kássen er sat óp på lóftet* and *han fándt en kande vánd i køkkenet* (from the 1961 material) *óp* is not only higher than *sat*, and *vánd* than *kande* for four out of five CSC and DSC speakers. They are also higher than the preceding stressed words (*kássen* and *fándt*, respectively) for three of the five speakers.

3. INTENSITY

The difference in intensity contributes somewhat more in the case of reduction by unit accentuation than in compounds with secondary stress. In the latter case there were very great individual differences, the grand mean for all speakers being 55% positive cases, i.e. cases where the intensity difference parallels the stress difference, but with only 4% significantly different cases. In the former case, i.e. unit accentuation, all speakers except NR had reduction of intensity in the majority of cases, the grand mean being 65% positive and 23% significant cases. Thus there is a somewhat stronger tendency to decrease the intensity in the stress reduced member of a unit accentuation group, and this cannot be explained by the Fo contour, but it is only a tendency, and it cannot be of much importance as a perceptual cue.

4. CONCLUSION

Reduced stress due to unit accentuation differs from secondary stress in compounds by loss of stød and (at least in monosyllabics) of phonological vowel length, by a more pronounced tendency to reduction in duration and a slight tendency to more reduction in intensity.

D. SECONDARY STRESS IN COMPOUNDS COMPARED TO WEAK STRESS

As mentioned in section II, Rischel has suggested that in syllables with short vowels there is hardly any difference between secondary stress and weak stress. In an oral discussion he gave the example *útỳske* 'monster' (compound) - *emfátiske* 'emphatic' (derivative). This has been tested in a single pair. Instead of *emfátiske* which has stød, I chose *polítiske* for three informants to be compared with *útỳske*; then, as this might be said to be a very common word with more reduction than the more rare word *ùtỳske*, I chose *herrenhútiske* (which certainly cannot be said to be a common word) for three more informants.

All speakers had a very clear and highly significant difference in the duration of *tỳsk* vs. *tisk* (9.7 cs on the average). The vowels were only significantly different for three speakers (two examples of *politiske* and one of *herrenhútiske*), and the closure of the *t* for two speakers. The main difference was in the aspiration, which was 8.6 cs in *útỳske* and 2.7 cs before the derivative *isk*, thus practically reduced to [d] (with the exception of SR's pronunciation of *polítiske*), and this lack of aspiration may be interpreted as indicating a following weak syllable, since it is a general rule that Danish stops are unaspirated before [ə] and the weak ending *-ig*. Moreover, four of the six informants had a clear difference in the Fo contour of the two words, and two had an inconsistent difference. OT, SR and IE have a much higher pitch in *-tisk* than in *-tỳsk*, and NR has a rise in *-tisk* and a fall in *-tỳsk*, whereas PH and EF sometimes have a rising-falling contour in *-tisk* but never in *-tỳsk*. This difference evidently depends to a large extent on the difference in the duration of the preceding consonant, which is much longer in *-tỳsk*, so that the Fo contour has come farther down when the vowel starts. However, this does not explain the whole difference. If the curves are superposed they fit completely for NR, who has a rise in *tisk* and a (higher starting) fall in *tỳsk*. But for PH and EF *tỳsk* starts at a too high level and, what is more important, for OT, SR and IE, who generally have a relatively low Fo on second members of compounds but otherwise a higher unstressed syllable, *tỳsk* is considerably lower than it should be according to the Fo wave of *polítiske*. Finally, NR, PH and OT have a significant difference in the intensity which is lower in *-tisk*, although Fo is higher, and EF and IE have a tendency in the same direction.

Thus, in this example there was no merging of secondary stress and weak stress. But more examples are needed, also examples with other consonants than stops, before any safe conclusion can be drawn about the general tendency.

VI. GENERAL CONCLUSIONS

The present investigation is in many respects very preliminary, and the conclusions must be taken with some reservations. I can only hope that it will provoke some more intensive studies of restricted problems combined with perceptual tests.

The preliminary conclusions are the following:

(1) Weak stress is distinguished from main stress in Danish by loss of stød and phonological vowel length, by duration, by the Fo contour and, for weak syllables with [ə] and [ɒ/ɔ] also - though less consistently - by intensity, in the sense that lower intensity is often combined with higher Fo. The shortening of unstressed syllables may be considerable; for the word pairs where syllables of the same structure could be compared it was 6.3 cs on the average. This must be of perceptual importance. The same must be true of the Fo differences, which are very clear and stable within the same norm, but of different type in different dialects.

The data for Advanced Standard Danish (ASC) is in complete conformity with Nina Thorsen's description of the ASC stress contour, i.e. a low predominantly falling, stressed syllable plus a high posttonic with a gradual fall of following, unstressed syllables. (This is here called type 1.) Speakers of a slightly more conservative norm (CSC speakers) follow the same pattern in principle, but with small deviations: somewhat more frequently low rising stressed syllables and a tendency to lower an unstressed syllable which belongs syntactically to the following stressed syllable and after stød. (This is here called type 2.) A North-East Jutlandish speaker (PM) and a speaker of the Århus regional standard had generally rising Fo in stressed syllables and a quicker fall, so that posttonic vowels after single consonants are higher than the stressed vowels, whereas they are lower after clusters (type 3). Finally, West Jutlandish speakers have high and rising stressed syllables and a steeper fall, all unstressed syllables being relatively low (type 4). The North Zealandish speakers follow type 3 in principle, but approach type 4 after obstruents. East Jutlandish speakers (apart from PM) follow type 4 in principle, but approach type 3 after sonorants. Copenhagen standard speakers with a Jutlandish background compromise between type 2 and type 4, and thus come close to type 3.

The relative importance of duration and Fo should be investigated by perceptual tests. The importance of intensity is more problematic. For full vowels it generally follows the Fo contour and increased intensity can hardly be expected to support the impression of weak stress, unless the increase is smaller than should be expected, if intensity is expected to follow Fo. It remains to be investigated what type of compensation the listener will make. But for the time being I assume that intensity is less important as a perceptual cue than both Fo and duration.

(2) Secondary stress in second members of compounds is distinguished from main stress by a reduction of syllable duration (on the average 2.9 cs in the standard language, thus less than in unstressed syllables). There is also a shortening of the distance between the start of the vowel in the syllables with main and secondary stress (2.6 cs), and a very weak tendency to shorten the first member of the compound (1.4 cs on the average). The total shortening, i.e. the sum of the shortening of first and second member of a compound thus amounts to 4.3 cs on the average, but there are large variations according to words and speakers, and there are countercases without shortening, which are nevertheless perceived correctly.

The Fo contour of syllables with secondary stress is the same as for unstressed syllables and very stable. The Fo contour of a compound thus differs clearly from the Fo contour of a sequence of words with main stress. 85% of the individual pairs were consistently distinguished, but there was some overlapping in certain word types.

The intensity is very variable in Standard Danish and only significantly lower for one informant, who has low Fo on secondary stress. In Jutlandish secondary stress has lower intensity, following Fo. Its perceptual importance is thus dubious.

(3) Reduced stress due to unit accentuation differs from secondary stress in compounds by loss of stød and (at least in monosyllabics) of phonological vowel length, by a more pronounced tendency to reduction in duration, and a slight tendency to more reduction in intensity. The Fo contour is practically the same, but there may be a difference in relation to the following stressed syllable. It approaches weak stress, but it may have overlapping both with secondary stress and with weak stress.

(4) The very limited material (one pair spoken by six informants) shows a clear distinction between secondary stress in a compound and weak stress in the derivative *-isk*.

ACKNOWLEDGEMENTS

I am grateful to the Institute of Phonetics at the University of Copenhagen for giving me access to the facilities of their laboratory, and particularly to Svend-Erik Lystlund for competent technical assistance. I also want to thank Niels-Jørn Dyhr and Lilian Andersen for reliable help with measurements, computations, graphs and proofreading, and Else Parkmann for very competent typing. - I am indebted to the staff of the Institute of Dialectology at the University of Copenhagen for advice and help, particularly for finding informants in Zealand and East Jutland. My sincere thanks are also due to the staff of the Institute for the Study of Jutlandish Language and Culture in Århus for efficient help with the recordings of Jutlandish dialects. - I am particularly grateful to Nina Thorsen and Jørgen Rischel, who have read the manuscript in an earlier version and given very valuable suggestions for improvements.

REFERENCES

Andersen, P. 1949: Review of Kroman, E. *Musikalsk Akcent i Dansk* (1947), *Danske Folkemaal XV*, p. 102-113

Andersen, P. 1954: "Dansk Fonetik", *Nordisk Lærebog for Talepædagoger I*, p. 308-353 (Rosenkilde & Bagger, Copenhagen)

Arnholtz, A. 1939-40: "Talekunstens og talelærens udvikling og stilling i Danmark", *Nordisk Tidsskrift for Tale og Stemme 4*, p. 82-98

Basbøll, H. 1971-73: "A commentary on Hjelmslev's Outline of the Danish expression system", *Acta Linguistica Hafniensia XIII*, p. 174-211, and II, *ibid. XIV*, p. 1-24

Basbøll, H. 1972: "Some remarks concerning the stød in a generative grammar of Danish", *Proceedings of the KVAL Seminar, Stockholm*, p. 5-30

Basbøll, H. 1978: "A note on boundaries and stress rules in Danish phonology", *Nordic Prosody* (eds. Gårding, E., Bruce, G., and Bannert, R.), *Travaux de l'Institut de Linguistique de Lund XIII*, p. 65-71

Benguerel, A. P. 1971: "Duration of French vowels in unemphatic stress", *Language and Speech 14*, p. 383-391

Bergsveinsson, S. 1941: "Sprog og tale", *Nordisk Tidsskrift for Tale og Stemme V*, p. 93-104

Berinstein, A. E. 1979: "A cross-linguistic study of perception and production of stress", *Working Papers in Phonetics, UCLA*, p. 1-59

Bertinetto, P. M. 1980: "The perception of stress by Italian speakers", *J. Phonetics 8*, p. 385-395

Bleakley, D. 1973: "The effect of fundamental frequency variations on the perception of stress in German", *Phonetica 28*, p. 42-59

Bo, A. 1933: *Tonegangen i Dansk Rigsmaal, Studier fra Sprog- og Oldtidsforskning 164* (Branner, Copenhagen)

Bolinger, D. L. 1958a: "A theory of pitch accent in English", *Word 14*, p. 109-149

Bolinger, D. L. 1958b: "On intensity as a qualitative improvement of pitch accent", *Lingua VII*, p. 175-182

Bolinger, D. L. and Gerstman, L. J. 1957: "Disjuncture as a cue to constructs", *Word 13*, p. 246-255

Botinis, A. 1982: "Stress in modern Greek: An acoustic study", *Working Papers, Lund 22*, p. 27-38

Chomsky, N. and Halle, M. 1968: *The Sound Pattern of English* (Harper and Row, New York)

Cohen, A. and t'Hart, J. 1967: "On the anatomy of intonation", *Lingua 19*, p. 177-192

Ejskjær, I. 1954: *Brøndum-Målet, Lydsystemet i en Sallingdialekt, Udvalg for Folkemaals Publikationer, Serie A, Nr. 10* (Schultz, Copenhagen)

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

Fischer-Jørgensen, E. 1961: "Some remarks on the function of stress with special reference to the Germanic languages", *Congrès international des sciences anthropologiques et ethnologiques, Bruxelles 1948, Comptes-Rendus, IIIe Session*, p. 86-88

Fischer-Jørgensen, E. 1964: "Sound duration and place of articulation", *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung 17*, p. 175-207

Fischer-Jørgensen, E. 1982: "Segment duration in Danish words in dependency on higher level phonological units", *Ann. Rep. Inst. Phon. Univ. Cph. 16*, p. 137-189

Fischer-Jørgensen, E. 1983a: "The acoustic manifestation of stress in Danish with particular reference to the reduction of stress in compounds", *Sound Structures* (eds. M.P.R. van den Broecke, V. van Heuven and W. Zonneveld) (Foris, Dordrecht), p. 81-104

Fischer-Jørgensen, E. 1983b: "Supplementary notes on vowel lengthening in Danish", *Ann. Rep. Inst. Phon. Univ. Cph. 17*, p. 21-53

Fónagy, I. 1958: "Elektrophysiologische Beiträge zur Akzentfrage", *Phonetica 2*, p. 12-58

Fónagy, I. 1966: "Electrophysiological and acoustic correlates of stress and stress perception", *J. Speech and Hearing Research 9*, p. 231-244

Forchhammer, V. 1954: "De Sieverske klangtyper og kurver" *Nordisk Lærebog for Talepædagoger I*, p. 475-483 (Rosenkilde og Bagger, Copenhagen)

Fry, D. B. 1955: "Duration and intensity as physical correlates of linguistic stress", *J. Acoust. Soc. Am. 27*, p. 765-768

Fry, D. B. 1958: "Experiments in the perception of stress", *Language and Speech 14*, p. 126-152

Gårding, E. 1977: *The Scandinavian word accents* (Gleerup, Lund)

Hansen, Aa. 1943: *Stødet i Dansk, Det Kgl. Danske Videnskabernes Selskab, Historisk-filologiske Meddelelser XXIX, Nr. 5* (Munksgaard, Copenhagen)

't Hart, J. and Cohen, A. 1973: "Intonation by rule: A perceptual quest", *J. Phonetics 1*, p. 309-327

Hirano, M. and Ohala, J. 1969: "The function of laryngeal muscles regulating fundamental frequency and intensity of phonation", *J. Speech and Hearing Research 12*, p. 616-628

Hjelmslev, L. 1951: "Det danske udtrykssystem med særligt henblik på stødet", *Selskab for Nordisk Filologis Årsberetning 1948-49-50*, p. 12-24 (English translation in L. Hjelmslev: *Essais Linguistiques II 1971*)

Hyman, L. 1977: "On the nature of linguistic stress", *Studies in Stress and Accent* (ed. L. Hyman), *Southern Californian Occasional Papers in Linguistics 4*, p. 37-82

Isačenko, A. and Schädlich, H. J. 1970: *A Model of Standard German Intonation* (Mouton, the Hague)

Janota, P. 1967: "Perception of stress by Czech listeners", *Proc. VI Int. Congr. Phon. Sc.* Prague, p. 457-461

Janota, P. 1979: "Some observations on the perception of stress in Czech", *Proc. IX Int. Congr. Phon. Sc.* Copenhagen II, p. 403-409

Jassem, W. 1959: "The phonology of Polish stress", *Word 15*, p. 252-269

Jensen, E. 1944: *Houlbjergmålet, Bidrag til en beskrivelse af en østjysk dialekt, Udvalg for Folkemaals Publikationer, Serie A, Nr. 6* (Schultz, Copenhagen)

Jespersen, O. 1897-99: *Fonetik* (Det Schuboteske Forlag, Copenhagen)

Jespersen, O. 1914: *Lehrbuch der Phonetik* (Teubner, Leipzig)

Jespersen, O. 1922: *Modersmålets Fonetik* (2nd ed.) (Gyldendal, Copenhagen)

Katwijk, A. 1969: "The perception of stress", *Inst. Perc. Res. Ann. Progr. Rep. 4*, (Eindhoven), p. 69-73

Katwijk, A. 1974: *Accentuation in Dutch* (Mouton, the Hague)

Kroman, E. 1947: *Musikalsk Akcent i Dansk* (Munksgaard, Copenhagen)

Ladefoged, P. 1967: "Stress and respiratory activity", *Three Areas of Experimental Phonetics* (Oxford University Press)

Lea, W.A. 1977: "Acoustic correlates of stress and juncture", *Studies in Stress and Accent* (ed. Hyman, L.), *Southern California Occasional Papers in Linguistics 4*, p. 83-119

Lehiste, I. 1980: "Interaction between test word and length of utterance", *The Melody of Language*, (eds. Waugh and van Schooneveld) (Baltimore), p. 169-176

Liberman, M. and Prince, A. 1977: "On stress and linguistic rhythm", *Linguistic Inquiry 8*, p. 249-336

Lieberman, P. 1960: "Some acoustic correlates of word stress in American English", *J. Acoust. Soc. Am. 32*, p. 451-454

Lieberman, P. 1967: *Intonation, perception and language*, (M.I.T. Press)

Marcus, S. M. 1979: "Perceptual centers (P-centers)", *Proc. IX Int. Congr. Phon. Sc. I*, p. 238, and discussion *ibid. III*, p. 264-265

Martinet, A. 1954: "Accents et tons", *Miscellanea Phonetica II*, p. 2-24

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

Ohala, J. 1970: *Aspects of the Control and Production of Speech, Working Papers in Phonetics, UCLA, 15*

Ohala, J. 1973: "The physiology of tone", *Consonant Types and Tone* (ed. Hyman, L.), *Southern California Occasional Papers in Linguistics 1*, p. 3-12

Ohala, J. 1977: "The physiology of stress", *Studies in Stress and Accent* (ed. Hyman, L.), *Southern California Occasional Papers in Linguistics 4*, p. 145-168

Ohala, J. and Ladefoged, P. 1970: "Further investigations of pitch regulation in speech", *Working Papers in Phonetics, UCLA, 14*, p. 12-24

Petersen, P. Riber 1973: "An instrumental investigation of the Danish 'stød'", *Ann. Rep. Inst. Phon. Univ. Cph. 7*, p. 195-234

Petersen, N. Reinholt 1976: "Identification and discrimination of vowel duration", *Ann. Rep. Inst. Phon. Univ. Cph. 10*, p. 57-84

Petersen, N. Reinholt 1978: "Intrinsic fundamental frequency of Danish vowels", *J. Phonetics 6*, p. 177-189

Rigault, A. R. 1962: "Rôle de la fréquence, de l'intensité et de la durée vocalique dans la perception de l'accent français", *Proc. IV Int. Congr. Phon. Sc.*, Helsinki, p. 735-748

Rischel, J. 1970: "Morpheme stress in Danish", *Ann. Rep. Inst. Phon. Univ. Cph. 4*, p. 111-144

Rischel, J. 1972: "Compound stress in Danish without a cycle", *Ann. Rep. Inst. Phon. Univ. Cph. 6*, p. 211-226

Rischel, J. 1980: "Phrasenakzent als Signal des Objekts ohne 'Determiner' im Dänischen", *Kopenhagener Beiträge zur Germanistischen Linguistik, Sonderband I*, p. 262-279

Rischel, J. 1982: "On unit accentuation in Danish - and the distinction between deep and surface phonology", *Folia Linguistica XVII*, (in print), also *Ann. Rep. Inst. Phon. Univ. Cph. 15*, p. 191-239

Rosenvold, E. 1981: "The role of intrinsic Fo and duration in the perception of stress", *Ann. Rep. Inst. Phon. Univ. Cph. 15*, p. 147-166

Rossi, M. 1970: "Sur la hiérarchie des paramètres de l'accent", *Proc. VI Int. Congr. Phon. Sc.*, Prague, p. 779-786

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

Rossi, M. 1971b: "L'intensité spécifique des voyelles", *Phonetica 24*, p. 129-161

Rossi, M., di Cristo, A., Hirst, D., Martin, Ph., and Nishinuma, Y. 1981: *L'intonation, de l'acoustique à la sémantique*, (Klincksieck, Paris)

Scholes, R. J. 1971: *Acoustic Cues for Constituent Structure*, (Mouton, the Hague)

Smith, S. 1944: *Bidrag til Løsning af Problemer vedrørende Trykbetoning i dansk Rigssprog, en eksperimentalfonetisk Studie* (Kaiser, Copenhagen)

Smith, S. and Thyme, K. 1978: *Accentmetoden*, (Herning)

Spang-Thomsen, B. 1963: "L'accent en français moderne", *Orbis Litterarum, Supplementum 3*, p. 181-208

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

Thorsen, N. 1979: "Lexical stress, emphasis for contrast, and sentence intonation in Advanced Standard Copenhagen Danish", *Proc. IX Int. Congr. Phon. Sc. II*, p. 417-423

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

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

Thorsen, N. 1981a: "Intonation contours and stress group patterns in declarative sentences of varying length in ASC Danish - supplementary data", *Ann. Rep. Inst. Phon. Univ. Cph. 15*, p. 13-47

Thorsen, N. 1981b: "The present state of the description of Danish utterance prosody", *Ann. Rep. Inst. Phon. Univ. Cph. 15*, p. 49-60

Thorsen, N. 1982a: "On the variability in Fo patterning and the function of Fo timing in languages where pitch cues stress", *Phonetica 39*, p. 302-316

Thorsen, N. 1982b: "Selected problems in the tonal manifestation of words containing assimilated or elided schwa", *Ann. Rep. Inst. Phon. Univ. Cph. 16*, p. 37-100

Thorsen, N. 1983a: "Standard Danish sentence intonation - phonetic data and their representation", *Folia Linguistica 17*, p. 187-220

Thorsen, N. 1983b: "Two experiments in the perception of Fo timing in Danish", *Ann. Rep. Inst. Phon. Univ. Cph. 17*, p. 1-20

Thorsen, N. and Jul Nielsen, B. 1981: "Fundamental frequency patterns and sentence intonation in two Jutlandish dialects - a first approach", *Ann. Rep. Inst. Phon. Univ. Cph. 15*, p. 1-12

APPENDIX I

The 1961 material.

1. a. Vi får sikkert sødsùppe/sød súppe til fórmåd.
   b. Hvis det víser sig at menúen står på sødsùppe/sød súppe, skal jeg íkke håve nóget.
   c. Kan du líde sødsùppe?/sød súppe?
   d. Néj, jeg bryder mig íkke om sødsùppe/sød súppe.

2. a. Har du nógensìnde sét en sórtmèjse /sórt méjse?
   b. Jå, jeg så en sórtmèjse/sórt méjse på túren igår.
   c. Hvis man skal ópnå at få øje på en sórtmèjse/sórt méjse, må man forhólde sig méget rólig.
   d. Jeg har áldrig ópnået at sé en sórtmèjse/sórt méjse.

3A. a. Dórthe vil méget gérne håve en sórt kjóle.
   b. Hun vil have en sórt kjóle på til fésten.
   c. Har Dórthe fået en sórt kjóle?
   d. Hvis Dórthe íkke får lóv at få en sórt kjóle, bliver hun fornærmet og går íkke méd.

3B. a. En katólsk præst blev tídligere ófte kaldt for en sórtkjòle.
   b. Når man kaldte den katólske præst for en sórtkjòle, blev han fornærmet og gík sin véj.
   c. Kaldte mánden præsten for en sórtkjòle?
   d. Ja, men han kalder ham kún for en sórtkjòle når han er vréd.

4A. a. Enhvér déltàger i øvelsen må sélv médbrìnge en gásmàske.
   b. Man skal have gásmàske på héle áftenen.
   c. Har du fået ánskàffet en gásmàske?
   d. Hvis du íkke har fået fát i en gásmàske, kan du íkke komme méd til øvelsen.

4B. a. Han havde en hvás máske på héle áftenen.
   b. Han går áltid rúndt og víser en hvás máske.
   c. Går han áltid rúndt med sádan en hvás máske?
   d. Hvis han áltid víser sádan en hvás máske, vil han íkke égne sig til stíllingen.

5. Kássen blev sát óppe / sat óp på lóftet.
   Han sátte kássen óppe / satte kássen óp / på lóftet.

6. Han fándt en kánde, en vándkànde / en kande vánd / en tóm kánde i køkkenet.

The sentences a,b,c,d within each group (1,2,3,etc.) were read in the given order, a and b (and c and d) in immediate succession, but otherwise the order was random, and other sentences were interspersed. - 3A was read by some speakers, 3B by others.

APPENDIX II

The 1982 material

The numbers of the sentences correspond to the numbers in Appendices III and IV and in the tables.

In some cases the sentences varied in details for different speakers, the variants are given in parentheses.

A. Main stress vs. weak stress

1a Blómsten hávde nogle gárnàgtige støvdrägere.
b Søren kóm med nogle barnágtige údtàlelser.

2a Jeg fík (købte) en (en gámmel) mát lámpe til sóvekàmmeret (igár)
b Jeg fík (købte) en (en gámmel) mát lampét i fórgårs (igár).

3a Han passérede Sánddàlen sídste sómmer (i fórgårs)
b Han kassérede sandálen sídste sómmer (i fórgårs)

3a Det er Sánddàlen de síger (NR only)
b Det er sandálen de síger

4a Jórdbær er bílligst om sómmeren (NR only)
b Jéns er bilíst om en háls

4a Det er bílligst de siger (EMG material)
b Det er bilíst de siger

5a De sagde mámam
b De sagde mamám

B. Main stress (a) vs. secondary stress in compounds (b) and unit accentuation reduction (c)

1a Óle spiser sød súppe til míddag (dáglig)
b Óle spiser sødsùppe til míddag (dáglig)
c Det var en sød suppedás

2a Jeg fík (købte) en (en gámmel) mát lámpe til sóvekàmmeret (igár)
b Jeg fík (købte) en (en gámmel) nátlàmpe til sóvekàmmeret (igár)
c Jeg fík (købte) en (en gámmel) mát lampét til sóvekàmmeret (igár)

3a Han lod Spíes réjse ságen
b Han tog Spíesrèjsen méd (EF only)
c Han lod Spíes rejse méd

APPENDIX II (continued)

4a Péters vén skåber sig
b Nú giver vénskàbet problémer (EF only)
c Péters vén skaber problémer

5a Hendes gás brænder dårligt
b Hendes gásbrænder söder
c Hendes gás brænder néd

6a Han vil måle kássen nú
b Han tog málekàssen méd
c Han vil måle Casanóva

7a Ótto vil måle bøtten stråks i mórgen (hér i gården)
b Ótto tager målebøtten frém i mórgen
c Ótto vil måle bøtte fém i mórgen
d Ótto vil male bøtten grøn i mórgen

8a Hendes lámme lår sygner hén
b Hendes lámmelår smager gódt

9a Péter smed det várme apparát néd på køkkengùlvet
b Péter tog vármeapparàtet méd til sómmerhùset
c Péter vil varme apparátet óp i fórvèjen

10a Sánnas (Hánnes) mís forstår spøgen
b Sánna (Hánne) mísforstår spøg
c Sánnas (Hánnes) mís forstår spøg

C. Secondary stress in compounds vs. weak stress

a Han var et útỳske mod sine børn.
b I det politiske kom han til kórt.
c Det Herrenhútiske var ham imód.

APPENDICES III-V

Introduction

In Appendix III syllables with main stress are compared to syllables with weak stress of the same (or almost the same) segmental structure; in Appendix IV syllables with main stress (a) are compared to syllables with reduced stress as second member of compounds (b), or weakened by unit accentuation (c), and the latter two are compared with each other. In Appendix V syllables with main stress are compared to syllables with reduced stress as second member of compounds, read by some dialectal speakers.

The duration measurements are averages of 8 tokens (for the dialect speakers sometimes only 6). In Appendix III the difference in duration between the first and the second vowel of the same word is also given if they had the same quality.

In Appendix IV-V the distance in cs between the start of the vowels of the first and second syllable (with stress or reduced stress) is also given.

In the graphs for Fo and intensity the time relations are not maintained in details. The Fo graphs are based on 3-4 superposed tracings and on sketches of all eight tokens. The tokens were generally very similar. Vowels with main stress are represented by thick lines, vowels with secondary or weak stress by thin lines, and sonorant consonants by dots. There may be dips in Fo for stød which are not indicated because the curves were not sufficiently clear. The absolute vertical dimensions are not comparable between individuals since the scales differed. In the row for differences (below the graphs) a + indicates that there is a consistent difference in all eight tokens, a + in parentheses that there are some overlappings, and a 0 that there is no clear difference. The difference may be either in the direction or in the level of the syllables, or both, or it may be in the intervening or following unstressed syllable. But a following stressed syllable, e.g. *(mát lam)pét* or *(mále bøtte) fém* is not included in the comparison. It is only included in the graph in order to show the difference in height from the preceding stressed syllable.

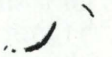
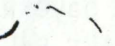
The intensity graphs are based on measurements of 5-7 tokens. The horizontal lines indicate the relative height of the peak measured in mm (because the calibration curves could not be used). The dB scale was not linear, but the upper part has been somewhat compressed in the graphs, though not sufficiently to show dB distances. Based on a few dB measurements and on a comparison with an earlier calibration curve it can be said that very roughly 0.5 mm in the graph corresponds to 1 dB in the lower part of the square, and to about 2 dB in the upper part. The indications of higher or lower within the same graph, or larger or smaller difference between pairs are, however, correct, also seen as dB differences. There

were only a few cases with displacement of the differences on the vertical scale and slightly larger differences for the higher pair where this was uncertain. They have been marked with a ?. For a few later recordings by NRP (of *bílligst - bilíst* and *Sánddàlen - sandálen*) both Fo measurements in Hz and intensity measurements in dB have been given. The Fo in Hz has been measured at the 67% point of the vowel duration (in [da:ʔlɳ] only the part until the abrupt fall or start of the irregular vibrations has been included in the vowel duration). A + in the row for differences means that the intensity relations parallel the stress relations, i.e. a syllable with weaker stress shows a larger decrease (or a smaller increase) of intensity in relation to the syllable with main stress than was found in the sequence of two main stresses (or, in Appendix II,1 and 3-5, than in the reversed sequence: weaker plus main stress), e.g. *sødsùppe* vs. *sød súppe* ‾ _ vs. - ‾ or - ‾ vs. _ ‾ . A - means that intensity and stress relations are reversed, and a 0 that there is no difference in the relations. (In the comparison between secondary stress in compounds and stress reduced by unit accentuation + and - are based on the preliminary experience that the latter is weaker.) One might also give the interpretation that a + means that the intensity relations support the perception of the stress differences, a - that it counteracts this perception, and a 0 that intensity does not play any role. However, as it is rather uncertain how the listener perceives intensity combined with sometimes higher, sometimes lower Fo on the more weakly stressed syllable, this may be a premature interpretation.

* means that the difference is significant at the 1% level according to the Mann Whitney test. (In the dialect recordings, which often comprised 6 readings only, significance at the 5% level is also indicated (by a * in parentheses).) Since the Fo curves have not been measured, no stars for significance are added to the plusses for Fo differences. But since a + means that there is a consistent difference, the weight of a + for Fo is the same as a +* for intensity and a (+) for Fo has the same weight as a + without star for intensity.

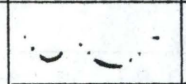
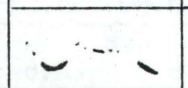
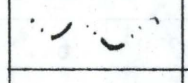
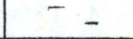
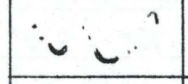
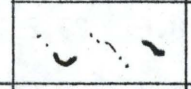
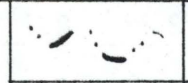
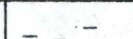
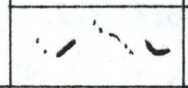
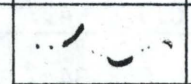
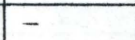
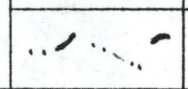
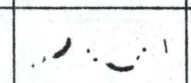
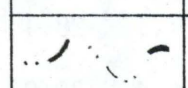
Appendix III,1

Differences in duration (in cs), Fo, and peak intensity between (a) gárnàgtig and (b) barnágtig in a medial stress group in Standard Danish. See further the introduction to App. III-V.

		Duration						Fo	Intensity	
		g/b	a:	n	(agt) a	g/ba:n	diff. a:-a	a:na i	a: a	á:-à > a:-á
NR	a	8.8	17.3	4.7	15.2	30.8	2.1*		— —	
	b	7.1	8.9	5.0	14.6	20.0	-5.7		— —	
a-b		1.7*	8.4*	-0.3	0.6	10.8*	7.8	+		-
PH	a	9.8	15.4	5.4	10.1	30.6	5.3*		— —	
	b	7.8	9.0	5.9	9.9	22.7	-0.9*		— —	
a-b		2.0*	6.4*	-0.5	0.2	7.9*	6.2*	+		0
EF	a	8.4	15.4	5.9	12.9	29.7	2.5*		— —	
	b	6.1	7.8	3.9	11.9	17.8	-4.1*		— —	
a-b		2.3*	7.6*	2.0	1.0	11.9*	6.6*	+		-

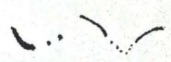
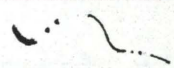
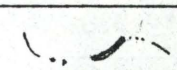
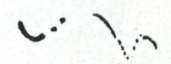
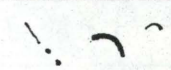
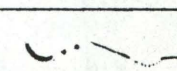
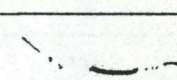
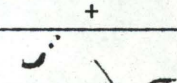
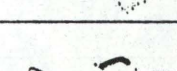
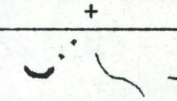
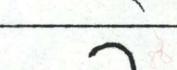
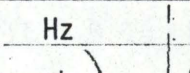
Appendix III,2

Differences in duration (in cs), Fo, and peak intensity between (a) mát lámpe and (b) mát lampét in a medial stress group in Standard Danish. See further the introduction to App. III-V.

	Duration						Fo	Intensity	
	(mat) a	l	ɑ	m	lɑm	diff. a-ɑ	ma lɑm^{ə}/ɛ	a ɑ ə/ɛ	á-ɑ > á-ɑ́
NR a	10.7	7.9	11.4	6.3	25.6	-0.7			
b	10.3	5.6	5.4	5.5	16.5	4.9*			
a-b	0.4	2.3*	6.0*	0.8	9.1*	-5.6*	+		0
PH a	8.9	5.9	10.2	6.8	22.9	-1.3			
b	9.1	4.9	6.2	5.8	16.9	2.9*			
a-b	-0.2	1.0	4.0*	1.0	6.0*	-4.2*	+		0
EF a	8.5	6.9	7.8	6.8	21.5	0.7			
b	8.9	5.1	4.6	5.1	14.8	4.3*			
a-b	-0.4	1.8	3.2*	1.7*	6.7*	-3.6*	+		0
OT a	11.1	7.1	11.2	6.9	25.2	-0.1			
b	10.1	5.2	7.3	4.9	17.4	2.8*			
a-b	1.0	1.9*	3.9*	2.0*	7.8*	-2.9*	+		-
SR a	8.3	6.5	9.9	6.1	22.5	-1.6*			
b	8.0	4.6	5.9	6.1	16.6	2.1*			
a-b	0.3	1.8*	4.0*	0	5.9*	-3.7*	+		-
IE a	12.4	7.9	13.4	10.8	32.1	-1.0			
b	12.1	6.6	8.7	7.8	23.1	3.4*			
a-b	0.3	1.3	4.7*	3.0*	9.0*	-4.4*	+		+ *

Appendix III,3

Differences in duration (in cs), Fo, and peak intensity between (a) Sánddàlen and (b) sandálen in a medial stress group in Standard Danish. See further the introduction to App. III-V.

						Fo		Intensity		á > à:ʔ	
	s	a	n	san	da:ʔlɒn	an	a:ʔlɒn	a	a:ʔ	a	á:ʔ
NR a	13.7	8.1	6.9	28.7	35.9			—	—		
b	8.3	5.1	6.1	19.5	38.1			—	—		
a-b	5.4*	3.0*	0.8	9.2*	-2.2	+				+	
PH a	12.4	7.2	9.0	28.6	37.3			—	—		
b	10.4	5.3	6.2	21.9	35.9			—	—		
a-b	2.0*	1.9*	2.8*	6.7*	1.4	+				0	
EF a	9.8	7.6	6.2	23.6	31.0			—	—		
b	7.4	6.3	5.0	18.7	32.6			—	—		
a-b	2.4*	1.3	1.2*	4.9*	-1.6	+				0	
OT a	10.8	9.8	8.6	29.2	36.1			—	—		
b	8.8	6.9	4.8	20.5	38.2			—	—		
a-b	2.0*	2.9*	3.8*	8.7*	-2.1	+				- *	
SR a	11.7	9.1	6.2	27.0	36.3			—	—		
b	10.6	6.4	5.3	22.3	37.3			—	—		
a-b	1.1	2.7*	0.9	4.7*	-1.0	+				+ *	
IE a	13.1	10.8	10.8	34.7	40.5			—	—		
b	12.1	8.2	6.9	27.2	40.4			—	—		
a-b	1.0	2.6*	3.9*	7.5*	0.1	+				+ *	
(one-stress utterance)						Hz		Hz	dB		dB
NR a	13.1	7.9	7.1	28.1	37.4			-26.3*	— —		0
b	9.5	4.9	6.4	20.8	41.7			-4.9*	— —		-5.7
a-b	3.6*	3.0*	0.7	7.3*	-4.3*	-1.1	20.3*	21.4*	-5.9*	-0.2	5.7
							+				+ *

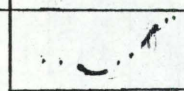
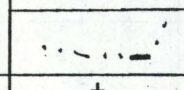
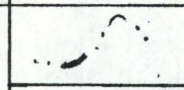
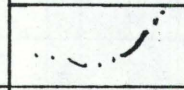
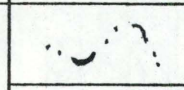
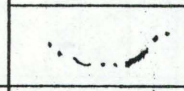
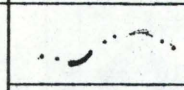
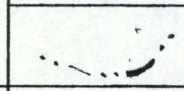
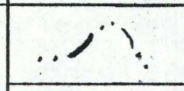
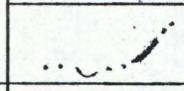
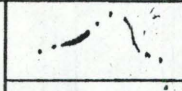
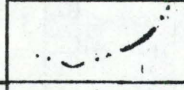
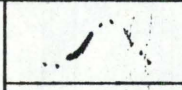
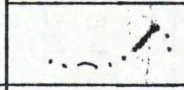
Appendix III,4

Differences in duration (in cs), Fo, and peak intensity between (a) bílligst and (b) bilíst in a medial stress group (= NR_1), and in a one stress group (in all other cases) in Standard Danish. For NR - BHo all averages comprise 11-15 tokens. See further the introduction to App. III-V.

		Duration						diff.	Fo	Hz		Intensity dB		Hz	
		b	i	l	i	bi	li	i-i	i i	i-i		i	i	i-i	
NR_1	a	12.0	7.9	6.1	7.6	19.9	13.7	0.3		-8.0*		—	—	3.2*	
	b	10.1	4.8	8.5	8.6	14.9	17.1	-3.8*		11.0*		—	—	0	
a-b		1.9*	3.1*	-2.4*	-1.0	5.0*	-3.4*	4.1*	-12.0* 7.0*	-19.0*	+	2.9*	0.3	3.2*	+
NR_2	a	13.9	8.4	7.1	8.2	22.3	15.3	0.2		-26.1*		—	—	1.7*	
	b	10.7	5.6	9.5	9.1	16.3	18.6	-3.5*		-9.4*		—	—	-1.2*	
a-b		3.2*	2.8*	-2.4*	-0.9*	6.0*	-3.3*	3.7*	2.4 19.1*	-17.3*	+	2.8*	-0.1	2.9*	+
NR_3	a	17.8	8.4	7.8	8.6	26.2	16.4	-0.2							
	b	12.4	6.5	11.7	8.2	18.9	19.9	-1.7*							
a-b		5.4*	1.9*	-3.9*	0.4	7.3*	-3.5*	1.5	+						
JJ	a	18.9	10.5	6.7	7.4	29.4	14.1	3.1*							
	b	12.1	7.2	11.4	10.5	19.3	21.9	-3.3*							
a-b		6.8*	3.3*	-4.7*	-3.1*	10.1*	-7.8*	6.4*	+						
BHu	a	11.9	11.6	6.9	10.2	23.5	17.1	1.4							
	b	9.6	8.2	10.1	11.2	17.8	21.3	-3.0*							
a-b		2.3*	3.4*	-3.2*	-1.0	5.7*	-4.2*	4.4*	+						
BM	a	14.2	8.9	6.1	7.5	23.1	13.6	1.4*							
	b	10.8	7.1	8.8	9.2	17.9	18.0	-2.1*							
a-b		3.4*	1.8*	-2.7*	-1.7*	5.2	-4.4	3.5*	+						
BHo	a	16.2	8.9	7.2	8.0	25.1	15.2	0.9							
	b	12.6	6.3	11.4	9.9	18.9	21.3	-3.6*							
a-b		3.6*	2.6*	3.8*	-1.9	6.2*	-6.1*	4.5*	+						

Appendix III,5

Differences in duration (in cs), Fo, and peak intensity between (a) mámam and (b) mamám in a one stress utterance in Standard Danish. See further the introduction to App. III-V.

		Duration								Fo	Intensity	
		m	a	m	a	ma	ma	diff. a-a	diff. ma-ma	ma mam	a a	á-a > a-á
NR	a	11.3	12.7	4.5	10.5	24.0	15.0	2.2*	9.0*		– –	
	b	7.8	8.1	9.9	11.2	15.9	21.1	-3.1*	-5.2*		– –	
	a-b	3.5*	4.6*	-5.4*	-0.7	8.1*	-6.1*	5.3*	14.2*	+		- *
PH	a	9.5	9.8	6.6	9.4	19.3	16.0	0.4	3.3*		– –	
	b	7.8	6.4	7.7	9.3	14.2	17.0	-2.9*	-2.8*		– –	
	a-b	1.7*	3.4*	-1.1	-0.1	5.1*	-1.0	3.3*	6.1*	+		+
EF	a	8.9	9.5	6.8	8.6	18.4	15.4	0.9	3.0*		– –	
	b	7.9	6.3	8.5	8.3	14.2	16.8	-2.0*	-2.6*		– –	
	a-b	1.0	3.2*	-1.7*	0.3	4.2*	-1.4	2.9*	5.6*	+		- *
OT	a	9.2	13.1	6.8	10.1	22.3	16.9	3.0*	5.4*		– –	
	b	7.5	8.9	7.1	12.0	16.4	19.1	-3.1*	-2.7*		– –	
	a-b	1.7*	4.2*	-0.3	-1.9*	5.9*	-2.2*	6.1*	8.1*	+		-
BJ	a	9.4	12.0	6.2	9.6	21.4	15.8	2.4*	5.6*		– –	
	b	8.5	10.3	8.5	11.3	18.8	19.8	-1.0	-1.0		– –	
	a-b	0.9	1.7*	-2.3*	-1.7*	2.6*	-4.0*	3.4*	6.6*	+		0
IE	a	10.1	13.6	12.8	16.5	23.7	29.3	-2.9*	-5.6*		– –	
	b	9.1	11.8	10.4	13.5	20.9	23.9	-1.7*	-3.0*		– –	
	a-b	1.0	1.8	2.4*	3.0*	2.8*	5.4*	-1.2	-2.6	+		+
NK	a	12.1	10.0	7.9	8.6	22.1	16.5	1.4*	5.6*		– –	
	b	9.6	9.5	11.0	9.5	19.1	20.5	0	-1.4		– –	
	a-b	2.5*	0.5	-3.1*	-0.9	3.0*	-4.0*	1.4*	7.0*	+		+

Appendix III,6

West and North Jutlandish

Differences in duration (in cs), Fo, and peak intensity between (a) Sánddàlen and (b) sandálen in a medial stress group. D means dialect, RSD regional Standard Danish. See further the introduction to App. III-V.

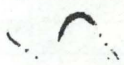
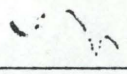
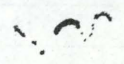
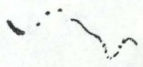
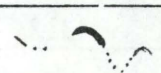
		Duration					Fo	Intensity	
		s	a	n	san	da:ʔl(n)	an a:ʔlŋ[1]	á a:ʔ	á-à: ≷ a-á:
BT D	a	9.3	7.8	9.0	26.1	-		– –	
	b	8.5	5.1	6.6	20.2	-		– –	
	a-b	0.8	2.7*	2.4*	5.9*		+		+ *
BT RSD	a	9.8	8.9	7.9	26.6	28.9		– –	
	b	9.0	4.6	5.7	19.3	28.9		– –	
	a-b	0.8	4.3*	2.2*	7.3*	0	+		-
TA RSD	a	9.4	7.7	8.1	25.2	39.7		– –	
	b	9.8	5.4	6.8	22.0	39.9		– –	
	a-b	-0.4	2.3*	1.3*	3.2*	-0.2	+		+*
EA RSD	a	12.5	9.6	12.9	35.0	38.9		– –	
	b	11.2	5.9	9.1	26.2	40.8		– –	
	a-b	1.3*	3.7*	3.8*	8.8*	-1.9	+		+ *
JD[1] D	a	9.8	9.8	11.8	31.4	27.4		– –	
	b	9.6	6.4	6.3	22.3	30.9		– –	
	a-b	0.2	3.4*	5.5*	9.1*	-3.5	+		+ *
JD RSD	a	11.6	9.1	8.3	29.0	39.2		– –	
	b	11.8	6.2	5.7	23.7	41.1		– –	
	a-b	-0.2	2.9*	2.6*	5.3*	-1.9	+		+*
PN RSD	a	15.1	8.4	9.8	33.3	46.9		– –	
	b	11.2	7.8	9.0	28.0	41.6		– –	
	a-b	3.9*	0.6	0.8	5.3(*)	5.3	+		+

[1]) JD(D) has [ɔ] [ɔ:ʔ] for [a] [a:ʔ] and no [n̩].

Appendix III,7

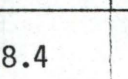
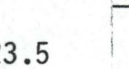
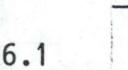
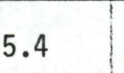
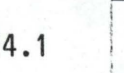
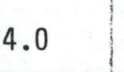
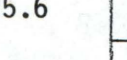
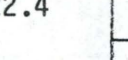
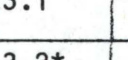
East Jutlandish

Differences in duration (in cs), Fo, and peak intensity between (a) Sánddàlen and (b) sandálen in a medial stress group. See further the introduction to App. III-V.

		Duration					Fo		Intensity	
		s	a	n	san	da:ʔln	an	a:ʔln	a a:ʔ	á-à: > a-á:
PM D	a	9.8	6.0	9.1	24.9	28.3			– –	
	b	9.5	4.1	7.7	21.3	28.3			– –	
a-b		0.3	1.9*	1.4*	3.6*	0	+			+ *
PM RSD	a	10.5	5.4	8.1	24.0	27.1			– –	
	b	8.6	4.0	6.7	19.3	27.0			– –	
a-b		1.9*	1.4*	1.4*	4.7*	0.1	+			+ *
LH RSD	a	9.8	8.5	9.9	28.2	34.6			– –	
	b	10.7	5.7	6.9	23.3	37.9			– –	
a-b		-0.9	2.8*	3.0*	4.9*	-3.3	+			+

Appendix IV,1

Duration, distance (in cs), Fo, and peak intensity in (a) sød súppe, (b) sødsùppe, and (c) sød suppedás, and the differences a-b, a-c, and b-c in Standard Danish. See further Introduction to App. III-V.

	Duration			Distance	Fo	Intensity
NR	søðʔ	sobə	o	ø — o	ø o ə	ø o
a	29.5	34.9	12.5	28.4		
b	26.5	28.7	10.0	23.5		
c	30.1	28.7	8.6	26.1		
a-b	3.0*	6.2*	2.5*	4.9*	+	-
a-c	-0.6	6.2*	3.9*	2.3	+	-
b-c	-3.6*	0	1.4*	-2.6*	+	0
PH						
a	26.9	30.1	9.6	25.4		
b	26.4	25.8	7.2	24.1		
c	27.7	24.5	6.0	24.0		
a-b	0.5	4.3*	2.4*	1.3	+	-
a-c	-0.8	5.6*	3.6*	1.4	+	+
b-c	-1.3	1.3	1.2	0.1	(+)	+
EF						
a	24.5	29.3	7.9	25.6		
b	24.7	24.1	5.6	22.4		
c	23.6	24.0	4.1	23.1		
a-b	-0.2	5.2*	2.3*	3.2*	+	+
a-c	0.9	5.3*	3.8*	2.5*	+	+
b-c	1.1	0.1	1.5	-0.7	(+)	+

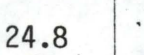
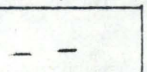
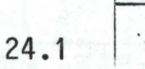
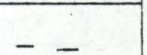
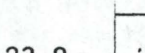
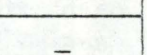
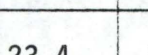
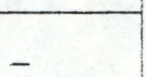
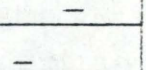
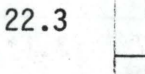
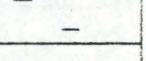
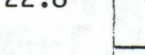
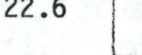
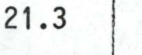
continued

(Appendix IV,1, continued)

	Duration			Distance	Fo			Intensity	
OT	søðʔ	sobə	o	ø - o	ø	o	ə	ø	o
a	27.9	35.4	13.3	28.4				—	—
b	26.3	30.1	10.2	25.3				—	—
c	26.3	30.1	7.9	25.2				—	—
a-b	1.6*	5.3*	3.1*	3.1*		+		+	
a-c	1.6	5.3*	5.4*	3.2*		+		+ *	
b-c	0	0	2.3*	0.1		(+)		+ *	
SR									
a	28.9	34.1	11.0	28.9				—	—
b	25.3	29.8	8.8	24.5				—	—
c	26.9	25.6	6.4	24.9				—	—
a-b	3.6*	4.3*	2.2*	4.4*		+		+	
a-c	2.0	8.5*	4.6*	4.0*		+		+	
b-c	-1.6	4.2*	2.4*	-0.5		(+)		0	
IE									
a	39.4	43.1	15.5	37.6				—	—
b	34.5	36.3	13.8	32.4				—	—
c	35.7	32.7	8.1	32.6				—	—
a-b	4.9*	6.8*	1.7	5.2*		+		+ *	
a-c	3.7*	10.4*	7.4*	5.0*		+		+ *	
b-c	-1.2	3.6*	5.7*	-0.2		0		+	

Appendix IV,2

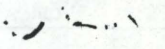
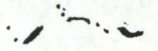
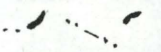
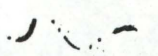
Duration and distance (in cs), Fo, and peak intensity in (a) mát lámpe, (b) nátlàmpe, and (c) mát lampét, and the differences a-b, a-c, and b-c in Standard Danish. See further introduction to App. III-V. [1]

	Duration					Distance	Fo	Intensity
NR	at	a	lɑm	lɑmb	ɑ	a — ɑ	ma lɑm ə/ɛ	a ɑ ə/ɛ
a	16.8	10.7	25.6	29.5	11.4	24.8		
b	16.4	11.2	24.3	27.5	10.3	24.1		
c	17.7	10.3	16.5		5.4	23.8		
a-b	0.4	-0.5	1.3	2.0	1.1	0.7	+	0
a-c	-0.9.	0.4	9.1*		6.0*	1.0	+	0
b-c	-1.3	0.9	7.8*		4.9*	0.3	+	0
PH								
a	17.5	8.9	22.9	27.7	10.2	23.4		
b	17.2	8.8	21.7	26.0	9.4	22.3		
c	17.4	9.1	16.9		6.2	22.3		
a-b	0.3	0.1	1.2	1.7	0.8	1.1	+	-
a-c	0.1	-0.2	6.0*		4.0*	1.1	+	-
b-c	-0.2	-0.3	4.8*		3.2*	0	0	+
EF								
a	16.0	8.5	21.5	27.4	7.8	22.8		
b	17.3	9.0	18.3	22.5	7.8	22.6		
c	16.1	8.9	14.8		4.6	21.3		
a-b	-1.3	-0.5	3.2*	4.9*	0	0.2	+	+
a-c	-0.1	-0.4	6.7*		3.2*	1.5	+	0
b-c	1.2	0.1	3.5*		3.2*	1.3	0	-

[1]) The difference ə-ɛ is not included in the comparisons a-c and b-c, only *mat lam*.

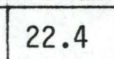
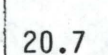
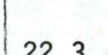
continued

(Appendix IV,2, continued)

	Duration					Distance	Fo	Intensity
OT	at	a	lɑm	lɑmb	ɑ	a — ɑ	ma lɑm ə/ɛ	a ɑ ə/ɛ
a	18.2	11.1	25.2	31.7	11.2	25.2		— — —
b	18.0	11.1	23.7	31.3	11.6	24.4		— — —
c	19.0	10.1	17.4		7.3	24.0		— — —
a-b	0.2	0	1.5	0.4	-0.4	0.8	+	-
a-c	-0.8	1.0	7.8*		3.9*	1.2	+	-
b-c	-1.0	1.0	6.3*		4.3*	0.4	(+)	0
SR								
a	18.7	8.3	22.5	28.6	9.9	25.1		— — —
b	18.5	8.8	20.8	25.5	10.1	23.6		— — —
c	18.3	8.0	16.6		5.9	22.9		— — —
a-b	0.2	-0.5	1.7*	3.1*	-0.2	1.5	+	-
a-c	0.4	0.3	5.9*		4.0*	2.2	+	-
b-c	0.2	0.8	4.2		4.2*	0.7	(+)	+
IE								
a	26.8	12.4	32.1	43.8	13.4	34.8		— — —
b	24.4	11.6	29.6	38.6	11.9	31.2		— — —
c	24.4	12.1	23.1		8.7	30.9		— — —
a-b	2.4	0.8	2.5*	5.2*	1.5	3.6*	+	+
a-c	2.4	0.3	9.0*			3.9*	+	+ *
b-c	0	-0.5	6.5*		3.2*	0.3	0	+ *

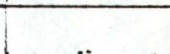
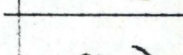
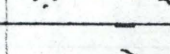
Appendix IV,3

Duration and distance (in cs), Fo, and peak intensity in (a) Spíes réjse, (b) Spíesrèjsen, and (c) Spíes rejse (méd), and the differences a-b, a-c and b-c in Standard Danish. See further introduction to App. III-V.

	Duration				Distance	Fo	Intensity
EF	sbi:ʔs	i:ʔ	ʁais͡ə	ɑi	i: ɑi	i:ʔ ɑi ə/n	i:ʔ ɑi
a	30.6	8.9	26.4	15.1	22.4		
b	29.3	9.1	25.4	14.3	20.7		
c	32.0	9.7	21.1	11.2	22.3		
a-b	1.3	-0.2	1.0	0.8	1.7	+	+
a-c	-1.4	-0.8	5.3*	3.9*	0.1	+	+
b-c	-0.7	-0.6	4.3*	3.1*	-1.6	0	-

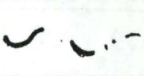
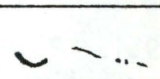
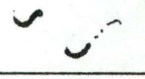
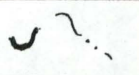
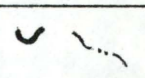
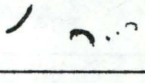
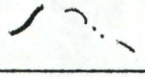
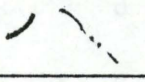
Appendix IV,4

Duration and distance (in cs), Fo, and peak intensity in (a) vén skáber, (b) vénskàbet, and (c) vén skaber (problémer), and the differences a-b, a-c, and b-c in Standard Danish. See further introduction to App. III-V.

	Duration				Distance	Fo	Int.
EF	vɛn	ɛ	sga:(ʔ)b	a:(ʔ)	ɛ-a:	vɛn a: ɒ/ð	
a	16.8	5.8	30.7	11.8	25.7		
b	15.2	5.6	25.3	8.7	22.7		
c	16.7	5.7	26.7	8.6	24.7		
a-b	1.6*	0.2	5.4*	3.1*	3.0*	+	-*
a-c	0.1	0.1	4.0	3.2*	1.0	+	-*
b-c	-1.5*	-0.1	-1.4	0.1	-2.0	0	+

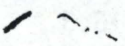
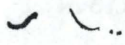
Appendix IV,5

Duration and distance (in cs), Fo, and peak intensity in (a) gás brænder, (b) gásbrænder, and (c) gás brænder (néd), and the differences a-b, a-c and b-c in Standard Danish. See further introduction to App. III-V.

	Duration			Distance	Fo	Intensity
NR	gas	a	bʁæn(ʔ)ɒ	a — ʁæ	a ʁænɒ	a æ
a	26.3	11.7	33.7	26.9		— —
b	25.1	11.3	34.0	26.0		— —
c	26.4	11.4	32.1	27.1		— —
a-b	1.2	0.4	-0.3	0.9	+	- *
a-c	-0.1	0.3	1.6	-0.2	+	0
b-c	-1.3	-0.1	1.9*	1.1	0	+ *
PH						
a	25.1	11.7	30.9	25.4		— —
b	23.6	11.1	29.5	24.8		— —
c	24.8	11.7	28.9	25.8		— —
a-b	1.5	0.6	1.4*	0.6	+	-
a-c	0.3	0	2.0*	-0.4	+	+ *
b-c	-1.2	-0.6	0.6	-1.0	+	+
EF(A)						
a	22.4	9.9	30.8	24.3		— —
b	22.0	10.4	30.8	23.6		— —
c	23.1	10.3	29.3	24.6		— —
a-b	0.4	-0.5	0	0.7	+	+
a-c	1.3	-0.4	1.5	-0.3	+	-
b-c	-1.1	0.1	1.5	-1.0	0	-

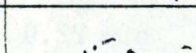
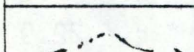
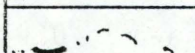
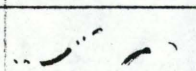
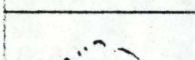
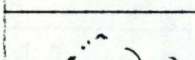
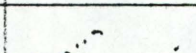
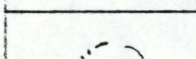
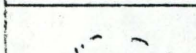
continued

(Appendix IV,5, continued)

	Duration			Distance	Fo	Intensity
EF(B)	gas	a	bʁæn(ʔ)ɒ	a - ʁæ	a ʁænɒ	a æ
a	22.1	8.8	29.3	24.1		— —
b	22.0	8.9	28.7	23.1		— —
c	22.9	9.6	29.4	24.6		— —
a-b	0.1	-0.1	0.6	1.0	+	+
a-c	-0.8	-0.8	-0.1	-0.4	+	+
b-c	-0.9	-0.7	-0.7	-1.4	0	-
OT						
a	26.9	11.8	36.1	29.1		— —
b	26.4	11.6	36.1	27.8		— —
c	26.8	11.6	36.3	29.2		— —
a-b	0.5	0.2	0	1.3	+	0
a-c	0.1	0.2	-0.2	-0.1	+	-
b-c	-0.4	0	-0.2	-0.4	0	-

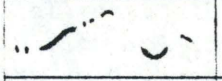
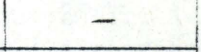
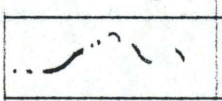
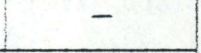
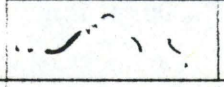
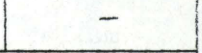
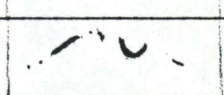
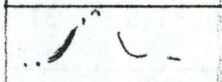
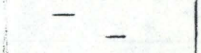
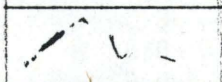
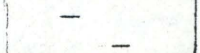
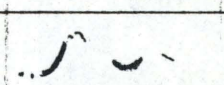
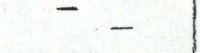
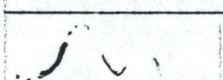
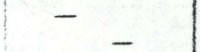
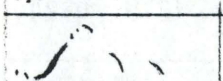
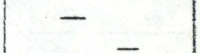
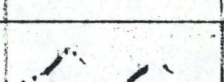
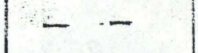
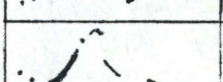
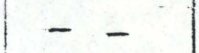
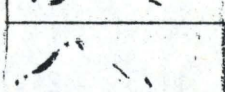
Appendix IV,6

Duration and distance (in cs), Fo, and peak intensity in (a) mále kássen, (b) málekàssen, and (c) mále Casanóva, and the differences a-b, a-c and b-c in Standard Danish. See further introduction to App. III-V.

	Duration				Distance	Fo	Intensity
NR	ma:lə	a:	kʰas	a	a: — a	ma:lə a n/a	a: a
a	36.6	16.3	38.3	16.4	39.9		— —
b	35.2	15.9	33.9	14.4	37.2		— —
c	36.6	16.4	25.6	9.8	38.1		— —
a-b	1.4	0.4	4.4*	2.0*	2.7*	+	-
a-c	0	-0.1	12.7*	6.6*	1.8	+	-
b-c	-1.4	-0.5	8.3*	4.6*	-0.9	0	+
PH							
a	28.0	12.8	28.8	10.7	32.0		— —
b	28.3	13.4	25.4	9.9	30.3		— —
c	29.5	14.0	21.6	7.0	30.9		— —
a-b	-0.3	-0.6	3.4*	0.8	1.7	+	0
a-c	-1.5	-1.2	7.2*	3.7*	1.1	+	+
b-c	-1.2	-0.6	3.8*	2.9*	-0.6	0	+
EF							
a	28.9	14.9	27.3	9.4	33.5		— —
b	25.8	13.0	23.7	8.9	28.3		— —
c	28.9	14.9	22.0	7.3	30.2		— —
a-b	3.1	1.9	3.6*	0.5	5.2*	+	+
a-c	0	0	5.3*	2.1*	3.3*	+	+
b-c	-3.1	-1.9	1.7*	1.6*	-1.9*	+	+

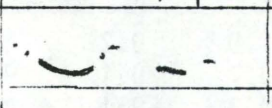
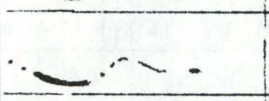
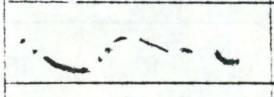
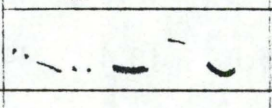
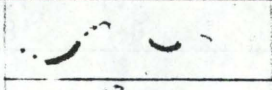
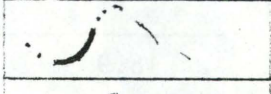
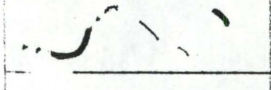
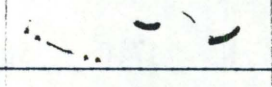
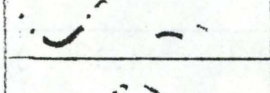
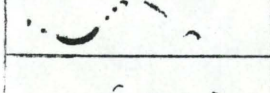
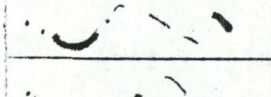
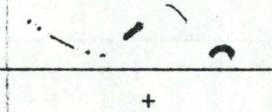
continued

(Appendix IV,6, continued)

	Duration				Distance	Fo	Intensity
OT	ma:lə	a:	k^has	a	a: – a	ma:lə a ə/a	ə: a
a	31.1	15.5	38.2	13.5	37.3		
b	30.1	15.3	33.6	12.1	35.2		
c	30.4	15.4	28.9	7.4	35.0		
a-b	1.0	0.2	4.6*	1.4	2.1*	+	0
a-c	0.7	0.1	9.3*	6.1*	2.3	+	-
b-c	-0.3	-0.1	4.7*	4.7*	0.2	+	-
IE							
a	34.4	18.5	46.8	16.0	41.9		
b	35.0	18.4	44.6	14.8	42.2		
c	35.5	18.4	35.9	10.4	40.7		
a-b	-0.6	0.1	2.2	1.2	-0.3	+	0
a-c	-1.1	0.1	10.9*	5.6*	1.2	+	+
b-c	-0.5	0	8.7*	4.4*	1.5	(+)	0
BJ							
a	31.1	16.9	38.1	14.7	38.2		
b	31.4	17.1	36.8	15.3	37.4		
c	31.1	17.0	30.8	9.6	37.9		
a-b	-0.3	-0.2	1.3	-0.6	0.8	+	0
a-c	0	-0.1	7.3*	5.1*	0.3	+	+
b-c	0.3	0.1	6.0*	5.7*	-0.5	0	+
NK							
a	34.1	14.3	34.6	11.8	36.7		
b	29.8	12.7	32.6	11.4	31.7		
c	34.4	14.7	31.1	10.3	36.4		
a-b	4.3*	1.6*	2.0	0.4	5.0*	+	+
a-c	-0.3	-0.4	3.5*	1.5*	0.3	+	+
b-c	-4.6*	-2.0*	1.5	1.1	-4.7*	(+)	+

Appendix IV,7

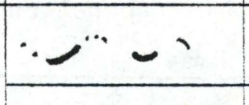
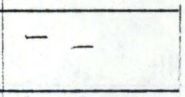
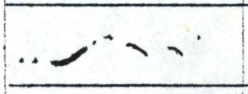
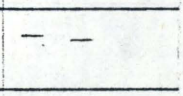
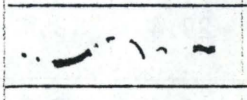
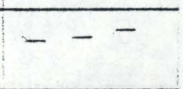
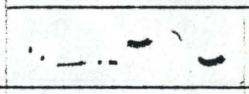
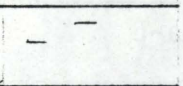
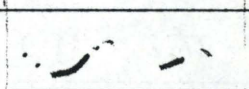
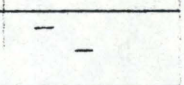
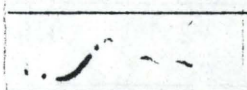
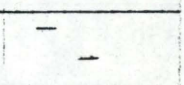
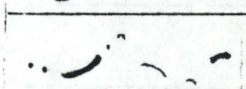
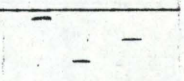
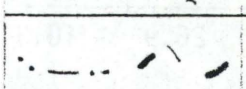
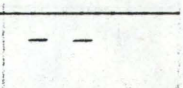
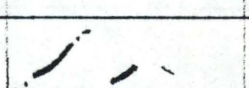
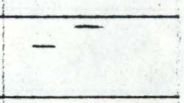
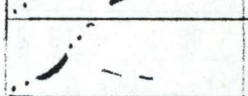
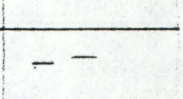
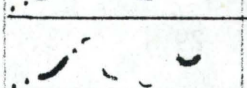
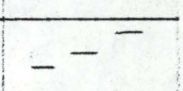
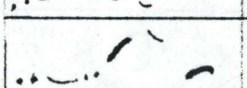
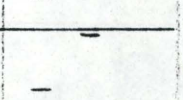
Duration and distance (in cs), Fo, and peak intensity of (a) mále bøtten, (b) málebøtten, (c) mále bøtte (fém), and (d) male bøtten (grøn), and the differences a-b, a-c, b-c, and a-d in Standard Danish. See further introduction to App. III-V. [1]

	Duration				Distance	Fo	Intensity
	ma:lə	a:	bød	ø	a:–ø	ma:lə ø ə n ɛ œ	a ø ɛ
NR a	32.6	15.6	30.1	10.5	35.7		— —
b	29.6	15.3	25.1	8.7	32.6		— —
c	32.2	15.9	21.2	7.8	33.3		— — —
d	26.5	11.9	31.6	10.5	31.3		— —
a-b	3.0*	0.3	5.0*	1.8*	3.1*	+	+
a-c	0.4	-0.3	8.9*	2.7*	2.4*	+	+?
b-c	-2.6*	-0.6	3.9	0.9	-0.7	0	-
a-d	6.1*	3.7*	-1.5	0	4.4*	+	+?
PH a	26.9	13.4	25.1	8.3	30.8		— —
b	27.3	13.2	20.5	7.2	27.6		— —
c	26.3	13.0	22.3	7.2	29.8		— —
d	24.1	12.9	24.0	8.0	28.6		— —
a-b	-0.4	0.2	4.6*	1.1	3.2*	+	+
a-c	0.6	0.4	2.8*	1.1	1.0	+	+*
b-c	1.0	0.2	1.8*	0	-1.2	0	+
a-d	2.8*	0.5	1.1	0.3	2.2*	+	+*
EF a	28.8	14.6	24.6	7.4	30.9		— —
b	26.2	12.6	18.8	5.3	27.5		— —
c	28.1	13.9	17.4	5.2	29.1		— — —
d	24.9	11.5	24.7	7.0	28.4		— —
a-b	2.6*	2.0*	5.8*	2.1*	3.4*	+	+
a-c	0.7	0.7	7.2*	2.2*	1.8	+	+
b-c	-1.9	-1.3	1.4	0.1	-1.6	(+)	+
a-d	3.9*	3.1*	-0.1	0.4	2.5*	+	+

[1]) ɛ and œ(in *fem* and *grøn*) are not included in the comparison.

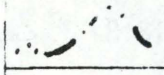
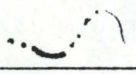
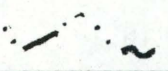
continued

(Appendix IV,7, continued)

	Duration				Distance	Fo	Intensity
	ma:lə	a:	bød	ø	a: – ø	ma:lə ø ə́ ɛ́ ɲ́ œ́	a ø ɛ
OT a	30.7	16.1	30.6	11.0	35.4		
b	28.9	13.8	24.9	8.4	31.6		
c	29.4	15.9	24.4	8.5	33.8		
d	25.6	12.6	29.1	9.7	30.4		
a-b	1.8	2.3*	5.7*	2.6*	3.8*	+	-
a-c	1.3	0.2	6.2*	2.5*	1.6	+	-
b-c	-0.5	-2.1*	0.5	-0.1	-2.2	0	-
a-d	5.1	3.5*	1.5	1.3*	5.0*	+	+
SR a	28.8	14.1	26.6	8.7	32.6		
b	29.3	13.6	25.9	9.7	31.6		
c	29.4	14.1	21.3	7.6	30.8		
d	26.4	12.1	24.4	7.7	30.1		
a-b	-0.5	0.5	0.7	-1.0	1.0	+	+
a-c	-0.6	0	5.3*	1.1	1.8*	+	+*
b-c	-0.1	-0.5	4.6*	2.1*	0.8	+	+*
a-d	2.4*	2.0*	2.2	1.0	2.5*	+	+*
IE a	30.1	15.7	37.7	14.3	34.8		
b	31.0	16.1	30.1	11.3	34.1		
c	31.5	16.8	27.1	8.6	34.9		
d	26.8	13.8	31.9	11.1	31.0		
a-b	-0.9	-0.4	7.6*	3.0*	0.7	+	+
a-c	-1.4	-1.1	10.6*	5.7*	-0.1	+	+ ?
b-c	-0.5	-0.7	3.0*	2.7*	-0.8	(+)	-
a-d	3.3*	1.9	5.8*	3.2*	3.8*	+	+

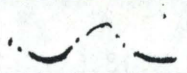
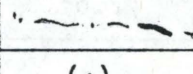
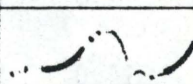
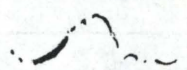
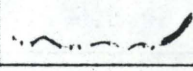
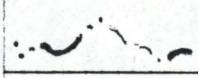
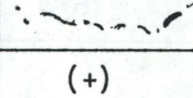
Appendix IV,8

Duration and distance (in cs), Fo, and peak intensity in (a) lámme lår and (b) lámmelàr, and the difference a-b, in Standard Danish. See further introduction to App. III-V.

	Duration				Distance	Fo	Intensity	
NR	lɑmə	ɑ	lɒ:ʔ	ɒ:ʔ	ɑ – ɒ:ʔ	lɑməlɒ:ʔ	ɑ	ɒ:ʔ
a	29.3	13.6	28.1	16.4	20.1		–	–
b	29.4	13.4	23.6	15.6	17.3		–	–
a-b	-0.1	0.2	4.5*	0.8	2.8*	+	-	
PH								
a	29.8	12.3	25.3	17.4	16.9		–	–
b	27.7	11.9	21.9	15.8	14.6		–	–
a-b	2.1	0.6	3.4*	1.6*	2.3*	+	+	*
EF								
a	26.9	10.3	24.1	18.4	16.6		–	–
b	25.1	9.6	19.8	15.5	14.9		–	–
a-b	1.8	0.7	4.3*	2.9*	1.7*	+	+	
OT								
a	32.5	13.2	27.1	18.5	20.3		–	–
b	29.4	12.1	23.2	16.0	17.5		–	–
a-b	3.1*	1.1	3.9*	2.5*	2.8*	+	+	

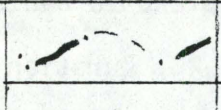
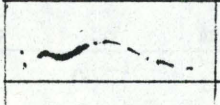
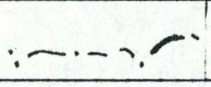
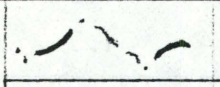
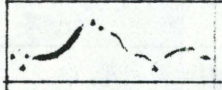
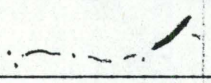
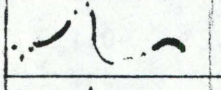
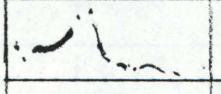
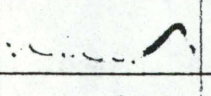
Appendix IV,9

Duration and distance (in cs), Fo, and peak intensity in (a) várme apparát, (b) vármeapparàtet, and (c) varme apparátet (óp), and the differences a-b and a-c in Standard Danish. See further Introduction to App. III-V.

	Duration					Distance	Fo	Intensity
NR	vɑ:m	ɑ:	əɑbɑ	ʁɑ:ʔ	ɑ:ʔ	ɑ:-ɑ:ʔ	vɑ:məɑ ɑʁɑ:ʔð	ɑ: ɑ ɑ ɑ:ʔ
a	26.2	17.9	14.4	19.6	15.3	40.2		— - —
b	24.2	16.7	11.2	16.5	12.8	37.7		— - —
c	24.8	16.4	13.1	20.1	15.9	37.8		— - —
a-b	2.0*	1.2	3.2*	3.1*	2.5*	5.5*	(+)	+
a-c	1.4	1.5	-1.9	-0.7	-0.6	2.4	+	-
PH								
a	27.9	14.7	18.6	18.2	12.8	45.1		— - — —
b	26.4	14.5	18.8	17.3	12.4	43.9		— - - —
c	25.7	14.8	19.9	20.1	14.4	45.6		— - — —
a-b	1.5	0.2	-0.2	0.9	0.4	1.2	+	+
a-c	2.2*	-0.1	-1.3	-1.9*	-1.6	-0.5	+	+ ?
EF								
a	25.3	15.1	19.8	17.5	13.1	44.5		— - - —
b	25.8	15.0	21.7	15.8	11.6	47.0		— - - —
c	22.3	13.1	23.3	18.3	13.7	45.6		— - - —
a-b	-0.5	0.1	-1.9	1.7	1.5	-2.5	(+)	+
a-c	3.0	2.0*	-3.5*	-0.8	-0.6	-1.1	+	0

continued

(Appendix IV,9, continued)

	Duration					Distance	Fo	Intensity
OT	vɑ:m	ɑ:	əɑbɑ	ʁɑ:ʔ	ɑ:ʔ	ɑ:-ɑ:ʔ	vɑ:məɑ ɑʁɑ:ʔð	ɑ: ɑ ɑ ɑ:ʔ
a	26.3	16.2	23.1	20.9	16.6	47.9		— - _ —
b	24.5	14.4	19.1	17.0	11.6	43.6		— - - —
c	25.1	13.6	19.6	22.6	16.8	44.1		— _ - —
a-b	1.8	1.8	4.0	3.9*	5.0*	4.3*	(+)	-
a-c	1.2	2.6*	3.5*	-1.7	-0.2	3.8*	+	+
SR								
a	24.9	14.6	19.0	17.4	13.4	42.9		— - _ —
b	25.8	14.8	18.8	16.6	13.2	42.6		— - _ —
c	25.7	14.0	19.4	19.1	14.8	43.4		— _ —
a-b	-0.9	-0.2	0.2	0.8	0.2	0.3	0	+
a-c	-0.8	0.6	-0.4	-1.7	-1.4	-0.5	+	+
IE								
a	33.8	17.9	24.1	27.2	20.8	56.8		— - _ —
b	32.3	16.9	22.1	19.6	16.2	51.5		— - _ —
c	30.9	16.4	23.8	25.4	19.2	54.1		— _ - —
a-b	1.5	1.0	2.0	7.6*	4.6*	5.3*	+	+*
a-c	2.9*	1.5	0.3	1.8	1.6	2.7*	+	+*

Appendix IV,10

Duration and distance (in cs), Fo, and peak intensity in (a) mís forstår, (b) mísforstår, and (c) mís forstår (spøg), and the differences a-b, a-c and b-c in Standard Danish. See further Introduction to App. III-V.

	Duration					Distance		Fo	Intensity
NR	is	i	fɒ	sdɒ:?	ɒ:?	i-ɒ	i-ɒ:?	mi ɒ ɒ:?	i ɒ ɒ:?
a	13.6	6.0	13.4	28.3	16.3	22.0	39.0		
b	14.1	7.4	12.8	25.4	14.1	21.9	38.3		
c	13.5	6.4	14.4	18.3	6.6	22.5	39.7		
a-b	-0.5	-1.4*	0.6	2.9*	2.2*	0.1	0.7	0	-
a-c	0.1	-0.4	-1.0	10.0*	9.7*	-0.5	-0.7	(+)	-
b-c	0.6	1.0	-1.6*	7.1*	7.5*	-0.6	-1.4	(+)	+?
PH									
a	13.5	6.3	11.6	27.7	14.8	21.6	38.0		
b	13.3	6.2	10.7	22.9	12.1	20.0	34.8		
c	13.9	5.8	11.8	23.5	12.1	20.9	36.2		
a-b	0.2	0.1	0.9	4.8*	2.7*	1.6	3.2*	(+)	+?
a-c	-0.4	0.5	-0.2	4.2*	2.7*	0.7	1.8	(+)	+
b-c	-0.6	0.4	-1.1*	-0.6	0	-0.9	-1.4	0	+
EF (A)									
a	13.3	5.6	12.0	24.1	12.7	20.8	37.3		
b	11.9	5.3	11.8	20.8	10.1	18.6	34.3		
c	13.4	5.6	12.4	17.9	6.4	20.9	37.3		
a-b	1.4*	0.3	0.2	3.3*	2.6*	2.2*	3.0*	(+)	+*
a-c	-0.1	0	-0.4	6.2*	6.3*	-0.1	0	0	+*
b-c	-1.5*	-0.3	-0.6	2.9*	3.2*	-2.3*	-3.0*	+	+

continued

(Appendix IV,10, continued)

	Duration					Distance		Fo			Intensity		
EF (B)	is	i	fɒ	sdɒ:?	ɒ:?	i-ɒ	i-ɒ:?	mi	ɒ	ɒ:?	i	ɒ	ɒ:?
b	12.8	5.3	11.9	20.7	11.1	21.1	34.3						
c	13.8	5.5	13.2	18.3	8.3	23.7	37.0						
b-c	-1.0	-0.2	-1.3	2.4*	2.8*	-2.6*	-2.7*		+			+	
OT													
a	17.7	8.6	13.4	28.6	15.8	24.7	44.1						
b	13.7	7.5	13.3	24.4	11.8	20.5	39.6						
c	17.3	8.3	13.4	22.1	9.8	24.5	42.9						
a-b	4.0*	1.1	0.1	4.2*	4.0*	4.2*	4.5*		(+)			0	
a-c	0.3	0.3	0	6.5*	6.0*	0.2	1.2		+			+ ?	
b-c	-3.6	-0.8	-0.1	2.3*	2.0	-4.0*	-3.3*		0			+	
SR													
a	16.6	6.9	13.9	25.6	14.3	25.2	41.8						
b	14.3	7.4	12.3	27.3	16.0	21.6	37.5						
c	17.0	6.1	15.0	26.5	14.5	26.3	44.2						
a-b	2.3*	-0.5	1.6*	-1.7	-1.7*	3.6*	4.3*		(+)			+*	
a-c	-0.4	0.8	-1.1	-0.9	-0.2	-1.1	-2.4		+			+*	
b-c	-2.7	1.3	-2.7*	0.8	1.5	-4.7*	-6.7*		(+)			+*	
IE													
a	20.4	9.1	15.2	38.0	20.3	20.5	29.6						
b	20.4	8.6	15.3	38.3	20.7	19.6	28.1						
c	21.2	8.9	15.7	33.1	15.7	20.9	29.7						
a-b	0	0.5	-0.1	-0.3	-0.4	0.9	1.5		+			+*	
a-c	-0.8	0.2	-0.5	4.9*	4.6*	-0.4	-0.1		(+)			+*	
b-c	-0.8	-0.3	-0.4	5.2*	5.0*	-1.3	-1.6		+			- ?	

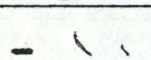
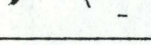
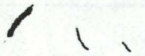
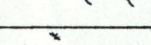
Appendix IV,11

Duration (in cs), Fo, and peak intensity in (a) útỳske and (b) Herrenhútiske, and the difference a-b in Standard Danish. See further Introduction to App. III-V.

	Duration						Fo			Intensity		
NR	u	t	h	y/i	sg	thysg/tisg	u	y/i	ə	n	y/i	ə
a	6.7	3.3	11.4	7.7	5.4	34.4						
b	6.6	2.7	4.6	5.3	5.1	24.5						
a-b	0.1	0.6	6.8*	2.4*	0.3	9.9*		+			+*	
PH												
a	7.3	4.4	8.1	6.9	12.4	31.8						
b	6.9	3.6	2.6	5.7	12.6	24.4						
a-b	0.4	0.8	5.5*	1.2	-0.2	7.4*		(+)			+*	
EF												
a	5.0	5.0	7.4	6.4	11.9	30.7						
b	5.8	4.6	2.8	6.2	12.6	26.1						
a-b	-0.8	0.4	4.6*	0.2	-0.7	4.6*		(+)			+	

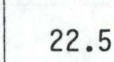
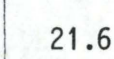
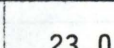
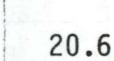
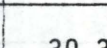
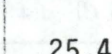
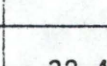
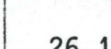
Appendix IV,12

Duration (in cs), Fo, and peak intensity in (a) útỳske and (b) polítiske, and the difference a-b in Standard Danish. See further Introduction to App. III-V.

	Duration						Fo	Intensity
OT	u/i	t	h	y/i	sg	thysg/tisg	u/i y/i ə	u/i y/i ə
a	8.0	6.1	7.3	9.4	12.9	35.7		
b	8.9	5.9	0	5.8	13.5	25.2		
a-b	-0.9	0.2	7.3*	3.6*	-0.6	10.5*	+	+*
SR								
a	8.4	7.2	6.9	7.0	13.1	34.4		
b	6.7	4.4	4.5	6.5	11.7	25.5		
a-b	1.7	2.8*	2.4*	0.5	1.4	8.9*	+	0
IE								
a	-	6.8	10.4	10.9	17.5	45.7		
b	9.8	3.8	1.8	8.4	14.4	28.3		
a-b		3.0*	8.6*	2.5*	3.1*	17.4*	+	-

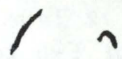
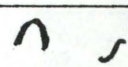
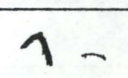
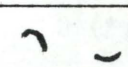
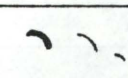
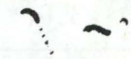
Appendix V,1

*Duration and distance (in cs), Fo, and peak intensity in (a) sød súppe and (b) sødsùppe, read by Jutlandish dialect speakers. D = dialect, RSD = Regional Standard Danish. * means significance at the 1% level, (*) at the 5% level. Each measure is the average of 6-8 tokens. See further Introduction to App. III-V.*

West Jutl.		Duration			Distance	Fo	Intensity
		søð?	sob	o	o – ø	ø o	ø o
BT D	a	21.1	25.0	9.4	22.5		– _
	b	22.3	21.9	9.2	21.6		– _
a-b		-1.2	3.1*	0.2	0.9	(+)	+
BT RSD	a	23.1	25.7	9.8	23.0		– _
	b	21.3	25.7	9.4	20.6		– _
a-b		1.8(*)	0	0.4	2.4*	+	-
JD D	a	31.5	36.3	9.4	30.3		– –
	b	27.5	33.3	9.8	25.4		– _
a-b		4.0(*)	3.0(*)	-0.4	4.9(*)	+	+*
JD RSD	a	32.3	29.7	10.0	32.1		– –
	b	25.6	25.4	8.4	26.1		– _
a-b		6.7*	4.3*	1.6	6.0*	+	+*

continued

(Appendix V,1, continued)

		Duration			Distance	Fo	Intensity
North Jutl.		søð?	sob	o	o – ø	ø o	ø o
PN D	a	39.8	32.3*	13.3	39.9		– –
	b	34.3	25.9	12.6	31.1		– –
a-b		5.5*	6.4*	0.7	8.8*	+	+*
East Jutl.							
PM D	a	23.9	19.1[1])	8.2	25.0		– –
	b	22.8	15.9	6.7	22.2		– –
a-b		1.1	3.2*	1.5*	2.8*	+	-
PM RSD	a	22.3	25.1	10.3	22.0		– –
	b	22.1	21.4	8.9	19.9		– –
a-b		0.2	3.7*	1.4	2.1*	+	+
LH RSD	a	25.2	20.8[1])	11.0	25.5		– –
	b	25.9	18.4	10.2	23.2		– –
a-b		-0.7	2.4	0.8	2.3(*)	0	+(*)

[1]) Only *so* could be measured.

Appendix V,2

Duration and distance (in cs), Fo, and peak intensity in (a) mát lámpe[1] and (b) nátlàmpe, read by Jutlandish dialect speakers. See further caption to App. V,1.

West Jutl.		Duration at[2]	lam(b)[3]	a	ɑ	Distance a – ɑ	Fo na	Fo lam	Intensity a	Intensity ɑ
BT D	a	15.2	18.3	8.2	6.7	21.6			–	–
	b	14.6	19.1	9.3	7.2	22.0			–	–
a-b		0.6	-0.8	-1.1	-0.5	-0.4	(+)		+	
BT RSD	a	15.2	27.6	9.5	11.4	21.4			–	–
	b	13.9	27.6	9.8	10.6	20.7			–	–
a-b		1.3	0	-0.3	0.8	0.7	+		-	
TA D	a	20.0	27.8	-	8.0	27.4			–	–
	b	20.8	26.6	-	8.2	28.6			–	–
a-b		-0.8	1.2		-0.2	-1.2	(+)		+(*)	
TA RSD	a	17.3	28.1	8.8	9.7	26.5			–	–
	b	16.3	28.3	8.3	9.6	25.1			–	–
a-b		1.0	-0.2	0.5	0.1	1.4	(+)		+	
EA D	a	24.6	25.4	-	9.6	32.8			–	–
	b	24.9	23.2	-	8.9	31.8			–	–
a-b		-0.3	2.2		0.7	1.0(*)	0		+	
EA RSD	a	17.8	38.3	11.6	11.5	30.1			–	–
	b	17.2	36.0	10.8	10.9	28.2			–	–
a-b		0.6	2.3	0.8	0.6	1.9	0		+	

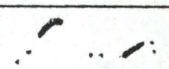
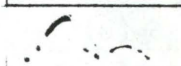
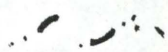
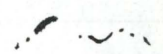
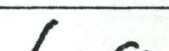
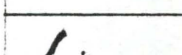
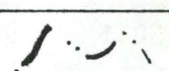
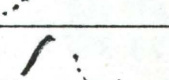
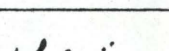
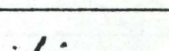
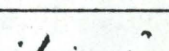
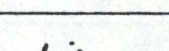
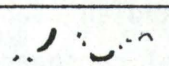
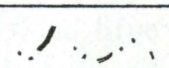
[1]) In the dialect the comparison was made between *den mátte lámpe* and *nátlàmpe*, since *matte* [maˀd̥] and *nat* [naˀd̥] have West Jutlandish stød, but not *mat*. There is also ˀ in [lɑmˀb̥].

[2]) The initial consonant of *nat* could not be delimited.

[3]) Only *lam* could be measured for EA,D.

continued

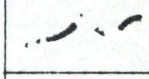
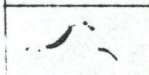
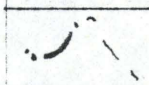
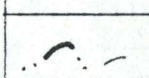
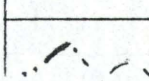
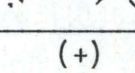
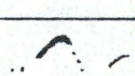
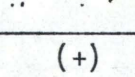
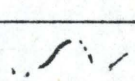
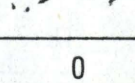
(Appendix V,2, continued)

		Duration				Distance	Fo	Intensity	
		at	lɑm(b)[1]	a	ɑ	a − ɑ	na lɑm	a	ɑ
JD D	a	22.8	34.4	9.6	7.3	32.4		—	—
	b	16.4	33.6	8.8	8.2	26.7		—	—
	a-b	6.4*	0.8	0.8	-0.9	5.7*	+		+
JD RSD	a	21.4	29.6	9.6	9.1	28.6		—	—
	b	22.1	27.6	9.9	9.3	28.1		—	—
	a-b	-0.7	2.0	-0.3	-0.2	0.5	+		+
North Jutl.									
PN D	a	22.1	34.8	11.1	12.9	36.1		—	—
	b	21.5	31.4	11.2	12.8	32.9		—	—
	a-b	0.6	3.4	-0.1	0.1	3.2	+		+*
PN RSD	a	18.5	32.5	10.3	10.4	28.8		—	—
	b	19.8	33.3	10.4	9.6	29.8		—	—
	a-b	-1.3	-0.8	-0.1	0.8	-1.0	+		+
East Jutl.									
PM D	a	13.3	18.8	6.6	6.0	20.3		—	—
	b	14.1	18.2	7.5	6.3	20.6		—	—
	a-b	-0.8	0.6	-0.9	-0.3	-0.3	+		-
PM RSD	a	13.6	25.9	6.9	7.8	20.3		—	—
	b	14.4	24.5	7.6	7.8	20.1		—	—
	a-b	-0.8	1.4(*)	-0.7	0	0.2	+		-
LH RSD	a	14.7	27.2	8.1	9.2	23.9		—	—
	b	15.4	26.7	8.0	9.7	23.6		—	—
	a-b	-0.7	0.5	0.1	-0.5	0.3	(+)		+

[1]) Only *lɑm* could be measured for PN(D) and LH.

Appendix V,3

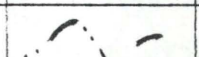
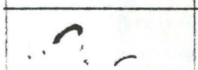
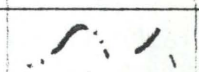
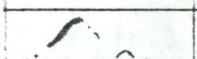
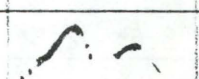
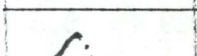
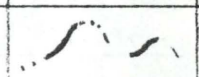
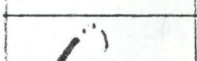
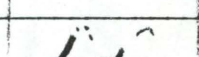
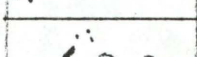
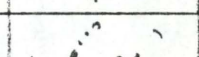
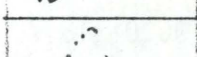
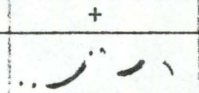
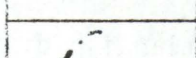
Duration and distance (in cs), Fo, and peak intensity in (a) mâle bøtten and (b) mâlebøtten, read by Jutlandish dialect speakers. See further caption to App. V,1.

West Jutl.		Duration ma:l(ə)[1]	bød	a:	ø	Distance a: – ø	Fo ma:l(ə)ø(n)	Intensity a: ø
BT D	a	26.4	22.6	14.5	7.5	28.7		– –
	b	22.8	16.0	11.5	6.1	24.3		– –
a-b		3.6*	6.6*	3.0*	1.3	4.4*	+	+
BT RSD	a	28.3	25.0	14.8	8.7	30.8		– –
	b	24.3	19.4	12.3	6.0	26.6		– –
a-b		4.0*	5.6*	2.5*	2.7*	4.2*	+	-
TA D	a	28.4	29.4	15.6	7.8	29.6		– –
	b	27.4	24.8	14.4	8.8	26.1		– –
a-b		1.0	4.6*	1.2	-1.0	3.5(*)	(+)	+*
TA RSD	a	30.1	34.3	13.1	9.3	32.3		– –
	b	27.9	31.7	12.9	9.3	29.6		– –
a-b		2.2*	2.6*	0.2	0	2.7*	(+)	-
EA D	a	36.0	30.9	-	9.6	36.3		– –
	b	35.7	30.9	-	8.0	35.2		– –
a-b		0.3	0		1.6	1.1	(+)	+*
EA RSD	a	38.6	31.7	20.4	9.7	41.9		– –
	b	36.1	30.0	18.7	9.2	38.6		– –
a-b		2.5*	1.7(*)	1.7(*)	0.5	3.3*	0	+*

[1]) In West Jutlandish dialect the definite article is a proclitic [ɛ]: ['mɔ:l ɛ 'bød] [ɛ 'mɔ:l,bød].

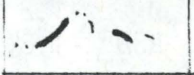
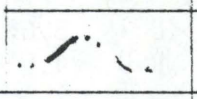
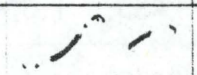
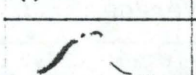
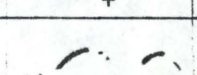
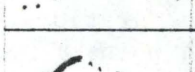
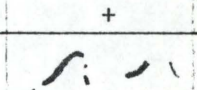
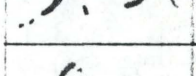
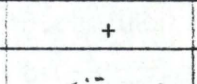
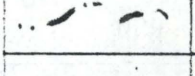
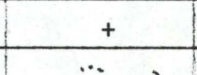
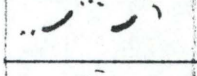
continued

(Appendix V,3, continued)

		Duration				Distance	Fo	Intensity
		ma:l	bød	a:	ø	a: – ø	ma:l(ə)ød(n)	a: ø
JD D	a	31.8	31.2	-	9.4	32.1		— —
	b	29.9	24.7	-	8.0	28.2		— —
a-b		1.9	6.5*		1.4	3.9*	+	+*
JD RSD	a	33.4	30.8	16.4	8.4	30.9		— —
	b	32.3	26.3	16.5	7.5	26.3		— —
a-b		1.1	4.5*	-0.1	0.9	4.6*	+	+*
North Jutl.								
PN D	a	35.6	28.8	-	10.1	36.1		— —
	b	35.8	28.1	-	10.5	34.0		— —
a-b		-0.2	0.7		-0.4	2.1	+	+*
PN RSD	a	38.9	32.4	18.6	10.3	41.9		— —
	b	39.4	26.4	17.1	8.9	37.4		— —
a-b		-0.5	6.0(*)	1.5	1.4(*)	4.5(*)	+	+
East Jutl.								
PM D	a	32.4	21.9	10.6	7.9	23.7		— —
	b	31.5	19.2	9.7	7.4	21.6		— —
a-b		0.9	2.7*	0.9	0.5	2.1*	+	0
PM RSD	a	25.9	22.1	11.0	8.4	27.4		— —
	b	25.0	19.8	10.5	6.9	25.7		— —
a-b		0.9	2.3*	0.5	1.5*	1.7*	+	0
LH RSD	a	31.6	29.0	15.9	10.0	35.3		— —
	b	30.8	25.6	14.9	10.3	33.0		— —
a-b		0.8	3.4*	1.0	-0.3	2.3*	+	+(*)

Appendix V,4

Duration and distance (in cs), Fo, and peak intensity in (a) mále bøtten and (b) málebøtten, read by Funish dialect speakers. See further caption to App. V,1.

Funish		Duration				Distance	Fo	Intensity
		ma:l	bød	a:	ø	a: - ø	ma:l(ə)ø ɳ	
MA	a	22.8	17.2	10.3	5.5	22.9		– –
	b	23.9	17.0	9.5	5.3	26.3		– –
a-b		-1.1	0.2	0.8	0.2	-3.4	+	+*
LA	a	27.4	22.3	12.0	6.5	22.3		– –
	b	26.9	21.8	12.9	6.0	21.8		– –
a-b		0.5	0.5	-0.9	0.5	0.5	+	+*
HV	a	38.1	30.9	19.2	7.6	40.2		– –
	b	35.4	30.3	16.6	9.8	34.3		– –
a-b		2.7	0.6	2.6(*)	-2.2	5.9*	+	+*
IP	a	40.0	35.9	19.5	8.8	45.8		– –
	b	40.6	29.1	17.0	9.6	43.1		– –
a-b		-0.6	6.8	2.5(*)	-0.8	2.7	+	+
EK	a	33.9	27.9	14.6	7.4	35.3		– –
	b	30.1	24.7	13.4	6.3	33.0		– –
a-b		3.8*	3.2*	1.2	1.1	2.3(*)	+	+(*)
EH	a	28.6	21.3	11.6	7.1	28.1		– –
	b	26.9	19.6	10.5	7.2	27.0		– –
a-b		1.7	1.7(*)	1.1	-0.1	1.1	+	0

Appendix V,5

*Reduction in duration (in cs) of the first member of compounds compared to the first word of a sequence of two separate words with main stress and the same segmental structure, read by Jutlandish dialect speakers. D = dialect, RSD = Regional Standard Danish. * indicates significance at the 1% level, (*) at the 5% level. There were 6-8 readings of each word.*

		BT	TA	EA	JD	PN	PM	LH	average
'søð?ˌsobə	D	-1.2	-	-	4.0(*)	5.5*	1.1	-	2.4
	RSD	1.8(*)	-	-	6.7(*)		0.2	-0.7	2.0
'ma:ləˌbødə	D	3.6*	1.0	0.3	1.9	-0.2	0.9	-	1.3
	RSD	4.0*	2.2*	2.5*	1.1	-0.5	0.9	0.8	1.6
'nadlambə	D	0.6	-0.8	-0.3	6.4*	0.6	-0.8	-	1.0
	RSD	1.3	1.0	0.6	-0.7	-1.3	-0.8	-0.7	-0.1
	average	1.7	0.9	0.8	3.2	0.8	0.3	-0.2	1.1\1.3

Appendix V,6

Reduction in duration (in cs) of the second member of compounds compared to the second word of a sequence of separate words with main stress and the same segmental structure, read by Jutlandish speakers. See further caption to App. V,5.

		BT	TA	EA	JD	PN	PM	LH	average
'søð?ˌsobə	D	3.1*	-	-	3.0(*)	6.4*	3.2*		3.9
	RSD	0			4.3*		3.7*	2.4	2.6
'ma:ləˌbødə	D	6.6*	4.6*	0	6.5*	0.7	2.7*		3.5
	RSD	5.6*	2.6*	1.7(*)	4.5*	6.0*	2.3*	3.4*	3.7
'nadlambə	D	-0.8	1.2	2.2	0.8	3.4	0.6		1.2
	RSD	0	-0.2	2.3	2.0	-0.8	1.4	0.5	0.6
	average	2.4	2.0	1.6	3.5	3.1	2.3	2.1	2.4\2.1

Appendix V,7

Reduction of the distance (in cs) between the start of the first vowel and the vowel with secondary stress in compounds compared to the distance between the vowels in a sequence of two words with main stress and (practically) the same segmental structure, read by Jutlandish speakers.

		BT	TA	EA	JD	PN	PM	LH	average
'søð?ˌsobə	D	0.9			4.9*	8.8*	2.8*		4.4
	RSD	2.4*			6.0*		2.1*	2.3	3.2
'ma:ləˌbødə	D	4.4*	3.5(*)	1.1	3.9(*)	2.1	2.1*		2.9
	RSD	4.2*	2.7*	3.3*	4.6*	4.5(*)	1.7*	2.3*	3.3
'nadˌlɑmbə	D	-0.4	-1.2	1.0(*)	5.7*	3.2	-0.3		1.3
	RSD	0.7	1.4	1.9	0.5	-1.0	0.2	0.3	0.6
average		2.0	2.0	1.8	4.3	3.5	1.4	1.6	2.4\2.6

ARIPUC 18, 1984, p. 163-184

THE PERCEPTION OF VOICE ONSET TIME: A CROSS-LANGUAGE STUDY OF AMERICAN ENGLISH AND DANISH

JENS B. CHRISTENSEN

This paper examines the relationship between the production and perception of the voicing distinction for American English and Danish stop consonants in initial position. In a comparison of the production of /p/ and /k/ for the two languages the Danish stops were found to have longer aspiration. It was therefore hypothesized that perceptually, Danish listeners would have a later cross-over point than American listeners. This was tested in a labelling experiment, using computer-edited, naturally produced stimuli. The aspiration was shortened from right to left to produce a series of stimuli ranging in VOT from +10 to +70 msec of the syllables pi, pu, ki, and ku. The listening tests showed a statistically significant difference in the expected direction for the labial stimuli but not for the velars. This may be due to the stimulus range which proved to be less appropriate for Danish listeners than for American listeners.

I. INTRODUCTION

Voice onset time, or VOT, the temporal relation between the release of a stop consonant and the onset of vocal fold vibration has been shown to characterize the phonological contrasts of homorganic stops in initial position for most languages. On the basis of spectrographic analyses of naturally produced syllables Lisker and Abramson (1964) proposed three categories into which these languages seem to group their stops: 1) voicing lead, assigned negative values, 2) coincident and short lag, assigned zero or low positive values, and 3) long lag, assigned high positive values. The stops in this last category have traditionally been called aspirated.

Their main objective for setting up these categories was to find "some single best measure" by which to separate the phoneme categories, and do away with the traditional concept of voicing, aspiration, and force of articulation as three mutually independent dimensions. They note that they have not been able to find any language where force of articulation would stand as one single feature, separating phoneme categories of stop consonants. They consider force of articulation, or fortis/lenis, to be closely connected with aspiration. One disconcerting factor which supports Lisker and Abramson's attempt to eliminate fortis/lenis is that phoneticians are still searching to find the physical, or acoustic correlate which will adequately describe this feature. See e.g. Fujimura and Miller (1979) and Kohler (1983).

It is well documented that VOT is a sufficient cue for listeners to differentiate between the phoneme categories found in their native languages according to the categories proposed by Lisker and Abramson: for American English and Thai by Lisker and Abramson (1970), for Spanish, Abramson and Lisker (1973) and Williams (1977), for Polish by Mikős, Keating and Moslin (1976) and Keating, Mikős and Ganong III. (1981).

Other cues have been shown to be operative in the perception of the contrast between stop categories; a cutback of the first formant, Liberman, Delattre and Cooper (1958), and the transition of the first formant, Stevens and Klatt (1974), (see also Lisker, 1975), as well as the fundamental frequency at the onset of voicing, (Haggard, Ambler and Callow, 1970), (Abramson and Lisker 1983).

Common to almost all cross-language studies is that they have compared languages which in their voicing contrasts differ across VOT categories. English has been contrasted with Thai, which shows a three-way contrast, with Spanish and Polish, both contrasting voicing lead with voicing lag. The theory of VOT has influenced work in other areas of linguistics, e.g. child language studies, where VOT was used to investigate possible innate feature detectors in infants. In a study Eimas, Siqueland,Jusczyk, and Vigorito (1971) found that one and four month old babies were able to discriminate between pairs of stimuli of +20 and +40 msec, but were not able to discriminate between pairs of stimuli that both had VOT values of either <20 or >40 msec. Eimas et al. therefore concluded that the 20-40 msec of voicing lag constituted a natural boundary, which incidentally is found in American adult studies. Support for this claim was found in a study on chinchillas (Kuhl and Miller, 1978) where these in a labelling test showed nearly identical category boundaries as those found in humans.

The cross-language studies clearly show that it is possible for humans to learn various sets of categories. Spanish and Polish listeners learn to distinguish between lead and lag, rather than short and long lag, and Thai listeners learn to distinguish among three categories, lead, coincident, and lag for the labial and alveolar places of articulation, as a function of

the linguistic input they receive in the acquisition of their native language. Spanish infants of 6-9 months seem to be able to discriminate between the voicing contrast found in Spanish.

A comparison of American English and Danish shows that in terms of VOT categories the same contrast is found in both languages (if the voicing lead is ignored in American English), namely short lag contrasting with long lag. As will be shown below, Danish aspirated stops differ from the American stops in that they show considerably longer VOT values. In the acquisition of the stop system for the two languages it is not necessary for the listeners to learn a completely new contrast. They will have to learn to distinguish between the same categories in both languages, both being near those found in the infant studies, and in the studies on chinchillas.

The question to be addressed in this paper is, whether Danish listeners, because of the longer aspiration, will show a later phoneme boundary than American listeners do, or whether a new contrast is learned in terms of these VOT categories. The question of the effect of formal phonetic training will also be tested to see whether phonetically trained subjects will access a special phonetic mode as participants in perception experiments. Furthermore, the universal difference found in both the production and perception of stop consonants will be considered; whether these differences are still present as reflected in different cross-over boundaries for different place of articulation, even though the stimuli have been constructed in such a way that most of the cues which might account for the perceptual differences have been eliminated.

II. PRODUCTION

In both languages there is a contrast between /*b*,*p*/, /*d*,*t*/, and /*g*,*k*/ in prestressed position. Danish /*d*/ and /*t*/ are often pronounced with a considerable degree of affrication following the release (Fischer-Jørgensen, 1980): [d̥^{s}] and [d̥^{sh}]. This feature is not found in American English, and it is therefore not relevant in a cross-language study of this kind, to compare the production and perception aspects of the alveolar stops.

The Danish /*p*/ and /*k*/ only contrast with /*b*/ and /*g*/ in syllable initial position, when followed by a (sonorant +) full vowel. Danish thus differs from English in that there is no phonological contrast of the kind *rapid / rabid*, *bagging / backing*. The Danish /*p*/ and /*k*/ are pronounced as voiceless aspirated: [b̥^{h}] and [g̊^{h}]. /*b*/ and /*g*/ are pronounced as voiceless unaspirated: [b̥], [g̊]. The American /*p*/ and /*k*/ are pronounced like the Danish, but are not quite as aspirated as their Danish counterparts. /*b*/ and /*g*/ are sometimes pronounced like in Danish, voiceless unaspirated, and sometimes as fully voiced, having voicing lead. This variant is most often found when preceded by a voiced sound. Since labelling tests have shown that American listeners group stimuli which have voicing lead

with those having short lag, it is safe to say that in terms of VOT, Danish and American English both contrast short lag with long lag. In order to see in what way the two languages differ in their production of the stops within the short lag and long lag categories VOT measurements for the two languages will be presented below.

A. DANISH PRODUCTION DATA AND POINTS OF DELIMITATION

The Danish data were obtained from the recordings of two male speakers: JR and NR. The test material consisted of the labials and velars followed by seven Danish vowels: /*i*, *e*, *æ*, *a*, *å*, *o*, *u*/ which were all pronounced long. The test word was inserted in a carrier sentence: *De ska(l) sige* .... (They will say ....). Speaker JR read the list three times and NR four times, giving a total of 49 tokens for each of the four stops.

When measuring VOT, the point of consonantal release usually does not constitute any problems. Voice onset is a different matter, and various points have been used. Fischer-Jørgensen and Hutters (1981) consider three different possibilities: A) the start of vocal fold vibration, B) the point at which F1 sets in, and C) the point at which the upper formants begin. In VOT studies there seems to be some disagreement among investigators as to which point to use. Lisker and Abramson (1964) do not include "edge vibrations", the vibrations of the vocal cords before they are fully adducted. They are thus considering point B to be the starting point. Point C has been used by e.g. Klatt (1975). Others use the word voice onset in its strictest sense, e.g. Keating et al. (1981) (Keating, personal communication), and thus consider A to be the starting point. This point is used in this study as well, for several reasons, one being that the measurements were made from digitized oscillograms.

In Table I are shown the results for the two Danish speakers. Since the material is limited, the individual values were checked against results found by Fischer-Jørgensen (1980). They were found to differ only slightly from her results, which may be due to e.g. different carrier sentences, and the different position of the test word in the utterance (Lisker and Abramson, 1967), and that Fischer-Jørgensen used point B for voice onset, or vowel onset.

Table I

Danish stops followed by 7 vowels

	Msec VOT			Msec VOT	
b	+	14.9	p	+	102.0
g	+	29.4	k	+	110.8

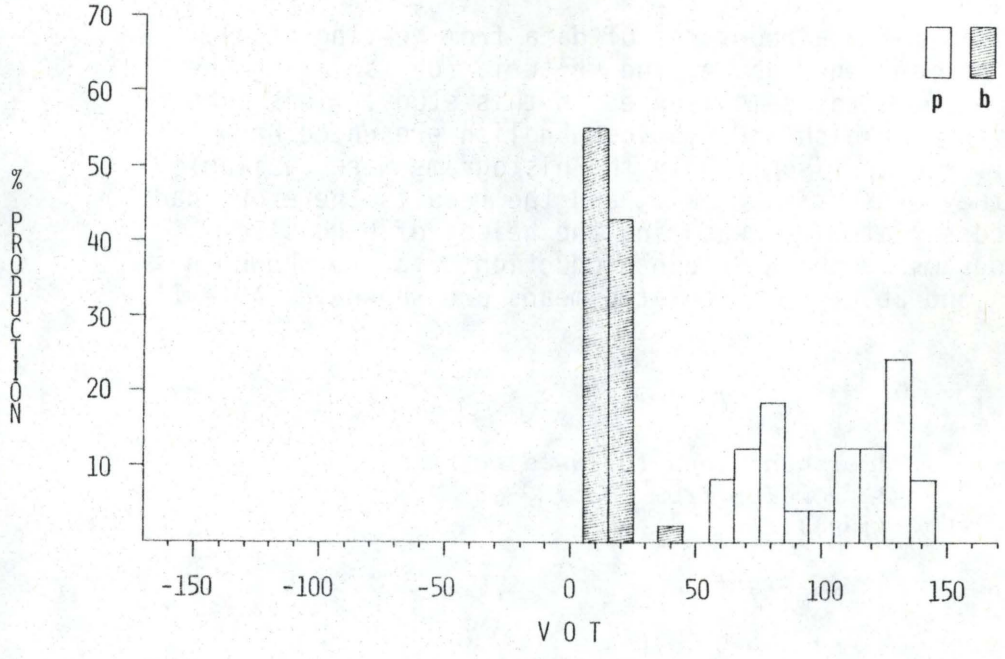


Figure 1a

Danish /b/ and /p/ followed by 7 vowels.

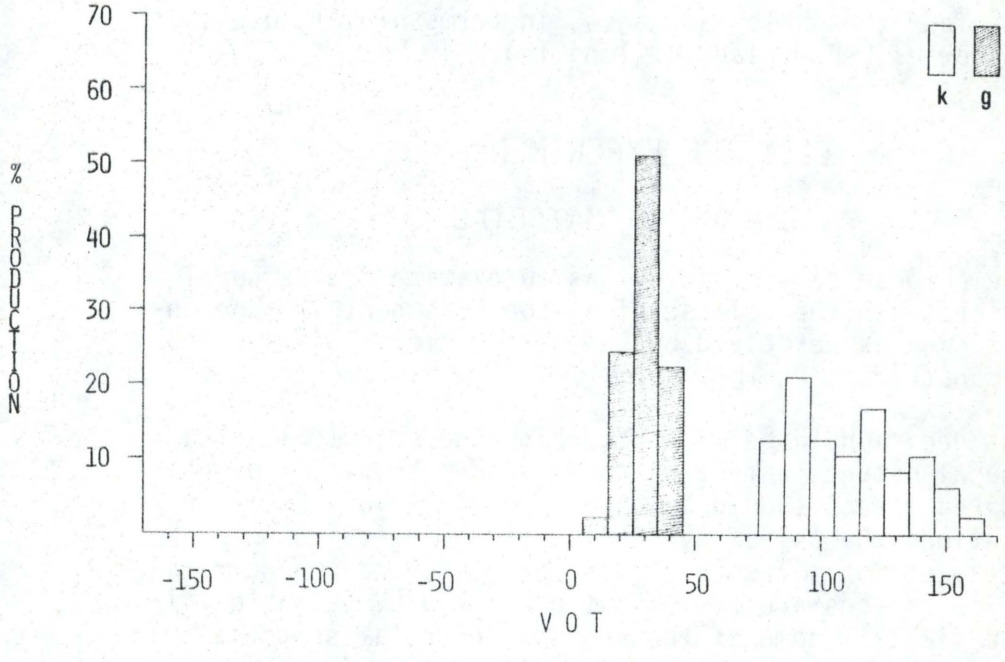


Figure 1b

Danish /g/ and /k/ followed by 7 vowels.

B. AMERICAN PRODUCTION DATA

The American material consists of data from Keating et al. (1981). As mentioned above, the criteria for isolating the voice onset were the same as used in this study. This makes the results of Danish and American English presented here suitable for comparison. Only the histograms were available in the study by Keating et al., and the results therefore had to be reconstructed by measuring the height of each block. The histograms of the American production data are shown in figure 2a and 2b. The calculated means are shown in Table II.

Table II

American stops followed by 12 vowels (from Keating et al. (1981).

	Msec VOT		
b	+ 5.6	p	+ 57.6
g	+ 15.5	k	+ 71.7

When comparing the data for the two languages it is clear that Danish aspirated stops are considerably longer than the corresponding American ones. In the following it will be shown how much this affects the perception of VOT by Danish and American listeners, reflected in different cross-over boundaries, despite the fact that both languages, in terms of VOT, use the same categories, short lag and long lag.

III. THE EXPERIMENT

A. CHOICE OF MATERIAL

Since the purpose of this study was to examine the temporal relations between the release of a stop consonant and the onset of voicing as perceived by Danish and American listeners, certain constraints limit the choice of test material.

The first decision to be made regarding the stimuli is of a more general nature, namely whether synthetic stimuli or natural edited speech should be used. In almost all American investigations the use of synthetic stimuli seems to be prevalent, except for a few (e.g. Winitz, LaRiviere and Herriman 1975). The reasons are obvious: synthesized CV syllables allow us to simulate some of the acoustic correlates of the gradual change of the vocal tract configuration from the stop to the vowel. Synthesized stimuli, however, have up till now had some buzzing quality to them, and it is therefore desirable to use real speech whenever possible.

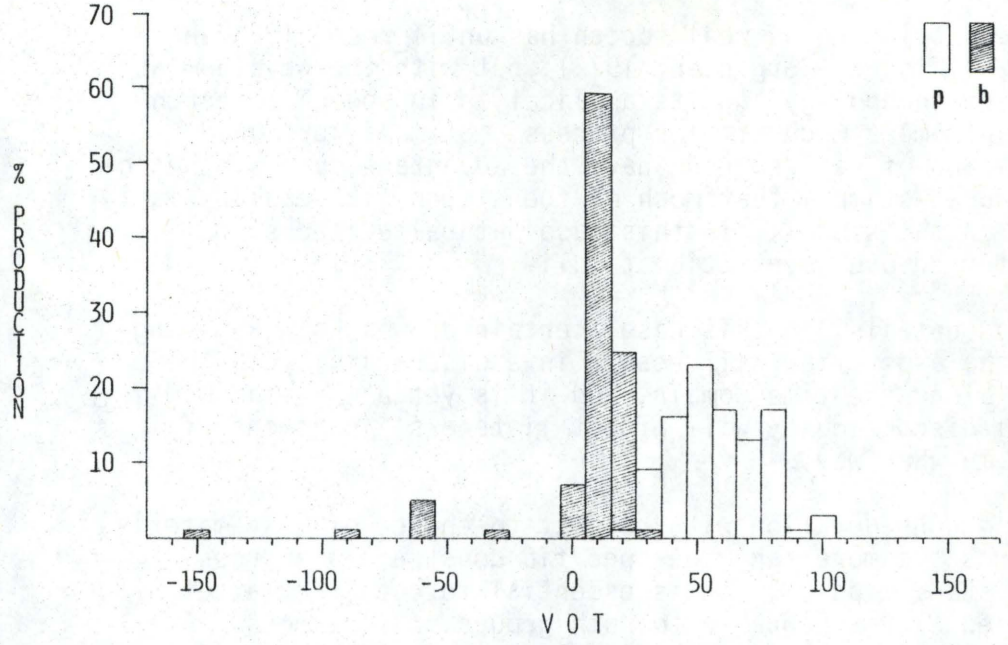


Figure 2a

American /b/ and /p/ followed by 12 vowels. From Keating et al. (1981). Permission by Patricia Keating.

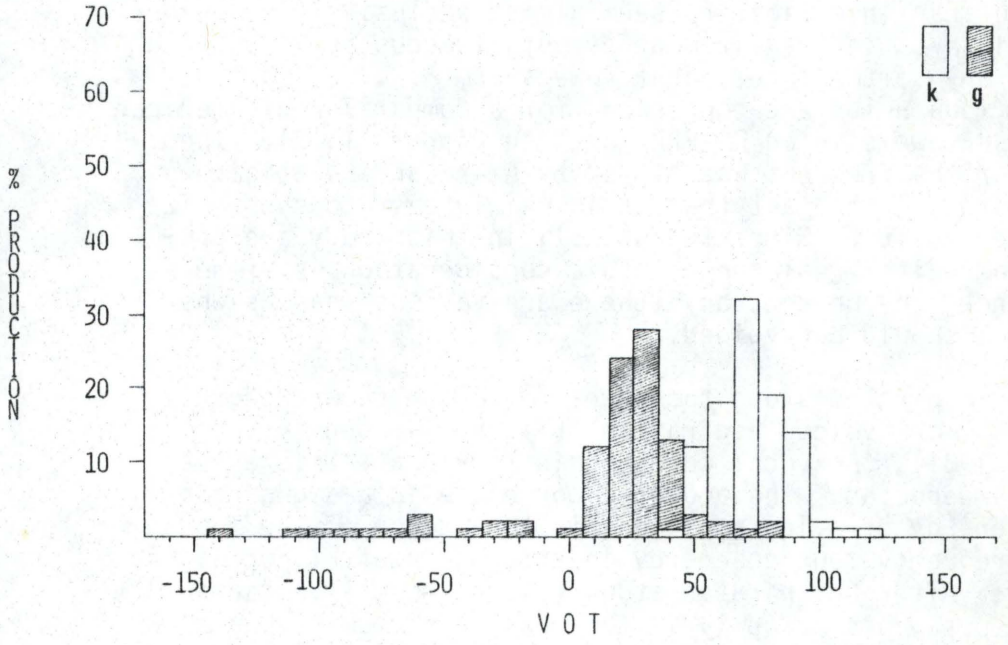


Figure 2b

American /g/ and /k/ followed by 12 vowels. From Keating et al. (1981). Permission by Patricia Keating.

Editing and splicing of real speech has until recently been cumbersome (Fischer-Jørgensen, 1972), but with the development of computer technology and its application in speech research this has become a much simpler process. Stimuli, produced from splicing of real speech have the advantage over synthetically produced stimuli that much of the speech-like quality is retained in the signal. In this case natural edited stimuli were preferred over synthetic stimuli.

The method entails, in this case, certain drawbacks. Removing part of the aspiration will result in a discontinuity in the spectrum along the time domain, and it is yet a question how much this discontinuity will affect listeners' judgement of VOT, and in what way.

The second consideration with respect to choice of test material concerns the more language-specific constraints a cross-language study imposes. It is essential to construct a set of stimuli equally acceptable to both groups of listeners. As mentioned earlier, the Danish alveolar stops differ significantly from the American ones in that the release of this type of stop in Danish is accompanied by a considerable degree of affrication, varying from greatest in most urban dialects to being absent in the rural dialects of especially the west coast of Jutland. The test material therefore includes only labial and velar stops.

The vowels following the consonant require some consideration too. American investigators seem almost exclusively to favour the vowel /*a*/. The requirement of mutual acceptability would hold for any of the three point vowels /*i*, *a*, *u*/ as far as formant frequencies are concerned (for a comparison of American and Danish vowels in an F1-F2 plot, see Disner, 1980). The reason for the frequent use of /*a*/ by American investigators may be found in the fact that it is fairly easy to synthesize with good results. Since the stimuli in this study are produced from edited real speech, this consideration is of no consequence; on the contrary, there are various reasons why this vowel should be avoided.

Following aspirated stops the vowel /*a*/ often shows "edge vibrations" or "voiced aspiration" (Fischer-Jørgensen and Hutters, 1981), the vocal cords begin to vibrate before they are fully adducted. The acoustic correlate is a signal containing energy of a low amplitude in the region of the fundamental frequency, but no energy in the region of F1, or higher formants. After unaspirated stops voicing normally starts simultaneously and abruptly in the whole spectrum. This also seems to be the case for the high vowels /*i*/ and /*u*/ when following aspirated stops. Using /*a*/ might therefore increase the risk that editing aspirated stimuli might produce unaspirated ones with a vowel onset characteristic of a vowel following aspirated stops. This would consequently result in stimuli containing possibly conflicting cues.

By choosing /*i*/ and /*u*/ we get a further advantage in the bargain. One of the arguments against the salience of VOT as a perceptual cue has been that listeners include F1 transitions in their judgements (Stevens and Klatt, 1974). Since both these vowels have the lowest F1 of all vowels, we will only get a slight change of this formant during the transition from stop to vowel, and thus minimize the spectral discontinuity in this range.

The stimuli were therefore constructed from real tokens of /*p*/ and /*k*/ followed by the vowels /*i*/ and /*u*/.

B. STIMULI

From the recordings of the two Danish male speakers, JR and NR one token of each of the test syllables [b̥ʰi, b̥ʰu, g̊ʰɪ, g̊ʰu] was selected, viz. the first occurrence on the tape which had been measured to have a VOT value greater than 90 msec. Some of speaker JR's tokens did not quite show VOT values of this duration, and in those cases the token with the highest measured value was then used. The original VOT values thus ranged between 80 msec and 136 msec (JR's [b̥ʰu] and NR's [g̊ʰu]).

The entire CV syllable was stored in the buffer of a PDP-8 Digital Equipment Corporation computer and the point of voice onset identified on a digitized oscillogram. The part of the syllable containing glottal periodicity was stored in a separate file. In the process of isolating the voiced portion of the syllable (i.e. the vowel) care was taken to ensure a cut at a zero-crossing in order to avoid offset clicks. The cursor was then positioned 70 msec from the point of consonantal release, and the remaining part of the aspiration was deleted from the buffer. A copy of the file containing the voiced portion could then be read down into the buffer and added to the 70 msec of burst + aspiration. This procedure was repeated, each time shortening the aspiration from right to left in steps of 10 msec, resulting in seven stimuli ranging in length of aspiration from 10 - 70 msec.

This was done with the tokens of both speakers, giving a total of 28 labials and 28 velars. The stimuli were recorded on magnetic tape, using a REVOX A77 open reel tape deck for subsequent tape generation for the listening test.

C. PRODUCTION OF TAPE FOR THE LISTENING TEST

The stimuli were transferred from the magnetic tape and stored in separate files on the disk of a PDP-11/34 Digital Equipment Corporation computer, using a sampling rate of 10.000 Hz and a high-pass filtering of 80 Hz. This sampling rate is, in effect, a low-pass filtering at 5.000 Hz, and it was not assumed that acoustic information above 5.000 Hz would in any way be crucial to the listeners in this experiment.

Four series were generated, the first and second series containing the labials and velars of speaker JR, respectively, the third and fourth series containing the same tokens spoken by NR. In each series the stimuli were duplicated ten times each and recorded on magnetic tape in a randomized order in blocks of 10 with an inter-stimulus interval of 2.5 seconds and an inter-block interval of 4 seconds. They were all recorded on the same tape in the order mentioned above, with a pause of approximately two minutes between series.

D. THE LISTENING TEST

The tape was presented to the subjects for labelling B or P and G or K. Immediately before the test the subjects were given a set of written instructions, explaining to them the inter-stimulus and inter-block intervals. They were also instructed not necessarily to expect an equal number of B's and P's or G's and K's. Orally they were asked not to introduce a third category, but to guess in those cases where they felt unable to decide.

The tape was presented binaurally over headphones. All the subjects took the test individually and had available to them a volume control. The total duration of the test was 28 minutes.

E. SUBJECTS

A total of 16 subjects participated in the experiment, 8 Danish and 8 American listeners. A further subdivision can be made of the two groups, into phonetically trained and phonetically naive listeners, 3 and 5 respectively in the two groups. The criteria according to which the subjects were classified as phonetically trained must almost inevitably differ for the Danish and the American group, for obvious reasons.

The Danish listeners were considered to be phonetically trained if they had taken an intensive three-semester course in ear-training and narrow phonetic transcription at the Institute of Phonetics. In this course, among other things, the students are trained to distinguish between degrees of aspiration and voicing during occlusion, and it could therefore be expected that this group would be more consistent in their responses, i.e. have sharper category boundaries. The rest of the Danish subjects were all students at various levels at the University of Copenhagen.

The American subjects were all graduate students in the linguistics department at Brown University, USA, except one, who was an undergraduate majoring in linguistics. Part of the graduate program in linguistics at Brown University is a one-semester course in ear-training and phonetic transcription (Linguistics 0121). But since it also includes an introduction to basic concepts in phonology, it is not comparable to the intensive course the Danish students take.

Three of the American subjects also worked as research assistants in the linguistics department's speech laboratory. Their duties as such included phonetic transcription of natural discourse between young children and their parents, and it could therefore be expected that their skills in phonetic transcription were beyond the ear-training course. However, the transcription they were required to use is not as narrow as the transcription required by the Danish students. Therefore, in the strictest sense, the American and the Danish phonetically trained groups are not comparable (for the role of formal training in vowel transcription, see Laver, 1965).

All Danish and American subjects were unpaid volunteers. None of the Danish naive listeners had ever participated in perception tests before, whereas all the American listeners had participated in - and were familiar with - such tests. No subject had a known history of hearing loss.

IV. RESULTS AND DISCUSSION

Initial calculations did not reveal any consistent difference between stimuli recorded by the two speakers with respect to cross-over point (i.e. that cross-over points occurred later for one speaker than for the other). Identification results for stimuli from the two speakers were therefore pooled.

A. EFFECTS OF PHONETIC TRAINING

In order to determine whether to treat phonetically trained and naive listeners as one group in the final cross-language comparison, the cross-over points were found for each of these groups. This was done by converting the identification scores for each stimulus to z-scores which could then be fitted to a straight line, using the method of "least sum of squares" (Gilford 1954:123f). The results for the American listeners are shown in Table III.

Table III

Means for identification functions of American phonetically trained and naive listeners.

	American					
	phonetically trained			naive		
stimuli	$\bar{X}$	sd.	N	$\bar{X}$	sd.	N
pi	17.98	8.77	60	19.06	6.55	100
pu	28.85	11.46	60	25.21	7.34	100
ki	61.10	13.19	60	45.59	11.81	100
ku	47.19	12.70	60	42.70	9.45	100

It is seen clearly that the American phonetically trained listeners show later cross-over boundaries than the naive listeners for three out of the four series of stimuli. However, a criterion was set up to determine whether the phonetically trained group should be included: The groups were considered to behave differently if, and only if all of the four cross-over points were found to occur later for one group than for the other, regardless of whether the individual results were statistically significant or not. This did not apply to the American group since *pi* showed an earlier cross-over point for the phonetically trained listeners, whereas for the rest it occurred later. They were therefore included in the final cross-language comparison.

In the case of the Danish listeners, there seems to be a clear tendency for the phoneticians to require shorter VOT before their percept begins to change from one phoneme category to the other. As shown in Table IV, their cross-over points occur earlier for all four series of stimuli than for the naive listeners. A non-directional t-test gave $p < .001$ for all, with the exception of *pi* which proved to be non-significant. The group of Danish phonetically trained listeners were therefore omitted for the cross-language comparison.

Table IV

Means for identification functions of Danish phonetically trained and naive listeners.

	Danish					
	phonetically trained			naive		
stimuli	$\bar{X}$	sd.	N	$\bar{X}$	sd.	N
pi	21.30	7.82	60	22.59	5.84	100
pu	32.42	5.86	60	39.72	7.72	100
ki	47.82	10.28	60	53.37	9.58	100
ku	36.06	10.10	60	45.66	7.83	100

It may be argued that the reasons for excluding the Danish but including the American phonetically trained listeners may be somewhat vague or arbitrary. One alternative would be to exclude both groups (rather than including the Danish listeners as well). However, the number of subjects in this study is small anyway, and including the American phonetically trained listeners only moves the cross-over points of the total group towards that expected for the Danish listeners. Consequently, the results influence the cross-language comparison in such a way that the expected difference between Danish and American listeners is reduced. The reason why phonetically trained persons were used as subjects in the first place in addition to

naive listeners was that I was interested in looking into the effects of formal phonetic training.

When comparing the results obtained from the groups of Danish listeners it is interesting to see that the expected effects were not found, viz. that trained phoneticians would show sharper category boundaries than naive listeners.

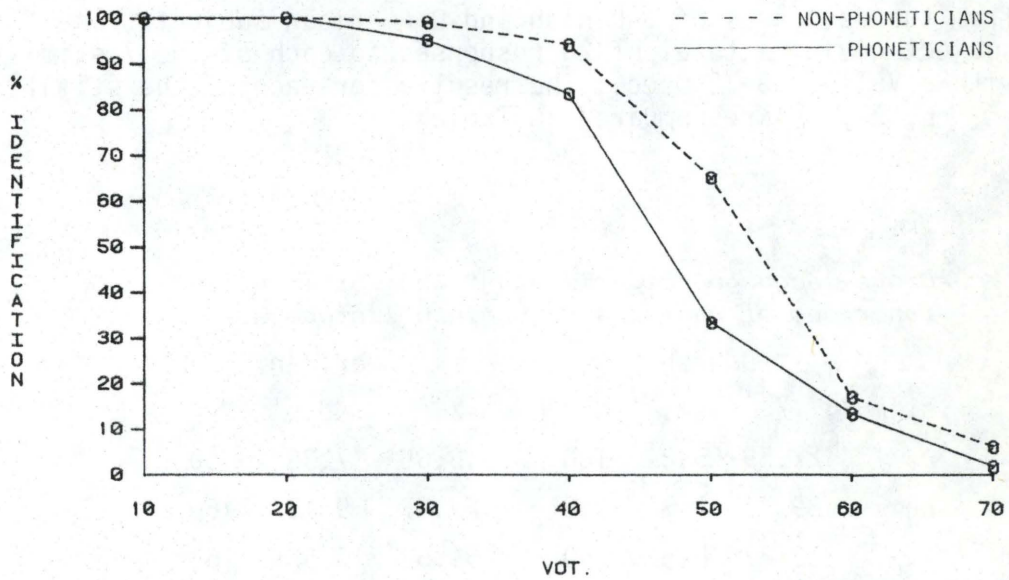


Figure 3

Identification functions for phonetically trained and naive listeners of the ki stimuli.

In Figure 3 the percentages of responses from the phonetically trained and naive listeners are plotted against the stimuli of *ki*. It is clear from this figure, as was seen in Table IV, that the cross-over point occurs much earlier for the phonetically trained listeners, which must be ascribed to the effect of formal training. What it does not tell us is how the same phonetically trained listeners would perceive VOT in normal discourse. One possible hypothesis may be that their category boundary would be identical to that of the naive listeners, and that the nature of a perception test will trigger a special "phonetic mode" of listening in the trained group of subjects.

If the identification functions for the two groups are compared with the standard deviations given in Table IV, it is interesting to note, since the standard deviation is reflected in the steepness of the curve (the slope), that the phonetically trained listeners do not show sharper category boundaries than those found for the naive listeners. This is in a sense surprising since it could be expected that phonetic training, among other things, would teach the listeners to "latch on" to one specific acoustic variable and in that way be able to identify the individual stimuli more consistently.

B. CROSS-LANGUAGE COMPARISON

Since the Danish phonetically trained listeners were omitted from the study, the final cross-language comparison was made from the responses of 5 Danish and 8 American subjects, each subject giving a total of 20 responses to each of the 7 stimuli with a VOT of 10-70 msec. The results for each of the stimuli *pi*, *pu*, *ki*, *ku* are compared in Table V.

Table V

Cross-language comparison. Means for identification functions of Danish and American listeners.

	Danish			American		
stimuli	$\bar{X}$	sd.	N	$\bar{X}$	sd.	N
pi	22.59	5.84	100	18.65	7.38	160
pu	39.72	7.72	100	27.67	9.15	160
ki	53.37	9.58	100	51.53	13.55	160
ku	45.66	7.83	100	45.49	10.48	160

Looking first at the means, i.e. cross-over points given in Table V, it is clear that the perceptual shift from one phoneme to the other takes place earlier for the American listeners, as was expected according to the original hypothesis. If the boundary between phonemes in the American production data is estimated by visual inspection of the histograms, figure 2a and 2b, it will be - for the labials - approximately 20-30 msec and for the velars approximately 40-45 msec. The corresponding phoneme boundaries would be, for the Danish data, figure 1a and 1b, 40-45 for the labials and 55-65 msec for the velars. For both places of articulation the Danish perceptual cross-over boundary would therefore be expected to occur much later.

For the labial place of articulation this is quite clearly confirmed in Table V. Although the Danish cross-over point occurs much earlier than predicted from the production data, it still differs from the American one as expected: a t-test gives $t = 4.528$ $p < .001$ for *pi*, and for *pu* $t = 10.952$ $p < .001$ when compared to the American cross-over points. In figure 4a and 4b

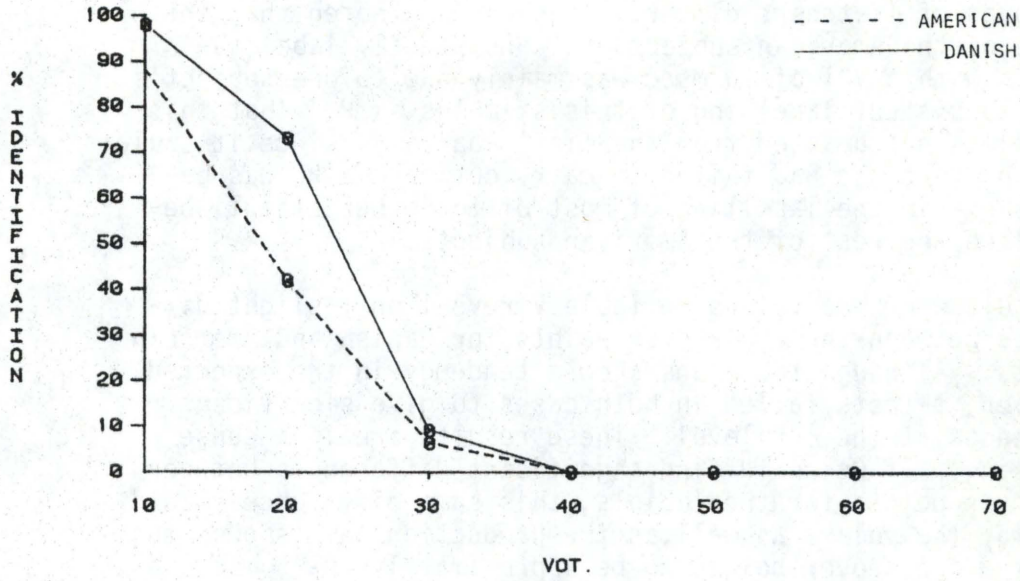


Figure 4a

Identification functions for American and Danish listeners of the pi stimuli.

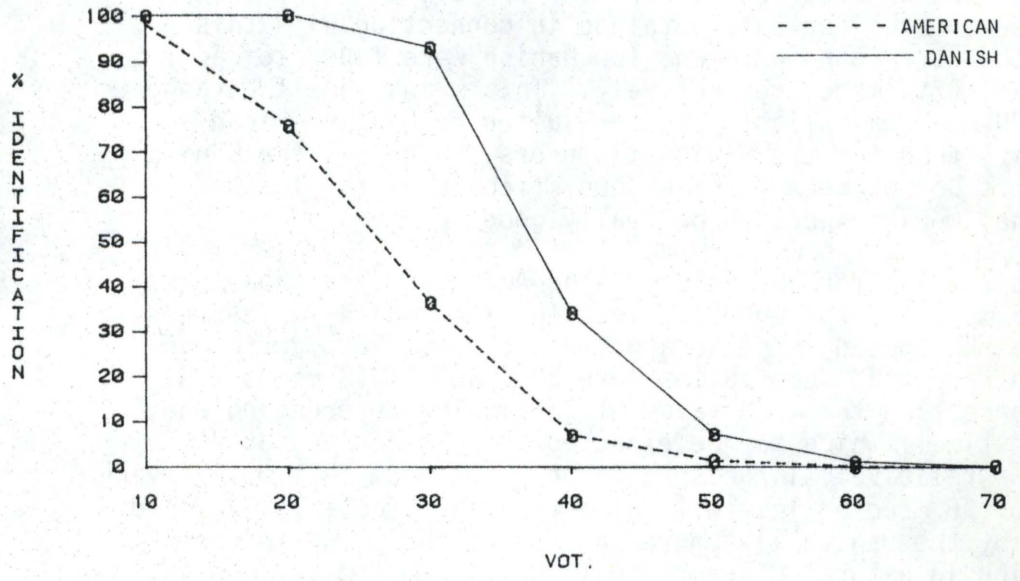


Figure 4b

Identification functions for American and Danish listeners of the pu stimuli.

the percentages of b-responses for the individual stimuli are plotted, and it is perhaps clearer from this figure how the two groups of listeners differ. It should be noted that the failure of the American subjects to consistently label the *pi* stimulus with a VOT of 10 msec was mainly due to one subject's almost consistent labelling of this stimulus as P. That this subject was not omitted from the final analysis, since it could be argued that she had failed to carry out the task, was because she - in the labelling of most of the other stimuli behaved like the rest of the American subjects.

The results for the velars in Table V reveal only slight differences between the cross-over points for Danish and American subjects. Although the means show a tendency in the expected direction, t-tests failed in both cases to give significant differences at the .05 level. These results are in a sense surprising. If we could find significant differences between cross-over points for the labials, this same difference should exist for the velars as well,as the production data showed the estimated cross-over points to be approximately 40-45 msec for the Americans, and 55-65 msec for the Danish data.

The question of range effects and listeners' strategies may help explain the failure to show any significant difference of cross-over points. That category boundaries may be shifted around due to range effects has been shown for English by e.g. Brady and Darwin (1978), but the shifts they found were relatively small, on the order of 5-7 msec. This led to a re-examination of the ranges employed in this experiment to see if they were in any way inappropriate for one or both groups of listeners. The computed means, on the basis of results from Keating et al. (1981), were for */b/* 5.6 msec and for */p/* 57.6 msec. From the data obtained in connection with this study the corresponding means for Danish were found to be 14.9 and 102.4 msec, respectively. This means that the range of 10-70 msec was slightly better suited for the American listeners than for the Danish listeners. However, the Danish listeners do not seem to have been affected by the lack of what they would expect to be really good */p/*'s.

It is different for the velars. The American data show means of 15.5 msec for */g/* and 71.7 msec for */k/*, which again seems to make the chosen range an appropriate one. For Danish, on the contrary, the same values were 29.4 and 110.8 msec. In this connection the mean value of */g/* is the interesting one, as the stimuli which may be expected to be perceived as */g/* occupy a relatively large span in the continuum. Stimuli having a VOT of 30 msec or less will almost without doubt be labelled as */g/* by the Danish listeners, and since they, due to normal variation in natural discourse will hear intended */g/*'s with longer VOT than the 30 msec, they are more likely to label them as G than American listeners. The 10-70 msec range must therefore be said to be less appropriate for Danish listeners.

In figure 5a and 5b are shown the identification functions for *ki* and *ku*. It is interesting that the Danish listeners con-

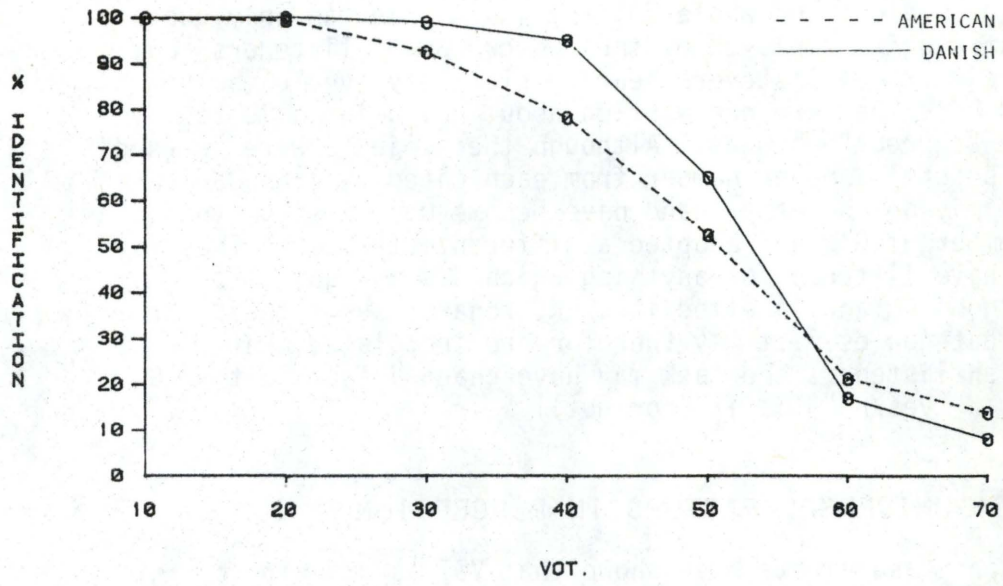


Figure 5a

Identification functions for American and Danish listeners of the ki stimuli.

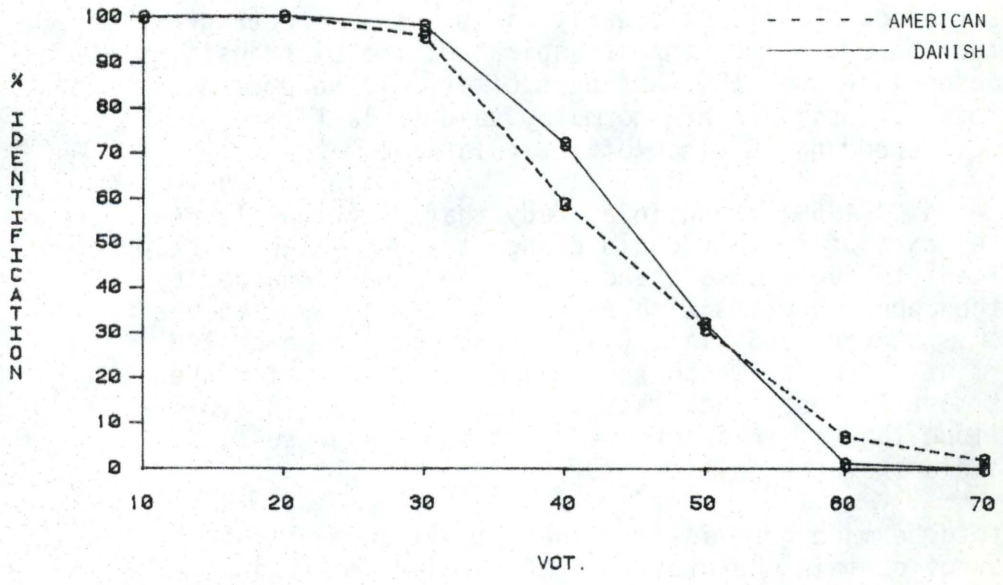


Figure 5b

Identification functions for American and Danish listeners of the ku stimuli.

sistently label stimuli, especially *ki* (figure 5a) of 10-40 msec, as G's. The Americans, on the other hand, show a more gradual drop along the whole VOT continuum. We can only guess at the strategies employed by the two groups of listeners, but since the American listeners hear a relatively even number of *gi*'s and *ki*'s they are not worried about not being able to give the "correct" answers. Although the subjects were "warned" about a possibly uneven number from each category, the Danish subjects may on the other hand have become uneasy about the large number of G's and adopted a different strategy. They may in fact have listened for anything which did not quite sound like a "good" G and labelled it as K, regardless of their normal phoneme categories. It may therefore be speculated that for the Danish listeners the task may have changed from "either G or K" to a "yes/no" (is it G or not).

C. UNIVERSAL FACTORS IN PERCEPTION

Production measurements have shown that VOT is affected by place of articulation (Lisker and Abramson, 1964, Fischer-Jørgensen, 1980). VOT has been found generally to be longer following the release of a velar than a labial stop. The explanation given to this phenomenon has been that the movement of the tongue body away from the passive articulator is slower, resulting in a delayed drop in oral air pressure relative to the release, which is necessary for vocal fold vibration to begin.

The acoustic correlate is consequently slower formant movements during the transition from the stop to the vowel. If glottal pulsing starts immediately after the release (i.e. short lag), the transition of F1 will be clearly visible on a spectrogram. On the other hand, if the stop is aspirated, the F1 transition will be absent, or only the last part of it will be present. These formant transitions are normally said to last for about 40-80 msec, depending on place of articulation.

Liberman et al. (1958) found in a study that by gradually delaying the start of F1 they could change the percept of a stop from "voiced" to "voiceless", and that effect was enhanced by filling the upper formants with noise simultaneously with the F1 cutback. Stevens and Klatt (1974) proposed the so-called First Formant Detector, which they claim can account for the fact that most listeners hesitate to identify stops as "voiceless" or long lag, if this formant transition is present, almost regardless of the physical VOT.

The point to be made here is that the stimuli used in this study did not contain measurable first formant transitions at all. This was checked on spectrograms of [g^hi], recorded by both speakers. Both the Danish and American listeners required longer VOT for the velar stimuli, before their percept changed from /*g*/ to /*k*/. This is shown in figure 6 where the responses for the two groups are compared for place of articulation.

The results therefore suggest that the perception of VOT is influenced by place of articulation, even though the stimuli have been stripped of the cues which are normally used to explain the later cross-over boundary for velars. Listeners must therefore have some expectation about the longer VOT for velars which is quite consistently reflected in the responses of both Danish and American listeners.

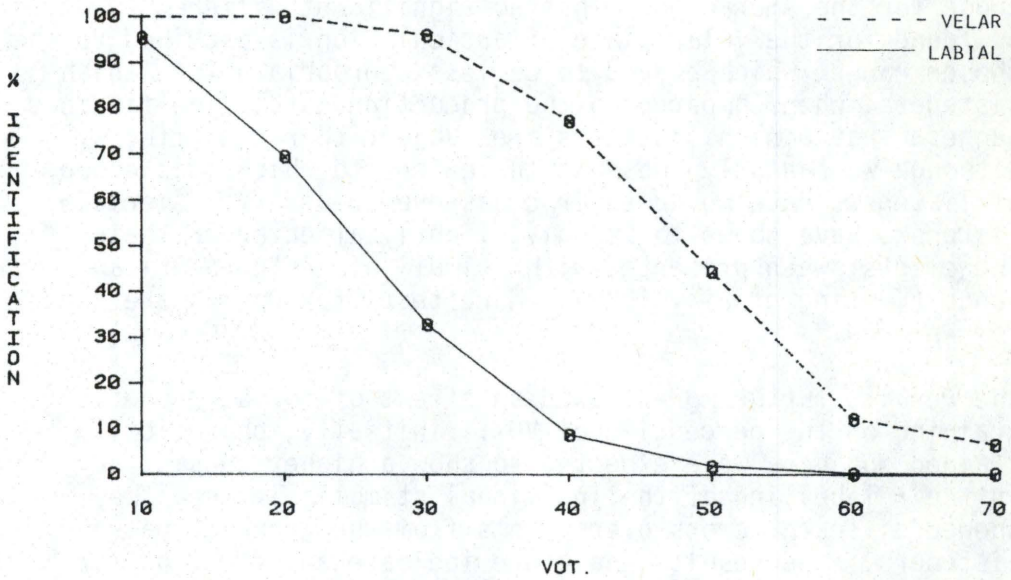


Figure 6

Identification function for Danish and American listeners, describing universal difference in the perception of VOT for place of articulation.

That this is not only the case for place of articulation of the stop consonant, but also for rounding of the vowel following the stop, may be seen from Table V. In her study of Danish production Fischer-Jørgensen (1980) found that rounded vowels following labial stops had a longer "open interval" (which she defines as the interval from the release of the consonant to the onset of the vowel, point B according to Fischer-Jørgensen and Hutters, 1981). Since the tongue is already in position for the vowel at the release of a labial stop, the place of oral constriction for the vowel cannot affect the pressure drop in the oral cavity, and the longer VOT in /*bu*/ and /*pu*/ must therefore be due to the coarticulation of the vowel. The cross-over points (in Table V) for labials followed by /*i*/ were compared to those followed by /*u*/ for the Danish and the American listeners in a t-test. Both groups were found to have later cross-over points for *pu*, $p < .001$, which again indicates that listeners expect rounded vowels to delay the onset of voicing

and therefore use this knowledge in their responses, even though the stimuli did not contain information which would effect a later cross-over boundary.

V. CONCLUSION

The results found in this pilot study confirm the original hypothesis in showing a clear correspondence between the production and perception of VOT. It has been shown that the cross-over points for Danish listeners did, on the whole, occur later than those for the Americans. That no significant difference could be found for the velar place of articulation is ascribed to the chosen range, which proved to be less appropriate for Danish listeners, when compared to the production data. The findings suggest that Danish listeners are subject to range effects, although we can only guess at the extent to which this group of listeners have moved their cross-over boundary. American listeners have shown to be only slightly affected in their judgements, when presented with stimuli of an "un-American" range (Keating et al., 1981). More tests for Danish are therefore needed.

The results furthermore showed an effect of formal phonetic training on the perception of VOT. Initially, phonetically trained subjects were expected to show a higher consistency in their labelling of the individual stimuli; instead they showed different cross-over points from the group of naive listeners. The results therefore indicate an effect of formal training which is different from what was expected. This calls for further experiments of phoneticians' perception of VOT since in other studies they have been shown not to behave differently when compared to non-phoneticians.

Differences in VOT which are physiologically and aero-dynamically determined (i.e. the differences found for labial and velar stops) are quite clearly reflected in the responses both by Danish and American listeners, even if the stimuli contained none of the cues (vowel formant transitions, etc.) that are normally used to explain these perceptual differences. The results therefore support the hypothesis that VOT and place of articulation are not processed as two independent cues (see e.g. Sawusch and Pisoni, 1974). In this experiment place of articulation was given on the answer sheet and the listeners may have assigned different "expected" VOT values accordingly.

ACKNOWLEDGEMENTS

This paper is a revised version of a thesis submitted in partial fulfilment of the requirements for the degree of Exam. art. in the Institute of Phonetics at the University of Copenhagen. I would like to thank Niels Reinholt Petersen who served as my advisor, and whose helpful suggestions have been of invaluable help to me, Jørgen Rischel for helpful comments, and the staff for moral support. Special thanks to Philip Lieberman and

Sheila Blumstein for the use of computer facilities in the linguistics laboratory at Brown University, USA, and to Christian Boysen, Tectronics, and Carsten Henriksen for their help with the graphs.

REFERENCES

Abramson, Arthur S. and Lisker, Leigh 1973: "Voice time perception in Spanish word-initial stops", *J. Phonetics 1*, p. 1-8

Abramson, Arthur S. and Lisker, Leigh 1983: "Relative power of cues: F0 shifts versus voice onset time", *Abstracts 10th Int. Congr. Phon. Sc.*, p. 505

Brady, S.A. and Darwin, C.J. 1978: "A range effect in the perception of voicing", *J. Acoust. Soc. Am. 63*, p. 1556-1558

Disner, Sandra F. 1980: "Evaluation of vowel normalization procedures", *J. Acoust. Soc. Am. 67/1*, p. 253-261

Eimas, P.D., Siqueland, E., Jusczyk, P. and Vigorito, J. 1971: "Speech perception in infants", *Science 171*, p. 303-306

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

Fischer-Jørgensen, E. 1980: "Temporal relations in Danish tautosyllabic CV sequences with stop consonants", *Ann. Rep. Inst. Phon. Univ. Cph. 14*, p. 207-261

Fischer-Jørgensen, E. and Hutters, B. 1981: "Unaspirated stop consonants before low vowels, a problem of delimitation - its causes and consequences", *Ann. Rep. Inst. Phon. Univ. Cph. 15*, p. 77-102

Fujimura, O. and Miller, J.E. 1979: "Mandible height and syllable-final tenseness", *Phonetica 36*, p. 263-272

Gilford, J.P. 1954: *Psychometric Methods*, 2nd ed. (McGraw-Hill)

Haggard, M.S., Ambler, S. and Callow, M. 1970: "Pitch as a voicing cue", *J. Acoust. Soc. Am. 47*, p. 613-617

Keating, P.A., Mikós, M.J. and Ganong III., W.F. 1981: "A cross-language study of voice onset time in the perception of initial stop voicing", *J. Acoust. Soc. Am. 70*, p. 1261-1271

Klatt, D.H. 1975: "Voice Onset Time, frication and aspiration in word-initial consonant clusters", *J. Speech and Hearing Research 18/4*, p. 686-706

Kohler, K. 1983: "Phonetic explanation in phonology", *Abstracts 10th Int. Congr. Phon. Sc.*, p. 275-283

Kuhl, P.A. and Miller, J.D. 1978: "Speech perception by the Chinchilla: Identification for synthetic VOT stimuli", *J. Acoust. Soc. Am. 63*, p. 905-917

Laver, J. 1965: "Variability in vowel perception", *Language and Speech 8*, p. 95-121

Liberman, A.M., Delattre, P.C. and Cooper, F.S. 1958: "Some cues for the distinction between voiced and unvoiced stops in initial position", *Language and Speech 1*, p. 153-166

Lisker, L. 1975: "Is it VOT or a first-formant detector?", *J. Acoust. Soc. Am. 57,6, part II.*

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. 1968 (1965): "Some effects of context on voice onset time in English stops", *Language and Speech 10,1*, p. 1-28

Mikős, M.J., Keating, P.A. and Moslin, B.J. 1976: "The perception of voice onset time in Polish", *J. Acoust. Soc. Am. Suppl. I*, p. S19

Sawusch, J.R. and Pisoni, D.B. 1974: "On the identification of place and voicing features in synthetic stop consonants", *J. Phonetics 2/3*, p. 181-194

Stevens, K.N. and Klatt, D. 1974: "Role of formant transitions in the voiced-voiceless distinction for stops", *J. Acoust. Soc. Am. 56/3*, p. 653-659

Williams, L. 1977: "The voicing contrast in Spanish", *J. Phonetics 5*, p. 169-184

Winitz, H., LaRiviera, C., and Herriman, E. 1974: "Variations in VOT for English stops", *J. Phonetics 3*, p. 41-52.

ARIPUC 18, 1984, p. 185-242

INTONATION AND TEXT IN STANDARD DANISH

NINA GRØNNUM THORSEN

Acoustic analysis of recordings by four Standard Danish speakers shows that each declarative sentence in a text is associated with its own declining intonation contour, but together two or three such contours describe an overall falling slope. Individual sentence intonation contours are steeper and demonstrate greater amounts of resetting between them in a succession of declarative terminal sentences than in a corresponding string of coordinate main clauses. In other words, the closer relation between coordinate structures is reflected in a more coherent or less segregated intonational structure. The results are compared with other languages, and the implications for the abstract representation of Danish intonation are discussed.

I. INTRODUCTION

Lehiste (1975, p. 195) hypothesizes

"*... that paragraphs possess a suprasegmental structure that indicates the beginning and end of paragraphs and characterize the body of the paragraph. For example, it may be that the intonation contour applied to a sentence produced in isolation (constituting a one-sentence paragraph) will differ from the intonation contour applied to the same sentence in the beginning, middle and end of a paragraph. In other words, it is possible that a paragraph is characterized by an overall intonation structure to which the intonation contours of its constituent sentences are subordinated. ... A corollary hypothesis is that since speakers and listeners share the same code, listeners are capable of deciding whether a sentence has been produced in isolation or as part of a larger structure.*"

Both hypotheses are verified in Lehiste's (1975) experiments. Acoustic analysis of one recording of paragraphs containing one, two, and three sentences (of greatly varying length) by one American English speaker, shows that an isolated sentence is longer than the same sentence in a paragraph context. Fundamental frequency (Fø) in the beginning of a sentence is lower in an isolated sentence than in a paragraph initial one, and higher in paragraph initial position than in medial and final positions. - In a listening experiment subjects were to judge what position a sentence - cut out from its paragraph context - had occupied in the paragraph. Their judgment was significantly better than chance, though not always in accordance with the speaker's intention.

Bruce's (1982) point of departure is an informal listening experiment where two sentences, produced in isolation and in two-sentence texts, are spliced together in various ways. Isolated sentences spliced together, and sentences from texts spliced together in reverse order sound odd, and neither combination constitute a perceptually coherent text. Acoustic (Fø) analysis of the original recordings shows that two sentences in a text are characterized by a continuous overall downdrift, though the sentence boundary is marked by a low, sentence final Fø minimum succeeded by a resetting to the first high Fø peak in the second sentence. This peak is lower than the corresponding Fø peak in the first sentence. Likewise, the first Fø peak is lower in a sentence produced in isolation than in the first sentence of two in a paragraph, whereas its last Fø minimum value is higher in isolation than in the second of two sentences in a paragraph. The difference between a single sentence and a combination of two seems to be signalled globally [i.e. over a temporal scope corresponding to the whole text, NT], except that each sentence in a text also has a final, very low Fø value. - These results are confirmed in acoustic analyses of a larger material with a Southern Swedish speaker. Three sentences (each containing two stress groups) are produced in isolation and in the nine possible two- and three-sentence combinations. They are separated by periods, but each begins with "*Å*" ('And'). The very first and very last Fø values (which occur in unstressed syllables) are low and constant across texts of different durations. Apart from that, local Fø maxima and minima are higher in the beginning of the text than at its end, and they are higher in the beginning of a longer text than in the beginning of a shorter text, whereas the final values are constant across texts of different length. Multi-sentence texts were produced with internal pause(s). Before each such pause Fø drops to the same low Fø value as found in absolute initial and final position (the "floor") and resets at the onset of the succeeding sentence to smoothly continue the overall downdrift of local maxima and minima. When sentences in different positions are compared, isolated sentences turn out to resemble text initial sentences more than text final ones. As Bruce notes (1982, p. 284) this corresponds to Lehiste's (1975) observation that listeners confuse isolated sentences with initial sentences more often than with final ones.

Cooper and Sorensen (1981, p. 85ff) analysed four different two-clause sentences (each clause containing four stressed words), among them a coordinate main clause construction. They found that individual clauses had separate declinations associated with them, but together these individual declinations describe an overall downdrift. They note specifically that the resetting from the lower value at the end of the first clause to the higher value of the beginning of the second clause is independent of any accompanying breathing pause. They hypothesize that resetting between clauses in a complex sentence is more likely to occur in the environment of longer clauses, slower rates of speech, and lower semantic relatedness between the two clauses (p. 97).

Uyeno et al. (1979) investigated Fø in declarative complex sentences in Japanese: a construction with an embedded relative clause, a coordinate main clause construction, as well as a sequence of two simple sentences in succession. Their data show that all three types are characterized by an overall declination with a local resetting at the onset of the second clause or sentence, except when the relative clause is utterance initial, in which case the utterance declines smoothly from the initial high rise. The resetting is greatest before a center embedded relative clause. The point relevant to the present investigation is that coordinate main clause constructions and successions of simple terminal sentences did not differ among themselves.

These results (especially those of Lehiste, 1975, and Bruce, 1982) are all in sharp contradistinction to Nakatani (1975) who reports that listeners, presented with stories read normally and stories spliced together from isolated sentences, had difficulty in distinguishing the stories as normal or spliced. He concludes that speech features observed in isolated sentences can be reasonably generalized to sentences in a coherent context.

The present study is to a large extent similar to those of Lehiste (1975) and Bruce (1982) to whom I owe the idea and general outline of my own procedures. Its aim was to establish how and to what extent sentences in a short text are coupled intonationally, and furthermore to see whether this presumed coupling is dependent upon the (syntactic) relation between the sentences in the text. I shall mainly be concerned here with fundamental frequency, though there is evidence from American English and Japanese that other factors such as length, laryngealization and pause duration are also involved in the production and perception of sentence and clause boundaries, cf. Lehiste (1975, 1979, 1980), Lehiste and Wang (1977), Cooper and Sorensen (1981), and Uyeno et al. (1981).

II. MATERIAL, SUBJECTS, AND PROCEDURES

A. MATERIAL

The base is the following three declarative sentences (stressed vowels are indicated orthographically with acute accents here):

(1) *Amánda skal afstéd på cámping.*

'Amanda is going away camping.'

(2) *Hendes mór skal på kúrsus i Týskland.*

'Her mother is taking a course in Germany.'

(3) *Hendes fár skal vándre i Lápland.*

'Her father is going to hike in Lapland.'

In isolation as well as combined in texts of two and three, these sentences could all be uttered in answer to a question about what Amanda and her family are going to do during the summer holidays.

The total number of possible texts that can be constructed from these three sentences amounts to fifteen, which doubles to thirty when two different boundary conditions are introduced. This was an unmanageably large number for subjects to record, if the material was to be mixed with filler utterances of a different syntactic and semantic make-up, which was highly desirable. I decided, therefore, to concentrate on (1) above, i.e. the *Amánda*-sentence, and be sure that one occurred in all the relevant positions.

The following combinations of sentences constitute the material:

1

1 3

2 1

1 3 2

2 1 3

3 2 1

The *Amánda*-sentence is the only one to occur in isolation. Note, however, that both sentence (2) (referred to as the *mór*-sentence in the following) and sentence (3) (referred to as the *fár*-sentence in the following) occur in all three positions in three-sentence texts.

Two different types of text result when the sentences in multi-sentence texts occur in sequences of declarative terminal main clauses versus coordinate main clause constructions, joined

by *og* ('and'). Examples are: *Amánda skal afstéd på cámping. Hendes fár skal vándre i Lápland. Hendes mór skal på kúrsus i Týskland.* versus *Hendes fár skal vándre i Lápland, og hendes mór skal på kúrsus i Týskland, og Amánda skal afstéd på cámping.*

The five terminal sentence texts and the isolated sentence were randomized three times, and the total of 18 sentences/texts were mixed with two materials recorded for different analysis purposes, so they were evenly spread over six pages of reading material. The same procedure was applied to the five coordinate clause texts and the isolated sentence. To avoid any direct comparison of the two types of text they were recorded in separate sessions at least one day apart. In this way the isolated sentence was recorded six times by each speaker, but the texts only three times each. This is a compromise between the demand for as comprehensive a material as possible, a sufficiently large number of recordings of each item by each speaker, and the wish to avoid fatigue effects on the part of the subjects. On a fair number of previous occasions (among them recordings of rather long simple sentences, cf. Thorsen 1983a) subjects' repetitions of the same item have always been rather remarkably constant, with standard deviations on calculated Fø mean values generally below 5% of the mean. I therefore decided that the demand for (the standard, minimum) six recordings of each item would have to yield to the possibility of presenting the texts under two different boundary conditions.

My material differs from Lehiste's (1975) texts ("paragraphs" in her terminology) in that the sentences are all equally long (in terms of the number of prosodic stress groups in each) and in the addition of coordinate main clause constructions. It differs from Bruce's (1982) material in the greater length of each sentence (three versus two prosodic stress groups). And again, where Bruce (1982) so to speak combined a terminal with a coordinate clause construction (by separation of the sentences with periods but commencing each sentence with 'and'), these two conditions are differentiated in the present material.

There is no conflict in Lehiste's (1975) choice of term ("paragraph") and mine (and Bruce's, 1982) ("text"). The present sentence combinations are both paragraphs and texts; at least if one does not claim that the total six pages of reading material constitutes a text. In that case my "texts" would only have the status of paragraphs. However, the six pages could lay no claim on any semantic or pragmatic coherence, and hardly deserve the designation as a text properly speaking.

B. SUBJECTS

Four phoneticians, two males (NRP and JBC) and two females (GB and NT (the author)) read the material, in two sessions on separate days, as mentioned above. They all speak a form

of Standard Copenhagen Danish. Their style of speech during the recordings can be characterized as fluent and conversational but distinct.

C. TECHNICAL PROCEDURES

The recordings were made with semi-professional equipment (Revox A-77 tape recorder, Sennheiser MD21 microphone) in a quasi-damped room at the Institute of Phonetics on Agfa PE39 tape, at 7½ i.p.s.

The tapes were processed by hardware intensity and fundamental frequency meters (F-J Electronics) and registered on a mingograph (Elema 800) at a paper speed of 100 mm/s. By adjustment of the Fø meter zero-line to the lower limit of the subject's voice range and full exploitation of the record space (about 80 mm) of the mingograph galvanometer, a measuring accuracy of 1 Hz for the males and 2 Hz for the female speakers is attained.

In unidirectional Fø movements in the stressed vowels, the beginning and end point were measured, according to a procedure outlined in Thorsen (1979, p. 63-66). In bidirectional movements, the turning point was measured as well. The unstressed vowels in the sentences were represented by a single Fø point, the midpoint (in time as well as frequency), which was an uncontroversial procedure since most of these vowels have monotonically falling Fø movements. (The first post-tonic syllable may be rising-falling, in which case Fø is measured at the peak.) The Fø movements are so short and slight in the unstressed syllables that it does not much matter which point you choose as the one measuring point: the initial, medial, or final (or any other) value. Changing the location of the measuring point will transpose all the unstressed vowels by very nearly the same (negligible) amount upwards or downwards, relative to the stressed vowels in the figures presented below. There is one instance where two unstressed vowels cannot be segmented. *skal af(stéd)* is pronounced without any intervocalic consonant [sg̊a a], and the measuring points are assigned time coordinates one quarter of the distance from the onset and offset of voicing in the long vowel sound. The distance in time of each measuring point from the first Fø value in the sentence or text was likewise measured. In the terminal sentence texts, time was set to zero at the first measuring point in each sentence, and the pause between sentences (defined as the time interval where no acoustic energy is registered) measured separately. In the coordinate main clause texts, time was measured cumulatively from the onset of the first sentence. Fø and time measurements were averaged over the three recordings by each subject (six recordings in the case of the isolated sentence). Average Fø values were converted to semitones (re 100 Hz) and average tracings drawn. No correction was attempted for intrinsic Fø level differences between the high stressed vowels of the *mór*-sentences and the low stressed vowels of the *fár*- and *Amánda*-sentences.

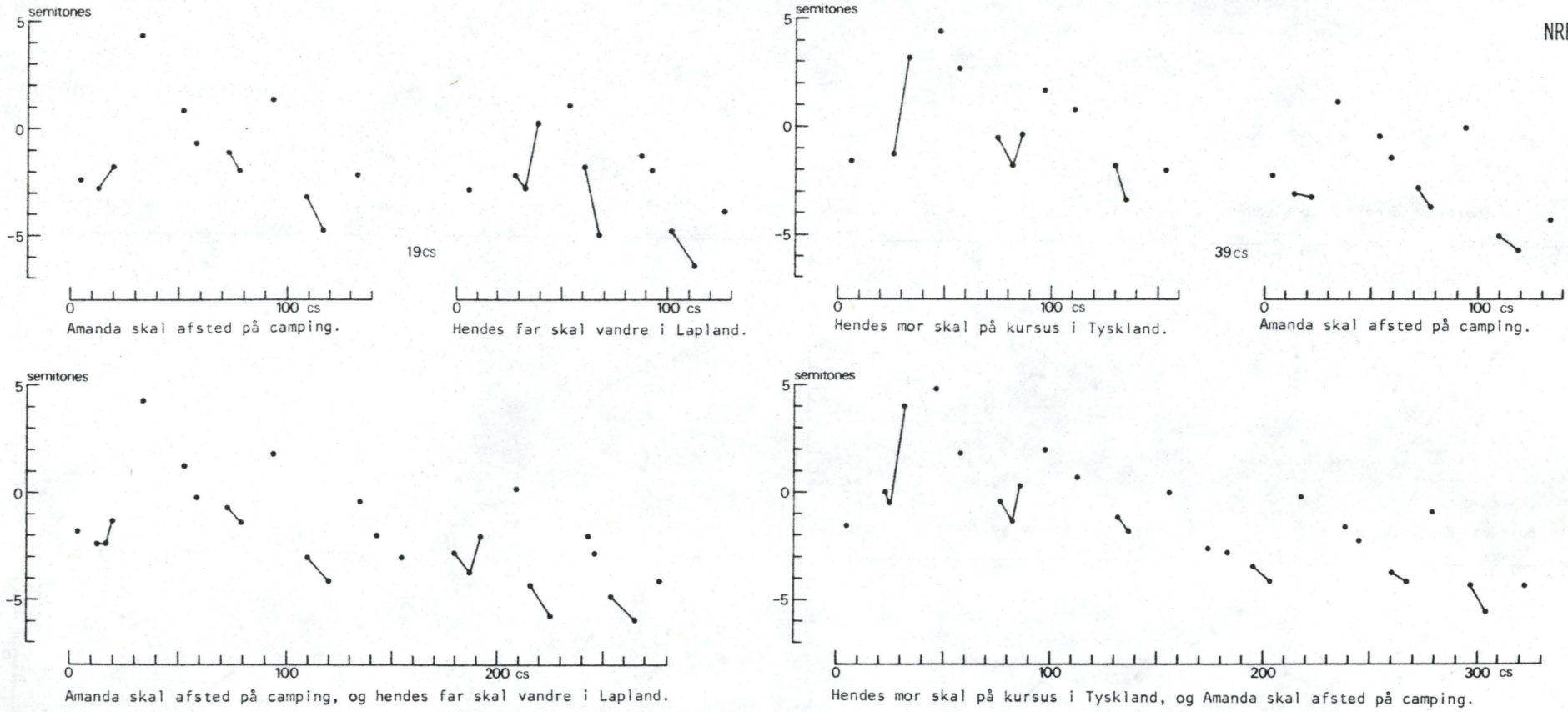
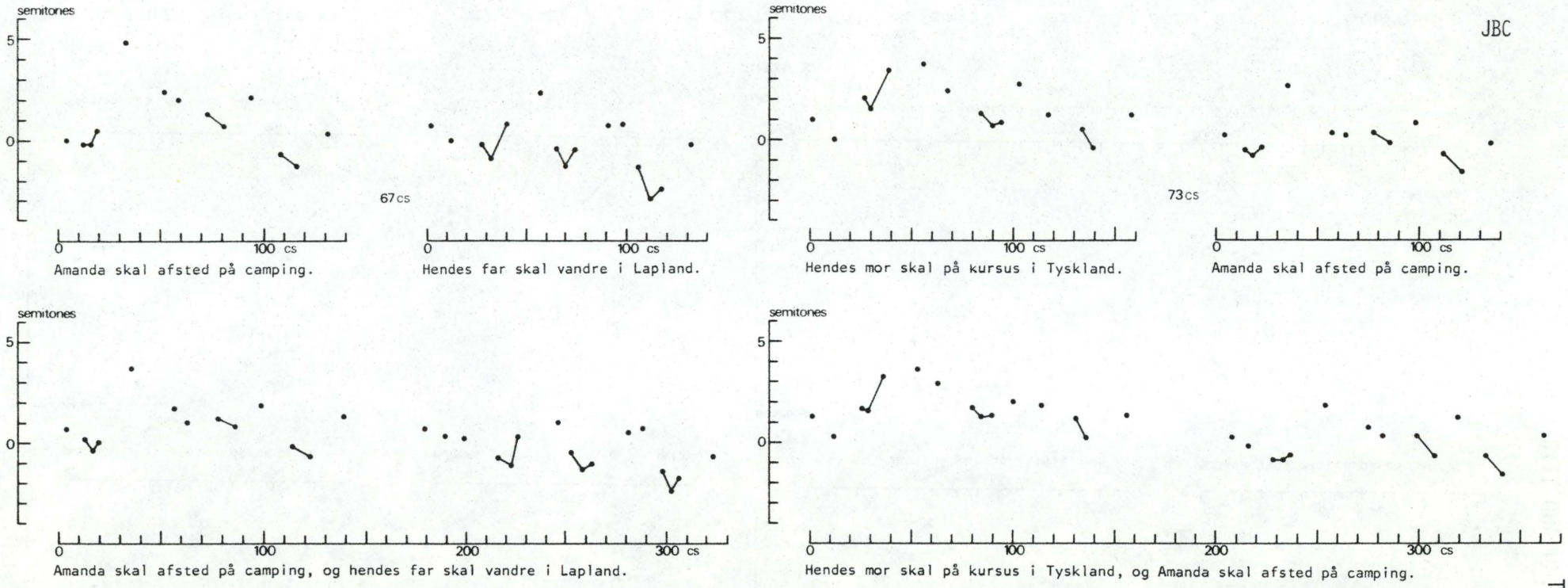
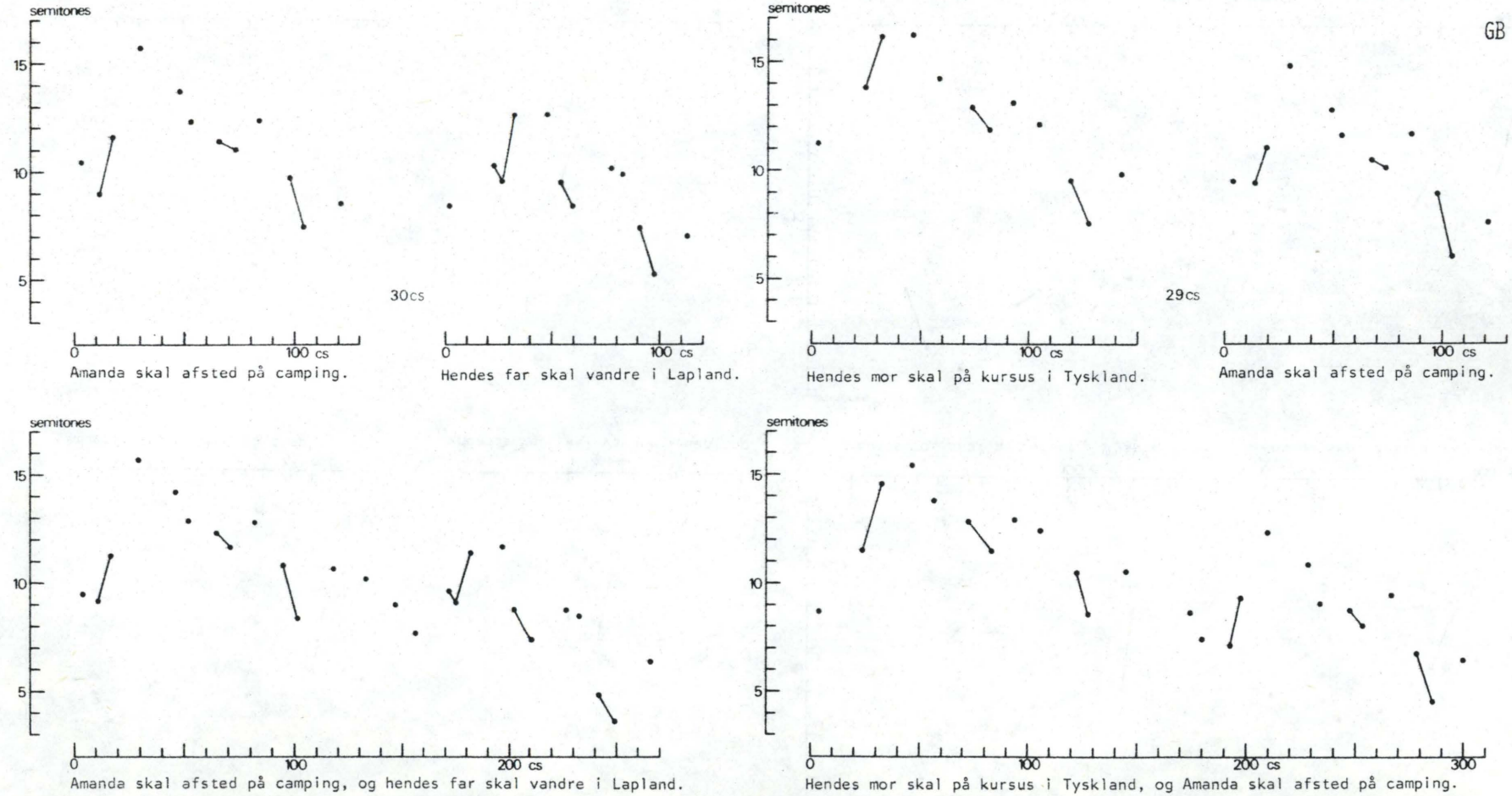


Figure 1

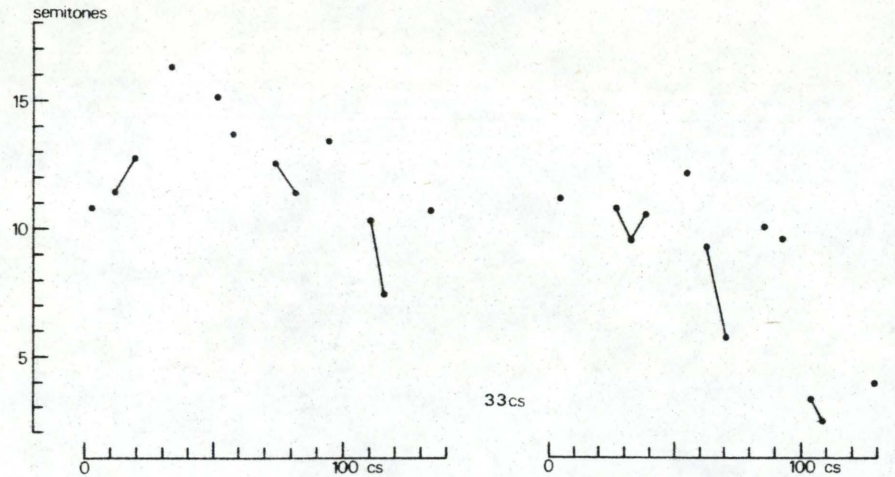
Fundamental frequency tracings (average over three recordings) of two texts, where the same sentence/clause occurs in text initial position (left side) and text final position (right hand side). Sequences of declarative terminal sentence above (with indication of the duration of the pause between sentences) and coordinate clause constructions below. Stressed vowels are indicated with full lines in the graphs, unstressed vowels are represented by points. Zero on the logarithmic frequency scale corresponds to 100 Hz. Four speakers, two males (NRP and JBC) and two females (GB and NT) - identified at the top right of each set of graphs.



(figure 1 continued)



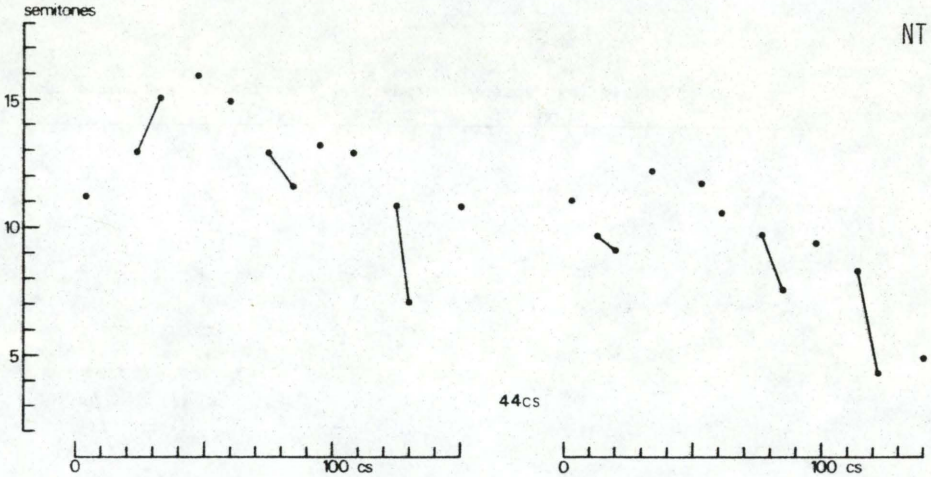
(figure 1 continued)



Amanda skal afsted på camping.

33cs

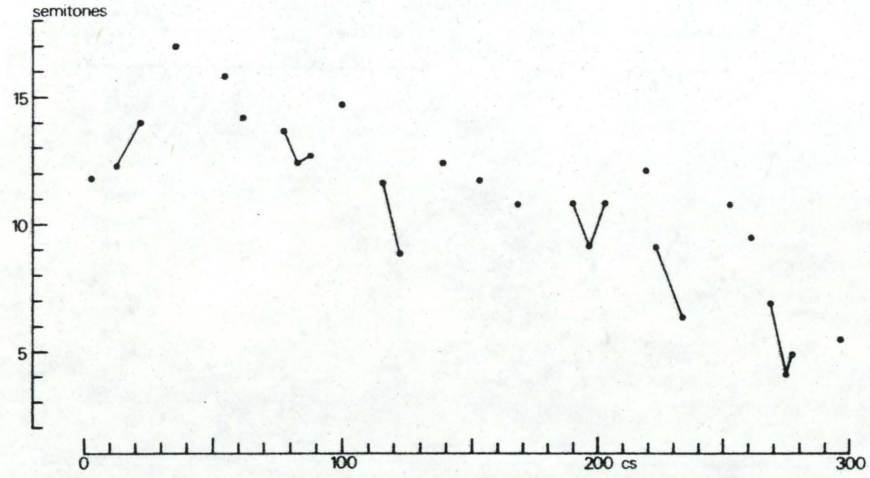
Hendes far skal vandre i Lapland.



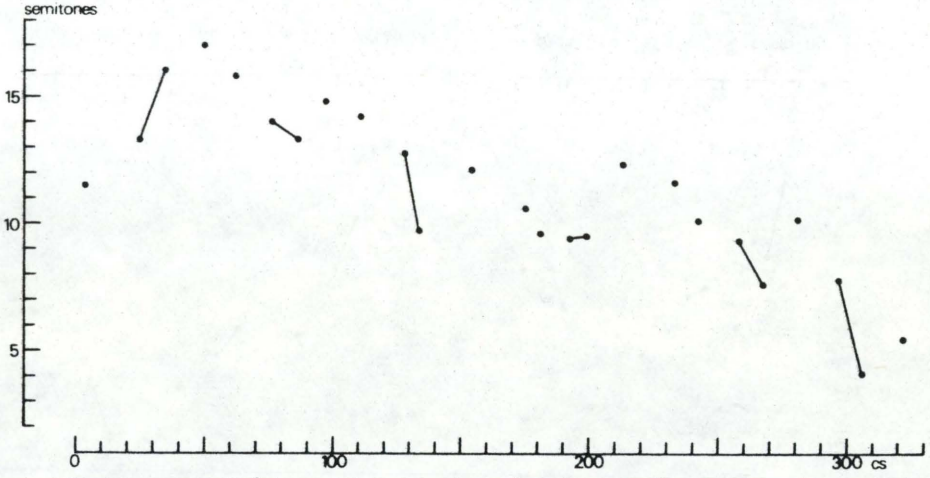
Hendes mor skal på kursus i Tyskland.

44cs

Amanda skal afsted på camping.



Amanda skal afsted på camping, og hendes far skal vandre i Lapland.



Hendes mor skal på kursus i Tyskland, og Amanda skal afsted på camping.

(figure 1 continued)

III. RESULTS

A. PRESENTATION OF THE DATA

In figure 1 is shown average tracings of two pairs of two-sentence texts, where the *Amánda*-sentence occurs in initial and final position, respectively. This is to illustrate the relation between the averaged "raw" data and the data reduced simplified representations in figure 2 (upon which the account of the results and the conclusions are based) and to give a more complete picture of similarities across (and differences between) individual speakers. The stressed vowels are indicated in figure 1 by lines connecting two or three points.

Although the main trends in the data are apparent in figure 1 (and of course also in the tracings of the averaged "raw" three-sentence texts, which are not shown here), a simplification is called for for a number of reasons. The three different sentences are not immediately comparable, due to the distribution of stressed vowel qualities, i.e. rather high vowels in the *mór*-sentence and rather low vowels in the *Amánda*- and *fár*-sentences, which makes the intonation contour too high in the *mór*-sentence relative to the other two, ceteris paribus. Different initial consonants in the stressed syllables within each sentence also obscure the picture. The [m] in *mór* and *Amánda* lowers the initial Fø value compared to the vowēls in thē following two stressed syllables which begin with unvoiced obstruents. Contrarily, the [f] in *fár* raises the initial Fø value considerably compared to the other two stressed vowels in that sentence. (See further Jeel, 1975, and the references there.) Both these difficulties could of course have been avoided in the construction of the material, but only - I feared - at the expense of the naturalness of the sentences and texts (to the extent that they can lay any claim to naturalness at all). As it is, these "extraneous" factors are alleviated if I cut and splice the texts (in the illustrations) so that initial, medial and final *Amánda*-sentences are presented together in pseudo-texts, and likewise for the *mór*- and *fár*-sentences. In this way we get a clearer impression of the relation between sentence intonation contours (a) in different positions in a text, (b) in texts of different length, and (c) in texts of different types, and it is the relations rather than the absolute slope shapes and values which are interesting in this connection. This last contention also justifies the rather drastic reduction of each sentence into two straight lines, an upper and a lower one. The upper line in figure 2 is the connection between the first and last post-tonic syllable in each sentence. The lower line is the connection of the Fø minimum in the first stressed vowel (the initial value in purely rising movements, the turning point in falling-rising movements) and the final value in the last stressed vowel. This lower line is a reflexion of, and related to, the intonation contour as I have chosen to define it, namely as the line or figure described by the stressed syllables of an utterance. It will be too steep in the *fár*-sentence, and not steep enough

in the *Amánda*- and *mór*-sentences, due to influence from the initial consonant in the first stressed syllable. - When I say about vowels that their initial Fø value is "too high" or "too low" and therefore intonation contours will seem "too steep" or "not steep enough", I mean that their physical manifestation does not, presumably, correspond to the way they are perceived. We know that differences in duration and Fø level between stressed vowels of different tongue height, ceteris paribus, are overheard or compensated for by the listener, see e.g. Reinholt Petersen (1974) and Hombert (1977). This compensation is explained with reference to the listener's "knowledge" of those constraints inherent in the speech production apparatus which are responsible for physical differences in sounds which are intended by the speaker to be the "same". From this knowledge presumably derives an expectation to find certain acoustic differences coupled with certain different spectral events, and these differences therefore carry no independent significance and they are disregarded by the listener in the interpretation of what (s)he hears. If the reference to the listener's double function as both speaker and listener (which in a more general perspective is what the motor theory of speech perception is founded on, cf. Liberman et al., 1967) explains the (non) perception of Fø and duration differences between stressed vowels of different tongue height, I think it reasonable to assume that it will also make the listener disregard differences in vowel onset Fø which are caused by the different nature of the preceding consonant. It is certainly my own impression from the recordings that the three different sentences do indeed sound "the same" as far as steepness and shape of the intonation contours are concerned when everything else is equal. This is of course no evidence, and formal experiments should be undertaken to verify or falsify my assumption, but note again that the conclusions to be drawn from the data do not hinge upon its verification since I am more concerned here with relations between sentence intonation contours than with their absolute gestalt.

In figure 2 the simplified data are presented which describe the intonational structure of three pairs of pseudo texts (as defined above), for each of the four speakers and their mean (mean of means). Sentences are identified by their original position; e.g. "1-1" is the first sentence of one, "2-2" is the second sentence of two, etc. The isolated sentence ("1-1") and the two-sentence text ("1-2+2-2") are both presented together with the three-sentence text (the set is complete only in the case of *Amánda*). For ease of comparison "2-2" is shown twice: directly after "1-2" together with the medial sentence ("2-3") and also together with the final one ("3-3").

One type of information is lost in the process of piecing together the figures from parts deriving from different actual texts, namely pause duration. All speakers paused between the terminal declarative sentences, cf. figure 1. Pauses range between 20 and 80 cs. One speaker (JBC) also paused between the coordinate main clauses, but his pauses were only about half as long there (20-50 cs) as between the terminals (50-80 cs).

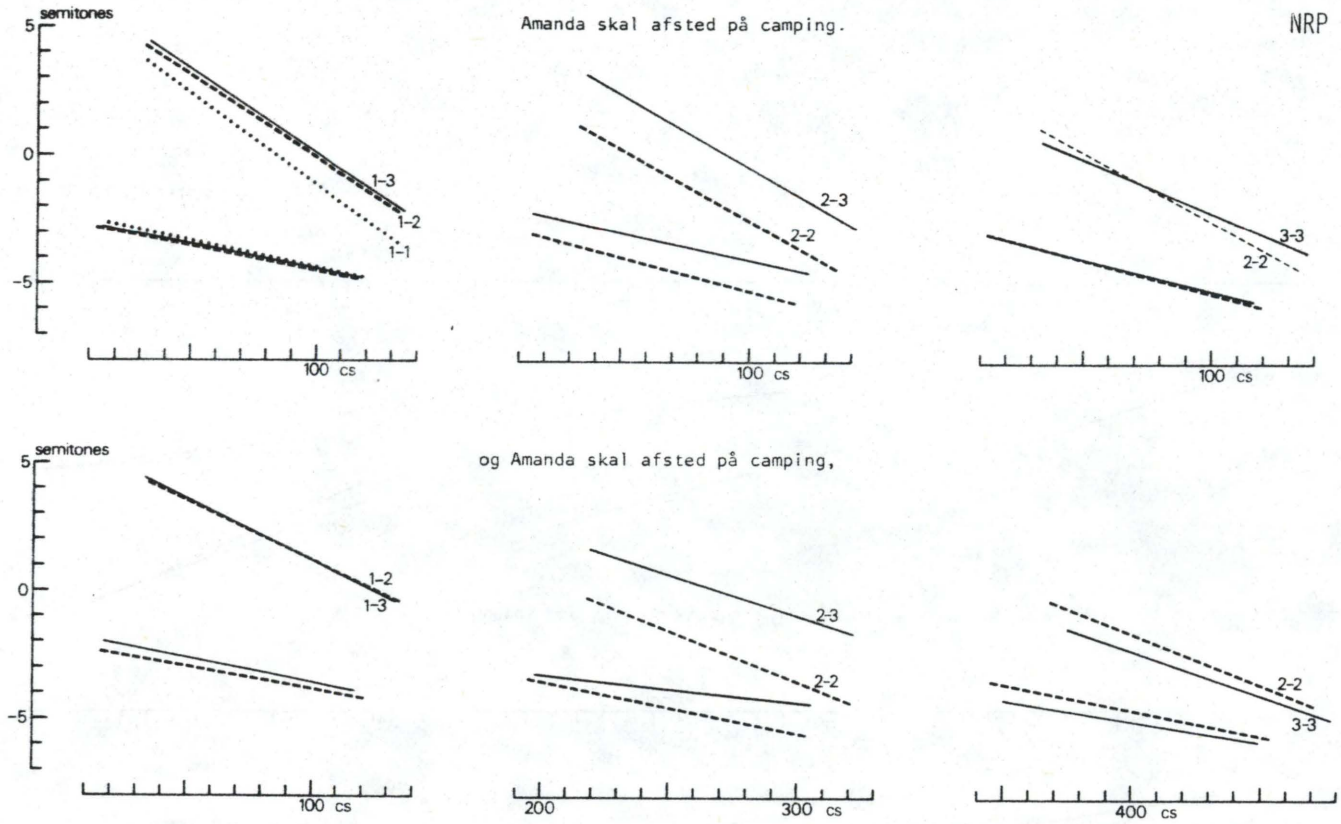
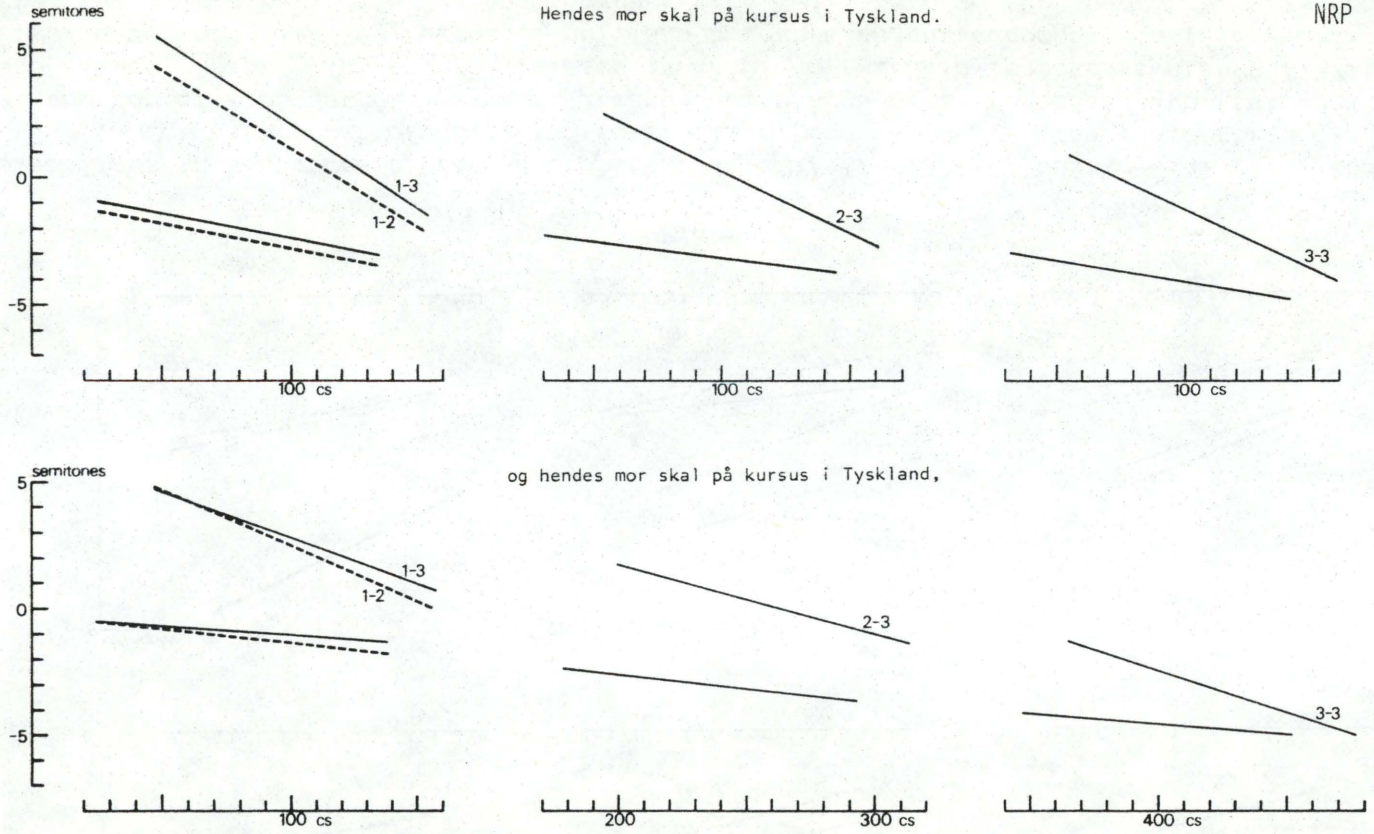
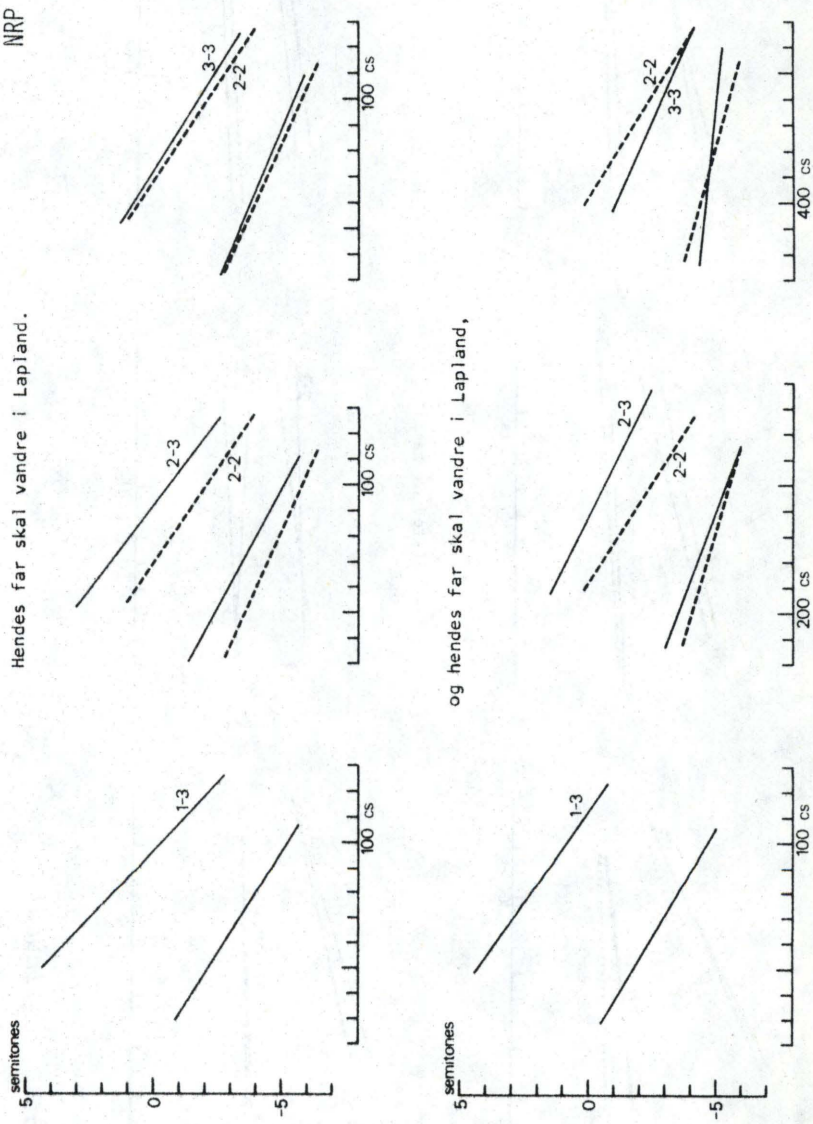


Figure 2

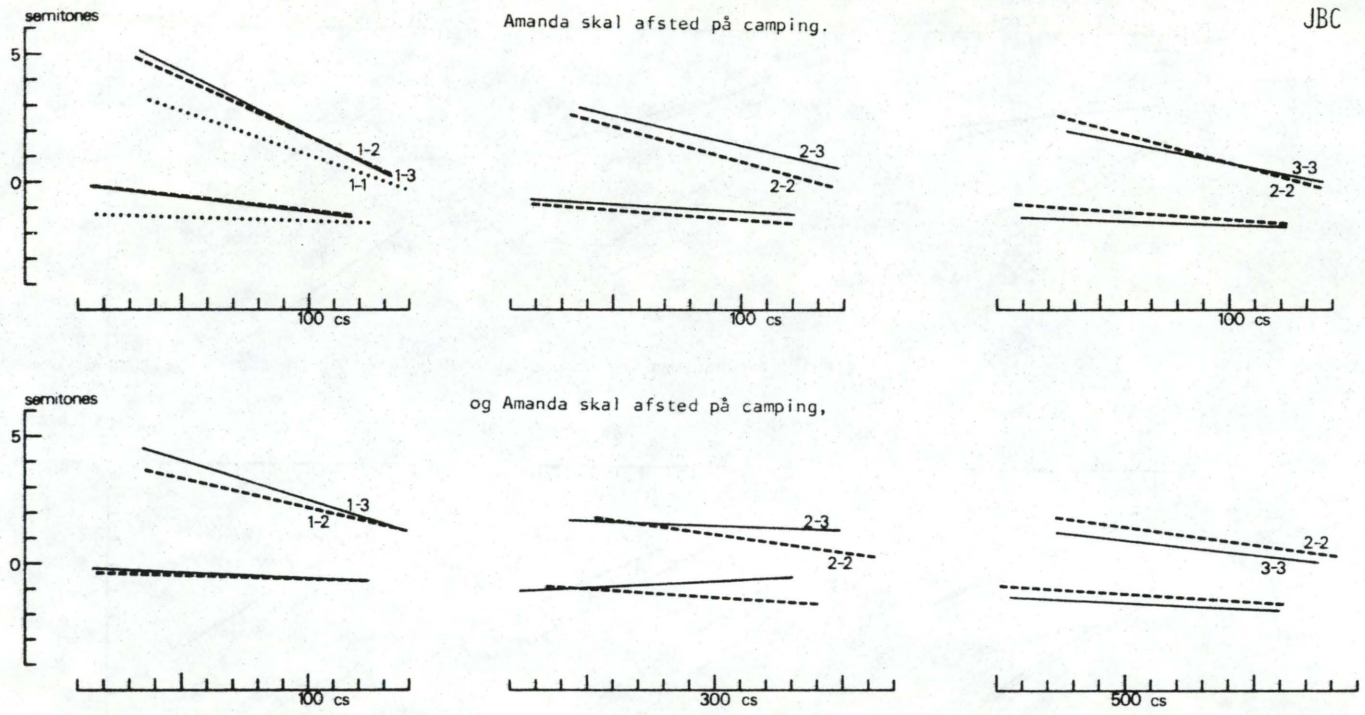
Upper and lower lines in sentences extracted from text initial (1-2, 1-3), text medial (2-3) and text final position (2-2, 3-3). For an account of upper and lower lines, see the text. There are three different sentences, identified at the top of each set of contours, arranged with sequences of declarative, terminal sentences above and coordinate clauses below. The Amánda .... set is complete, with an isolated sentence (1-1, dotted line), a text with two components (1-2+2-2, dashed line), and a text with three components (1-3+2-3+3-3, full line). 2-2 is depicted twice: together with 2-3 and together with 3-3. Zero on the logarithmic frequency scale corresponds to 100 Hz. There are four speakers, two males (NRP and JBC) and two females (GB and NT) and a grand mean (mean of means) - identified at the top right of each set of graphs.



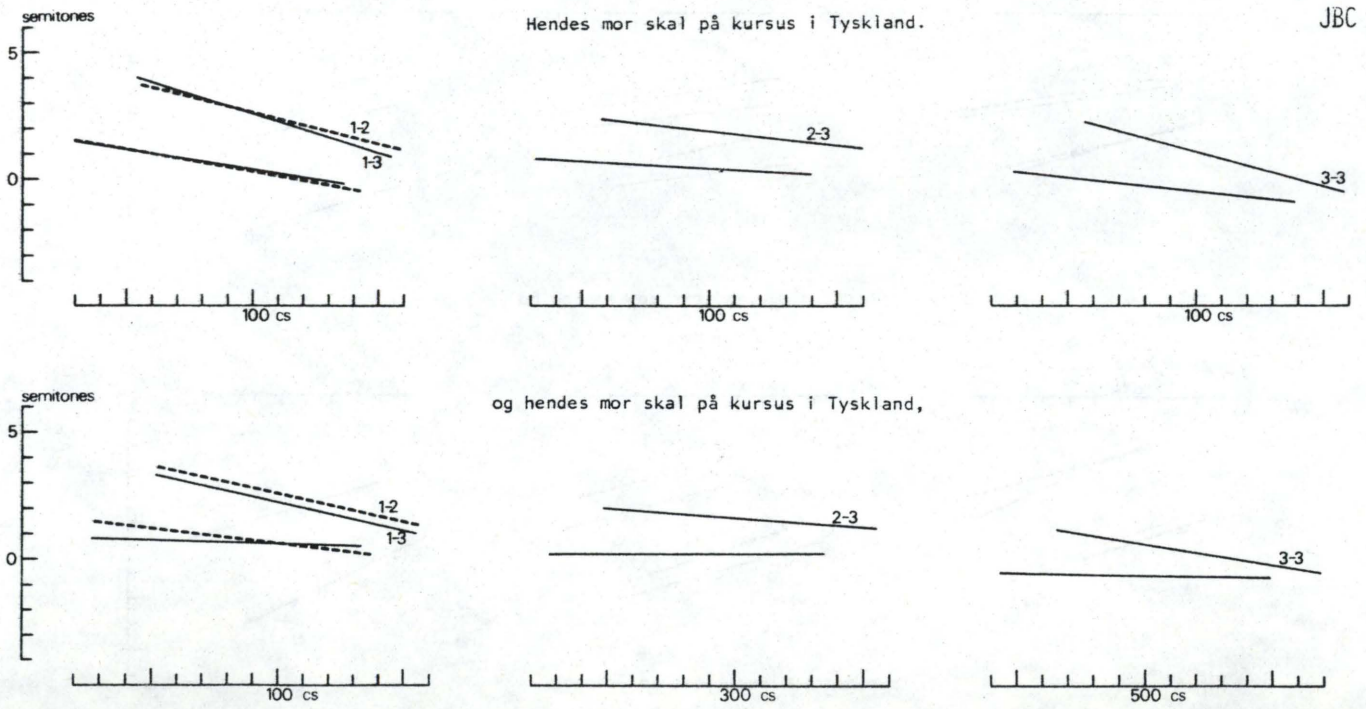
(fig. 2 continued)



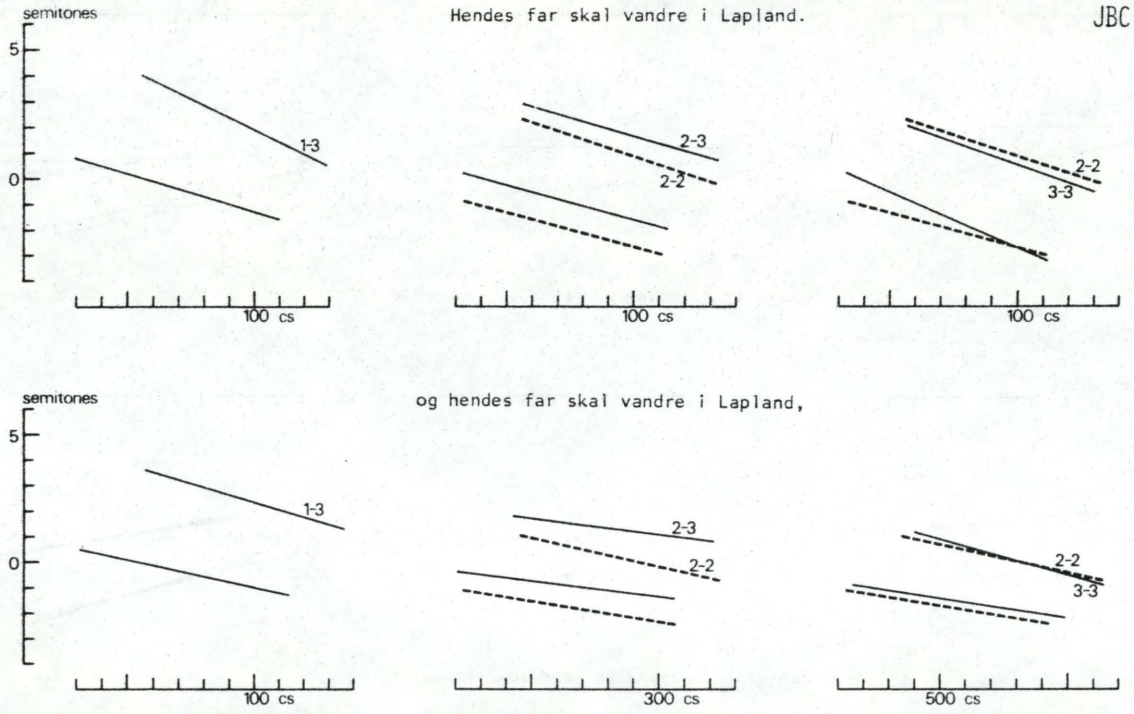
(fig. 2 continued)



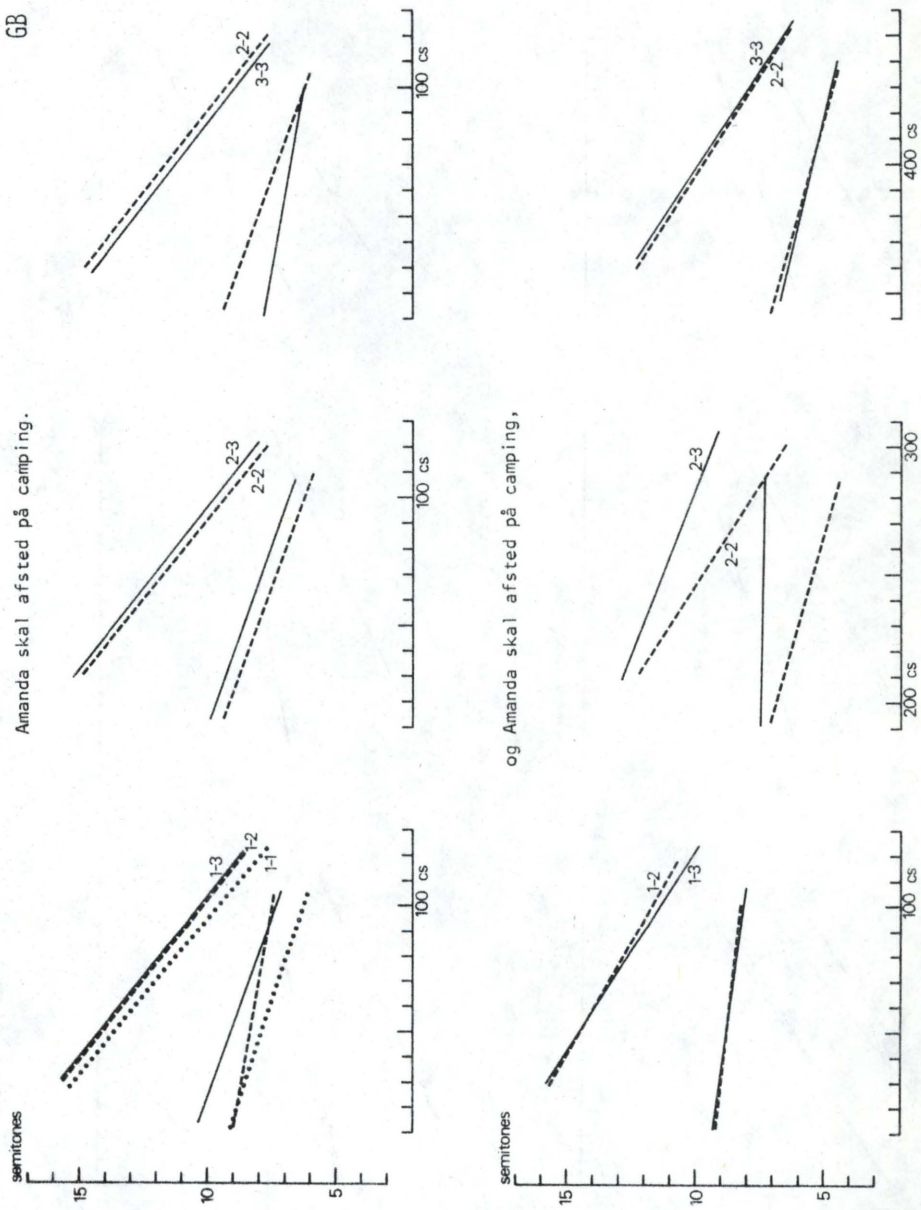
(fig. 2 continued)



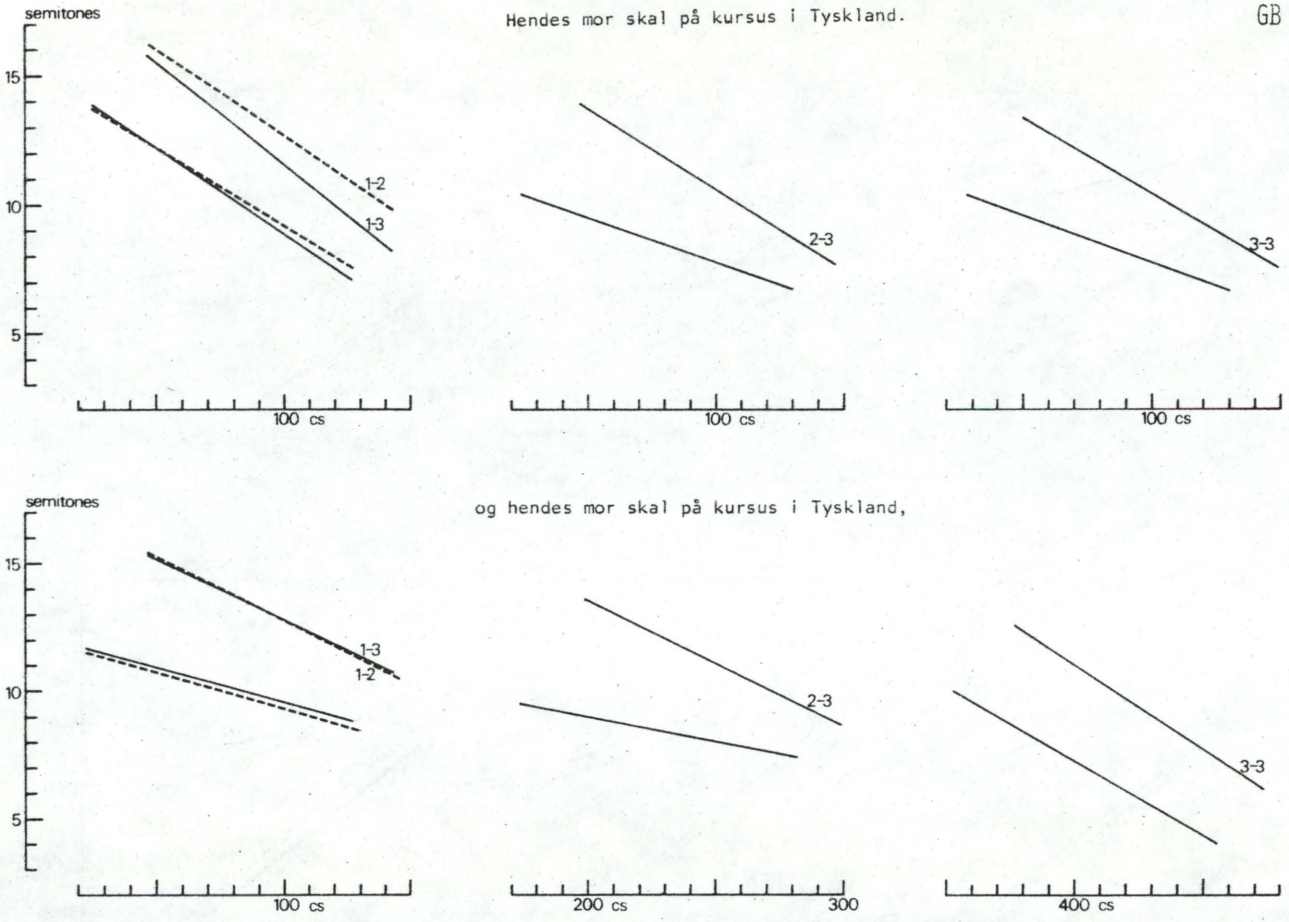
(fig. 2 continued)



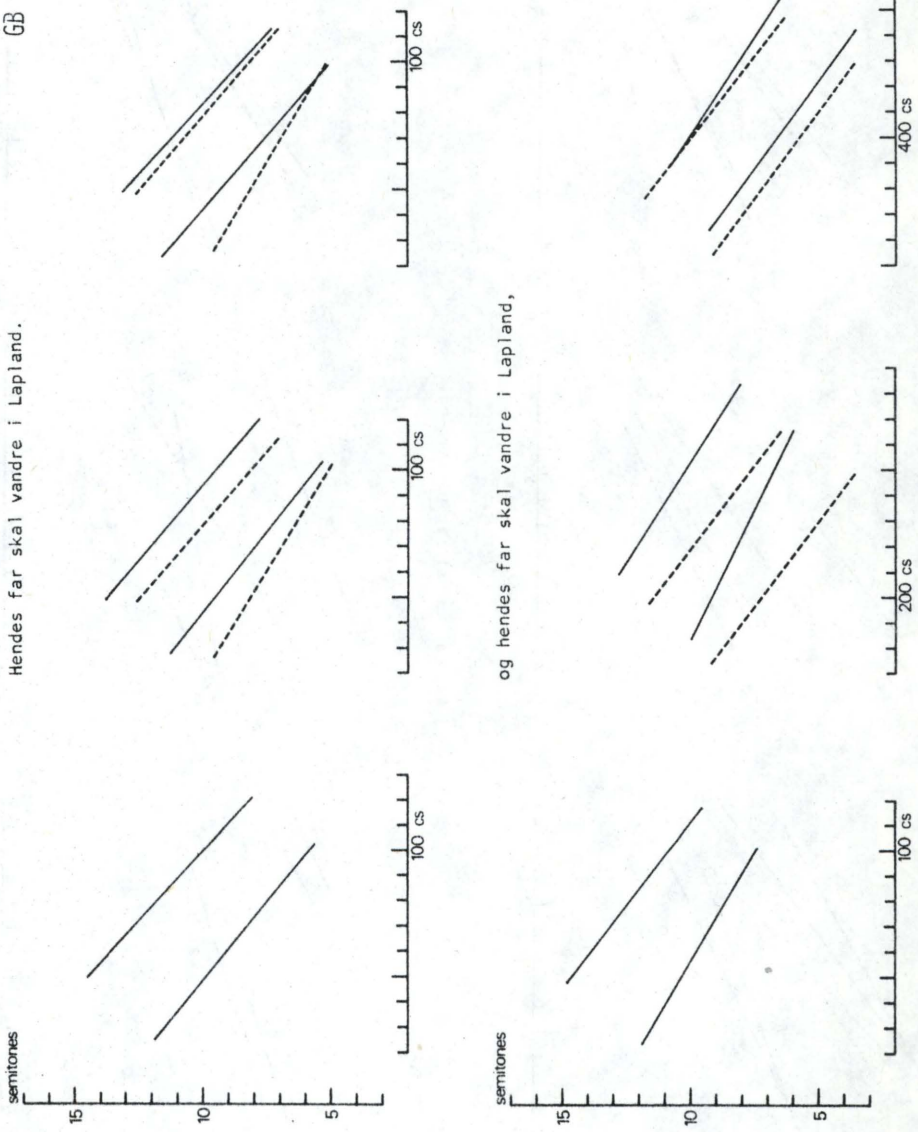
(fig. 2 continued)



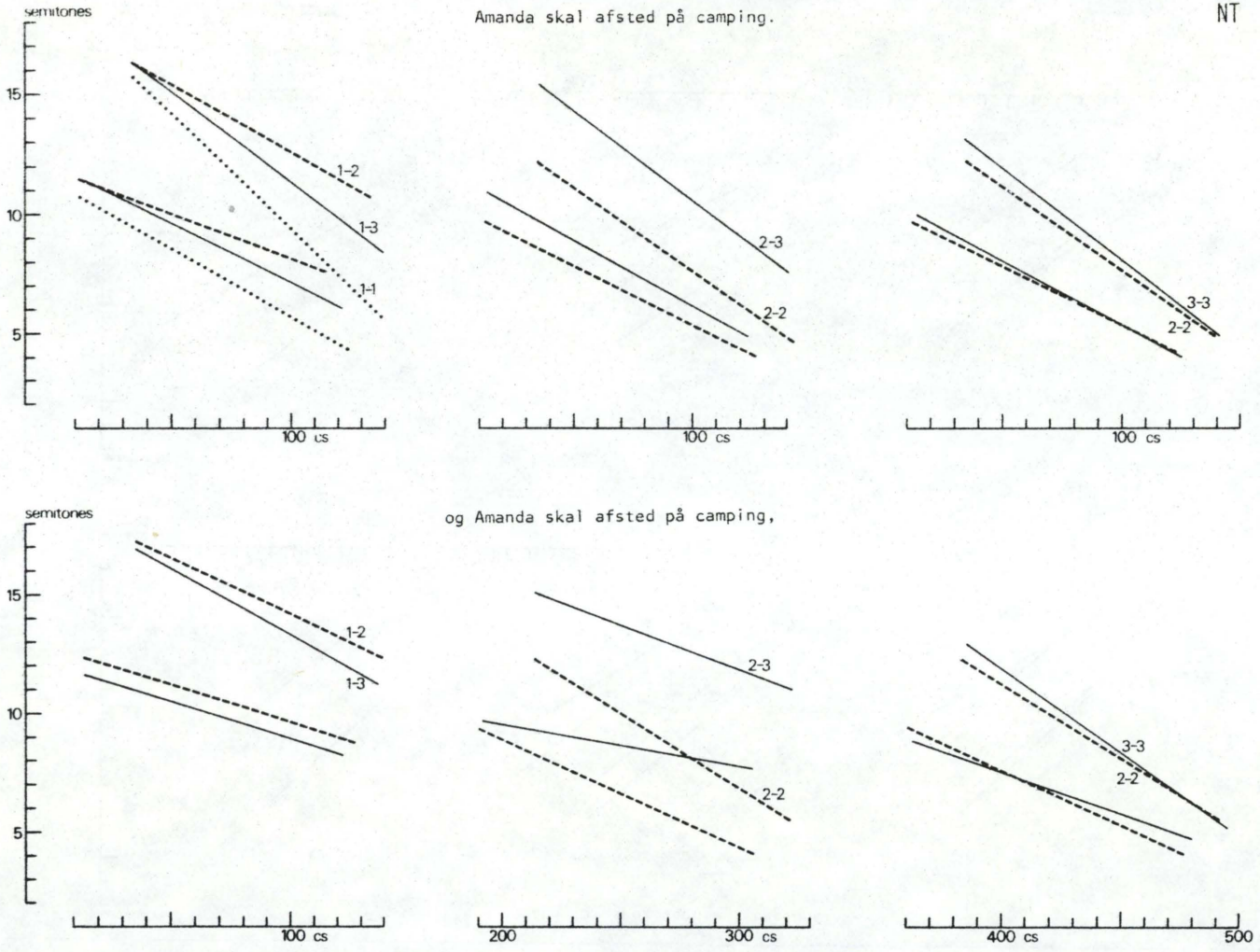
(fig. 2 continued)



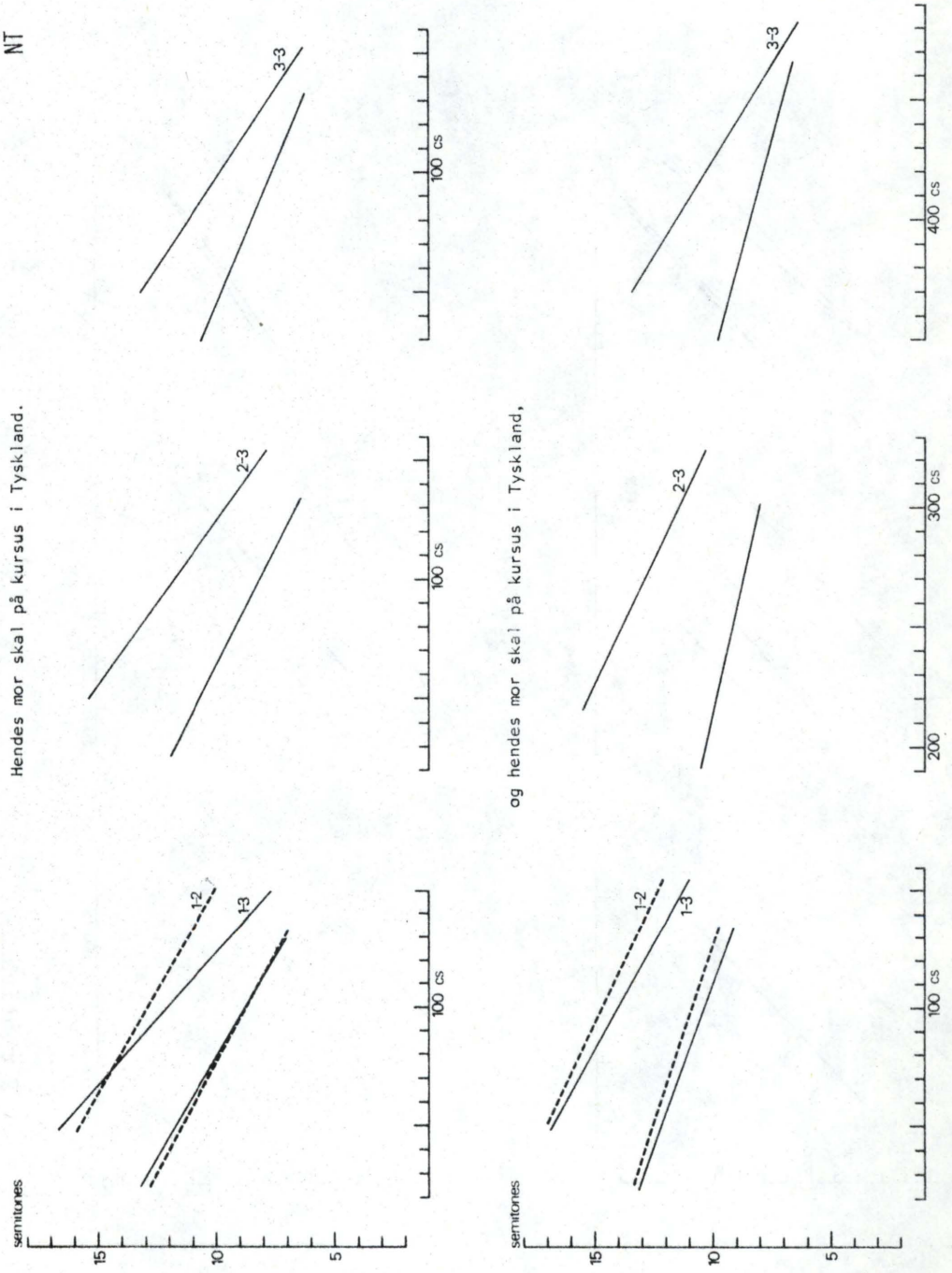
(fig. 2 continued)



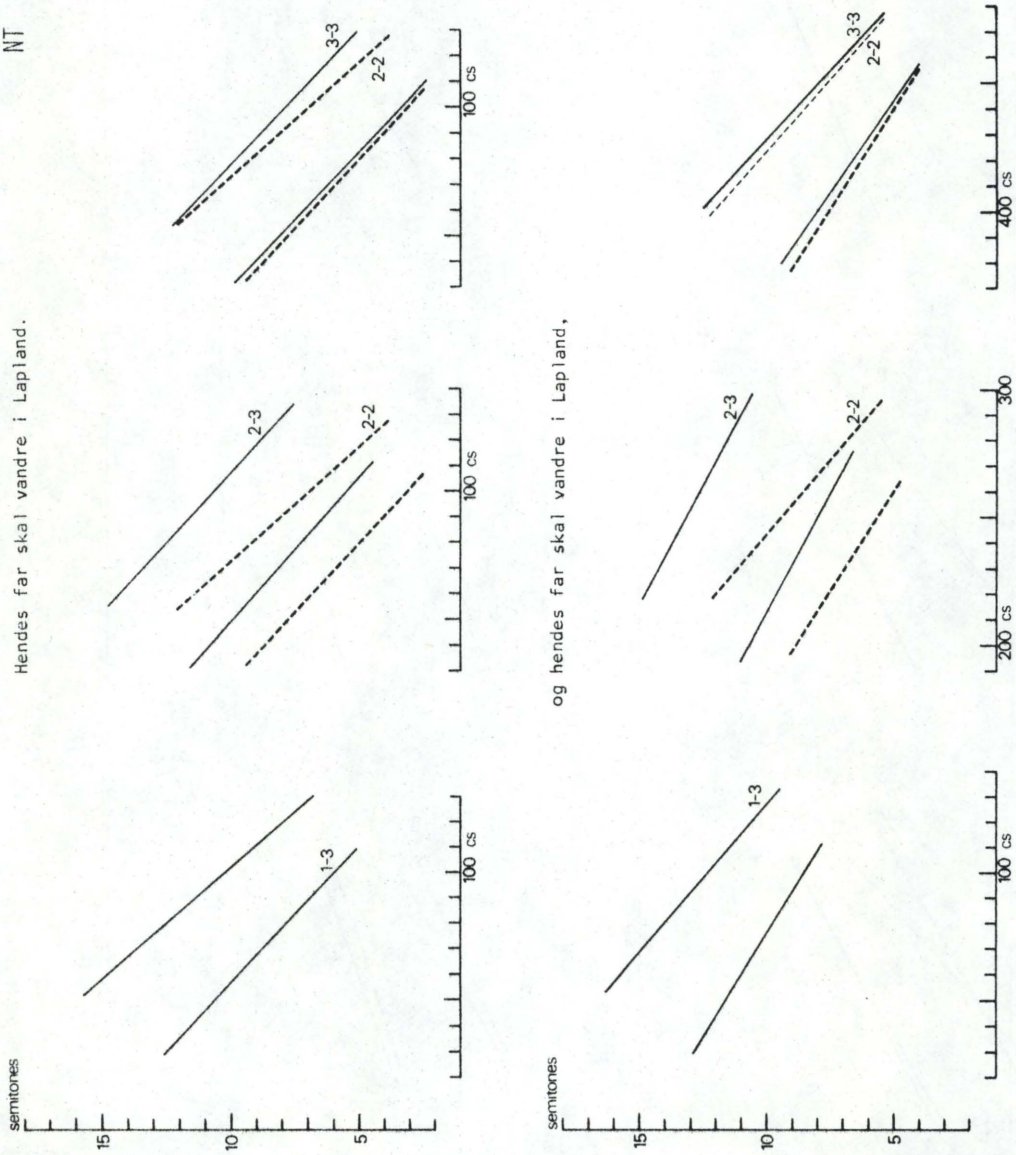
(fig. 2 continued)



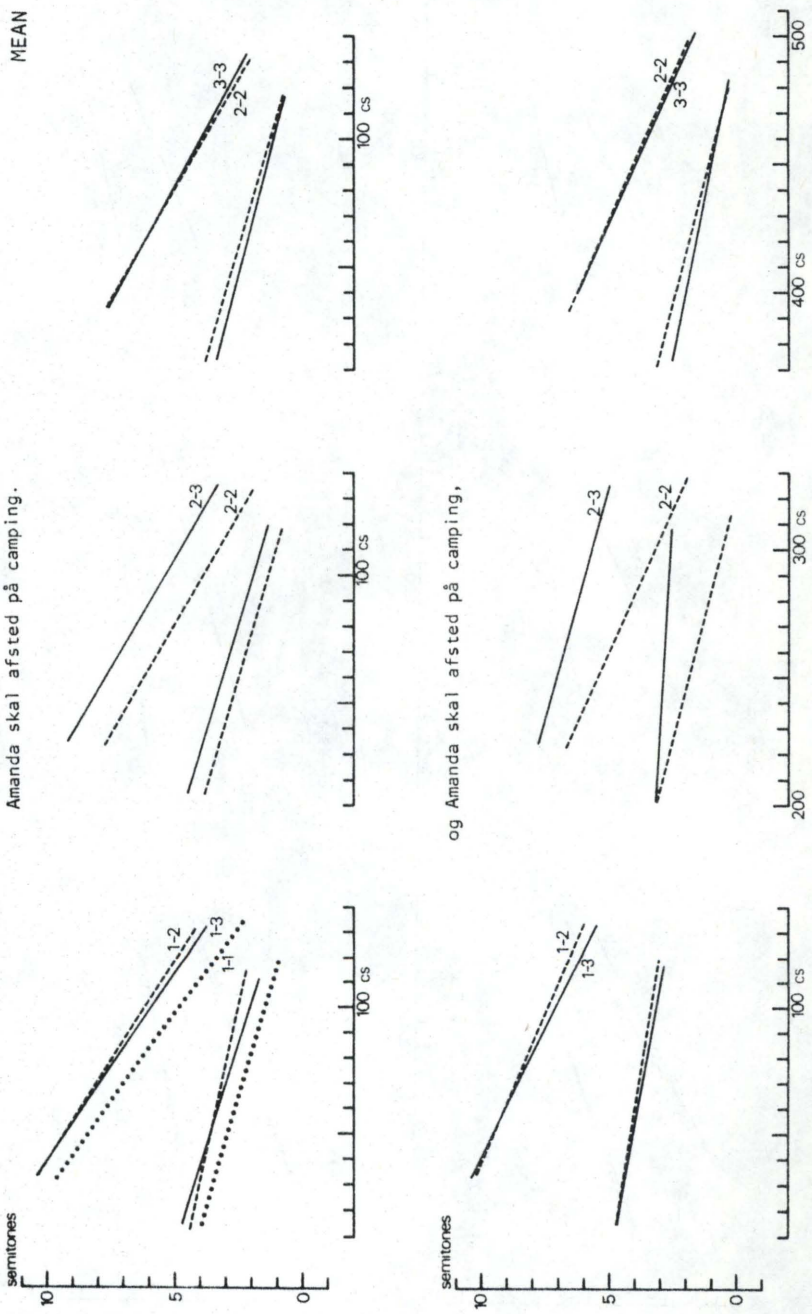
(fig. 2 continued)



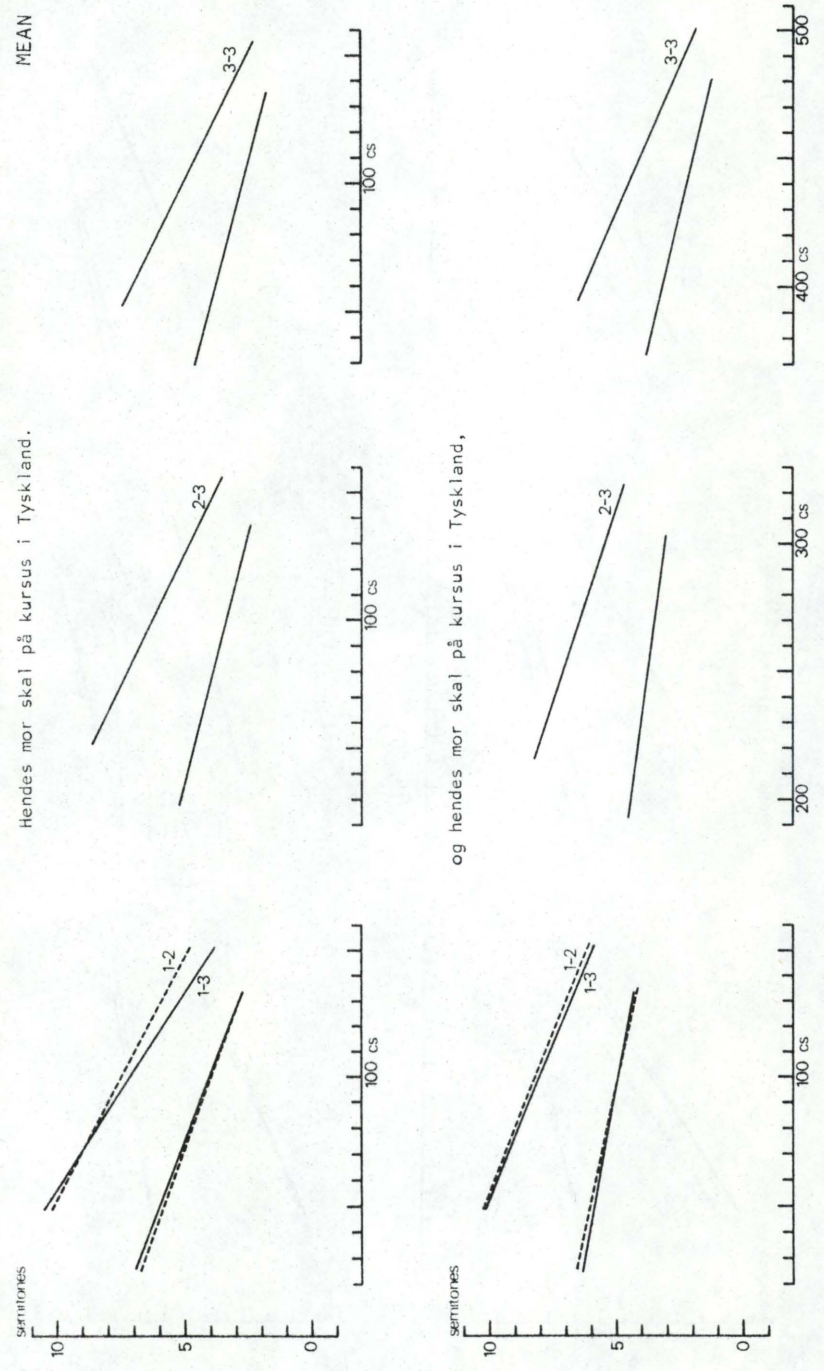
(fig. 2 continued)



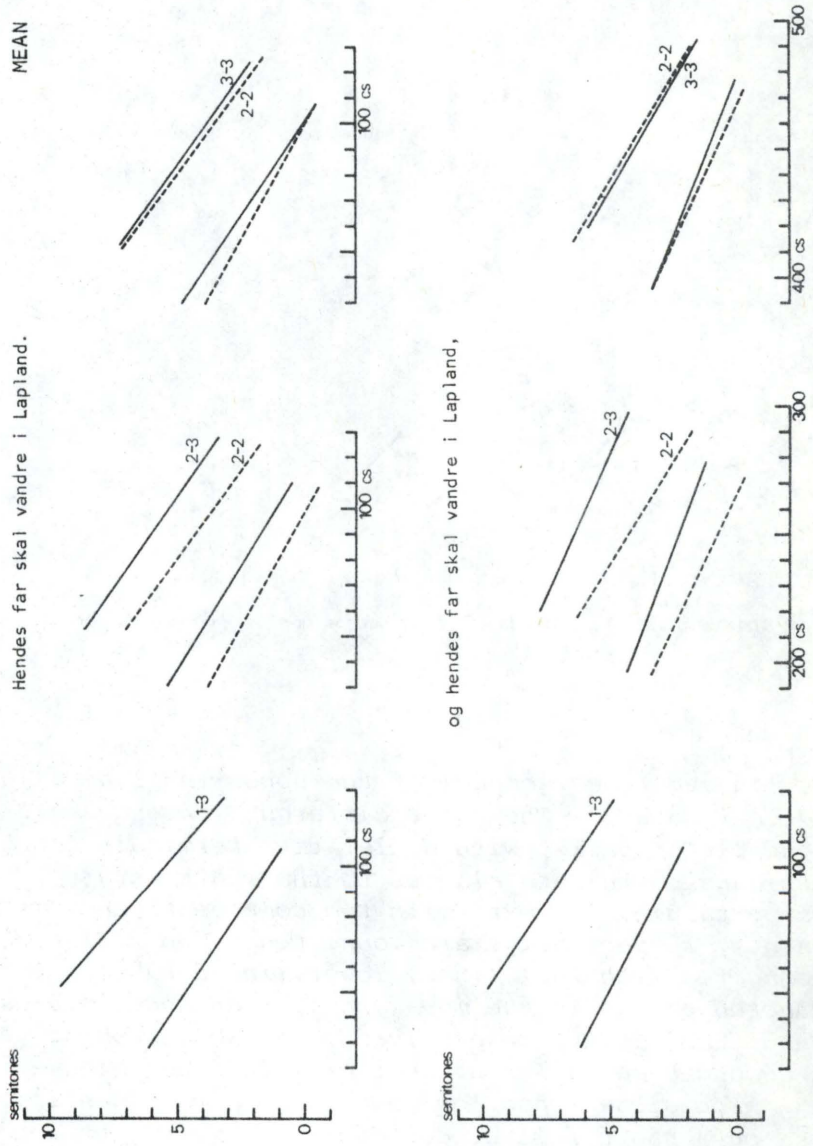
(fig. 2 continued)



(fig. 2 continued)



(fig. 2 continued)



(fig. 2 continued)

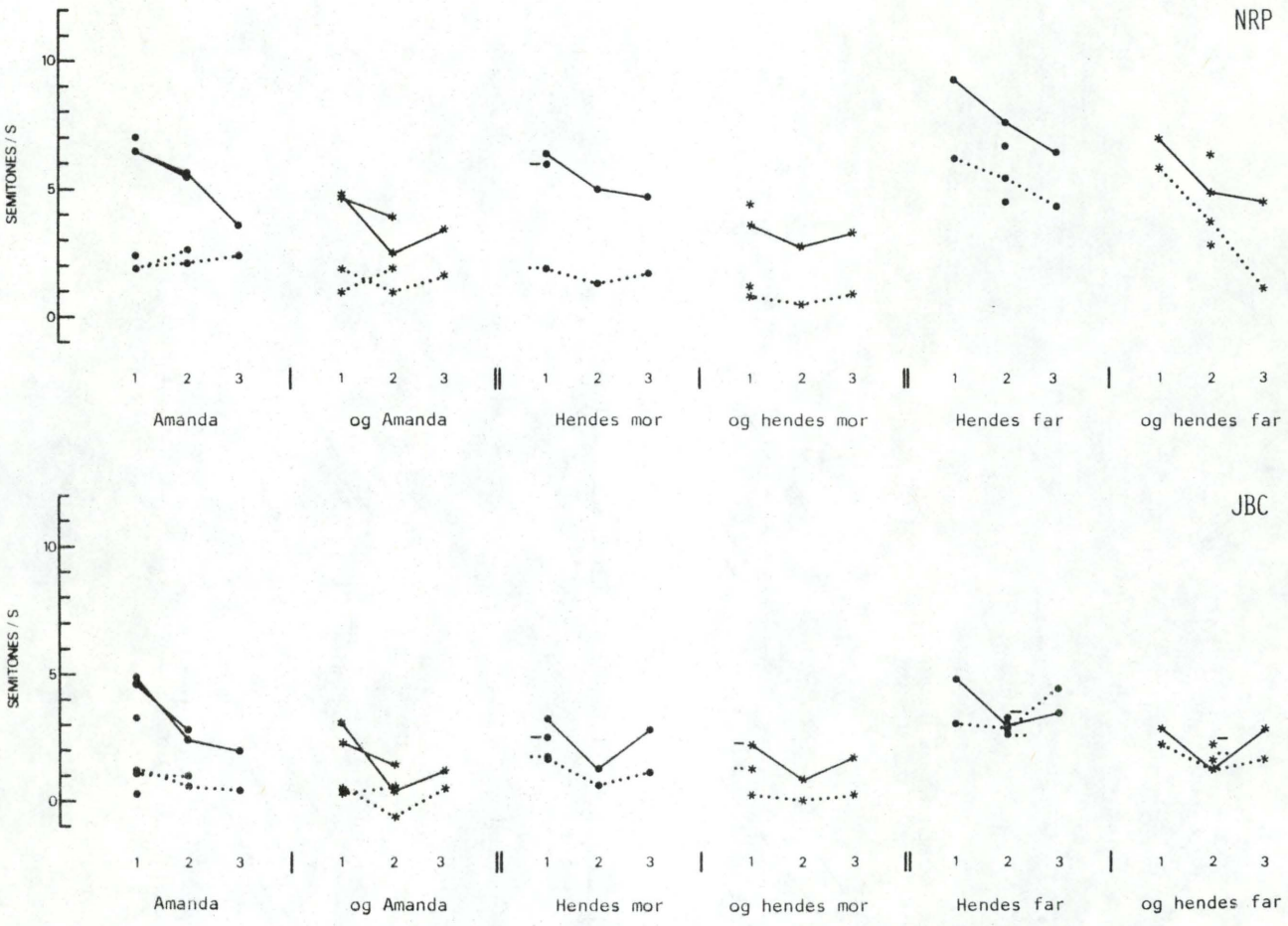
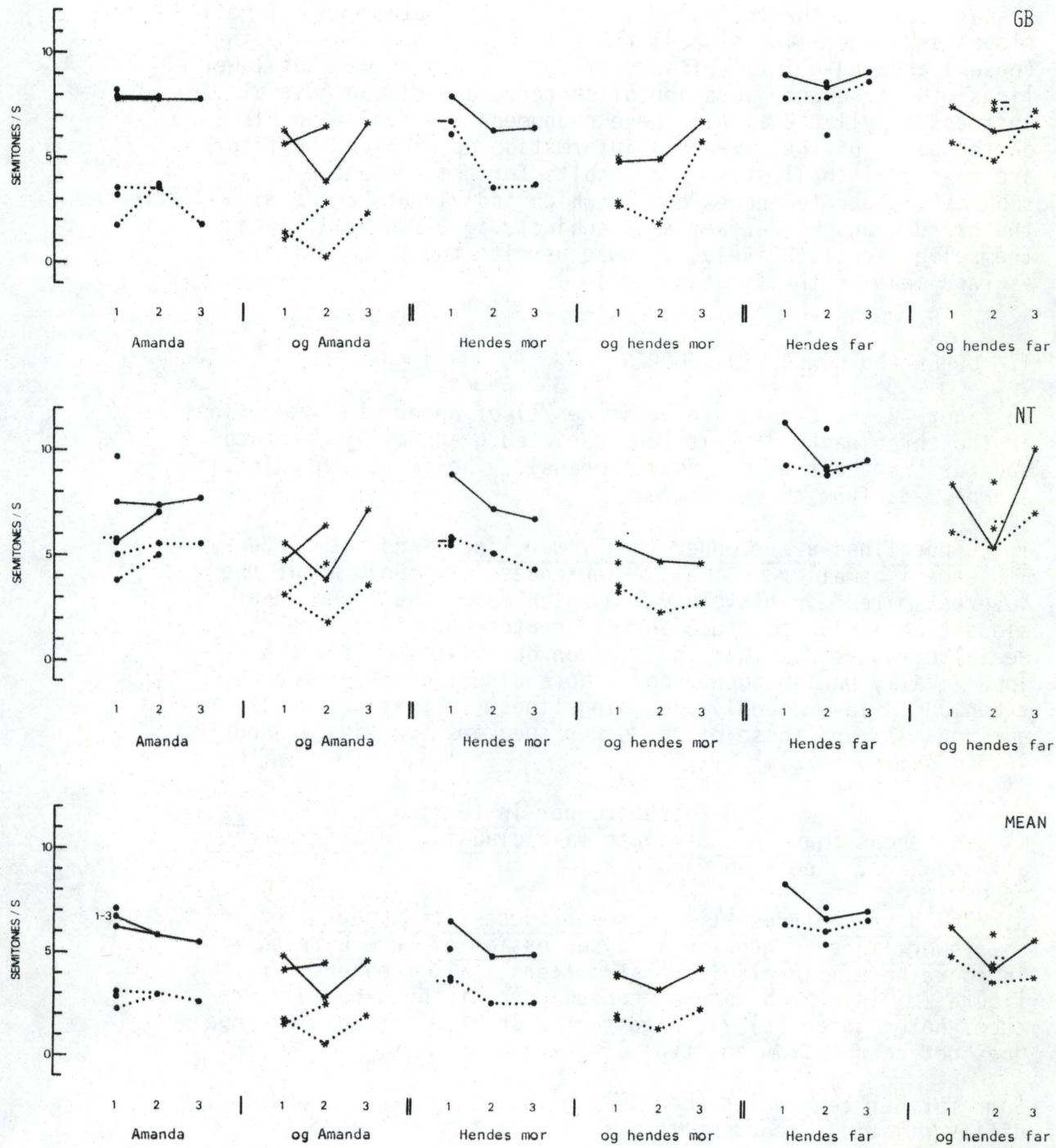


Figure 3

Slopes (in semitones/second) of the upper and lower lines depicted in figure 2. The three different sentences are arranged horizontally in pairs, with declarative terminals to the left (points) and coordinate clauses to the right (stars). Data points pertaining to texts with two components and with three components, respectively, are connected. Upper line data points are connected with full lines, lower line data with dotted lines. Unconnected points in the mór- and fár-sentences represent initial and final items, respectively, in two-component texts. Where unconnected points are not clearly associated - in the graph - with either upper or lower line data, identification is ensured by a short full or dotted line next to the dot or star. Four speakers, two males (NRP and JBC) and two females (GB and NT) and their grand mean (mean of means) - identified at the top right of each graph.



(figure 3 continued)

B. INTERPRETATION

I shall take up the following aspects of the data: upper line slopes and lower line slopes; the frequency of beginning points (onset) and end points (offset) of upper lines as well as lower lines; the frequency location of sentence and clause initial unstressed syllables. All these phenomena are evaluated also on the basis of the numerical information upon which the figures are based. I shall state the results for the average over all subjects, and note the extent to which individuals agree with the grand mean. (The fact that subjects agree so well among themselves, qualitatively, is what permits the calculation of a grand mean in the first place.)

1. UPPER AND LOWER LINE SLOPES

In figure 3 the slopes (in semitones/s) of upper and lower lines in the three pairs of texts are depicted graphically for each speaker and their mean (mean of means). The following points can be made from these figures:

1. Upper lines are steeper than lower lines, and the difference is smaller in the *fár*-sentences, cf. above about the "overestimated" initial Fø value which makes the lower line slope relatively too steep in this sentence: all speakers. See also figure 2. This is a common observation and not a specifically Danish phenomenon. Note also the difference between JBC's less steeply declining lines (varying around 2½ semitones/s) and those of the other speakers (varying around 4 to 6 semitones/s).

2. Upper and lower lines are steeper in terminal declarative sentences than in coordinate main clauses, ceteris paribus: all speakers. See also figure 2.

3. The degree of resetting between successive upper lines and lower lines is greater in a succession of terminals than in coordinate main clause constructions: all speakers, cf. figure 2. This is a natural consequence of the slope differences noted under (2), granted that a speaker's total Fø range does not change from one type of text to another.

Some further tendencies (but that is all they are) towards a differentiation according to position in the text appear:

4. The upper and lower line slopes in an isolated sentence are equal to or very slightly steeper than initially in multi-sentence texts (relevant only for terminal sentences): though with JBC upper and lower lines are less steep in the isolated sentence. Otherwise, upper and lower line slopes tend to be more nearly identical in initial sentences and in final sentences, irrespective of text length. I.e., 1-2 resembles 1-3 and 2-2 resembles 3-3 (rather than 2-3). See also figure 2.

5. Upper and lower line slopes in the terminal sentences tend to be steeper in initial than in non-initial position in three sentence texts: there are random exceptions with all speakers, typically where a final slope is as steep as the initial one. The two sentence text offers no uniform picture. See also figure 2.

6. Upper and lower line slopes in coordinate main clauses are less steep in medial position than initially and finally - where they are approximately equal - in three sentence texts, and slopes are steeper finally than initially in two sentence texts: there are a few slight and random exceptions with every speaker. See also figure 2.

The full and dotted lines in figure 3 run approximately parallel to each other, considered frame by frame, with a couple of apparently random exceptions. In other words, the difference in upper and lower line slope is nearly constant across a given multi-sentence text. I.e., given a certain change in lower line slope across a two- or three-sentence text, the upper lines follow suit, and can be deduced from the behaviour of the lower lines. (Arguments to the effect that the lower line - the stressed syllables of the utterance - is the independent variable are given in e.g. Thorsen 1980b and 1980c. Suffice it here to say that every utterance has a lower line (one or more stressed syllables), but an utterance does not invariably have an upper line which depends for its existence upon the presence of a post-tonic syllable in the prosodic stress group(s).) The steeper upper lines derive from the fact that the magnitude of the Fø interval between the minimum in the stressed syllable and the maximum in the post-tonic varies with position on the intonation contour (early - larger interval, later - smaller interval), cf. Thorsen (1980b, 1984).

From the figures and (1) to (6) above we learn that the individual components are subordinated differently to the superordinate declination which spans the whole text. Individual declinations are steeper and have greater amounts of resetting between them in a series of declarative terminal sentences than in a corresponding sequence of coordinate main clauses. Furthermore, in a series of three terminals the individual slopes tend to become successively less steep, whereas coordinate clauses level out in medial position and a final declination is always steeper than a preceding (initial or medial) one. This can be taken to indicate a more integrated and coherent intonational structure accompanying the closer relation between coordinate main clauses.

Finally, it appears that increasing text length from two to three sentences/clauses or, inversely, decreasing text length from three to two, will leave the initial and final components unaffected. A component is squeezed in or removed medially with no apparent changes in the surroundings. As a result, the second component in a text is different when it is simultaneously text final from when it is medial (compare 2-2 and 2-3 in figure 2). In other words, the production of the second

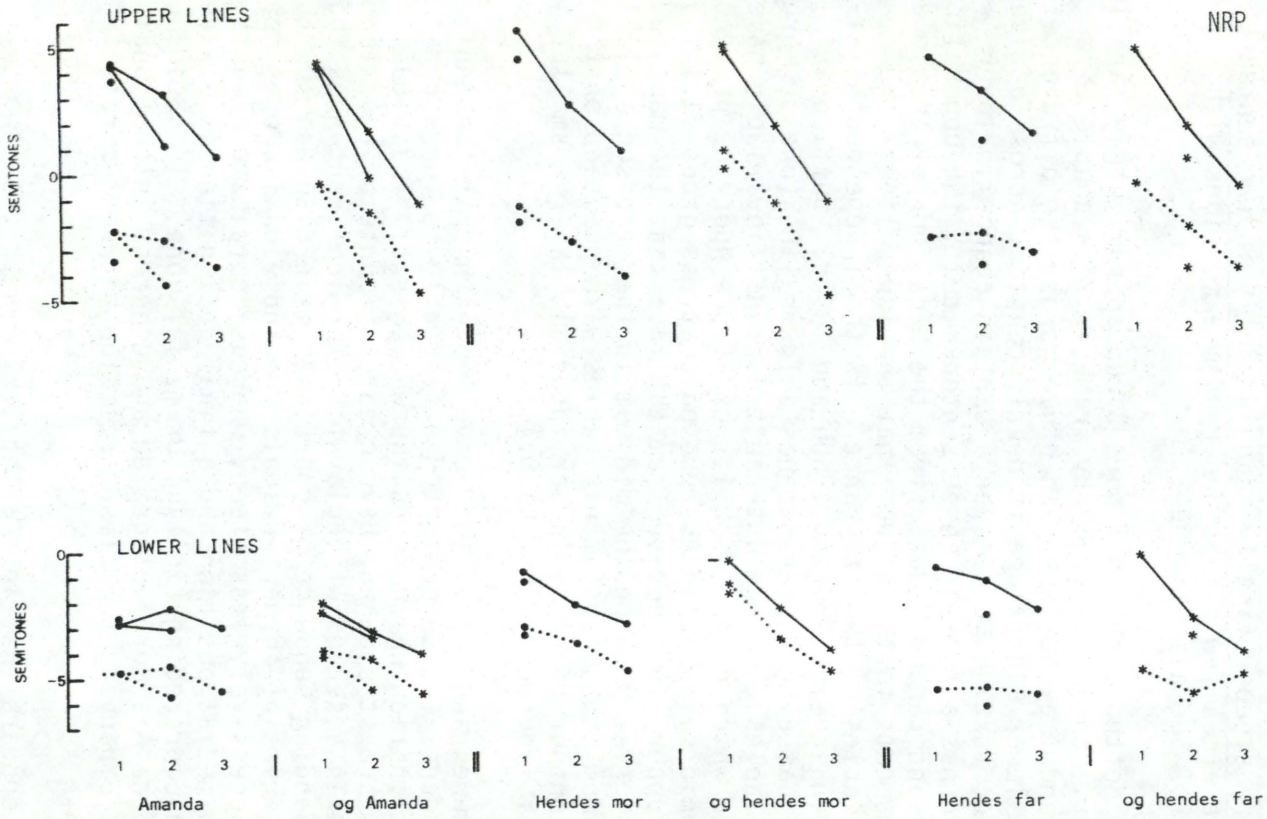
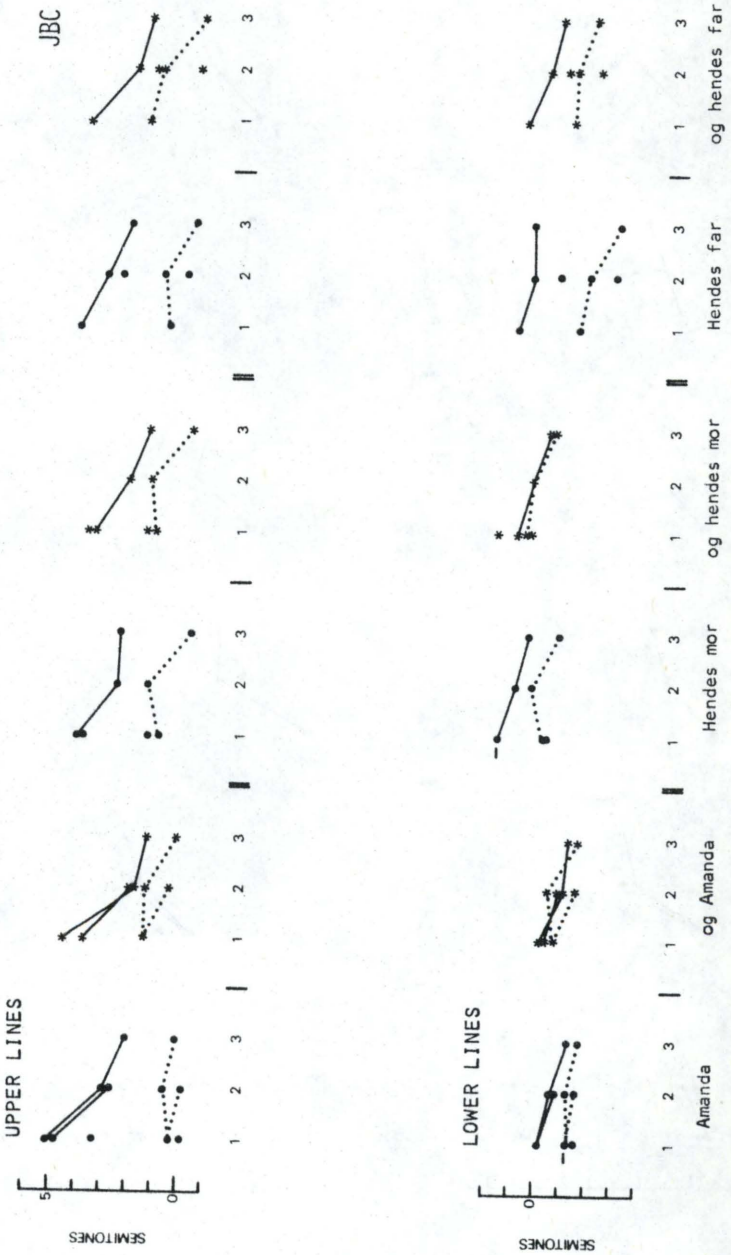
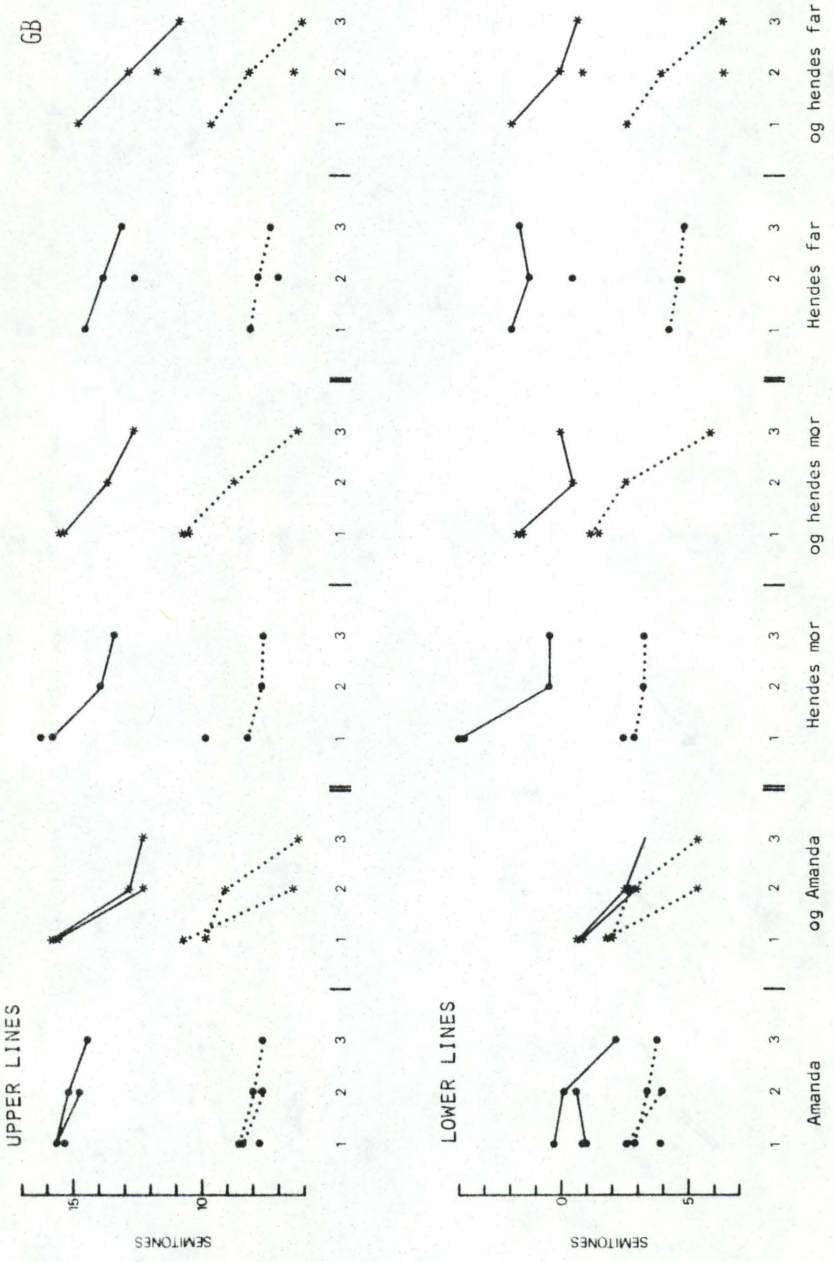


Figure 4

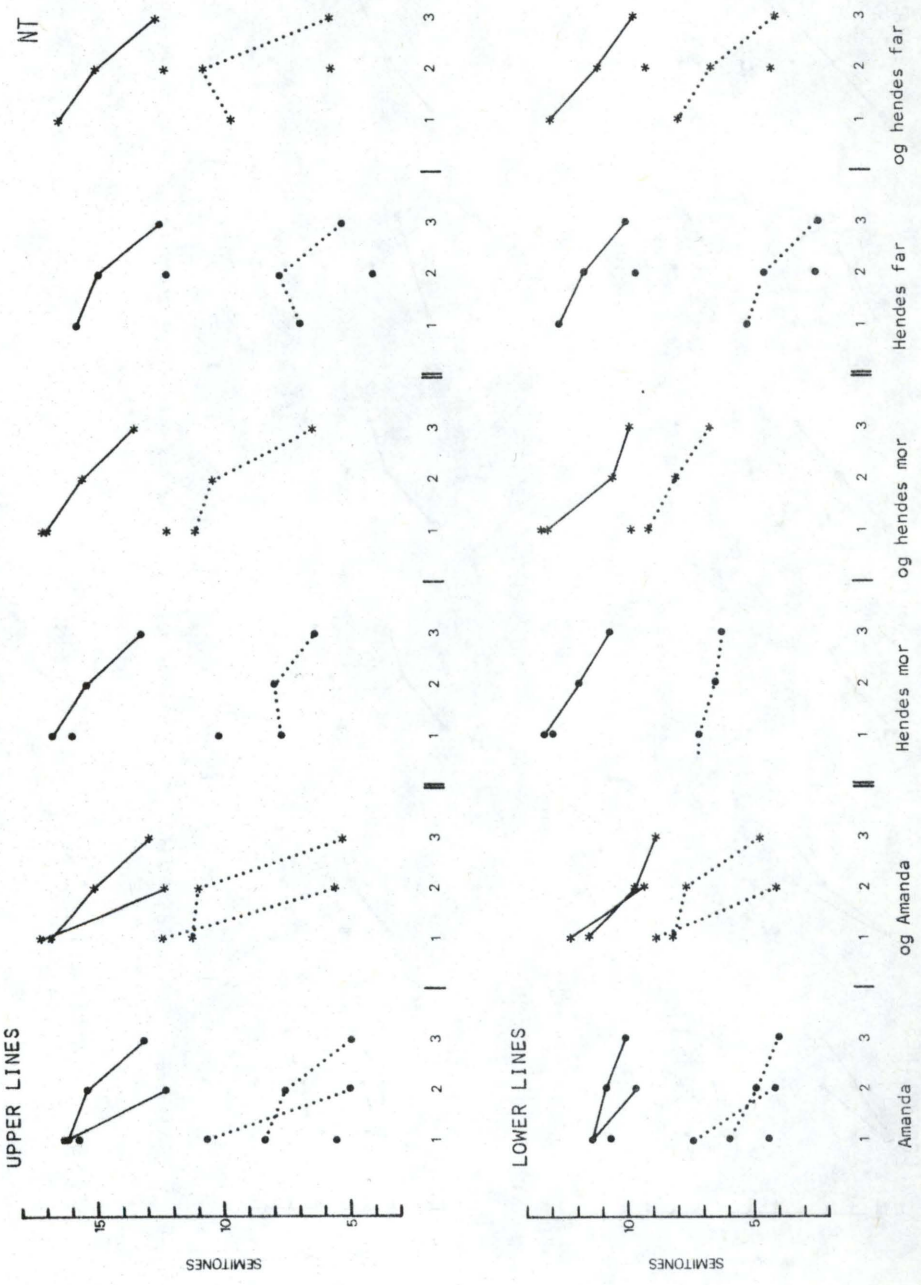
Frequency values of beginning and end points of the upper (above) and lower (below) lines depicted in figure 2. The three different sentences are arranged horizontally in pairs, with declarative terminals to the left (points) and coordinate clauses to the right (stars). Data points pertaining to texts with two components and with three components, respectively, are connected. Beginning points are connected with full lines, end point data with dotted lines. Unconnected points in the mór- and fár-sentences represent initial and final items, in two-component texts. Where unconnected points are not clearly associated - in the graph - with either beginning or with end point data, identification is ensured by a short full or dotted line next to the dot or star. Zero on the logarithmic frequency scale corresponds to 100 Hz. Four speakers, two males (NRP and JBC) and two females (GB and NT) and their grand mean (mean of means) - identified at the top right of each set of graphs.



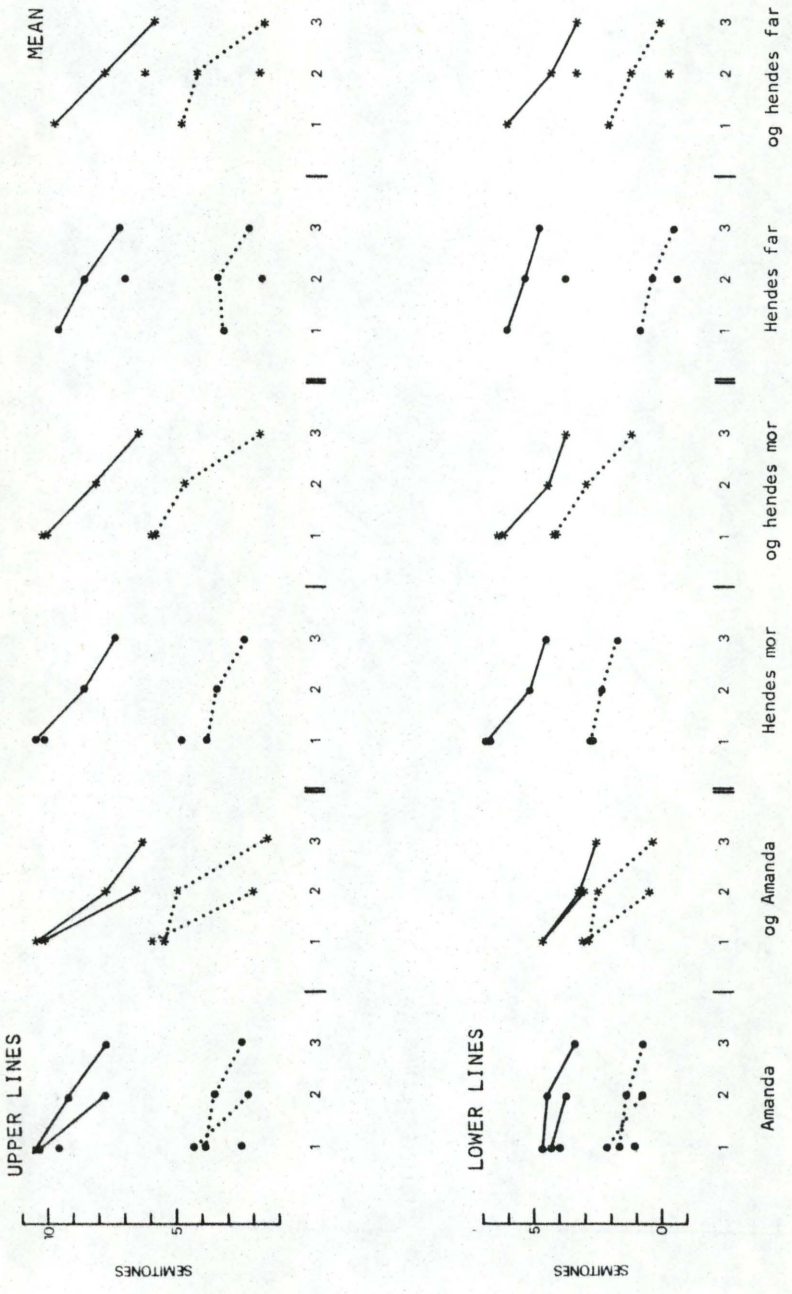
(fig. 4 continued)



(fig. 4 continued)



(fig. 4 continued)



(fig. 4 continued)

component is sensitive not only to what precedes it but also to upcoming events. This argument extends to isolated versus text initial sentences, of course. See further section III below.

2. UPPER AND LOWER LINE ONSET AND OFFSET

In figure 4 beginning and end point frequencies (full vs. dotted tracings) of upper and lower lines in the three pairs of texts are depicted graphically for each speaker and their grand mean (mean of means). From these figures the following points can be made:

First of all, there is no qualitative difference between texts with terminal sentences and texts with coordinate main clauses but

7. changes in onset and offset frequencies across a text are larger in coordinate main clause constructions: all speakers, but with NT the difference between terminal sentences and coordinate clauses is small in this respect. The general trend is a consequence of (2) and (3) above: clause contours are less slanted relative to the same overall decline, so onsets and offsets decline more through the text. See also figure 2.

With this difference in mind, "clause" can substitute for "sentence" at any point in the following four statements (8-11):

8. Upper line onsets decrease smoothly across each text: all speakers, with spurious exceptions to the smoothness by GB and NT. A final upper line onset is 3-4 semitones lower than an initial one, on the average. Upper line onsets are very nearly identical across initial sentences, including the isolated one, irrespective of text length, and across final sentences (excluding the isolated one), i.e. 1-1 resembles 1-2 which resembles 1-3, and 2-2 resembles 3-3: all speakers, except JBC where 1-1 resembles 2-2 (and 3-3). See also figure 2. Initial upper line onsets are also identical across the two boundary conditions: all speakers.

9. Upper line offsets decrease across each text but not quite smoothly, since the drop from 1-3 to 2-3 is smaller than between 2-3 and 3-3: with JBC and NT we even get a "rise-fall", i.e. the end point in 2-3 is higher than in 1-3 but lower in 3-3 than in both preceding ones, most notably in the coordinate clauses, where it is due to the way upper line slopes develop: a steep initial one, then a levelling out medially, and then again a steeper final slope, cf. (6) above. Offsets also decrease by 3-4 semitones (average) from initial to final position. Upper line offsets are very nearly identical across initial sentences (excluding the isolated one), irrespective of text length, and across final sentences, including the isolated one: all speakers. See also figure 2. Final upper line

offsets are identical across the two boundary conditions: all speakers, except GB where they are lower in coordinate main clause texts.

10. Lower line onsets decrease across each text: NRP has one instance of a step up between 1-3 and 2-3, and GB gives no uniform picture. A final lower line onset is 2 semitones lower than an initial one, on the average. In the lower line the onset of an isolated sentence is intermediate between onsets in initial and final sentences in multi-sentence texts (relevant only in the terminals): with NRP and NT it is much closer to text initial lower line onsets, and with JBC it is - conversely - closest to text final onsets. See also figure 2. (If we compare upper and lower lines it is apparent that lower line onsets and offsets span a smaller Fø range than in upper lines, and so the same slightly lower beginning of upper and lower lines in an isolated sentence compared with initial sentences in multi-sentence texts will bring isolated lower line onsets relatively nearer to the onset of lower lines in final sentences.) Otherwise, lower line onsets are very nearly identical across initial sentences (i.e. 1-2 equals 1-3) and across final sentences (i.e. 2-2 equals 3-3): all speakers, with a couple of spurious exceptions. See also figure 2. Initial lower line onsets are identical across the two boundary conditions: all speakers, except GB's *mór*-sentences.

11. Lower line offsets decrease across each text: no speaker is as regular as the grand mean. As with upper lines, lower line offsets are very nearly identical across initial sentences (excluding the isolated one), irrespective of text length, and across final sentences, including the isolated one: with NRP the isolated sentence resembles initial rather than final sentences. See also figure 2. Final lower line offsets are also identical across the two boundary conditions: all speakers, except GB where they are lower in final coordinate clauses (as are her upper line offsets, cf. (9) above).

With a reservation about the onset of the lower line in an isolated sentence (cf. (10)), the ensemble of upper and lower lines in a text span almost the same Fø range, irrespective of text length and boundary condition. The upper and lower lines, respectively, begin at the same frequency in text initial sentences/clauses and end at the same frequency in text final sentences/clauses. Thus, the superordinate declination varies with text length but it is the same in both boundary conditions. The fact that both onsets and offsets of upper and lower lines are identical in initial sentences (1-2 resembling 1-3) and in final sentences (2-2 resembling 3-3) ties up with the conclusion drawn from points (1)-(6) above. The second component in a text is different when it is simultaneously text final than when it is medial, and the same is true of an isolated versus a text initial sentence. Together the upper and lower lines bear testimony to a considerable amount (in terms of temporal scope) of preplanning and look-ahead in the production of intonational events. It is possible - though it remains to be tested - that only three different textual positions are

distinguished, as far as sentence intonation onset and slope are concerned: initial, final, and medial. That is to say, two or more medial sentences or clauses may not be further differentiated among themselves. Such an arrangement of individual text components would be less taxing on the speaker's pre-planning operations (cf. section III).

3. INITIAL UNSTRESSED SYLLABLES

In text initial position the *Amánda*-sentences have only one unstressed syllable, [a], the *mór*- and *fár*-sentences have two (the second of which may be carried by a syllabic nasal as a result of schwa-elision) but only the first one, [hen], will be dealt with here. In text medial position an unstressed *og*, [ʌ], is added to the beginning of each coordinate main clause. This syllable, [ʌ], is - with very few exceptions - higher than the succeeding [a] or [e], and by nearly the same amount for each subject (0.5 to 1.0 semitone with NRP and JBC; 1.0 to 2.0 semitones with GB and NT). Therefore, the conjunction can be derived from [a] and [e] and is not considered any further here.

In figure 5a,b,c the frequency of sentence or clause initial unstressed [a] *(Amánda)* and [e] *(hendes mór, hendes fár)* in initial sentences (5a), final sentences (5b) and across a three sentence text (5c) is depicted, for each speaker and speakers' grand mean (mean of means). Again, I shall state the results for the mean and note the extent to which each subject agrees with the overall tendencies.

12. The pretonic syllable(s) in text initial sentences are constant across texts of different length and different syntactic conditions: all subjects, with exceptions by GB, cf. figure 5a. This (near) constancy in text initial unstressed syllables is analogous to the constant onset of lower lines (i.e. to the constant frequency of the first stressed syllable) across text lengths and boundary conditions, cf. (10) above.

13. Sentence initial unstressed syllables in text final sentences are constant across different text lengths in terminal sentence texts: there are spurious exceptions by NRP, GB and NT, cf. figure 5b. This corresponds to the behaviour of the final lower line onset, i.e. the frequency of the first stressed syllable (cf. (10)). In coordinate main clause constructions the pretonic syllables in the last clause tend to be lower in the longer text: JBC is an exception; his final pretonics are constant in this condition as well. The general tendency is likely to be due to a higher offset of the declinations preceding 2-2 than the ones preceding 3-3, i.e. the final Fø minimum and the (derived) peak in the preceding clause is higher in the first clause of two (1-2) than in the second of three (2-3), see figure 4 ("lower lines"-data points connected by dotted line) and also figure 2.

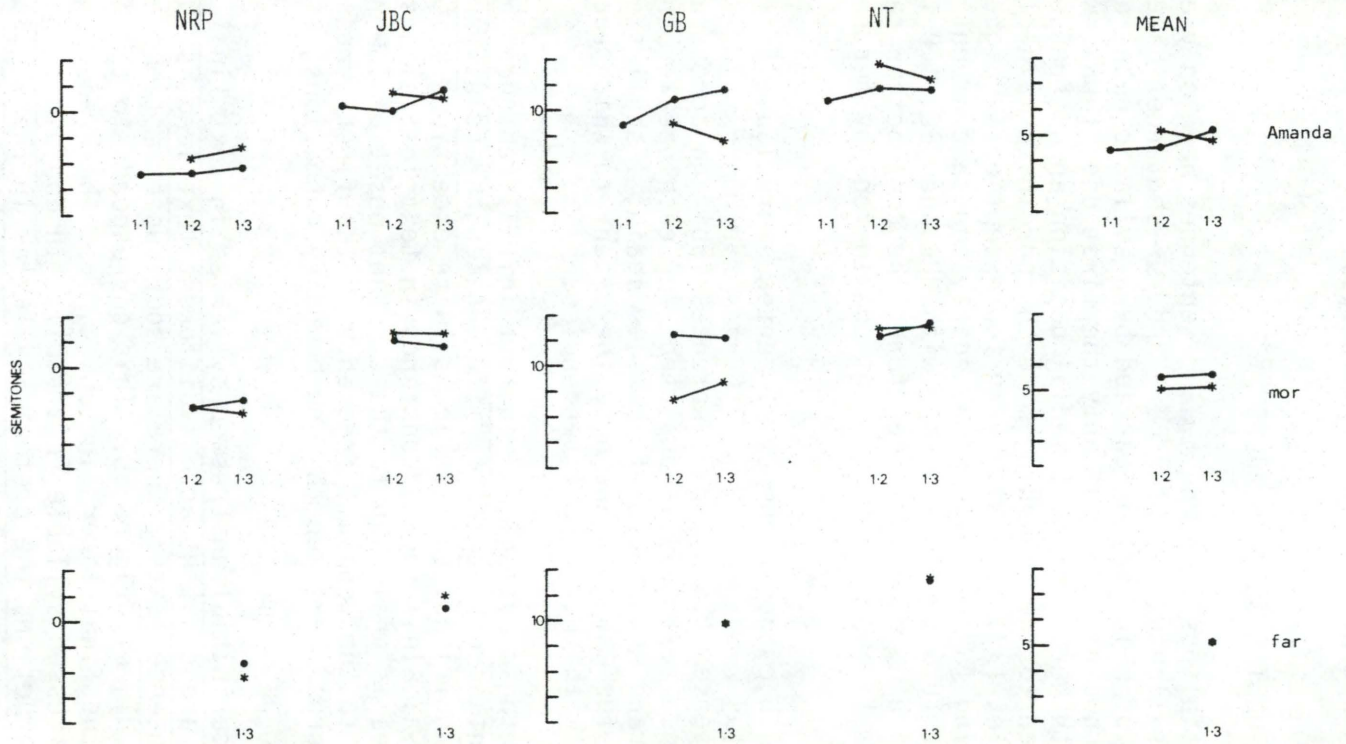


Figure 5a

Frequency values of sentence or clause initial unstressed syllables in the sentences whose upper and lower lines are depicted in figure 2. The three different sentences are arranged vertically, as indicated by key words to the right of each graph. Data pertaining to declarative terminal sentences are indicated with points, coordinate clause data with stars. Data from text initial sentences (including the isolated one) and clauses are plotted in figure 5a, data from text final sentences (including the isolated one) and clauses are plotted in figure 5b, and data from initial, medial and final items in three component texts are plotted in figure 5c. Zero on the logarithmic frequency scale corresponds to 100 Hz. Four speakers, two males (NRP and JBC) and two females (GB and NT) and their grand mean (mean of means) - identified at the top of each figure.

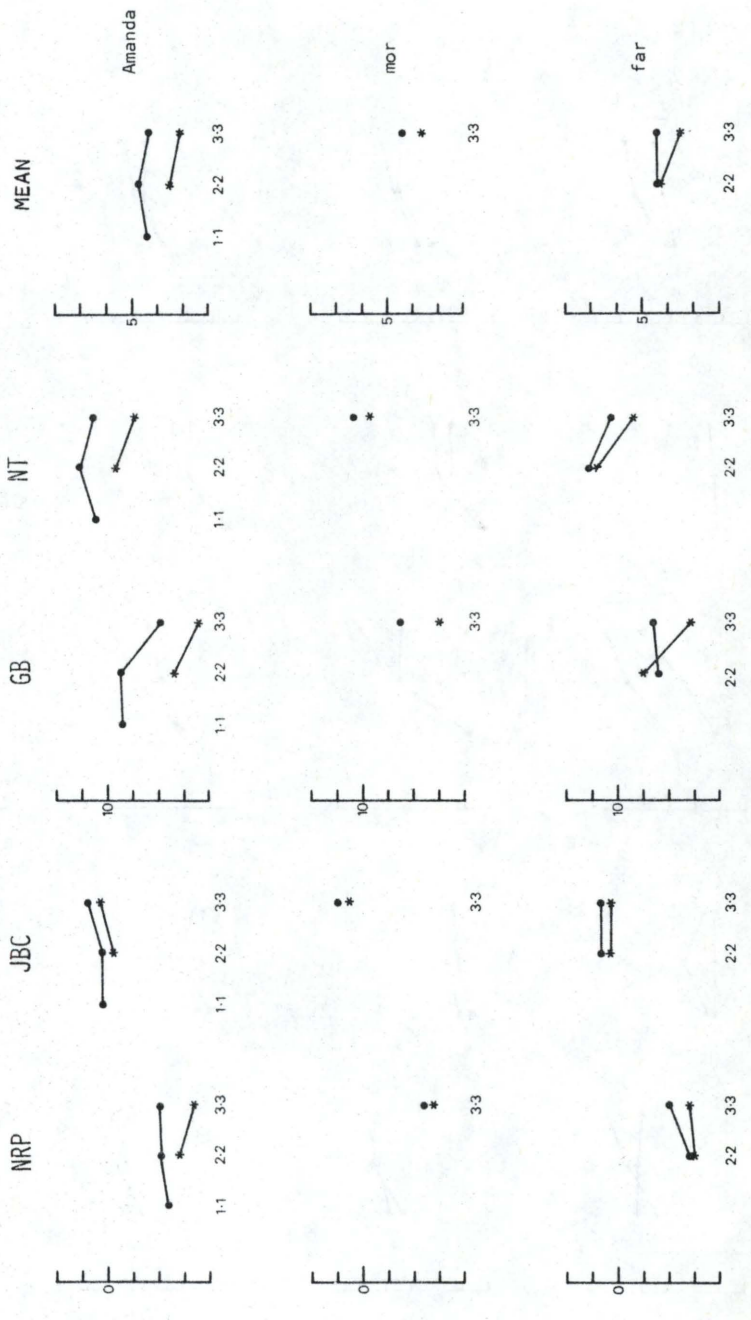


Figure 5b

See the legend to figure 5a

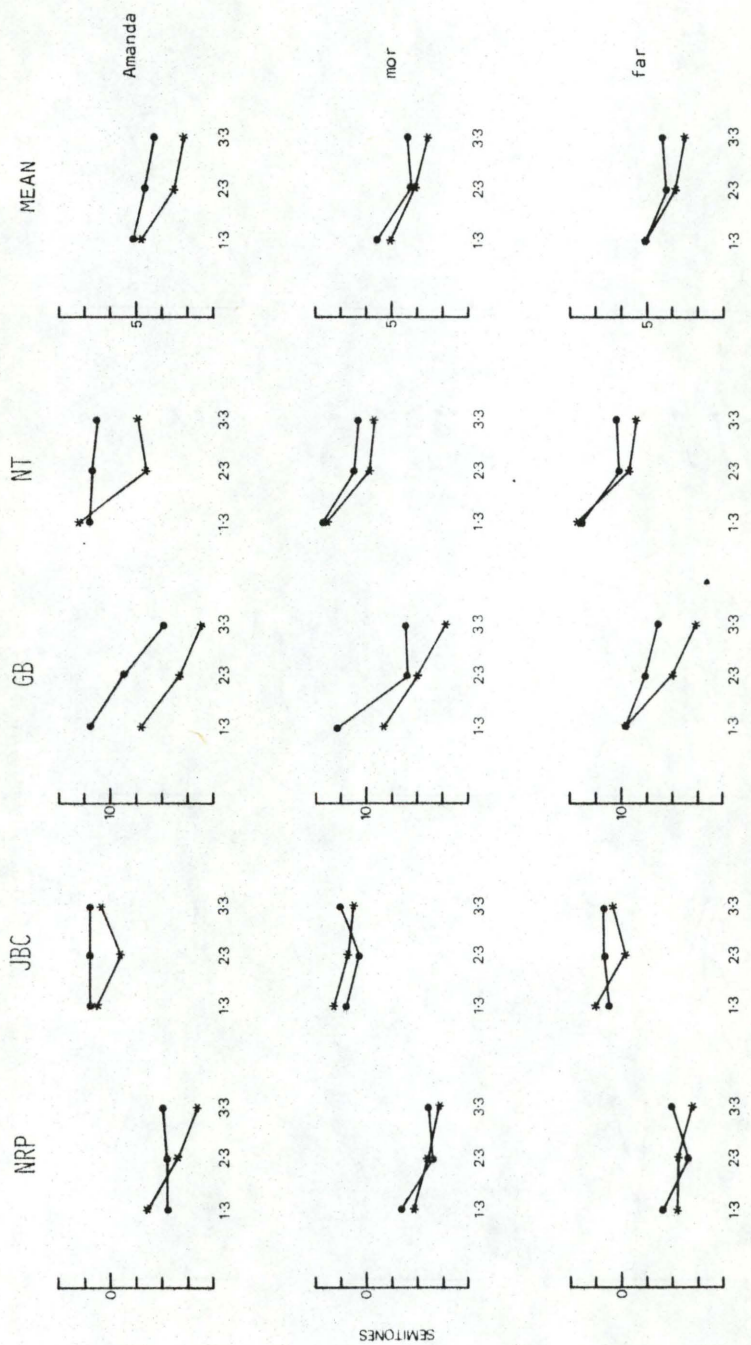


Figure 5c

See the legend to figure 5a

At least this difference is greater, and corresponds better to the difference observed in the clause initial unstressed syllables in figure 5b than does the slight or even reverse differences in the onset of lower lines in 2-2 and 3-3. This would mean that clause initial unstressed syllables in text medial position are governed tonally by the offset of the preceding clause, or - in other words - such syllables join up prosodically with the last stress group in the preceding clause. They are neither isolated tonally, nor do they attach themselves to the clause to which they belong syntactically. This interpretation gets a certain substantiation from figure 1 (and from the corresponding longer texts not reproduced here): there are a majority of instances (exceptionless with GB and nearly so with NT) where clause initial unstressed syllables perform a smooth and continuous fall together with the posttonic in the preceding clause; together with the last stressed syllable in the preceding clause the ensemble of unstressed syllables thus perform the low plus high-falling Fø pattern characteristic of prosodic stress groups in Standard Danish (cf. Thorsen 1980b, 1984). Only with JBC are there no such cases of prosodic association across clause boundaries, maybe because he (as the only one) paused between the coordinate clauses. (I do not think that a pause necessarily disrupts a tonal pattern (apart from the introduction of a silent interval), and clause boundaries per se apparently do not, but the combination: pause and clause boundary may be more than an Fø pattern can survive.)

14. Initial unstressed syllables through a multi-sentence text decrease slightly through a string of terminal sentences: the variation is small and non-uniform with NRP, nearly nil with JBC, apparent only in NT's *mór*- and *fár*-sentences. With GB the decrease is considerable in the *Amánda*- and *mór*-sentences, cf. figure 5c. Again, this is analogous to the onset of lower lines, i.e. to the behaviour of the first stressed syllable in each clause in the texts, cf. (10) above. In coordinate main clause texts the decrease is greater and generally more uniform: all speakers. Again, this is more in line with the way upper line offsets develop through a text than with lower line onsets (cf. figure 4), i.e. it seems to be caused by the tonal association of the clause initial unstressed syllables with the preceding clause final stress group, cf. above.

What points (12)-(14) demonstrate is merely a tonal dependency of initial unstressed syllables in the sentence or clause upon other parameters: In text initial position, initial unstressed syllables are constant (for a given speaker), as is the first stressed syllable, across different text lengths and different boundary conditions. After a text medial sentence boundary, initial unstressed syllables likewise associate with the succeeding stressed syllable, i.e. differences in frequency location across different text positions are analogous with and can be derived from differences in the first stressed syllable in the sentence. After a clause boundary (unaccompanied by any pause) clause initial unstressed syl-

lables (often but not invariably) associate tonally with the preceding prosodic stress group, i.e. differences across different text positions are analogous with the differences in the last tonal pattern in the preceding clause. In other words, sentence and clause initial unstressed syllables in text medial position may be extrametrical phonologically (and rhythmically as well, after a sentence boundary), but tonally they are not isolated entities; they vary in a principled fashion in association either with a preceding or a succeeding prosodic stress group.

C. DURATION

I have measured the duration of each sentence and clause, as the time interval between the first and last Fø measuring point, excluding the conjunction in the coordinate clauses. Average values for each speaker, and speakers' mean are given in Table I.

The differences in duration of a given sentence across text positions - in one or the other boundary condition - range between 16 cs (NRP's coordinate *fár*-clauses) and 3 cs (NRP's terminal *fár*-sentences), i.e. 13.7 and 2.5%, respectively, of the shortest sentence/clause. The most consistent trend in the durations is that an initial sentence or clause is shorter than the succeeding, medial or final one (true in 29 out of 32 instances). Secondly, a final sentence or clause tends to be shorter than a medial one in three sentence texts (true in 16 out of 24 instances). Isolated sentences are not systematically different from sentences in context: though with JBC the isolated sentence is clearly longer than initial sentences (6 and 7 cs), it is only 2 or 3 cs longer than medial and final sentences. With GB the isolated sentence is 7 cs longer than the final sentence of three but only 1 cs longer than initial sentences. Lehiste (1975) found that isolated sentences were longer than in three sentence texts. To judge from the present data, such a finding may not generalize to all speakers. Furthermore, coordinate main clauses tend unexpectedly to be longer (by around 6 cs) than terminal sentences (which is not due to the conjunction *og* which was excluded from the measurement). The opposite relation would perhaps have been less surprising: A terminal sentence develops freely, a non-terminal clause is shortened by the succeeding clause. I have no explanation to offer at the moment for the longer coordinate clauses.

A more thorough treatment of durational differences warrants a separate investigation. The only conclusion I wish to draw at present is that sentences or clauses are shorter text initially than in later positions but - somewhat surprisingly maybe - isolated and text final sentences or clauses are not systematically longer than sentences in other positions in the text.

Table I

Average durations in centiseconds of three different sentences in various contexts: isolation ("1-1"), initially ("1- "), medially ("2- "), and finally ("3- ") in texts consisting of two (" -2") or three (" -3") sentences, and in two different boundary conditions, as indicated to the left. See further the text.

	NRP	JBC	GB	NT	mean
Amánda					
terminal sentences					
1-1	128	135	119	137	129.8
1-2	128	128	118	131	126.3
2-2	130	132	117	136	128.8
1-3	131	129	118	136	128.5
2-3	132	133	118	137	130.0
3-3	133	133	112	143	130.3
coordinate clauses					
1-2	131	136	114	136	129.3
2-2	139	146	122	141	137.0
1-3	129	136	120	134	129.8
2-3	134	141	125	141	135.3
3-3	136	135	120	137	132.0
Hendes mór ...					
terminal sentences					
1-2	148	157	140	145	147.5
1-3	148	153	139	146	146.5
2-3	156	157	144	149	151.5
3-3	152	157	145	148	150.5
coordinate clauses					
1-2	151	154	141	150	149.0
1-3	152	153	140	150	148.8
2-3	159	160	145	156	155.0
3-3	159	160	141	153	153.3
Hendes fár ...					
terminal sentences					
2-2	121	138	111	124	123.5
1-3	121	128	118	127	123.5
2-3	120	132	116	131	124.8
3-3	118	129	109	126	120.5
coordinate clauses					
2-2	122	144	118	128	128.0
1-3	117	133	114	130	123.5
2-3	133	135	124	134	131.5
3-3	123	130	121	131	126.3

D. SUMMARY OF THE RESULTS ON INTONATION

A short text, composed of two or three simple declarative sentences or corresponding clauses, of which none is excessively long, will be associated with a superordinate Fø declination which supports individually slanting sentence/clause components. The most interesting and consistent further findings, i.e. where the inter subject agreement is best, are points (2, 3) and (8 - 11): Upper and lower lines - which characterize individual sentences or clauses - are steeper and with greater amounts of resetting between them in a succession of declarative terminal sentences than in a corresponding succession of coordinate main clauses. Concomitantly, the onset and offset Fø values of these upper and lower lines decrease across a text, more so through coordinate main clauses, because individual lines are less slanted relative to the same superordinate decline. The decrease is greater in upper than in lower lines, ceteris paribus, because upper lines - individually and as a whole - span a greater Fø range than do lower lines. These facts are depicted in the stylized graph in figure 6, which shows the course of upper and lower lines in two different three-sentence texts, one with declarative terminal sentences (full lines), and one with corresponding coordinate main clauses (dashed lines). (Differences in duration and pause length are obliterated.) The dotted lines reflect the overall downdrift which is identical in the two boundary conditions. They connect the (upper and lower) onset values in

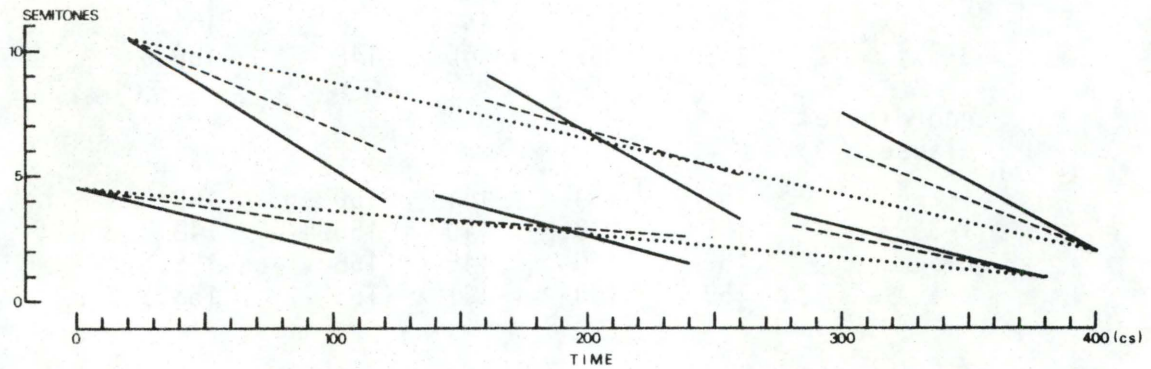


Figure 6

Stylized model of the course of upper and lower lines in texts with three declarative terminal sentences (full lines) and three coordinate main clauses (dashed lines). The dotted lines reflect the overall downdrift through the text, which is identical in the two boundary (sentence or clause) conditions. A text with two components is obtained by leaving out the middle section in the figure and moving together the initial and final components across the gap. An isolated sentence results if the dotted lines are suitably compressed in time.

the first sentence/clause with the (upper and lower) offset values of the last one. These values can be considered constant for texts of different length (and with different internal (sentence or clause) boundaries), which is another important aspect of the results. An isolated sentence results if these dotted lines are suitably compressed in time. A text with two sentences or clauses is obtained by leaving out the middle section of figure 6 and moving together the initial and final components across the gap. Since the frequency span is (assumed to be) constant, an isolated declarative sentence will decline more steeply than the ensemble of two or more sentences in a text: The results actually do show that isolated slopes are at least as steep as the steepest (initial) sentence in multi-sentence texts. Text initial unstressed syllables are likewise constant across different texts, whereas text medial, sentence or clause initial, unstressed syllables vary in a principled fashion in association with a succeeding and preceding, respectively, prosodic stress group.

It should be evident from the account of the results that the description is true of the overall, average picture exhibited by the four speakers in this investigation. If either one of the four speakers had been the only one in the experiment, the conclusions would have been slightly different on some points, because a certain amount of inter and intra individual variation or "noise" exists. However, this variation is random, i.e. one speaker may differ in one direction from the main trend in one aspect, and in another direction in another aspect; or the same speaker may deviate on a certain point from the main trend in only one of the three different sentences recorded here. Therefore, the presentation of a grand mean over four speakers is possible, on the one hand, while at the same time we can assert that no individual is as regular in his or her behaviour as the resulting model (figure 6) would predict. Speech production is a more plastic process than such a model can capture, also where the intonational structuring of multi-sentence/clause texts are concerned (cf. Thorsen 1984, where I reach a similar conclusion about the production of Fø patterns in prosodic stress groups).

The less steeply declining upper and lower line slopes in coordinate clauses vis-a-vis terminal sentences, ceteris paribus, corresponds well with the observations in Thorsen (1978) that isolated terminal declarative sentences have steeper sentence intonation contours than corresponding non-terminal (declarative and interrogative) clauses.

The fact that different text internal boundaries (clause boundary or sentence boundary) induce differences in intonational manifestation is not a reflection of a general, tight syntactic/prosodic interplay, cf. Thorsen (1980a) - which treats word boundaries and Fø patterning - and Thorsen (1983a) - where long simple declarative isolated sentences are analysed and the relation between intonation contour resettings and sentence internal syntactic boundaries are discussed.

E. COMPARISON WITH OTHER LANGUAGES

First of all, the present results - like Lehiste's (1975) and Bruce's (1982) - clearly contradict Nakatani's conclusion that speech features [including prosodic ones, NT] can be reasonably generalized to sentences in a coherent context. The upper and lower line(s) of a terminal declarative sentence span a greater frequency range in isolation than in a larger text context. Secondly, Standard Danish agrees well with the gross text intonation features of English and Swedish. All three languages (and Japanese as well, cf. Uyeno et al., 1979) exhibit a superordinate text intonation structure upon which are superposed individual sentence or clause components. There are a number of (minor) differences, however. I cannot assert, as did Lehiste (1975) that an isolated sentence is longer than the same sentence in context. Lehiste further found that Fø in the beginning of a sentence is lower in an isolated one than in paragraph initial position. This is not true here if we consider the very first unstressed syllables, which have a constant Fø value independently of the length of the succeeding text. If we consider the Fø value of the first stressed syllable and especially the first succeeding Fø peak (in the first posttonic syllable), then Lehiste's finding is true of one speaker but the isolated sentence is only slightly lower (less than one semitone) than paragraph initial ones with three speakers. Lehiste's observation that Fø in the beginning of a sentence is higher in paragraph initial than in medial and final position holds for Standard Danish as well, no matter what Fø point we consider to be the beginning, and this difference is not negligible.

Bruce's (1982) observations from his larger material, that there are initial and final Fø values which are constant across texts of different length, hold true of Standard Danish as well, though these constant points pertain to different units in the chain of syllables in the two languages. In Southern Swedish the very first and very last syllable lie at the floor of the range of normal Fø variation. But the first local Fø maximum and minimum (in the first stressed plus unstressed syllable) in Swedish is lower in an isolated sentence than in context (corresponding to Lehiste's (1975) observation), and it is higher in a longer than in a shorter text. The last local Fø maximum and minimum is constant across texts of different length. In the present material, not only extrametrical sentence initial unstressed syllables but the frequency location of the whole of the first prosodic stress group is almost insensitive to text length. Thus, the total range of variation varies proportionately (upwards) with text length in Southern Swedish, whereas it is constant in Danish. Concomitantly, superordinate text declination will vary more in Danish than in Swedish with the duration of the text.

Before the text internal pauses in Bruce's recordings Fø drops to the floor of the range (where also absolute initial and final syllables are constantly located). Such a marking of text medial boundaries is not replicated here. This may have

to do with the tonal relation between stressed and unstressed syllables in Standard Danish: the first post-tonic syllable rises above the preceding stressed one, succeeding post-tonics describe a more or less steep fall. In the present sentences there was only one sentence/clause final post-tonic, and a drop to the floor of the Fø range is therefore not possible. The high sentence final fall in Bruce's material may, however, also be due to a sentence accent, which evokes a particularly elaborate Fø movement in the posttonic syllables, and Danish lacks such a nuclear accent (cf. Thorsen 1980b). According to a personal communication by Gösta Bruce, Southern Swedish may also be produced without a final focal accent, but it is not clear to me whether this was in fact the case in his (1982) recordings.

In Bruce's texts, the Fø minimum in the first accent after a text medial sentence boundary continues at the same Fø value that the Fø minimum had in the last accent in the preceding sentence. He sees that as a tonal coupling or assimilation of the succeeding constituent sentence to the preceding one, see further below, section III. I am inclined to think, though, that this relation is an artifact of his material, due to the particular length of his sentences (two stressed words in each). With longer sentences he might have found the onset of a succeeding sentence to be higher than the offset of the preceding one, as it is in the present Standard Danish material, cf. figure 6. Incidentally, the Danish data confirm Cooper and Sorensen's (1981) observation, that the resetting between clauses in a complex sentence is independent of any accompanying pause.

Uyeno et al. (1979) found no difference between coordinate main clause constructions and successions of simple terminal sentences in Japanese. In Standard Danish, the closer relation between coordinate main clauses is accompanied by a more integrated and coherent intonational structure than a series of terminal sentences.

IV. ON THE ABSTRACT REPRESENTATION OF INTONATION

By "abstract representation" I mean the level where significant and distinctive choices are made and from which the phonetic implementation can be derived by rules. It contains no redundant specifications. This representation may correspond to a stage (early) in the actual speech production process. Thus, a representation of intonational data does not necessarily become more abstract by mere transformation of frequency scales, or by transcription of Fø curves in terms of high and low points or contour shapes. The statement may seem trivial, but such a concept of "abstract" is not a matter of course, at least not in some of the present day phonological schools.

In this section I shall briefly present two different, current models of intonation, the hierarchial and the tonal sequence representation; discuss two particular claims about the shortcomings of the hierarchical theory; and discuss the present data in the light of both representations.

A. THE HIERARCHICAL AND THE LINEAR INTONATION THEORIES

On a number of previous occasions, most recently in Thorsen (1983a), I have presented data and arguments in favour of a model of Standard Danish intonation in terms of a hierarchical, layered system of simultaneous, non-categorial intonational components of varying temporal scope, from sentences through prosodic phrases and prosodic stress groups to segments. I shall not repeat the reasoning here, but note that to the hierarchy of intonational components we may apparently add a textual contour. Similar views of the composition of fundamental frequency courses - with such modifications as language specific differences impose - are expressed about Japanese in Fujisaki et al. (1979), about French in Vaissière (1983), about Dutch in 't Hart and Collier (1979), and about Swedish in Gårding (1983) and Bruce (1977).[1] However, recently Bruce (1982) seems to be leaning more towards the basic concept of the linear theory of intonation as formulated by Pierrehumbert (1980).

Pierrehumbert (1980) analyses English intonation in terms of a linear sequence of categorially different, non-interacting pitch accents, which are associated with the stressed syllables of the utterance. They consist of either a high tone (H), a low tone (L), or bitonal combinations of the two. The inventory of pitch accents includes a (final) phrase accent and a boundary tone. The resulting description is very elaborate and very elegant, if occasionally somewhat complicated. Overall slope - if it is too steep to be accounted for by other factors - is handled by a downstep rule which lowers a H tone by a constant factor relative to the preceding H, in certain environments. Everything else being equal (such as the prominence relations between the stressed syllables of the utterance), this rule will create asymptotically declining slopes. Liberman and Pierrehumbert (1983) modify and expand the theory to include variations induced by changes in pitch range, as well as a lowering of the final pitch accent. Once the Fø value of the initial pitch accent is determined, such a theory requires no pre-planning, no look-ahead, on the part of the speaker for the execution of intonational phenomena, and "In general, we see no evidence that the Fo implementation for an entire phrase is necessarily laid out before speaking begins, even when the phrase is known in advance and fluently produced. All of our measurements can be modeled quite well on a left-to-right, plan-as-you-go basis." (Liberman and Pierrehumbert, 1983. The quotation is from the manuscript; I do not have the final version at the time of writing). This is also made explicit by Pierrehumbert (1983, p. 141):

"Superficially global trends arise from iterative application of these local rules." and "We also deny that intonation is built up in layers, by superposing local movements on a global component."

Pierrehumbert and Liberman see the data on Standard Danish intonation (specifically the fact that the slope of the declination varies with utterance type and function, cf. e.g. Thorsen 1978 and 1980c) as a challenge to the idea that intonation can be generated by local rules, i.e. rules with a narrow domain or temporal scope (extending backwards). They meet the challenge with a proposal for a downstep rule (whose specific formulation is different from the one they posit for English) where either the downstep factor or the reference line (which is part of the formula for target value calculation) is allowed to vary with sentence type. (Thorsen 1983a+b address this specific issue.)

The essential difference between the two types of intonation model, a hierarchical versus a linear one, is not merely formal or notational but conceptual or ideological, in a hackneyed term. One theory claims - or implies tacitly - that certain, gross aspects of the intonation of an utterance are anticipated and planned for, or laid out, at the moment the utterance is initiated. The other theory denies the existence of such preplanning and look-ahead, except that variation in total length may be staked out in the value of the very first pitch accent. One sees the fundamental frequency contours of utterances as a composite of contributions from, among others, an intonational and an accentual component. In the linear theory intonation is made up of, consists of, a sequence of pitch accents. These aspects, the layered versus simply sequential and the global versus local planning of tonal events, are what Pierrehumbert (1980) and Liberman and Pierrehumbert (1983) are most concerned with when they opt for the linear, tonal sequence model. They do not concern themselves so much with, e.g., descriptive adequacy. This and other problems are treated extensively by Ladd (1983a and b).

B. LADD'S CRITICISM OF THE HIERARCHICAL REPRESENTATION

Ladd (1983a and b) modifies Pierrehumbert's (1980) theory while staying within its general framework of linearity and locality. He restricts the inventory of pitch accents and replaces the downstep rule with a downstep feature. Ladd adds a couple of other features and claims (I think rightly) thereby to achieve a less complex description of intonational phenomena. Ladd (1983a) specifically treats the generation of overall declining intonation contours without recourse to preplanned grid-lines (cf. Gårding, 1983), or baselines and plateau's (cf. Vaissière, 1983) or other such abstract contour shapes, among which he counts the sentence intonation component which I have posited for Standard Danish (e.g. Thorsen, 1978 and 1983a). In his discussion of the relative merits of the

two types of intonation models, he makes a couple of statements about the hierarchical model which I would like to address, because they are likely to obscure the issue.

Ladd (1983a, p. 40) sees a difference in the two theories in their weighting of phonetic explicitness against functional relevance. Functional generalizations are conveniently expressed in descriptions of overall shape or slope of sentence contours in the hierarchical model, but phonetic detail is ironed out or lost, he says. Once such detail is admitted, Ladd claims, the number of different sentence intonation types multiplies rapidly, and he sees no basis for extracting further, explicit functional generalizations. Ladd refers here to the results on sentence intonation in long sentences in Danish (Thorsen 1983a): At and above five prosodic stress groups, sentence intonation contours are not straight lines but become bumpy, due to partial resetting of the declining slope. The theme is picked up again later (p. 50), where Ladd asserts that not all hierarchical models assume that the lines connecting accents should be straight, since "...*Thorsen..admits the existence of irregular intonation lines, though she is at a loss to account for them or fit them into her overall view of sentence intonation.*" This conclusion is Ladd's, not mine; moreover, it is incorrect. The fact that long intonation contours are bumpy does not exclude them from being globally preplanned, and it does not prohibit the extraction of explicit functional generalizations (about sentence type and the like). But the intonational realization of the same function, say terminal declarative, may come out in different shapes, or variants, depending - inter alia - on the length of the utterance. A complication arises from the fact that the location of the bumps (the resettings) in an intonation contour is determined in an intricate and - so far - nowhere near fully investigated interplay with the syntax and semantics of the utterance. This is (probably) what Ladd alludes to when he claims that "irregular" contours cannot be accounted for. However, the problems of prosodic, syntactic and semantic interplay are not specific to the hierarchical model of intonation, and they are problems that await a massive amount of empirical studies for their solution. Thus, I would still like to claim that the variation in sentence intonation contour shape is principled (allowing, of course, for inter as well as intra speaker variation or "noise"); i.e. it is context determined, bound variation, even though the governing principles are not all known to us yet.

C. PRE-PLANNING TEXTUAL CONTOURS

Thorsen (1983a+b) contain a discussion of some specific drawbacks that I think the tonal sequence theory faces in the description of Danish intonation. Suffice it here to repeat the conclusion: If the downstep rule or feature is to capture the amount of variability and complexity in intonation contours in simple sentences of varying length and syntactic and semantic make-up, it must be a scalar and continuous feature which will

often have to take different values within one and the same phrase or sentence. I.e. it must have at least some of the properties it presumably was intended to rid intonation analysis of. - I am of course fully aware that the hierarchical model faces just as serious problems where the formalization and description of fundamental frequency data are concerned, but I repeat that the real difference between the two theories is neither formal nor notational but lies in their assumption of hierarchical, globally preplanned versus sequential, locally determined tonal events.

Before I proceed, let me stress the fact that the texts accounted for in section III were composed solely of declarative sentences and clauses. I do not know how a series of questions, strung together into a text, would have behaved, or alternating declarative and interrogative sentences. Nor can I make any claims about longer texts or texts composed of shorter or longer constituent components. And last, but not least, the present analysis pertains to highly planned (read) speech. The production of an ensemble of utterances, constituting a coherent text, may be much less regular in free, spontaneous, non-monitored speech.

The discussion to follow can be limited to the behaviour of the lower lines, since the upper lines are closely correlated with and can be derived from them, cf. section III.B.1 above. In itself this fact may be taken as evidence for the layered composition of fundamental frequency: the lower line is the connection of the stressed syllables of an utterance, the intonation contour proper in my framework. The upper line connects the first post-tonic syllable in the prosodic stress groups. A stress group may or may not have post-tonics, i.e. there may or may not be centres of suspension for an upper line. However, the lower line is oblivious to this vacillation. The stressed syllables are frequency scaled in relation to each other, without regard to the presence or not of any "highs" in the surroundings. The lower line is a constant, ceteris paribus, upon which high-falling patterns are superposed whose range and extent and detailed shape are determined by the context (earlier or later, on a more or less sloping contour) and by the duration of the prosodic stress group (in terms of number of syllables). See further Thorsen (1980b and 1984).

First of all, I find it difficult to accept that patterns such as those in figure 6 could be obtained without a fair amount of pre-planning and look-ahead on the part of the speaker. More specifically, the different arrangement of the individual sloping (lower) lines according to the strength of the boundary between otherwise identical sentences in a text can only be the result of look-ahead and pre-planning, it seems to me. The step down between the first and second, and the second and third stressed syllable, i.e. the slope of the line connecting them, is different when the unit is bounded by a sentence boundary versus a clause boundary, in a manner which makes three successive coordinate clauses less differentiated and more co-

herent relative to the overall declination exhibited by the whole text than three terminal sentences in succession, cf. figure 6.

The other - as I see it - crucial point ties up directly with a very explicit hypothesis about the domain of phonetic implementation rules in Liberman and Pierrehumbert (1983 - once again the quotation is from their manuscript):

"Now comes the conjectural part: we insist that the computation of any parameter or object Y[i] can only depend on the "accessible" properties of Y[i] and Y[i-1], where Y[i-1] means the immediately previous object of the same type (if any). Thus, pitch accent can look back to previous pitch accent, phrase to previous phrase, etc.

... A restriction of this type has a certain functional value, for both speakers and hearers; speakers can get on with the task at hand without knowing all the details of what follows, while hearers can in principle complete the phonetic processing of what they have heard up to any given point in the stream of speech.

Many apparent instances of anticipatory effects are known. Unless our conjecture is wrong, all such cases must turn out to be explained either by feature spreading at the phonological level, or else by computation of some parameter of a higher level constituent."

I will not argue the functional value of this conjecture. But, functional or not, I think it is clearly contradicted by the following fact (cf. III.B.1. and B.2. above): the slope of a sentence or clause contour, occupying a given slot in the left-to-right order of components, takes different values according as it is succeeded by one or more components, or not. Thus, an isolated sentence has a steeper slope, the downstep is greater between successive stressed syllables, than when the same sentence is succeeded by one or more (that is immaterial) sentences in the same text. And sentence or clause no. 2 from the left is steeper when it is simultaneously text final than when it is text medial. In fact, at least three positions are distinguished in longer texts: initial, medial, and final. Note specifically that the variation with text length (isolated sentence versus multi and two versus three sentence text) is not primarily expressed in a lower or higher beginning of the whole contour. (Such a phenomenon could be accommodated in Liberman and Pierrehumbert's representation by manipulating the overall range and the derived reference line against which the initial pitch accent target value is computed.) On the contrary, beginning and end points are essentially constant across texts of different length. The overall textual contour slope varies, and so do the slopes of its individual components, according not only to preceding, but also to upcoming events.

This last observation differs from Bruce's (1982) conclusion. As mentioned in section III.E above he found that an Fø minimum in the first accent after a text medial sentence boundary continues at the same value that the corresponding Fø minimum had in the last accent before the boundary. He says (p. 285): *"The observed tonal adaptation of the earlier part of a succeeding sentence to the later part of a preceding one within the same textual unit can be taken as an indication that there is a partial rather than a total preplanning of the global Fø course in relation to the length of the actual text unit."* If - as I speculated above - this particular intonational configuration of successive sentence components is an artifact of the material, and if it would not be replicated with longer individual sentence components, then Bruce's hypothesis of partial rather than total preplanning is no longer so easily sustained.

Similar contradictory evidence against exclusively backwards extending tentacles was found in intonation contours in declarative terminal sentences of varying length (Thorsen 1983a). The second stressed syllable is lower, relative to the first stressed syllable, when it is simultaneously sentence final than when another stressed syllable follows, and it is lower succeeded by one than by two stressed syllables; likewise for the third stressed syllable. The phonetic implementation is demonstrably not independent of or insensitive to succeeding events.

If data of the kind presented in this paper is to be accommodated within the theory of locally determined tonal sequences, say Liberman and Pierrehumbert in the last paragraph of the quotation above, it must be explained either as feature spreading at the phonological level, or else by computation of some parameter of a higher level constituent. I cannot see that feature spreading at the phonological level is anything but the linguist's way of formally accounting for the preplanning of articulatory events; it does not "explain" these events. And I cannot see that referring anticipatory effects to a higher level constituent is anything but admitting the existence of hierarchically organized tonal events.

Experiments like the present one are the most favourable we can imagine if we want to know what speakers can achieve by way of precision and constancy in speech production tasks. Not only is the speech monitored and planned for them, but they get to repeat the same items over and over (though not in straight succession). Even so, the signals we pick up are "noisy". This intra speaker variation (as well as the more considerable variation among speakers) is perhaps less disturbing, less alien, to the representation of intonation in terms of gross, global events which are filled in with more detailed contributions (i.e. of lesser temporal scope). Precisely because the actual execution of tonal events is the result of interacting larger and smaller scale components (because they are expressed in the same medium), we should not expect these events to be performed with mathematical precision. It seems

to me that the tonal sequence theory, and the concept that gave birth to it, focus too heavily on just such a mathematical precision and thus ignores or at least underrates the plasticity involved in all aspects of speech production. With this last paragraph I am also implying that free, non-monitored texts will hardly be as regular as the ones investigated here. That is not to say that it would not be an interesting and worthy object of study - quite the contrary.

V. NOTE

1. I am aware that Ladd (1983a and b) sees 't Hart and Collier's description of Dutch as belonging within the tonal sequence theory of intonation, and Gårding's and Bruce's Swedish model as a compromise model between the hierarchical and linear theories. I disagree with Ladd's interpretation. This controversy is, however, not central to my present purpose, and will not be dealt with any further here.

ACKNOWLEDGEMENT

My thanks are due to Jørgen Rischel for valuable comments. Jeanette Holtse competently performed the measuring and graphic work.

REFERENCES

Bruce, G. 1977: "Swedish word accents in sentence perspective", *Travaux de l'Institut de Linguistique de Lund 12*, p. 1-155

Bruce, G. 1982: "Textual aspects of prosody in Swedish", *Phonetica 39*, p. 274-287

Cooper, W.E. and Sorensen, J.M. 1981: *Fundamental Frequency in Sentence Production*, (New York: Springer Verlag)

Fujisaki, H., Hirose, K. and Ohta, K. 1979: "Acoustic features of the fundamental frequency contours of declarative sentences in Japanese", *Ann. Bull. Res. Inst. Logopedics and Phoniatrics 13*, p. 163-173

Gårding, E. 1983: "A generative model for intonation", in: A. Cutler and D.R. Ladd (eds.) *Prosody - Models and Measurements*, (Berlin: Springer-Verlag), p. 11-25

't Hart, J. and Collier, R. 1979: "On the interaction of accentuation and intonation in Dutch", in: E. Fischer-Jørgensen, J. Rischel and N. Thorsen (eds.) *Proc. 9th Int. Cong. Phonetic Sciences, vol. II*, (Copenhagen: Institute of Phonetics, Copenhagen University), p. 395-402

Hombert, J.-M. 1977: "Development of tones from vowel height?", *J. Phonetics 5*, p. 9-16

Jeel, V. 1975: "An investigation of the fundamental frequency of vowels after various consonants, in particular stop consonants", *Ann. Rep. Inst. Phon. Univ. Copenhagen 9*, p. 191-211

Ladd, D.R. 1983a: "Peak features and overall slope", in: A. Cutler and D.R. Ladd (eds.) *Prosody - Models and Measurements*, (Berlin: Springer-Verlag), p. 39-52

Ladd, D.R. 1983b: "Phonological features of intonational peaks", *Language 59*, p. 721-759

Lehiste, I. 1975: "The phonetic structure of paragraphs", in: A. Cohen and S.G. Nooteboom (eds.) *Structure and Process in Speech Perception*, (Berlin: Springer-Verlag), p. 195-206

Lehiste, I. 1979: "Perception of sentence and paragraph boundaries", in: B. Lindblom and S. Öhman (eds.) *Frontiers of Speech Communication Research*, (London: Academic Press), p. 191-201

Lehiste, I. 1980: "Phonetic characteristics of discourse", *Ac. Soc. Japan, Transactions of the committee on speech research, April 1980*, p. 25-38

Lehiste, I. and Wang, W. S.-Y. 1977: "Perception of sentence boundaries with and without semantic information", in: W.U. Dressler and O.E. Pfeiffer (eds.) *Phonologica 1976, Innsbrucker Beiträge zur Sprachwissenschaft* (Institut für Sprachwissenschaft der Universität Innsbruck), p. 277-283

Liberman, A.M., Cooper, F.S., Shankweiler, D.S., and Studdert-Kennedy, M. 1967: "Perception of the speech code", *Psychological Review 74*, p. 431-461

Liberman, M. and Pierrehumbert, J. 1983: "Intonational invariance under changes in pitch range and length", in: M. Aronoff and R. Oehrle (eds.) *Language Sound Structure*, (Cambridge: MIT Press)

Nakatani, L. 1975: "Sentence prosody in isolated and in story contexts", *J. Acoust. Soc. Am., Supplement 1*, p. 120

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

Pierrehumbert, J.B. 1980: *The Phonology and Phonetics of English Intonation*, M.I.T. Doctoral dissertation

Pierrehumbert, J.B. 1983: "Linguistic units for Fo synthesis", *Abstracts of the 10th Int. Cong. Phonetic Sciences*, (Dordrecht: Foris Publications), p. 137-144

Thorsen, N. 1978: "An acoustical investigation 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: "Word boundaries and Fo patterns in Advanced Standard Copenhagen Danish", *Phonetica 37*, p. 121-133

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

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

Thorsen, N. 1983a: "Standard Danish sentence intonation - Phonetic data and their representation", *Folia Linguistica 17*, p. 187-220

Thorsen, N. 1983b: "Two issues in the prosody of Standard Danish", in: A. Cutler and D.R. Ladd (eds.) *Prosody - Models and Measurements*, (Berlin: Springer-Verlag), p. 27-38

Thorsen, N. 1984: "Variability and invariance in Danish stress group patterns", *Ann. Rep. Inst. Phon. Univ. Copenhagen 18*, p. 13-43. To appear in *Phonetica 41*

Uyeno, T., Hayashibe, H. and Imai, K. 1979: "On pitch contours of declarative, complex sentences in Japanese", *Ann. Bull. Res. Inst. Logopedics and Phoniatrics 13*, p. 175-187

Uyeno, T., Hayashibe, H. and Imai, K. 1981: "Syntactic structures and prosody in Japanese: A study on pitch contours and the pauses at phrase boundaries", *Ann. Bull. Res. Inst. Logopedics and Phoniatrics 15*, p. 91-108

Vaissière, J. 1983: "Language-independent prosodic features", in: A. Cutler and D.R. Ladd (eds.) *Prosody - Models and Measurements*, (Berlin: Springer-Verlag), p. 53-66.

ARIPUC 18, 1984, p. 243-254

A NOTE ON WORK IN PROGRESS: SECONDARY ARTICULATION IN THAI STOPS

JØRGEN RISCHEL
AND
AMON THAVISAK*

The present paper outlines the system of obstruents in Modern Central Thai and points out various controversial issues having to do with laryngeal control and supralaryngeal secondary articulation. Some very preliminary findings concerning Thai /p t/ are mentioned, and it is shown with fiberoptic illustrations how the epiglottis is crucially involved in an articulatory gesture found with these consonants in certain environments.

I. INTRODUCTION

There are certain phonetic properties associated with the stop consonants of (Central) Thai which are frequently mentioned in general phonetic literature because they have a bearing on central issues. Thus, over the years, evidence from Thai has been adduced repeatedly in connection with the concept of Voice Onset Time (VOT) in Consonant-Vowel sequences (cf. Lisker and Abramson 1964). Another subject of very general interest is constituted by the pitch perturbations associated with CV-sequences in tone languages, Thai giving clear evidence (Erickson 1975 and others) that the fundamental frequency starts lower after voiced than after voiceless stops, even in such a language in which tone in itself has a lexically distinctive function (and even though the tonal accents of Thai are to some extent reflexes of earlier manner differences among initial

---

*) *Mahidol University, Thailand. - State scholarship holder at the Institute of Phonetics in the winter 1982-83.*

consonants (cf. Brown 1962, Egerod 1971, Henderson 1982), so that the scenario repeats itself over time, as it were).

In the more specialized literature on Thai linguistics and phonetics the emphasis of interest is somewhat different, more controversial issues being the presence or absence of certain secondary articulatory features in unaspirated stops in this language. Before mentioning some of these controversies it may be expedient to tabulate the stop consonants of Thai (or rather the whole obstruent system, for reasons that will be apparent).

II. THE OBSTRUENT SYSTEM

Syllable initially the stops exhibit a contrast of three manners of articulation, viz. voiceless unaspirated, voiceless aspirated, and voiced (unaspirated), and a contrast of four oral points of articulation, which we may roughly label as follows: labial, alveolar, palatal (more precisely: alveolo-palatal, often with some affrication), and velar. In addition, there is a glottal stop (which of course goes with the voiceless unaspirated series if glottal closure is defined as a point of articulation on a par with the others). The system is "asymmetric" in a non-surprising way in that the voiced series is limited to the most advanced points of articulation. For these points of articulation there is also a series of voiceless fricatives, and if we include /h/ as an obstruent (classificatorily), the total system looks as follows:

Initial obstruents

ph	th	ch	kh		
p	t	c	k	and	ʔ
b	d				
f	s			and	h

Syllable finally there are no manner distinctions whatsoever within the obstruent system, the four series above being matched by only one series, which is variably rendered in the literature as voiceless unaspirated or voiced stops, although the former notation (with support from instrumental phonetic observation, cf. Abramson 1972) is now dominant. Moreover, there is (in Modern Thai) no palatal point of articulation (palatals having changed to alveolars). We thus get the following set:

Final obstruents

p	t	k	and	ʔ

Such a tabulation is, however, more controversial than it may look. Thai syllables behave differently in terms of termination depending on the quantity of the vowel. If the vowel is long (in a current phonemic notation: geminate) the syllable may or may not end in a consonant, but the only possible stop consonants then are /p t k/. The glottal stop, if it occurs, is not

contrastive with its absence. Syllables with a short vowel, on the other hand, must end in a consonant, and in this case all four possibilities above are utilized. What is not always fully recognized in authoritative phonemicizations is that syllables ending in /ʔ/ may drop this element in sandhi (and certain items regularly do this). It is more important, however, to note that there is even a marginally occurring though extremely frequent termination in [h], cf. the polite particle given as /khâ/ in Haas' dictionary (1964), which occurs utterance finally (or as a whole utterance in itself) and is often said with a very audible puff of air: [kʰɑh]. One must, then, consider whether this means that Thai has open syllables with a short vowel, or whether /h/ should be added to the inventory of final obstruents (the latter would seem somewhat far-fetched, perhaps).

As for the final inventory it may be noted, incidentally, that there is another collapsing of manners of articulation within the resonants. Initially there is a series of (three) nasals and moreover two types of liquids, but finally these are matched by nasals only. Historically this altogether impressive mismatch between initial and final consonant inventories is to some extent traceable to mergers (and both old and recent borrowings give evidence of substitutions such as /n/ for both /l/ and /r/).

Leaving aside the special problem of "khâ" and other final particles we may thus conclude that consonantal syllable terminations in the remaining Thai vocabulary involve oral or nasal stops or /ʔ/, that is, a syllable final consonant must involve oral and/or glottal closure (with or without concomitant nasalization). One of the standing issues in Thai phonetics is whether the oral stops in final position are (always or sometimes) glottalized or laryngealized. Harris (1972, p. 11ff) maintains that there is simultaneous oral and glottal closure in these consonants (both initially and) finally (for initials, see below). This question of manner of articulation in final stops is of course interesting in itself (also to satisfy the curiosity of language teachers who may wonder why foreigners often have such difficulty in hearing the place of articulation in final stops), but it is perhaps of particular interest in a historical and comparative perspective, viz. in connection with the general discussion of phonation type and tonogenesis in Southeast Asian languages.

Syllable initially, the series /ɓ ɗ/ are very strongly voiced, and according to Harris (1972, p. 14) "utterance initial voiced stops and approximants are usually preceded by weak glottal closure", though they are not assumed to exhibit implosive articulation. Historically, authoritative reconstructions derive /ɓ ɗ/ of Modern Thai from */ʔɓ ʔɗ/, so from this point of view the exact articulation of the modern stops is, of course, of considerable interest (although other Thai languages and dialects, as well as loanwords exchanged between Ancient Thai and Mon-Khmer languages, give strong evidence for the reconstruction */ʔb ʔd/, whatever the exact phonetic interpretation of these symbols).

Most controversial, however, is the series /p t c k ʔ/. If one series of stops is aspirated and another possibly laryngealized (in addition to being voiced), /p t c k ʔ/ would seem to be the truly "plain" series of stops. However, some authors claim that these consonants have simultaneous oral and glottal closure. This is said quite explicitly by, a.o., Marvin Brown (1965, p. 39), who refers to the "tenseness of pronunciation" of /p t c k/ in Bangkok in favour of the assumption that there is simultaneous oral and glottal release. Harris (1972, p. 11-13) states that "Siamese voiceless glottalized stops and affricates are pronounced with simultaneous oral and glottal closures. The release of the oral and glottal closure is usually simultaneous so that the glottal release is not heard. (..) In the pronunciation of Siamese glottalized stops and affricates the articulation is usually quite tense with a firm closure between active and passive articulators. In the release phase of these consonants there is usually a clear sharp onset of the following vowel." (p. 11). Interestingly enough, this contention is not particularly favourable in the context of Brown's diachronic account of the tonal developments in Thai, since he has some difficulty with register assignment under the assumption that /p t c k/ were always accompanied with glottal closure, but he finds the evidence for a change of articulation over time too weak and therefore assumes this feature even for Ancient Thai. Again, a precise analysis of the phonetic nature of Modern Thai stop consonants would seem very relevant to the historical and comparative debate.

A new dimension entered the discussion with the study of Gandour and Maddieson (1976). They found by measurement that the larynx is sharply raised in the production of /t/ and conclude that the glottis cannot be closed in this consonant, since this would produce an ejective quality, which is not found in Thai. (They refer to E.A. Henderson for the contention that /t/ is said with a closed glottis; also cf. the references above.)

The most intriguing feature of secondary articulation in Thai stops is the alleged velarization in the series /p t c k/: It is generally recognized that at least some of these stops have a special quality before certain vowels. This special quality has been interpreted by several authors as a matter of velarization. Miller (1956, p. 254) notices velarization in the speech of a Thai speaker born in Bangkok (but speaking another dialect for the purpose of his study); among later sources are Noss (1964, p. 9), Egerod (1961, p. 65), who speaks of velar pressure in the consonants in question and velarized quality in the following vowel, Noss (1964, p. 9), and Harris (1972, p. 13), who characterizes /t/ before close front vowels as both glottalized and velarized.

There is some disagreement among these authors as to the distribution of this feature of velarization. As for limitations on the stop consonants exhibiting this feature, Harris only mentions velarization in some allophones of /t/; Miller and Egerod note it for both /p/ and /t/; Noss even includes /k/,

which of course is velar in itself. Nobody posits velarization of the palatal stop or fricative. In addition to the stops, Harris (p. 17) also points to the existence of velarized variants of the voiceless fricatives /f/ and /s/; Noss (p. 9) speaks of both stops /p t k/, nasals /m n ŋ/, and spirants /f s h/ as being slightly velarized (before certain vowels). As for the vowels with which this alleged velarization is heard, Harris speaks only of "close front vowels" (i. e., in the phonemicization used in this paper, /i/ or /ii/); in the case of /s/ he refers to emphatic speech as a condition under which the velarized variant is common (before close front vowels). Egerod speaks of /ii/ as being velarized, and notes that a gliding quality of /uu/ is audible after the same consonants (viz. /p t/). Miller (p. 254) speaks more generally of vowels as being "heard slightly velarized as a consequence of the release of the velaric pressure", and he even includes the lateral consonant following /p t/ in clusters. He notes that velarization is most audible in the case of "the high front vowel phonemes, where the delay in raising the root of the tongue, due to the nature of the stop, produces a very clear velarization of the vowel, almost diphthongal in effect" (it should be remembered that his study does not deal with the Bangkok dialect, or Central Thai in general, but still its description seems highly relevant to the present paper). Noss, finally, speaks of slight velarization before all of the high vowels /i y u/. - It is not clear to what extent some of the above statements were supported by instrumental observation.

Gandour and Maddieson (1976) found raising of the larynx in /t/, as said above. This was measured externally, but they suggest the presence of pharyngeal constriction, rather than glottis closure, on this basis. This, in their view, can explain the much debated modification found with vowels after unaspirated voiceless stops, a modification which they refer to as "the commonly observed 'dark' quality of vowels, especially the high front vowel, following this stop series". (Incidentally, Nina Thorsen in a transcription worked out in 1969, as part of the requirements for the B.A. degree in phonetics, noted pharyngealization in some vowels, when transcribing Thai as an unknown language.)

III. PRELIMINARY OBSERVATIONS ON /p t/

In the winter 1982-83 the authors of this paper had the opportunity of entering a discussion of the nature of Thai stops in a weekly seminar held by Professor Søren Egerod, and partly as a corollary of this (and partly for educational purposes) we did some very preliminary experiments with fiberoptic observation of Thai syllables involving stops as articulated by Amon Thavisak. We found that the view was seriously hampered in the case of /p t/ before /i/, and according to the opinion of Harris (1972) and others referred to above one might speculate that this were due to the fiberscope being pushed out of position by a velarization gesture. However, with proper positioning of the fiberscope it became clear that the field

of vision was reasonably stable, but that the upper edge of the epiglottis moved backwards so that a greater or lesser part of the picture of the glottis was covered by the front side of the epiglottis.

Similar backward movement of the epiglottis has been reported for certain Semitic languages, partly with reference to emphatic consonants and partly with reference to pharyngeal fricatives (cf. Laufer and Condax 1979, El-Halees 1982), but it does not seem to be claimed anywhere that this feature typically accompanies velarization as such (unless "velarization" is taken as a broad term covering the features associated with emphatic articulation in Arabic). Indeed, there would seem to be no obvious physiological reason why these features of articulation should go together, whereas there is an obvious connection between tongue retraction and (passive) epiglottal movement, and also a clear connection between larynx raising and epiglottal movement (cf. Lindqvist 1972).

There are, then, two kinds of questions to be asked in articulatory terms: (a) is the backward movement of the epiglottis associated with a backward movement of the tongue root?, with some kind of laryngeal articulation?, with both?, with neither? (the last possibility is hardly applicable according to the generally accepted views on the physiological constraints on speech articulation), and (b) are /p t/ before /i/ characterized by both velarization and something going on at the level of the epiglottis, or is the alleged velarization not truly velarization but rather pharyngealization, as suggested on independent grounds by Gandour and Maddieson (1976)?

In linguistic terms one may ask why something special should happen to /p t/ before /i/ but not before all vowels, and why it should not comprise the consonants /c k/ as well. If there is in fact a laryngeal gesture involved one might speculate whether this has something to do with the skewness of the system, /p t/ being the only "plain" stops which are in contrast with voiced stops, and moreover represent the points of articulation most favourable to spontaneous voicing, so that some special feature helping to keep them distinct from /b d/ might be called for. However, the same "velarized" quality can be heard with the Thai fricatives /f s/ in similar environments (as pointed out in Harris 1972, p. 17), which certainly does not lend much support to such a functional explanation. Moreover, it has been argued (Søren Egerod, personal communication) that the restrictions on the occurrence of this strange feature of pronunciation seems reasonable enough if it is velarization that is involved: it is heard most clearly with the front vowel /i/, and it does not occur with consonants whose articulation already involves the dorsum of the tongue (viz. /c k/).

We have undertaken a series of videotape recordings using the fiberscope technique in order to study laryngeal and supralaryngeal aspects of Thai stop consonants and of /p t/ in particular. These must of course be supplemented by studies

aiming at confirming or disconfirming the contention that these consonants are velarized in certain environments. So far we have fiberscope recordings of four (female) informants plus an X-ray recording of a series of words spoken by one informant (ATh).

It would be entirely premature to report on the results in any detail. This must await further work including supplementary techniques (photo-glottograms) and additional X-ray recordings, and comparisons with acoustic displays (sound spectrograms, Fo tracings) of the syllables in question.

However, it can be safely claimed already that the backward movement of the epiglottis is a seemingly very constant feature of syllables involving initial /p t/ at least before the vowel /i/ (to what extent this feature appears before the remaining vowels of Thai, of which /e/ and /ɯ/ are - along two different dimensions - the first candidates for comparison, will be investigated in the future). This feature does not occur in our recordings with /c k/, nor with any of the other stops (e.g. aspirated voiceless /ph th/; voiced /b d/), so it is genuinely a feature characterizing /p t/, whatever the reason for this restriction on its occurrence within the stop consonants. We have observed a similar movement of the epiglottis with voiceless fricatives (and even with /r/, where the reason may be that the trill involves a complex articulation which may have nothing directly to do with the epiglottal feature in obstruents). The backward movement of the epiglottis is not equally strong in all cases; but its occurrence with /p t/, and its absence with other stop consonants, is a surprisingly stable phenomenon.

As for the question of the cases of this epiglottis movement, the limited X-ray evidence available so far suggests that it may be a consequence of tongue-root retraction. However, a passive effect by which the epiglottis is pushed backwards by the tongue-root may not be all there is to it, since the recording of a series of syllables spoken with very emphatic articulation showed stronger displacement of the epiglottis than usual, the latter moving way out from the tongue-root in a not very fast gesture. One might, then, speculate whether some laryngeal gesture is involved. On this point the evidence from our recordings is extremely clear as regards lateral constriction in the larynx. Firstly, the view of that part of the glottis which is left visible by the epiglottis is totally unhampered by constriction of the false vocal folds, so the consonants are at least not strongly laryngealized. Secondly, the glottis is typically not firmly closed in /p t/, as it should be if these consonants were glottalized, but it is rather slightly open, exactly as one expects with a plain voiceless unaspirated stop. The Appendix illustrates what the difference in epiglottis position looks like for certain types of syllables with various informants.

IV. CONCLUDING REMARKS

According to our preliminary findings it can be concluded that syllable initial Thai /p t/ are not glottalized, whereas certain sequences of these consonants with vowels exhibit a mechanism of pharyngeal constriction involving backward movement of the epiglottis (i.e. pharyngealization?, cf. the description of Arabic "emphatics" in terms of pharyngealization proper in Giannini and Pettorino 1982). Whether there is concomitant velarization remains to be studied.

Tongue-root retraction having been posited as a feature of certain "register type" languages of South-East Asia (but found to have no phonetic basis for at least Nyah Kur, see Thongkum 1982), the existence of a possibly similar feature in Central Thai is of some interest in an Area Linguistic perspective, no matter how this feature will eventually turn out to be conditioned (in terms of context and in terms of possibly concomitant features).

As will be obvious, the above remarks constitute an explanation of the raison d'être of, rather than a proper report on, a phonetic investigation in its beginnings. Our preliminary observations on /p t/ were presented in guest lectures by one of the authors (JR) at Chulalongkorn university and Mahidol university in August 1983, and the reactions seemed to warrant that these limited observations are non-trivial enough to be presented in more accessible form, although the bulk of phonetic research certainly remains to be done (and to be presented with full documentation), both as regards Central Thai proper, and as regards the possible later extension of such study to other, typologically (if not genetically) related languages and dialects of South-East Asia.

ACKNOWLEDGEMENTS

We are grateful to the Dentists' High School of Copenhagen for much helpful assistance including X-ray video recordings. - The authors also wish to thank Birgit Hutters and Svend-Erik Lystlund of the Institute of Phonetics for their interest in and assistance with the recordings made with the fiberscope technique.

REFERENCES

Abramson, A.S. 1972: "Word-final stops in Thai", *Thai Phonetics and Phonology* (Harris, J.G. and Noss, R.B. eds.), p. 1-7 (Bangkok)

Brown, M. 1965: *From ancient Thai to modern dialects* (Bangkok)

Egerod, S. 1961: "Studies in Thai dialectology", *Acta Orientalia 26*, p. 43-91

Egerod, S. 1971: "Phonation types in Chinese and South East Asian languages", *Acta Linguistica Hafniensia XIII*, p. 159-171

El-Halees, Y. 1982: "A fiberoptic and xeroradiographic study of emphasis in Arabic", *Abstracts of papers for the tenth international congress of phonetic sciences*, p. 466

Erickson, D. 1975: "Phonetic implications for an historical account of tonogenesis in Thai", *Studies in Tai linguistics in honor of William J. Gedney*, p. 100-111 (Bangkok)

Gandour, J. 1974a: "The glottal stop in Siamese: Predictability in phonological description", *Working Papers in Phonetics 27, UCLA*

Gandour, J. 1974b: "Consonant types and tone in Siamese", *Journal of Phonetics 2*, p. 337-350

Gandour, J. and Maddieson, I. 1976: "Measuring larynx movement in Standard Thai using the thyroumbrometer", *Phonetica 33*, p. 41-67

Giannini, A. and Pettorino, M. 1982: *The emphatic consonants in Arabic, SLR IV*, Instituto Universario Orientale, Laboratorio di Fonetica Sperimentale, Napoli (Giardini editori, Pisa)

Haas, M. 1964: *Thai-English student's dictionary* (Stanford University Press)

Harris, J. 1972: "Phonetic notes on some Siamese consonants", *Tai Phonetics and Phonology* (Harris, J.G. and Noss, R.B. eds.), p. 8-22 (Bangkok)

Henderson, E.J.A. 1982: "Tonogenesis: Some recent speculations on the development of tone", *Transactions of the Philological Society*, p. 1-24

Laufer, J. and Condax, I.D. 1979: "The epiglottis as an articulator", *Working Papers in Phonetics 45, UCLA*, p. 60-83

Lindqvist, J. 1972: "A descriptive model of laryngeal articulation in speech", *Speech Transm. Lab., Quart. Prog. and Status Rep. 2-3/1972, Royal Inst. Techn. Stockholm*, p. 1-9

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

Miller, Roy A. 1956: "Phonology of the Thai dialect of Nakhonsitthammarat", *Orbis 5*, p. 250-258

Noss, Richard B. 1964: *Thai reference grammar* (Washington D.C.: Foreign Service Institute)

Thongkum, Theraphan L. 1982: "Register without tongue-root in Nyah Kur (Chao Bon)", *Paper presented at the XVth Int. Conference on Sino-Tibetan Languages and Linguistics, Beijing, Aug. 17-19, 1982.*

APPENDIX

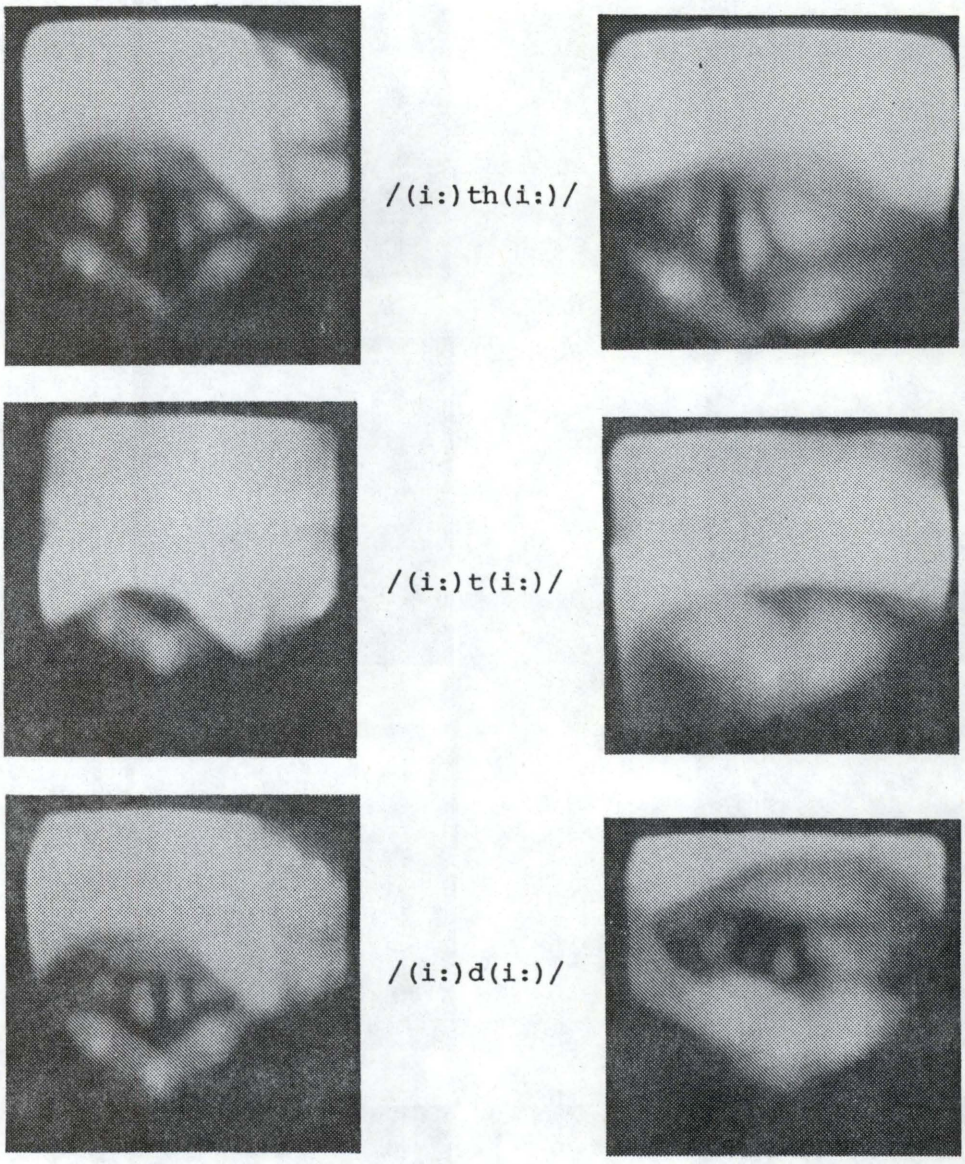


Figure 1

Fiberscope pictures of the glottis (photographed from the screen of a TV monitor, hence the rather poor quality) for three different consonants: aspirated, unaspirated voiceless, and voiced alveolar stop. The sequences /i:Ci:/ occurred in meaningful phrases of analogous structure, all pictures of each column (i.e., for each speaker) being from the same reading of a list to ensure minimum displacement of the fiberscope relatively to the glottis. With these recordings we had no synchronization pulses enabling us to select a specific moment during the articulatory events; the pictures were chosen so as to illustrate the (maximum) degree to which the epiglottis tilts back during the consonantal phase for each of the three consonants.

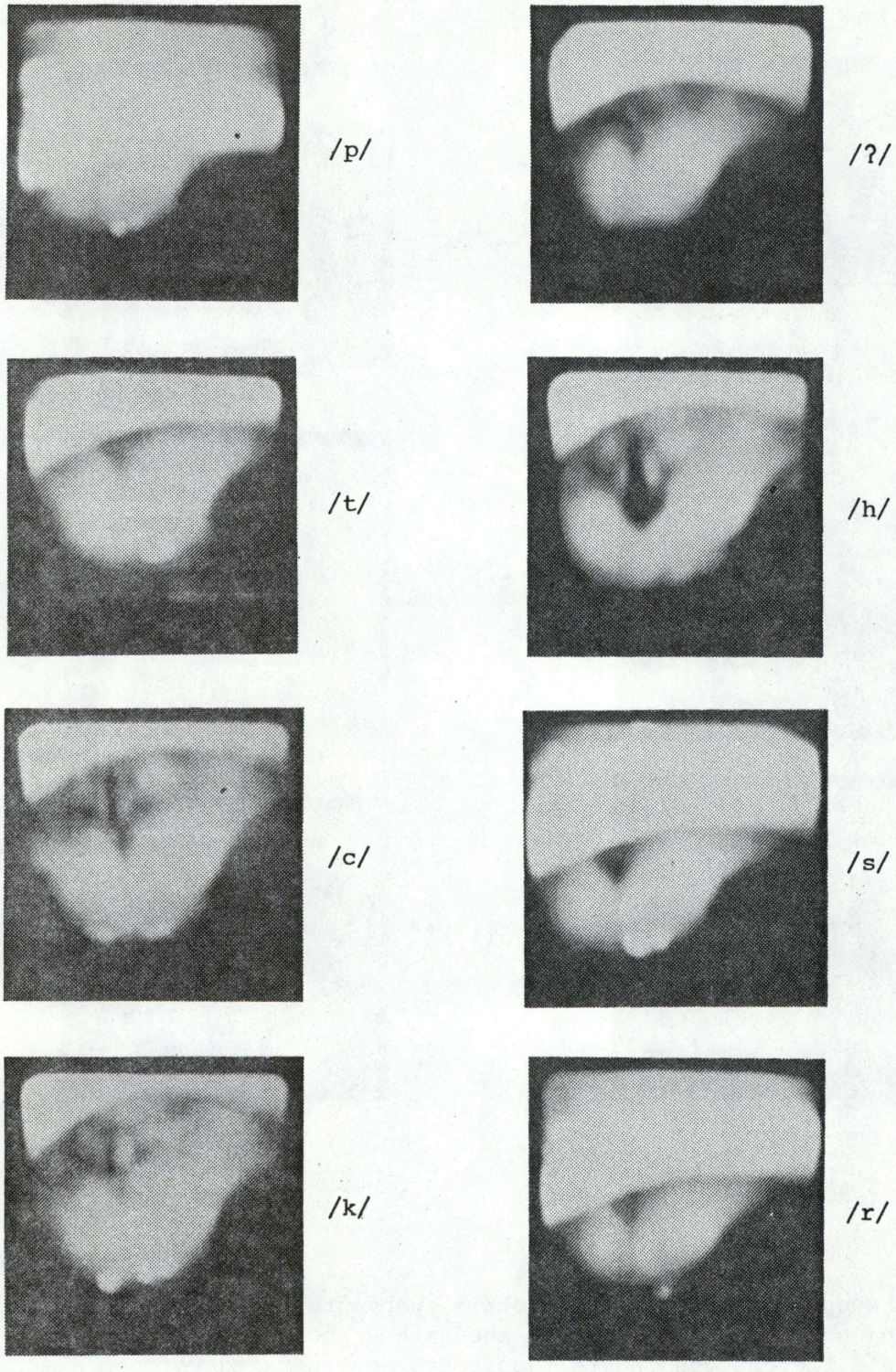


Figure 2

The four unaspirated voiceless stops (left) compared with the two glottal consonants and with /s/ and /r/ (right). Note the similarities among the epiglottis gestures for /p t s r/. - Speaker AT. (These pictures were taken from a reading of nonsense sequences of the structure /i:Ci:/.)

ARIPUC 18, 1984, p. 255-276

SOME BASIC VOWEL FEATURES, THEIR ARTICULATORY CORRELATES AND THEIR EXPLANATORY POWER IN PHONOLOGY*

ELI FISCHER-JØRGENSEN

The classical articulatory vowel features, vowel height (or degree of openness) and front-back, have been criticized as physiologically inexact and in reality based on a misinterpretation of auditory impressions. Some want to replace them by new features, partly based on constriction (Wood), others want to interpret them exclusively in auditory terms (Ladefoged). It is argued in this paper that the classical articulatory vowel features are not as inexact physiologically as maintained by their critics, and that they are indispensable in phonological descriptions and in this respect more useful than e.g. Wood's feature system.

I. INTRODUCTION

For the description and explanation of universal constraints and tendencies in phonological patterns and phonological change one needs detailed models of speech production and speech perception, but it is also necessary to have a general frame of reference in the form of a system of more abstract phonetic dimensions according to which the speech sounds of a language can be grouped into classes. These dimensions which, according to a now generally accepted but not quite unambiguous terminology, are also called features, and which can be considered potentially distinctive, must on one hand have correla-

---

*) *This is a revised and enlarged version of my contribution to the symposium "Phonetic Explanation in Phonology" at the Tenth International Congress of Phonetic Sciences, Utrecht 1983. Particularly the phonetic discussion is more detailed in the present paper.*

tion to speech production and speech perception, on the other be adequate for the description of phonological patterns and rules.

In the present paper I will consider vowel features only, and only some basic articulatory features which have been the subject of debate in recent years, i.e. vowel height (or degree of openness), front-back, and tenseness.

II. CRITICISM OF THE TRADITIONAL SYSTEM

Since the days of Bell and Sweet it has been the tradition to describe vowel systems by means of the basic features high/low, front/back and rounded/unrounded and, for some languages, tense/lax. This system, which was defined in articulatory terms, has been severely criticized for not covering the articulatory facts.

As early as 1910 E.A. Meyer found that lax [ɪ] may have a lower tongue position than generally assumed, e.g. North German [ɪ] is often lower than [e:]. G.O. Russell (1928 and 1936) attacked the traditional system of classical phonetics on the basis of a large number of X-ray photos, particularly of American English, but also of German and French. He found that often the tongue position was not as one should expect; [ɪ] might, e.g., be lower than [e], and for the back vowels there were great discrepancies. For [ʊ], [ɔ] and [ɑ] the point of articulation was rather in the pharynx, whereas [u] might almost be a front vowel. On the whole, the importance of the pharynx cavity for vowel sounds had not been realized. Russell did not have much influence, though, perhaps because of his very aggressive tone. He called it "a wildly unscientific absurdity" to listen to speakers' sounds and then record them in terms of their physiological character, and he characterized the traditional physiological dimensions (particularly high/low) as purely imaginary.

In the sixties the criticism was renewed by Peter Ladefoged (1962 (1967), 1971, 1975, 1976). He maintained, like Russell, that Bell, Sweet and their followers had described their auditory impressions but translated them into physiological terms, and that these did not correspond to a physiological reality. He showed, among other things, that in the X-ray photos of Stephen Jones' cardinal vowels [o] and [ɔ] were lower than front [a], whereas [ɑ] was still lower, as far as the highest point of the tongue is concerned. On the whole, high vowels do not all have the same height ([u] often being lower than [i]) and [u] is often advanced compared to [o]. Ladefoged et al. (1972) also found a rather large individual variation in the articulation of front tense and lax vowels in American English. Ladefoged (1976) finds a much closer agreement between traditional vowel descriptions and acoustic facts. As an example he compares Uldall's placement of the Danish vowels

in Jones' cardinal vowel chart and a plot of the formant frequencies of Danish vowels, height corresponding to the frequency of F1, and front-back to the difference between the first and second formant. Joos (1948) had made a similar observation, comparing Jones' cardinal vowel chart with Holbrook and Carmody's X-ray vowel quadrilateral on one hand, and with an acoustic chart of French vowels on the other, and finding a greater similarity in the latter case. - But this is not so astonishing since Jones' cardinal vowels (apart from [i] and [ɑ]) were based on auditory impressions of equal distances.

Nearey (1978), who has analysed the vowels of three American English speakers, also emphasizes that the individual variation is more pronounced in articulation than in the acoustic pattern, and that the traditional vowel features are more closely related to the acoustic aspect.

The most severe attack on the classical vowel system has been made by Sidney Wood (1975a and 1982), based on an analysis of 38 sets of X-ray photos of 15 different languages. He found that [ɪ] and [ʊ] are often lower than [e] and [o], and that the height relations between [o,ɔ] and [a,ɑ] are random. He also emphasizes that the classical system neglects the pharynx cavity completely, and thus cannot account adequately for the relations between vowel production and vowel acoustics, nor is there any clear relation between the dimensions front/back and high/low and the function of the muscles. He concludes that the model is not only inaccurate but irrelevant to the processes of speech production, that it is a complete illusion and must be rejected and replaced by a new model.

III. THE CLASSICAL SYSTEM IS NOT AS INACCURATE AS MAINTAINED BY THE CRITICS

I agree with Catford (1981) that the criticism of the classical system is exaggerated. In the first place, even when accepting the premises of the critics, e.g. that the classical system describes the position of the highest point of the tongue, many X-ray photos show quite a good agreement with the traditional description, as shown for instance by Mona Lindau (1978) for five American English speakers. Ladefoged's own vowels, which he quotes (1976) as an example of bad agreement, are not too bad either. Nobody has ever claimed that all high vowels should have the same height or that front and back vowels of different height should lie each on a straight (vertical) line. For front vowels it is evident that the lower vowels are normally gradually retracted, and the fact that many languages have a somewhat advanced [u] does not invalidate the general classification. Further, Ladefoged's worst example (Stephen Jones' cardinal vowels) (1971) should not be given too much weight. The X-ray photos were taken in 1929, at a time when the technique was much less advanced than today; the subject had a chain lying along the surface of his tongue and another chain through his nose hanging down behind the uvula (Catford

1981), and, moreover, I have been told that he had all the teeth on one side of his mouth extracted in order to get nice pictures! It is, of course, laudable that he suffered so much for science, but it has hardly contributed to the naturalness of his pronunciation.

In the second place, some of the more serious criticisms raised by Ladefoged and Wood lose their force when it is realized that they are mainly objections against Daniel Jones' cardinal vowel chart, which was meant as a practical device for field work and not as a theoretical vowel system, and which differs from the classical system on various points. As mentioned above, it was in the first place an auditory chart, but both Jones himself and his pupils also tended to interpret it physiologically in the sense that it indicated the position of the highest point of the tongue. Ladefoged (1971, p. 67) says that for the past hundred years vowels have been described in terms of the highest point of the tongue. But I do not think this is correct. Catford (1977 and 1981) states that neither Bell nor Sweet used the term. But he is not sure about their followers since Jespersen mentions it as a current term in 1889. But Jespersen does not use the term "highest point of the tongue" here. He talks about the highest part of the tongue. Thus, what he discusses is whether one should talk about tongue height or (as Jespersen and many others preferred) about the distance from the palate. As far as I can see, none of the founders of classical phonetics mention the highest point of the tongue, not, e.g., Jespersen, nor Sievers, nor Viëtor, nor Passy, nor Storm. They therefore often use the terms "degree of openness" or "distance" instead of "height". The same is true of their followers on the European continent with the exception of some more recent works which are influenced by Jones. It seems to be a particular British tradition introduced by Daniel Jones in "The Pronunciation of English" (1909) (not in his "Outline of English Phonetics" 1918, as I said in Fischer-Jørgensen 1983). It is an attempt at a more precise description, but the highest point of the tongue is rather variable and much too precise a concept to be used in a general vowel system. What is implied in the classical system is the distinction between an advanced and a retracted tongue body (sometimes including a central position), and the overall distance between the articulating part (front or back or central) and the palate.

Similarly, the fact that e.g. English [ɪ] and [ʊ] are often lower than [e] and [o] can be used as an argument against Jones' placement of the former vowels above [e] and [o] in the vowel chart, although, as Catford remarks (1981), nothing in the system prevents a placement of a relatively low American English [ɪ] below [e]. What is disturbing is, however, that they will be placed sometimes above and sometimes below [e] and [o] for different speakers. The cause of the trouble is that Jones does not like to recognize tenseness as a separate vowel dimension but prefers to describe [ɪ] as a lowered and retracted [i], and [ʊ] as a lowered and less rounded [u]. If tenseness is recognized as a separate dimen-

sion, then [ɪ] and [ʊ] are simply the highest lax vowels, and their exact relation to [e] and [o] is not relevant. This is the position taken by various phoneticians within the classical tradition. The height of [ɪ] and [ʊ] thus cannot be used to criticize the traditional vowel system as such. But the attitude to this dimension has been somewhat hesitating, also within classical phonetics. One reason is probably that Bell's description was evidently wrong. He used the terminology primary vs. wide and thought that the wide (e.g. lax) vowels have a retracted soft palate and an expansion of the pharynx. Sweet, who uses the terminology narrow/wide, explains the difference as depending on the shape of the tongue, which is more convex in narrow (tense) vowels and more flattened and relaxed in wide (lax) vowels. The narrowing is thus not the result of raising the whole body of the tongue with the help of the jaw as in [i] versus [e], but of bunching up the part with which the sound is formed. Thus, passing from [i] to [e] one does not go via [ɪ]. This description is much more to the point, but it was not generally accepted. Jespersen (e.g. 1897-99 and 1914) gave a somewhat different formulation; he described the difference as a thinner vs. broader channel between the articulating part of the tongue and the palate, which did not necessarily involve a generally lower tongue position. Others emphasized the muscle tension as the decisive difference (e.g. Sievers 1901), and the difference is now (therefore) generally called tense-lax. Almost all British phoneticians have followed Jones in not recognizing tense/lax as an independent dimension, whereas a good number of Scandinavian, German, and Dutch phoneticians have accepted the distinction. The difference is also clearer in German and Dutch than in English, where all vowels are relatively lax, but where the difference is clear at least for high vowels. But it is true that many phoneticians found E.A. Meyer's and Russell's findings that [ɪ ʏ ʊ] often had a lower tongue position than [e ø o] disturbing because they found them auditorily closer to [i y u]. However, this auditory impression may be caused partly by the fact that [ɪ ʏ ʊ], e.g. in German, are the highest lax vowels and thus systematically high, and also by orthography; cp. that Danish listeners described German [ɪ ʏ ʊ] - cut out of words and presented in isolation - as being close to Danish [ɛ: œ: ɔ:] (Fischer-Jørgensen 1973 (1975)).

The tense/lax distinction has been widely accepted in modern phonology. The definitions given by Jakobson and Halle (1956) and by Chomsky and Halle (1968) are based on the idea of a difference of muscular tension leading to a greater deviation from the neutral position of the vocal tract. - The terms tense/lax have also been applied to the difference which is the basis of vowel harmony in a number of West African languages. Ladefoged (1964) uses this terminology (though only as a tentative label); but at the same time he shows that the decisive difference, at least in Igbo, lies in an advancement or retraction of the tongue root. Halle and Stevens (1969) think that this is also the main characteristic of the tense/lax difference found in, e.g., the Germanic languages and they therefore propose to use the term "advanced tongue root"

in both cases. However, Stewart (1967), Ladefoged et al. (1972) and Lindau et al. (1972) showed that these features should not be confounded (cf. also Lindau 1978, who proposes to call the African feature "expansion"). Whereas the tongue root movement is really an independent feature in various African languages, it is not consistent in English, and it is closely related to tongue height differences. Acoustically the two features are also different (cp. the acoustic graphs of African languages in Lindau et al. 1972, and the graphs of German in Jørgensen 1969). Finally, expansion does not have the close relation to length generally found for tense/lax.

Since the articulatory correlates of the tense/lax difference in English seem to vary according to speakers (Ladefoged et al. 1972), Ladefoged (1975) proposes to define the difference in purely distributional terms (lax vowels being those which do not occur in open syllables), whereas Lindau (1978) gives an acoustic definition (peripheral versus central in the formant chart).

More recently Wood (1975b and 1982) has undertaken an extensive investigation of the articulatory characteristics of tense and lax vowels in a number of languages. He finds that in various languages [i] and [ɪ] have approximately the same jaw opening, whereas for [e] and [ɛ] it is lower, conversely [i] and [e] have approximately the same tongue bunching, whereas [ɪ] and [ɛ] have a flattened tongue. There are exceptions (cp. that the six informants investigated by Ladefoged et al., 1972, all have the distinction in jaw opening described by Wood, but only three have a clear difference in tongue lift). On the whole, however, the tendency is very clear (cp. also Fischer-Jørgensen 1973). It also appears from the measurements of jaw opening for Dutch vowels by Zwaardemaker and Eijkman (1928) and by Kaiser (1941), and, e.g., from X-ray photos of Telugu taken by Nagamana Reddy, which I have had occasion to see. - Thus, whether [ɪ] is lower or higher than [e] is a more accidental consequence of the relative extent of the two movements involved (tongue flattening and jaw opening), but at any rate [ɪ] has less constriction than [i] at the place of articulation, and Wood states that this is true of all other lax vowels compared to their tense counterparts, with the exception of [o/ɔ]. At the same time lax vowels have a narrower pharynx cavity (except for tense [ɑ] versus lax [a]), and less pronounced lip activity. Wood has shown, through synthesis, that these different factors may be of different acoustical importance. It is, however, clear that they can all be regarded as consequences of one single difference: tense versus lax articulation, involving a flattening of the tongue which is closely connected with less space in the pharynx as well as less jaw opening and less pronounced lip activity. There exist few EMG-recordings of the muscular tension involved in the production of tense and lax vowels, and only American English has been investigated. There seems to be a clear difference in the tension of the genioglossus (Smith and Hirano 1968, Raphael and Bell-Berti 1975, Alfonso and Baer 1981, and Alfonso

et al. 1982) and of the inferior longitudinal (Raphael and Bell-Berti 1975), whereas there are contradicting results for the styloglossus for [u/ʊ] (EMG-recordings of German and Dutch vowels would be an obvious task for the future).

It thus seems clear that tenseness is an independent vowel dimension (which is often - but not always - combined with length). Height should therefore be indicated separately for tense and lax vowels.

The most serious objection against the traditional height dimension seems to be the irregular height relations between back vowels, as described by Wood. It was, however, also Jones who, for practical reasons, placed the unrounded back vowel [ɑ] in the same column as the rounded vowels [u o ɔ] because, in many languages, [ɑ] is the only low back vowel, whereas the higher back vowels are generally rounded. But from the point of view of a general system of vowel features [ɑ] does not belong in this series but in the series of unrounded back vowels [ɯ ɤ ʌ ɑ]. Thus, the fact that [o] and [ɔ] may be lower is only crucial if height is taken to be an absolute property, not if it is taken to be a relative property within each series of rounded and unrounded, front and back vowels, a view which would be in accordance with Jakobson's conception of distinctive features.

Thus, if tenseness is considered a separate dimension and height is taken to mean the relative distance between the articulating part of the tongue and the palate within each series of rounded or unrounded, tense or lax, front or back vowels, most of the inconsistencies between these traditional labels and the articulatory facts disappear.

What remains of the criticism is that the pharynx cavity, which could not be observed at the time when the classical system was set up, has been neglected. The description does not take account of the fact that the most narrow constriction for [a] and [ɑ] and generally for [o ɔ] is in the pharynx, and as degree and place of the constriction in the total vocal tract are essential for the calculation of the acoustic output, the classical system is not the most adequate starting point for such calculations. It is true that since there are strong constraints on the possible positions of the tongue body (e.g. a low back vowel will necessarily have a narrow pharynx, and a high front vowel a wide pharynx), it has been possible to set up correlations between the features of the classical system and formant frequencies, but the causal relations are not clear if the pharynx is not taken into account.

Wood (1975a) also argues that the connection between vowel articulation and muscle activity becomes much clearer in a model based on place of constriction than in the traditional model. This is not quite so convincing. The relations are complicated in either case (for the relation between muscle activity and the traditional dimensions, see Catford 1977, p. 186 and Halle 1982).

IV. NEW FEATURE SYSTEMS

It is evident that models of vowel production must be based on these more recent insights, i.e., they must take account of the total vocal tract. And it is possible to construct new feature systems, which are connected more closely to these new models.

However, phoneticians do not agree on the number of places of constriction. Catford (1977) tentatively sets up six points, but he does not find this very useful for the description of vowel systems. Lindblom and Sundberg's model of vowel production (1969 and 1971) is based on place and degree of constriction, jaw position, width and height of lip opening, and larynx height. They set up three places of constriction: palatal [i e ɛ], velar [u], and pharyngeal [o a ɑ]. Tongue height is a derived parameter, controlled by means of both jaw position and tongue raising relative to the jaw. In optimal articulation jaw and tongue go together, but there may be compensations. In the 1969 paper a system of binary features is based on this model and applied to the Swedish vowels. It contains the features palatal, pharyngeal, velar, close, open, and labial. It is not, however, mentioned in later publications, and I will therefore not discuss it in detail.

Wood operates with four places of constriction: the hard palate, the soft palate, the upper pharynx, and the lower pharynx. There are thus four categories of vowels: palatals (all front vowels except [æ] and [a]), velars ([u] and [ʊ]), velopharyngeals ([o] and [ɔ]), and low pharyngeals ([æ a ɑ]). It is not a continuum, but four discrete places (he thus denies the existence of central vowels). His arguments are partly empirical (the analysis of X-ray photos from many languages), partly theoretical; these are the places where the spectrum is relatively insensitive to moderate tongue displacements according to Stevens' quantal theory (1972) and Fant's nomograms (1960). He also adduces sound typology (1982, p. 72), pointing to the fact that languages with only three vowel phonemes generally use the three places mentioned by Stevens (/i ɑ u/), and if there are five phonemes the normal type is /i ɛ ɑ ɔ u/. But he does not mention that in languages with four vowel phonemes it is extremely rare to find his four basic vowel types /i u o ɑ/, the most common types being /i u ɑ/ plus /ɛ/ or /ɨ/ (cp., e.g., Crothers 1978). The typological argument is thus very weak.

In Wood's system the tense vowels [i u o ɑ] may have lax counterparts with less constriction [ɪ ʊ ɔ a]. Moreover, the jaw may be close or open. This adds a distinction to palatal vowels: [e] and [ɛ] being the open counterparts to [i] and [ɪ]. The velar vowels are considered to be (redundantly) close, the pharyngeal vowels open. Finally, vowels may be rounded or unrounded.

On this basis he sets up a binary feature system (1982, p. 168):

constriction	palatal				palatovelar		pharyngo-velar		low pharyngeal	
vowel	i	ɪ	e	ɛ	u	ʊ	o	ɔ	ɑ	a
palatal	+	+	+	+	+	+	-	-	-	-
velar	-	-	-	-	+	+	+	+	-	-
pharyngeal	-	-	-	-	-	-	+	+	+	+
open	-	-	+	+	-	-	+	+	+	+
round	-	-	-	-	+	+	+	+	(-)	-
tense	+	-	+	-	+	-	+	-	+	-

Wood does not comment on this table. Evidently it has been a problem to translate a place dimension with four places into a binary feature system. This has the consequence that he has had to call his velar place palatovelar, and his upper pharyngeal place velopharyngeal, which blurs the quantal aspect. It would have been more in keeping with his own description to set up a multivalued dimension with four steps.

It may also be argued that the place of constriction is just as variable as the highest point of the tongue. Front vowels may have their maximal constriction at the alveolar ridge (Straka even calls the French palatals alveopalatals, or simply alveolars (1978)). An [o] may have its maximum constriction at the velum, at the uvula, or in the pharynx, depending on the shape of the individual palate, the position of the velum, etc. Wood is aware of this. 1982, p. 142 he says that "place of maximal constriction" is ambiguous in natural speech. What matters is the direction of the tongue body movement. Even with this precaution one may sometimes get into doubt. In Danish [o:] and [ɔ:] the main direction of the movement seems to be towards the soft palate, although they have at the same time a narrower pharynx compared to [u] (see figure 1)*. Danish [o:] is very close, and it might perhaps be described as an *u*-like vowel. This would only require a step "open" for velar vowels.

V. THE USEFULNESS OF THE TRADITIONAL FEATURE SYSTEM AND WOOD'S SYSTEM IN PHONOLOGICAL DESCRIPTIONS

A feature system based on place of constriction may come closer to a model for vowel production (but it seems rather difficult to translate it into a two- or three-dimensional figure resembling the formant chart). Now, as stated in the introduction, a feature system should also be applicable to phonological descriptions, and Wood expressly emphasizes that his

---

*) See p. 276.

model has more explanatory power in this respect than the traditional system. But for this purpose I find the classical system clearly superior.

Wood only allows for two steps in vowel height (or openness), and for the back vowels the traditional two degrees are reinterpreted as a difference of place. But there are many languages with three degrees of opening, and Danish even has four (/i e ɛ æ/). Wood considers [æ] to be a pharyngeal vowel, but in that case it has to be distinguished by a different feature from /ɑ/, which may be considered a separate phoneme in "Advanced Standard Copenhagen" (see Basbøll 1972). According to his practice for English, Wood must then describe /æ/ as lax. The same must be done for /ɛ œ ɔ/, both in Danish and in languages with similar systems. But there is no justification for describing Danish /ɛ œ ɔ/ as lax (apart from the fact that a certain laxing is normally combined with lowering). Danish has both long and short vowels with, in most cases, almost the same quality. In the series /i e ɛ æ/ there is a gradual lowering of the tongue height (as can be seen in X-ray photos and palatograms) and generally a gradual lowering of the jaw, which appeared in a recording of jaw opening for three Danish informants. With the exception of /e,ɛ/ for one of the informants all the differences were statistically significant (the data will be published later). This speaks against considering any of the vowels as lax. As for Swedish, Lindau (1978) quotes a case of diphthongization in the Swedish Skåne dialect which can only be formulated in a simple way on the basis of four degrees of openness. The English vowel shift also requires more degrees of openness to describe the diphthongization of /i: u:/, the development of /e: o:/ to /i: u:/, of /ɛ: ɔ:/ to /e: o:/, and of /a:/ via /æ:/ to /e:/.

Now this might be partly remedied by adding more steps to the close/open dimension without changing the rest of the system, but in phonological rules /ɑ/ does not go with /e ø o/ as it should according to Wood's feature system because it has the same degree of openness. - It behaves like a lower vowel. There are, e.g., languages where /e: ø: o:/ are diphthongized to, e.g., /iə yə uə/, whereas /ɑ:/ is not diphthongized, for instance Old High German.

There are also a number of well known universal phonetic tendencies connected with vowel height. Low vowels are longer than high vowels, they are pronounced on a lower pitch and have higher intensity. These differences are gradual, i.e. [a] is longer than [ɛ], [ɛ] than [e], [e] than [i], and similarly for pitch and intensity. It is not a difference between two categories high and low. These differences are generally not perceived, but when Danish speakers are asked to manipulate synthetic vowels in duration adjustment tasks, they make [ɛ] longer than [i], and [ɛ:] longer than [i:], and the cross-over value between /i/ and /i:/ is smaller than between /ɛ/ and /ɛ:/ (Petersen 1974). These are phonetic differences, but they may turn up in historical developments, so that long high vowels become short vowels (e.g. in Dutch), or short low vowels are lengthened (e.g. English /æ/).

However, it is possible to quote one example where /ɑ o ø e/ go together and behave as one class. In French there are common rules of lengthening and distribution for /ɑ o ø/ and (partly) /e/ vs. /a ɔ œ ɛ/. This induced Oluf Thorsen (see Jensen et al. 1971) to describe /a ɑ/ not as low vowels but as pharyngeal vowels. The common property for /ɑ o ø e/ versus /a ɔ œ ɛ/ was then a more "close timbre" vs. a more "open timbre", later called tense vs. lax. I wonder, however, whether it is correct phonetically to call French /a ɔ œ ɛ/ lax vowels. They do not sound lax to me. Wood (1982, p. 139) found the tongue to be lower relative to the mandible in French /ɛ/ than in [e] in three sets, but the jaw is also more open, and in Straka's X-ray pictures (1950) /ɛ œ ɔ/ have a larger jaw opening than /e ø o/, whereas /a/ has a slightly smaller opening than /ɑ/. In the case of the low pharyngeal vowels it is difficult to judge tenseness on the basis of the constriction, because two conflicting tendencies are at work: (1) a narrower constriction at the point of articulation (i.e. the pharynx) in tense vowels, and (2) a narrower pharynx in the lax vowels. Wood considers [ɑ] generally as tense. This is probably true of German, where it is long, as other tense vowels, but in Dutch it is the front [a] that is long.

Anyhow, the French example can be considered to support Wood. But in all the other cases mentioned we need the traditional height dimension, where [a ɑ] are the lowest vowels.

As for the use of four places of constriction, it makes the formulation of various phonological rules and developments rather complicated. In Germanic umlaut back vowels are changed to front vowels before an [i] in the following syllable (and [ɑ] is also raised). In Finnish and Turkish vowel harmony back vowels are changed to front vowels after front vowels. These rules can only be formulated very clumsily in terms of four places of articulation. The same is true of the allophonic fronting of back vowels in palatal surroundings in Russian.

There are, however, a few cases where the feature pharyngeal for vowels might perhaps give a simpler and more explanatory formulation, i.e. in the cases of assimilation of vowels to pharyngeal or uvular consonants, as found in Greenlandic before /ʁ/ and /q/ (mentioned by Wood). However, as /æ/, which becomes [ɑ] in this position, is considered by Wood to be pharyngeal already, whereas this is not true of /i/ and /u/, the formulation will not be simple. The same reasoning is valid for the ʁ-colouring of Danish vowels (described by Basbøll 1972, and by Basbøll and Kristensen 1975). Perhaps it is just as acceptable to say that vowels may be retracted and lowered before pharyngeal consonants.

This raises the more general problem of the utility of having common features for vowels and consonants, whereby assimilations between vowels and consonants could be formulated in a more explanatory way. Jakobson's features grave/acute and compact/diffuse were not satisfactory in this respect. But

instead of using traditional consonantal features for the vowels the goal may also be reached by using the traditional vocalic features for consonants, as it was done by Chomsky and Halle (1968) (see the discussion in Fischer-Jørgensen 1975, p. 230ff) and by Halle (1982). Ladefoged (1971) proposes to specify both consonants and vowels in high/low, front/back and in terms of place of articulation (palatal, velar, pharyngeal). This seems somewhat complicated. But it might be possible to call the vowel front-back dimension palatal/non-palatal, and use non-palatal as a cover term for velar, upper pharyngeal (or uvular), and lower pharyngeal. One might then use the specification in the (rather few) cases where it seems adequate.

The general conclusion of this discussion is that we cannot do without the traditional classical dimensions front-back and high-low.

VI. THE AUDITORY INTERPRETATION OF THE DIMENSIONS HIGH-LOW AND FRONT-BACK

Russell, Ladefoged and Wood agree in the assumption that the dimensions high/low and front/back as used by Bell and Sweet were in reality a translation of their auditory impressions into physiological terms and thus an illusion. Catford rejects this assumption. He says that there is ample internal evidence in the works of Bell and Sweet to show that they were really observing tongue positions. Particularly Sweet was very explicit on this point. He criticizes the German phoneticians for basing their vowel systems on auditory similarity instead of production, and he recommends whispering the vowels in order to better feel the muscular sensations and says that training of the vocal organs is a better way of learning sounds than doing it by ear. If they had built on auditory impressions they would also, as Catford remarks, have placed [ʏ] and [ɯ] between [i] and [u] and not set up rounding as a separate dimension. I think it is true that these old phoneticians, who did not have the possibility of looking at X-ray photos, worked really hard to train their muscular sensations, much more than we do nowadays (this is also true of Jespersen). Catford concludes that the classifications of Bell and Sweet were primarily based on "highly trained phoneticians' perception of proprioceptive and tactile sensations, not upon the misinterpretations of auditory sensations" (1981).

Ladefoged does not, however, draw the same conclusion from his criticism of the traditional system as Wood does. It is true that he also wants to set up a new model of speech production (Harshman, Ladefoged and Goldstein 1977). It is based on a factor analysis of 18 cross sections of the vocal tract for 10 English vowels, ending up with two factors: (1) a forward movement of the root of the tongue together with raising of the front part, approximately from [ɔ] to [i], and (2) raising of the back of the tongue, approximately from [ɑ] to [u]. But he does not want to set up a new feature system on this basis.

He finds that the traditional features have proved to be useful for the description of phonological sound patterns (in 1975 he only gives the supplementary terms palatal and velar for front and back), whereas for the purpose of describing the difference between languages we need a larger number of physiological and acoustic parameters. It must be possible to map the phonological features onto basic phonetic parameters, either physiological or acoustic, but this is not necessarily a one-to-one relation (1980). As for the traditional dimensions, he finds that rounding can be described simply in physiological terms, whereas the physiological correlates to front/back and high/low cannot be used in language description. They have, however, clear auditory and acoustic correlates, height corresponding to the frequency of F1, and front/back to F2-F1 (e.g. 1976). I find it difficult to accept this exclusively acoustic-auditory interpretation.

As for the height dimension it is probably true that it has a somewhat simpler connection with its physical than with its physiological correlates but, as argued above, when height is taken to be a relative dimension within each category of vowels, the correlation to the physiological facts is quite good. It is striking that in experiments intended to bring out the auditory dimensions it has not proved quite easy to factor out a height dimension. Many years ago I tried in various ways (Fischer-Jørgensen 1967). Phonetically naive subjects were asked, e.g., to group vowels according to auditory similarity, with the result that very often front vowels [i e ɛ æ] were put into one group, separated from back vowels, whereas high and low vowels were never sorted out, not even if only six vowels [i y u] and [ɛ œ ɔ] were presented. When asked directly if they did not find that [i y u] belonged together, most declared that [u] does not belong with [i] and [y]. When subjects were asked to group vowels in bright and dark, thin and thick, small and large, etc., or to place them on a scale from dark to bright etc., the same dominating dimension from dark to bright came out in almost all cases, irrespective of the pair of adjectives used. Only when subjects were asked about tight/loose and compact/diffuse did something that looked more like the vertical dimension appear, but [i y u] were designated as compact and [ɛ œ ɔ] as diffuse, in contradistinction to Roman Jakobson's terminology. I think this indicates that the subjects were guided mainly by tactile sensations in this case. When, in some cases, subjects were asked to pay more attention to their articulation, [u] got somewhat closer to [i] and [y]. In more recent experiments, based on more refined methods, when for instance subjects are asked to judge the similarity between sounds presented in triads and the results are factor-analysed, vowel height also turns up (see, e.g., Terbeek 1977). But, altogether, the evidence for "height" being a specifically auditory feature is not very strong.

Moreover, a number of phonetic and phonological rules and developments involving height are better understood when described in articulatory terms. The relatively longer intrinsic duration of lower vowels can hardly be explained from an auditory

point of view, whereas it may be assumed that a more extensive jaw movement takes longer time. Preliminary measurements of jaw opening show that the maximum aperture is reached later in open vowels. It may not be economical to accelerate the movement so much that full compensation is achieved. As for intrinsic Fo there are also some plausible articulatory explanations. In the case of palatalization of velars before front vowels, which Ladefoged (1972) also describes as an articulatory development, the difference in palatalizing power between [i e ɛ a] must also be explained on the basis of articulation. But the regularity of F1-correspondences point to an auditory adjustment.

I therefore think that the height feature should be considered to have clear physiological as well as auditory correlates, which both play a role in phonological systems and developments.

As for the dimension front/back it is more complicated. Here two articulatory dimensions, front/back and rounded/unrounded combine to form one auditory dimension: the dark/bright dimension, which was often used in pre-Bell vowel systems. This is, certainly, a dominating auditory dimension, which shows up in experiments with auditory similarity and in phonetic symbolism; it is also prevalent in the patterning of vowel systems (cp. Trubetzkoy's (1939) "Helligkeit" dimension) where /u/ and /i/ are extremely common because they are maximally different in a two-dimensional auditory vowel space, whereas /ɯ/ and /y/, although also distinguished by two articulatory features, are rare (cp. Crothers 1978 and Lindblom 1980), because they are auditorily intermediate between /i/ and /u/. As shown by Ladefoged (1967), even trained phoneticians have difficulty in distinguishing between front rounded and back unrounded vowels. It is also very difficult to elicit rounding as a separate feature in experiments with auditory similarity (cf. Terbeek 1977).

However, in many phonological rules and developments the articulatory dimensions front/back and rounded/unrounded are kept apart. For instance, in Russian there is an allophonic contextually conditioned variation between front [i] and mid [ɨ] and between back [u] and fronted [ü] with preservation of the rounding distinction. Further, in Finnish vowel harmony only the front/back dimension is at work, and in Turkish the front/back harmony and the rounded/unrounded harmony function separately according to different rules. Similarly in i-Umlaut back vowels become front vowels, but the rounding difference is preserved, and conversely, rounding of vowels in labial environment does not involve a change in place of articulation. It is possible that perception may play a role at the last stage of the i-Umlaut, where the [i] of the ending may have become so weak that the listener did not hear it and therefore perceived the front feature of the stem vowel as an independent feature (Ohala 1981), but in its origin it must have been a mainly articulatory process. On the whole, it seems more plausible to explain such processes of assimilation in motor terms as an anticipation of an articulatory position.

Thus I do not think we can do without a front/back dimension defined in articulatory terms. It is probably necessary to operate both with two articulatorily defined dimensions: front/back and rounded/unrounded, and with an auditorily defined dimension of brightness which has a causal relation to a combination of the two articulatory dimensions. They all seem to operate in phonology and thus to be in some sense psychologically real. The articulatory features seem to be at work in assimilatory rules and developments, whereas the auditory feature evidently plays a role in the structure of vowel systems. This is in good agreement with Lindblom's assumption that the necessity for sufficient auditory distance between phonemes is the most important determinant factor in the structure of vowel systems.

However, the relative importance of the horizontal and the vertical dimension has given some problems. In Liljencrants and Lindblom 1972 the possible acoustic vowel space was calculated on the basis of the speech production model set up by Lindblom and Sundberg (1969 and 1971). The calculation was made in terms of formant frequencies. In order to get a two-dimensional space the frequency of formant 2 corrected with respect to formant 3 was chosen as one dimension, and the other dimension was the frequency of formant 1. By transformation into the mel scale an approximation to an auditory space was obtained. By means of a computer program it was calculated where the vowels should be placed in this space if maximal perceptual distance between all vowels should be obtained for different numbers of vowel phonemes (from 3 to 12). The result was compared with known data on actual vowel systems. The prediction turned out to be quite good for vowel systems with 3 to 6 vowels, but above that limit the model generated too many high vowels, i.e., the horizontal dimension was utilized more than the vertical dimension, in contradiction to what is the case in natural vowel systems.

In order to bring the model in better agreement with actual vowel systems it was modified so that the function of the peripheral auditory system was taken more directly into consideration, i.e., instead of formant analysis a filter analysis based on critical bands was used, and masking and non-linear frequency response were taken into account. Moreover, the idea of "maximal" auditory distance was replaced by "sufficient" distance. This model is compared with the older model in Lindblom 1980. It produces a smaller number of high vowels. There are two different versions of the new model. In one of them, which should be closest to the auditory system, phons are transformed into sones, but this version gives less good results for vowel systems with a small number of vowel phonemes, whereas the other version operating on phon/Bark gives better results for systems with few vowels, but somewhat less reduction of the number of high vowels in systems with many vowels. None of the two versions generate the common seven vowel system [i e ɛ a ɔ o u]. Lindblom suggests that the vertical dimension may play a greater role in actual systems because F1 has high intensity and thus is more resistant to noise.

I should like to suggest a different explanation. Perhaps the auditory distance between [i] and [u] really is felt as relatively long compared to the vertical distance between [i] and [a]. The experiments with Danish subjects mentioned above seem to support this, and the same appears from the dominant use of the dark-bright opposition in sound symbolism. It may be objected that Danish [i e ɛ æ] are closer together than in most other languages. But why then do we keep them apart phonologically even though we feel them as auditorily related? - I suppose production plays a role in this connection. As demonstrated by Lindblom and Sundberg the simplest way to produce differences in vowel height is by raising and lowering the mandible. Now in the first place the proprioceptive sensitivity seems to be more developed for jaw movements than for advancing or retracting the tongue. This may have something to do with the fact that jaw opening and closing is used for other biological purposes, e.g. eating. Moreover, it is visible. (It may happen that a student starting a phonetics course believes that he produces an [e] by retracting his tongue, but he will not maintain that he produces an [a] by closing his mouth.) Finally, steps in jaw movement have a simple one-to-one correlation with steps in F1 and thus with steps in auditory impression of the series [i e ɛ a], whereas the series [i y ɯ u] requires a complicated interplay of tongue and lip movements. Vowel height is a physiologically simpler dimension, and therefore utilized more extensively.

VII. CONCLUDING REMARKS ON EXPLANATION IN PHONOLOGY

Finally I should like to admit that I have used the terms explanation and explanatory in a somewhat slipshod way. But there are many kinds and steps of explanation. When I argue that front-back is more explanatory than four places of articulation in i-Umlaut and vowel harmony rules, the point is that 'back' comprises more cases in one rule, and generalization is a first step in explanation. Distinguishing more than two heights also allows more generalizations, e.g. saying that mid vowels [e ø o] have diphthongized, which is not possible if [a] is considered to have the same height.

But when I argue for an articulatory interpretation of front-back in assimilatory developments, it is because the development can then be described by plausible production mechanisms. Those who, like Roger Lass (1980), require that explanations must be deductive cause-effect explanations which permit prediction would call the explanations mentioned above "understanding" and not "explanation". But that depends on how the word "explanation" is defined. Here it is used in a wider sense. You can never predict a concrete sound change, but you can sometimes explain it afterwards, and you may assume with high probability that if there is a change it will go in a certain direction. There are very strong phonetic constraints on phonological systems and on sound change, and I agree with

John Ohala (e.g. 1983) that it is an important task for phonetics to find these constraints.

ACKNOWLEDGEMENT

I am grateful to Nina Thorsen for improvements of various formulations.

REFERENCES

Alfonso, P. and Baer, Th. 1981: "An electromyographic-cinefluorographic-acoustic study of dynamic vowel production", *Status Rep. Speech Res., Haskins Labs. 65*, p. 109-124

Alfonso, P.J., Honda, K., Baer, T. and Harris, K.H. 1982: "Multichannel study of tongue EMG during vowel production", *Paper presented at the 103rd Meeting of the Acoustical Society of America, Chicago, April 26-30*

Basbøll, H. 1972: "Some conditioning phonological factors for the pronunciation of short vowels in Danish with special reference to syllabification", *Ann. Rep. Inst. Phon. Univ. Cph. 6*, p. 185-210

Basbøll, H. and Kristensen, K. 1975: "Further work on computer testing of a generative phonology of Danish", *Ann. Rep. Inst. Phon. Univ. Cph. 9*, p. 265-291

Catford, I. 1977: *Fundamental Problems in Phonetics* (Edinburgh University Press)

Catford, I. 1981: "Observations on the recent history of vowel classification", *Towards a History of Phonetics* (eds. Asher and Henderson) (Edinburgh University Press), p. 19-32

Chomsky, N. and Halle, M. 1968: *The Sound Pattern of English* (New York: Harper and Row)

Crothers, J. 1978: "Typology and universals of vowel systems", *Universals of Human Language* (eds. Greenberg, Ferguson and Moravcsik), p. 93-152

Fant, G. 1960: *The Acoustic Theory of Speech Production* (The Hague)

Fischer-Jørgensen, E. 1967: "Perceptual dimensions of vowels", *To Honor Roman Jakobson*, p. 667-671, Janua Linguarum, Series Major XXXI

Fischer-Jørgensen, E. 1973: "Perception of German and Danish vowels with special reference to the German lax vowels /I Y U/", *Ann. Rep. Inst. Phon. Univ. Cph. 7*, p. 143-194, and (a shorter version) in *Auditory Analysis and Perception of Speech* (eds. Fant and Tatham) 1975, p. 153-176

Fischer-Jørgensen, E. 1975: *Trends in Phonological Theory*, (Copenhagen: Akademisk forlag)

Fischer-Jørgensen, E. 1983: "Vowel features and their explanatory power in phonology", *Abstracts X Int. Congr. Phon. Sc.*, p. 259-265

Halle, M. 1982: "On distinctive features and their articulatory implementation", to be published in: *Natural Language and Linguistic Theory*, 1983

Halle, M. and Stevens K. 1969: "On the feature "advanced tongue root", *MIT Quart. Prog. Rep. 94*, p. 209-215

Harshman, R., Ladefoged, P. and Goldstein, L. 1977: "Factor analysis of tongue shape", *J. Acoust. Soc. Am. 62*, p. 693-707

Jakobson, R. and Halle, M. 1956: *Fundamentals of Language*, (The Hague: Mouton)

Jensen, O. Kongsdal, Landschultz, K. and Thorsen, O. 1971: *Fransk Fonetik* (University of Copenhagen)

Jespersen, O. 1889: *The Articulation of Speech Sounds Represented by Means of Analphabetic Symbols* (Marburg: Elwert)

Jespersen, O. 1897-1899: *Fonetik* (Copenhagen: Det Schubotheske Forlag)

Jespersen, O. 1914: *Lehrbuch der Phonetik* (Leipzig: Teubner)

Jones, D. 1909: *The Pronunciation of English* (Cambridge University Press)

Jones, D. 1918: *An Outline of English Phonetics* (Cambridge: Heffers and Sons)

Joos, M. 1948: *Acoustic Phonetics*, *Language Monographs 23*

Jørgensen, H.P. 1969: "Die gespannten und ungespannten Vokale in der norddeutschen Hochsprache mit einer spezifischen Untersuchung der Struktur ihrer Formantfrequenzen", *Phonetica 19*, p. 217-245

Kaiser, L. 1941: "Biological and statistical research concerning the speech of 216 Dutch students, III", *Archives néerlandaises de phonétique expérimentale XVII*, p. 92-118

Ladefoged, P. 1962: *The Nature of Vowel Quality*, *Revista do Laboratório de fonética experimental da faculdade de letras da universidade de Coimbra*, and *Three Areas of Experimental Phonetics*, 1967, p. 50-142

Ladefoged, P. 1964: *A Phonetic Study of West African Languages West African Language Monographs 1*, (Cambridge University Press)

Ladefoged, P. 1971: *Preliminaries to Linguistic Phonetics*, (University of Chicago Press)

Ladefoged, P. 1972: "Phonological features and their phonetic correlates", *Journal of the IPA 2*, p. 2-12

Ladefoged, P. 1975: *A Course in Phonetics* (New York: Harcourt Brace & Jovanovich)

Ladefoged, P. 1976: "The phonetic specification of the languages of the world", *Working Papers in Phonetics, UCLA 31*, p. 3-21

Ladefoged, P. 1980: "What are linguistic sounds made of?", *Language 55*, p. 485-502, also *Working Papers in Phonetics, UCLA 45*, 1979, p. 1-24

Ladefoged, P. 1983: "The limits of biological explanations in phonetics", *Abstracts X Int. Congr. Phon. Sc.*, p. 31-37

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

Ladefoged, P., Harshman, P., Goldstein, P. and Rice, L. 1977: "Vowel articulations and formant frequencies", *Working Papers in Phonetics, UCLA 38*, p. 16-40

Lass, R. 1980: *On explaining Language Change* (Cambridge University Press)

Liljencrants, J. and Lindblom, B. 1972: "Numerical simulation of vowel quality systems: the role of perceptual contrast", *Language 48*, p. 839-862

Lindau, M. 1978: "Vowel features", *Language 54*, p. 541-563, also *Working Papers in Phonetics, UCLA 38*, p. 49-81

Lindau, M., Jacobsen, L. and Ladefoged, P. 1972: "The feature advanced tongue root", *Working Papers in Phonetics, UCLA 22*, p. 76-94

Lindblom, B. 1980: "Phonetic universals in vowel systems", ms. to be published in *Experimental Phonology* (ed. J. Ohala) (Academic Press)

Lindblom, B. and Sundberg, J. 1969: "A quantitative model of vowel production and the distinctive features of Swedish vowels", *Speech Transm. Lab., Quart. Prog. and Status Rep. Royal Inst. Techn., Stockholm 1/1969*, p. 14-32

Lindblom, B. and Sundberg, J. 1971: "Acoustical consequences of lip, tongue and larynx movement", *J. Acoust. Soc. Am. 50*, p. 1166-1179

Meyer, E.A. 1910: "Untersuchungen über Lautbildung", *Die neueren Sprachen 18 (= Festschrift Viëtor)*, p. 166-248

Nearey, T.M. 1978: *Phonetic feature systems of vowels. Indiana Linguistics Club*

Ohala, J. 1981: "The listener as a source of sound change", *Papers from the Parasession on Language and Behaviour* (eds. Miller et al., Chicago)

Ohala, J. 1983: "The origin of sound patterns in vocal tract constraints", *The Production of Speech* (ed. P.F. MacNeilage) (New York: Springer), p. 189-216

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

Raphael, L.J. and Bell-Berti, F. 1975: "Tongue musculature and the feature of tension in English vowels", *Phonetica 32*, p. 61-63

Russell, G.O. 1928: *The Vowel* (Columbus, Ohio State University Press)

Russell, G.O. 1936: "Synchronized X-ray, oscillograph, sound and movie experiments showing the fallacy of the vowel triangle and open-closed theories", *Proc. II Int. Cong. Phon. Sc.*, p. 198-205

Sievers, E. 1901: *Grundzüge der Phonetik* (Leipzig: Breitkopf und Härtel)

Smith, T. and Hirano, M. 1968: "Experimental investigation of the muscular control of the tongue in speech", *Working Papers in Phonetics, UCLA 10*, p. 145-155

Stevens, K.N. 1972: "The quantal nature of speech: evidence from articulatory-acoustic data", *Human Communication, a Unified View* (eds. David and Denes)

Stewart, J.M. 1969: "Tongue root position in Akan vowel harmony", *Phonetica 16*, p. 185-204

Straka, G. 1950: "Système des voyelles du français moderne", *Bull. de la Faculté des Lettres de Strasbourg 28*, p. 172-180, 220-233, 275-284, 268-375

Straka, G. 1978: "Apropos du classement articulatoire des voyelles", *Hamburger phonetische Beiträge 25*, p. 432-460

Terbeek, D. 1977: *A Cross-Language Multidimensional Scaling Study of Vowel Perception, Working Papers in Phonetics, UCLA 37.*

Trubetzkoy, N.S. 1939: *Grundzüge der Phonologie, Travaux du Cercle Linguistique de Prague 7*

Wood, S. 1975a: "The weakness of the tongue arching model of articulation", *Working Papers, Dept. Ling. Lund 11*, p. 55-108

Wood, S. 1975b: "Tense and lax vowels - degree of constriction or pharyngeal volume?", *Working Papers, Dept. Ling. Lund 11*, p. 110-134 and *23*, 1982, p. 157-182

Wood, S. 1979: "A radiographic analysis of constriction location for vowels", *J. Phonetics 7*, p. 25-43, and *Working Papers, Dept. Ling. Lund 23*, 1982, p. 57-75

Wood, S. 1982: *X-ray and Model Studies of Vowel Articulation, Working Papers, Dept. Ling. Lund 23.*

Zwaardemaker, H. and Eijkman, L.P.H. 1928: *Leerboek der Phonetiek* (Haarlem: De Erven F. Bohn).

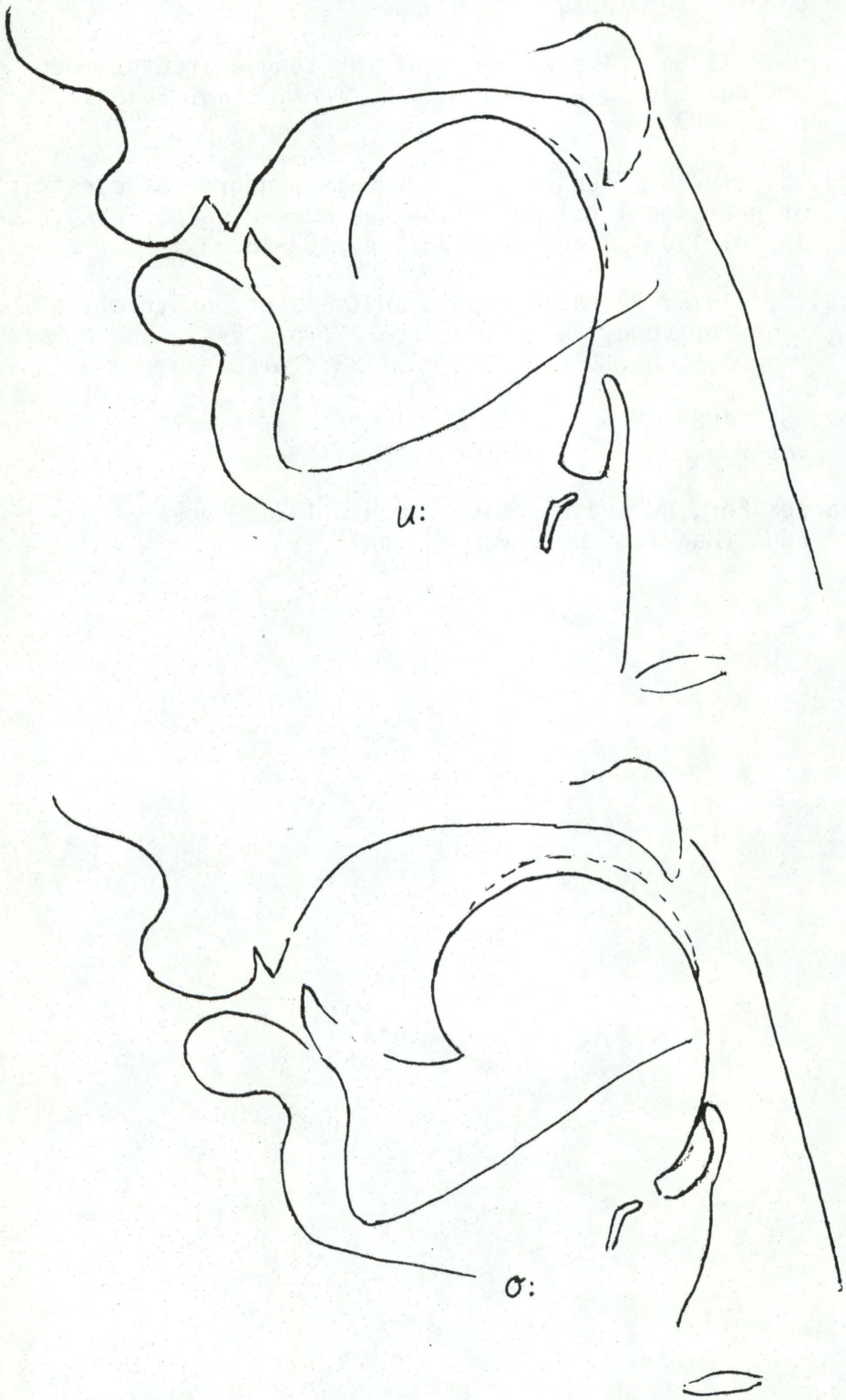


Figure 1

Tracings from X-ray photos of the Danish long vowels /u:/ and /o:/ pronounced by subject KS

ARIPUC 18, 1984, p. 277-284

ON THE RELATION BETWEEN VOWEL HEIGHT AND FRONT-BACK: A COMMENT ON ELI FISCHER-JØRGENSEN'S PAPER "SOME BASIC VOWEL FEATURES, THEIR ARTICULATORY CORRELATES AND THEIR EXPLANATORY POWER IN PHONOLOGY"

HANS BASBØLL*)

Eli Fischer-Jørgensen (1983) has given an interesting contribution to the symposium on "Phonetic Explanation in Phonology" during the Tenth International Congress of Phonetic Sciences, Utrecht, August 1-6, 1983. In this comment, I shall refer to the extended and revised version of her contribution, which is printed on the preceding pages of this volume.

Eli Fischer-Jørgensen in her contribution (to which the reader is referred for general references) argues in favour of what is basically the traditional front-back and high-low distinction in vowels, cf. the following quotation:

> *"if tenseness is considered a separate dimension and height is taken to mean the relative distance between the articulating part of the tongue and the palate within each series of rounded or unrounded, tense or lax, front or back vowels, most of the inconsistencies between these traditional labels and the articulatory facts disappear"* (this volume, p. 261).

While I agree on almost every point with Eli Fischer-Jørgensen's criticism of alternative proposals which have been advanced, and also find her general conclusions convincing, I shall in this comment briefly consider the relation between the two traditional articulatory features for the vowel space, viz. vowel height and front-back. I shall use (primarily) Danish *r*-colouring and (secondly) Nordic *i*-Umlaut as evidence for what I take to be a slightly different conception of the distinctive feature front-back, viz. for the way I have used it in a description of Modern Danish (cf. Basbøll and Kristensen 1975: 273ff, Basbøll 1981:48f).

---

*) *University of Odense, Denmark.*

Eli Fischer-Jørgensen argues:

> *"There are, however, a few cases where the feature pharyngeal for vowels might perhaps give a simpler and more explanatory formulation, i.e. in the cases of assimilation of vowels to pharyngeal or uvular consonants, as found in Greenlandic before /ʁ/ and /q/ (mentioned by Wood). However, as /æ/, which becomes* [ɑ] *in this position, is considered by Wood to be pharyngeal already, whereas this is not true of /i/ and /u/, the formulation will not be simple. The same reasoning is valid for the ʁ-colouring of Danish vowels ... Perhaps it is just as acceptable to say that vowels may be retracted and lowered before pharyngeal consonants"* (this volume, p. 265).

This formulation could be made even more precise: the feature pharyngeal must be multivalent if *r*-colouring in Modern Danish should be accounted for by means of this feature since this term in Danish covers a whole productive and systematic (and, of course, language specific) series of changes in vowel quality, both before and after /ʁ/ (cf. Basbøll 1972:202ff, and Basbøll and Wagner, forthcoming: ch. IV, sect. 7). The conception of the feature pharyngeal as multivalent seems inconsistent with Wood's proposals according to which it should denote articulation at a certain (although not completely invariable) point. On the other hand, nor do I see any reason to treat *r*-colouring of /æ/ (i.e. [æ] → [ɑ]) as backing, and all other cases of *r*-colouring (i.e. [e] → [ɛ], [ɛ] → [æ], [ø] → [œ], [œ] → [ɶ]) as lowering, as the normal use of the feature front-back would force one to do (cf. Colman and Anderson (1983:187): *"the proposals we are aware of would all assign the* [ɑ] → [æ] *and* [æ] → [e] *shifts to different dimensions (backness vs. height)"*).

The problem is, of course, whether [a] (or [æ]) should be considered as a maximally low and maximally front vowel (as expressed in the placement of [a] at the lower left corner of Daniel Jones' Cardinal vowel quadrangle, cf. the fact that just four X-ray pictures have determined the geometrical shape of the Cardinal Vowel Diagram). I do not think any phonetician would like to argue for a particular "corner" at [a]. Acoustically and perceptually this is rather clear, I think, and as far as articulation is concerned, [a] cannot be considered a phonetically "extreme" vowel - able to become a phonetic consonant in one little move, so to speak - like [i u ɑ] (this is not to deny that different configurations of muscles may be involved, however). I would describe the situation so that, phonetically speaking, the "horizontal aspect" of front-raising has much more relative weight, as compared to the "vertical aspect", when we go from [ɑ] to [æ], and that the relative weight of the "horizontal aspect" is continually diminishing when we go further towards [i] (so that the "vertical aspect" is clearly most important in the distinction between [e] and [i]). The important thing to me is that no "corner" at [æ] or [a] can be argued for on general grounds, in contradistinc-

tion to [i], [ɑ] and (partly) [u] (it should be remembered that Cardinal Vowels [i] and [u] are the highest possible vowels at their place of constriction, and Cardinal Vowel [ɑ] the lowest possible one, whereas the notion "lowest possible front vowel" has no precise meaning in phonetic terms). It seems to me that some consequences for the feature system must be drawn from this fact. I therefore use the feature label "front" as referring to vowels which lie on the (curved) (i.e. elliptic, cf. Jones' earlier drawings of the Cardinal Vowel Diagram) line from [i] to [ɑ] (and, similarly, "back" for the vowels from [u] to [ɑ], but this is less controversial, of course). After this slight re-definition of the distinctive feature term front-back, *r*-colouring can be described as simply consisting in moving one step in the direction towards [ɑ], and it seems to me that this account of the whole set of systematic and productive *r*-colouring effects is phonetically and phonologically more adequate than any alternative account one can think of. Notice that this description presupposes that the vowel-height feature is multivalent (I have used the feature "distance" instead (see Basbøll and Kristensen 1975:273ff with figure 1 (reproduced here)), expressing distance from (an extreme) [ɑ] towards either (an extreme) [i] (in "front"-vowels) or (an extreme) [u] (in "back"-vowels)).

What exactly is the difference, if any, between the traditional use (or uses) of the feature front-back, including Eli Fischer-Jørgensen's use, as far as I have understood her, and my use of it? Phonetically, I think there is none. Phonologically, the traditional use might be illustrated by the following quotation from Eli Fischer-Jørgensen (this volume, p. 261), a propos Jones' primary cardinal vowels:

> *"But from the point of view of a general system of vowel features,* [ɑ] *does not belong in this series* ((viz. [u o ɔ ɑ]/HB)), *but in the series of unrounded back vowels* [ɯ ɤ ʌ ɑ]".

I draw from this the (not logically necessary!) inference that [ɑ] does not, according to Eli Fischer-Jørgensen, belong to the series of unrounded front vowels [i e ɛ æ] (in any case, this is the view encoded in traditional feature systems). On the other hand, I would rather say that [ɑ] enters into two series of unrounded vowels, viz. both the "front" series (in a perhaps not too felicitous terminology) [i e ɛ æ a ɑ] and the back series [ɯ ɤ ʌ ɑ] (notice that only the dimensions are crucial, not the individual symbols for non-extreme vowels). According to the traditional conception, [ɑ] is phonologically a back vowel (not a front vowel), and the difference [ɑ]-[æ] is, phonologically, one of "backness" and not of height (although, phonetically, it is both). According to my conception, on the other hand, the phonological distinction front-back is in a broad sense "neutralized" in, i.e. does not apply to, [ɑ], and the difference [ɑ]-[æ] is, phonologically, one of "height" and not backness (although phonetically it is both). (By the term 'phonological' I do not refer to contrasts in any specific language but only to a certain level of abstractness encoded

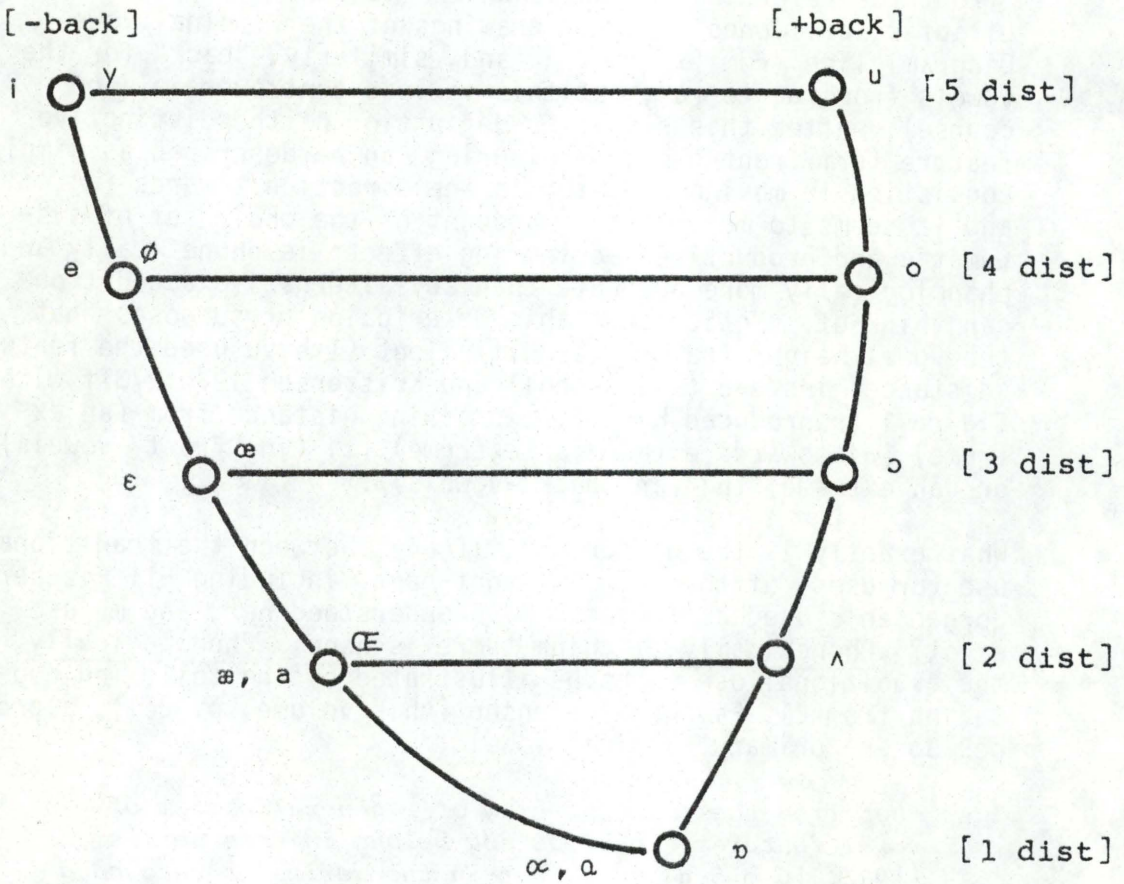


Figure 1

Schematic drawing showing the relationships between the vowels in terms of the features [back] and [distance] (...)

by means of distinctive features with discrete values; this cannot be discussed here, however.) In my view, one could even say that [ɑ] might be raised in two directions, viz. "front-raised" (towards [æ]-[i]) and "back-raised" (towards [ʌ]-[ɯ], cf. [ɔ]-[u]). (In a sense, this view corresponds to some types of "vowel triangle", so the tradition is still there, of course!) Notice that if the change [æ] → [ɑ] is seen as involving backing phonologically, the same reasoning could also be applied to the change [e] → [æ], for example.

At this point, a brief informative note on *r*-colouring in Danish might be useful. Diachronically, it appears at first sight to be the result of two different changes, viz. a lowering of certain front vowels adjacent to /*r*/, and a "front-raising" of *a* except when adjacent to /*r*/. This does not really change the unity of *r*-colouring, however, since there has been a general "front-raising" of long *a* (and of certain cases of short *a*), which has been prevented, however, by an adjacent /*r*/ (cf. Brink and Lund 1975:67ff, 96ff). Thus the adjacency of /*r*/ is in the former case of *r*-colouring a positive condition, in the second a negative one, but apart from this difference the effect of *r*-colouring is similar. Synchronically it is, as far as I can see, a productive process which - certainly without being innate - is normally unconscious, but which sometimes can become conscious, and be subject to hypercorrections and stigmatizations. In sum, I find the process of *r*-colouring in Modern Danish very well suited as an illustrative case of phonological vowel features.

I thus conclude that a unified and explanatory description of *r*-colouring in Modern Danish, by means of distinctive features, necessarily leads to considering [i e ɛ æ a ɑ] as one phonological dimension, one argument being that any account in terms of [æ] → [ɑ] as backing will necessarily lead to a "corner" at [æ] or [ɑ], as far as I can see; and this consequence I find phonetically as well as phonologically unsound. In the basic phonetic vowel system [ɑ] is, within the framework argued for here, necessarily the lowest unrounded vowel (in actual vowel systems this need not be the case, however, since a language-specific [ɑ] need not be so low as cardinal vowel [ɑ]). Furthermore (still within this system), the movement from [ɑ] to [æ] is (as already noted) just the first step of a "front-raising" and therefore of the same basic type as the "front-raising" from [ɛ] to [e], etc. According to such an interpretation, a quadrangular vowel system will thus be "less natural" than a triangular one, everything else being equal, since a triangular vowel system in the sense used here does not presuppose the specification of a "lowest front vowel", viz. of a "corner" at [æ] or [a]. Nevertheless, the two elliptic lines seem to allow such a specification (based upon phonological arguments), for a specific language, as a codification of the fact that the "front-raising line" becomes increasingly less vertical when we move from [i] towards [ɑ], and ends by being nearly horizontal.

The explicit rule for *r*-colouring in Modern Danish, within a framework such as the one which has been argued for here, is given in Basbøll and Wagner forthcoming (ch. VI, sect. 7; basically: "front-vowels are lowered" or, alternatively!: "vowels are "front-lowered"", i.e. lowered along the dimension [i]-[ɑ]). Its simple formulation is there compared to the much more complex formulations which are necessary within alternative frameworks of distinctive features-analysis, both the traditional use of front-back and an analysis by means of the features palatal, velar and pharyngeal (which are used in that work within a consistent binary framework related - but not identical - to Wood's use of them).

Now, it seems to me that this way of looking at the feature front-back may perhaps also contribute to an interesting account of Nordic *i*-Umlaut. I should emphasize that there are several crucial differences between the evolution of *i*-Umlaut in different Germanic languages/dialects, so what I suggest for Nordic is not necessarily relevant outside Scandinavia (which is of course not a homogeneous area either). Furthermore, the status of the so-called *i*-Umlaut of short /*e*/ is highly controversial. One reason to choose Nordic *i*-Umlaut as an example is that "Norse developed vowel mutation very extensively, virtually to the limit ..." (Prokosch 1938:110, cf. 107ff).

If we start with a purely descriptive statement, the main pattern of Nordic *i*-Umlaut seems to me to be the following: Rounded vowels are fronted, and unrounded vowels are "raised" in the sense that they move away from [ɑ] along the line on which the front vowels are placed, i.e. towards [i]. In this way the raising of [e] to [i] is placed on the same footing as [ɑ] to (say) [æ] or [ɛ]. (The change from [e] to [i] certainly also involves fronting, in the usual phonetic sense, but the traditional feature-analysis cannot capture its relation with other cases of *i*-Umlaut. Anderson and Jones (1977:53ff) argue in the same direction, although within another theoretical framework, viz. dependency phonology.) It should be emphasized, however, that rounding in itself has of course nothing to do with the underlying mechanism of *i*-Umlaut, but that it only serves to define the set of back vowels which do not lie on the "front-raising" line, viz. the one between [i] and [ɑ].

I think one clear reason can be given for the phonological naturalness of the description of *i*-Umlaut proposed here: *i*-Umlaut is a kind of assimilatory process between vowels, and the vowels which lie on the line from [i] and down - and that is exactly the unrounded vowels - move in the direction of [i], which is the conditioning factor. The rounded vowels, on the other hand, lie on the line between [ɑ] and [u], and if they follow that line this will not take them anywhere near [i]. So they move in the other - i.e. "horizontal" - dimension, which means that they are being fronted (both phonetically and phonologically). In short, the rule accounting for the effect of Nordic *i*-Umlaut may be stated informally like this: "front-raise" or front ("horizontally") the vowel in question! (since

the sets of vowels which are a priori eligible for "front-raising" and ("horizontal") fronting are non-overlapping, the rule does not have to be specified as to priority between its two parts). Chronologically, the order was probably front-raising first rather than last (i.e. *e* was not affected later than *a*, and *a* not later than rounded back vowels).

What is common to all the cases of *i*-Umlaut, under the present interpretation, is thus that the vowel to be changed is assimilated to [i], but only with respect to one single feature (and only one phonological degree, in the case of multivalent features). The vowels on the "front"-line must assimilate in the only feature in which they differ from [i] (i.e. in "distance", as I say, or in height, traditionally speaking). The rounded vowels, on the other hand, differ from [i] in two features: rounding and front-back. Rounding is a separate dimension which classifies the vowels into two, in the old Germanic languages at least, basic series, it seems, and the vowels thus adjust in vowel-space-position (or in place of articulation in its broadest sense) just like the unrounded vowels.

Of course, I do not claim that this in any way explains the mysteries of the *i*-Umlaut (for my personal view, see Basbøll 1982, further cf. Skomedal 1980 and Benediktsson 1982). It seems to me, however, that the conception of the feature front-back as presented here, indeed allows an interesting account of one aspect of its phonology, and that it is by no means inferior to other alternative accounts, by means of distinctive features, of which I am aware.

One final remark. I do not want to deny that e.g. Turkish and Finnish vowel harmony, as argued by Eli Fischer-Jørgensen, speak in favour of considering the distinction between [æ] and [ɑ] as one of front-back in those languages (although the treatment of loanwords suggests that these cases of vowel harmony are - at least to-day - not productive processes in the same strong sense as Danish *r*-colouring). How can this be conciliated with the conclusions of the present comment? If we are not to simply accept the weak (but not unreasonable!) position that two different sets of features may both be relevant in such cases, I shall again point to the fact that the back vowels which cannot be ("horizontally") fronted are exactly those which can be "front-raised", i.e. "front-raising" may be considered a complementary way of fronting which takes place instead of ("horizontal") fronting in certain well-defined cases, by convention. This is at present nothing but an airy suggestion, of course.

REFERENCES

Anderson, J.M. and Jones, Ch. 1977: *Phonological structure and the history of English (= North Holland Linguistic Series 33)* Amsterdam: North Holland

Basbøll, H. 1972: "Some conditioning phonological factors for the pronunciation of short vowels in Danish with special reference to syllabification", *Ann. Rep. Inst. Phon. Univ. Cph. 6*, p. 185-210

Basbøll, H. 1981: "Remarks on markedness and distinctive features in generative phonology", *Theory of markedness in generative grammar*, edited by L. Rizzi et al., p. 25-64, Pisa: Scuola Normale Superiore

Basbøll, H. 1982: "Nordic i-Umlaut once more: A variational view", *Folia Linguistica historica III/1*, p. 59-86

Basbøll, H. and Kristensen, Kj. 1975: "Further work on computer testing of a generative phonology of Danish", *Ann. Rep. Inst. Phon. Univ. Cph. 9*, p. 265-291

Basbøll, H. and Wagner, J. forthcoming: *Dänisch-deutsche kontrastive Phonologie: Segmentale Wortphonologie*

Benediktsson, Hr. 1982: "Nordic umlaut and breaking: Thirty years of research (1951-1980)", *Nordic Journal of Linguistics 5*, p. 1-60

Brink, L. and Lund, J. 1975: *Dansk Rigsmål 1-2*, Copenhagen: Gyldendal

Colman, F. and Anderson, J. 1983: "Front Umlaut: A celebration of 2nd fronting, i-Umlaut, life, food and sex", *Current Topics in English Historical Linguistics (= Odense University Studies in English 4)*, edited by M. Davenport et al., p. 165-190, Odense: Odense University Press

Fischer-Jørgensen, E. 1983: "Vowel features and their explanatory power in phonology", *Abstracts of the Tenth International Congress of Phonetic Sciences*, edited by A. Cohen and M.P.R. van den Broecke, p. 259-265, Dordrecht: Foris

Prokosch, E. 1938: *A comparative Germanic Grammar*, William Dwight Whitney Linguistic Series, Philadelphia: Linguistic Society of America

Skomedal, Tr. 1980: "Synkope, omlyd og bryting i nordisk", *The Nordic Languages and Modern Linguistics*, Proceedings of the Fourth Int. Conf. on Nordic and General Linguistics in Oslo 1980, edited by E. Hovdhaugen, p. 120-139, Oslo: Universitetsforlaget

ARIPUC 18, 1984, p. 285-291

A NOTE ON VOWEL TRIANGLES AND QUADRANGLES - AN ANSWER TO HANS BASBØLL

ELI FISCHER-JØRGENSEN

In his paper printed above Basbøll argues for regarding [ɑ] as a vowel which in the general distinctive feature system is neutral as far as the difference front-back is concerned, and which can thus be considered a common starting point for a front raising and a back raising line.

(1) I agree, of course, with Basbøll that it seems to be a general fact that front vowels get more retracted the lower they are. This also appears in a gradually lower F2, and it seems to be due to the structure of our vocal tract. Thus the difference front-back is smaller for low vowels. This is generally implied when we talk of front and back vowels, but it may be relevant for the explanation of some developments.

(2) I also agree with Basbøll in his interpretation of *r*-colouring in Advanced Standard Copenhagen Danish (ASC-Danish). It is certainly simpler to consider the modern Danish vowel system as triangular with /ɑ/ as the most open vowel. It gives a simple formulation of the *r*-colouring rule, and it is also in agreement with the phonetic manifestation. It is even possible to adduce some more phonetic evidence for this interpretation: The measurement of jaw opening for two ASC speakers (8-12 examples of the nonsense words /dæ:də/ and /dɑ:də/ in a frame sentence) showed a larger jaw opening for /ɑ:/ than for /æ:/ (for one speaker the difference is statistically significant at the 1% level, for the other at the 5% level). Moreover, /ɑ/ behaves as a very low vowel with respect to intrinsic duration and Fo. Figures 1 and 2 give a graphic display of intrinsic duration and Fo for ASC speakers based on measurements by Bundgaard (1980) and by Reinholt Petersen (1976).

Since we are discussing articulatory features (with rounding as a separate dimension) I should, however, find a three-dimensional graph more adequate. Basbøll here follows the tradition from Daniel Jones, which has also given difficulties for the use of the cardinal vowel chart. In Basbøll's figure the common point for /ɑ/ and /ɒ/ seems to indicate that /ɒ/ is also neutral in the back-front dimension. But is that really in-

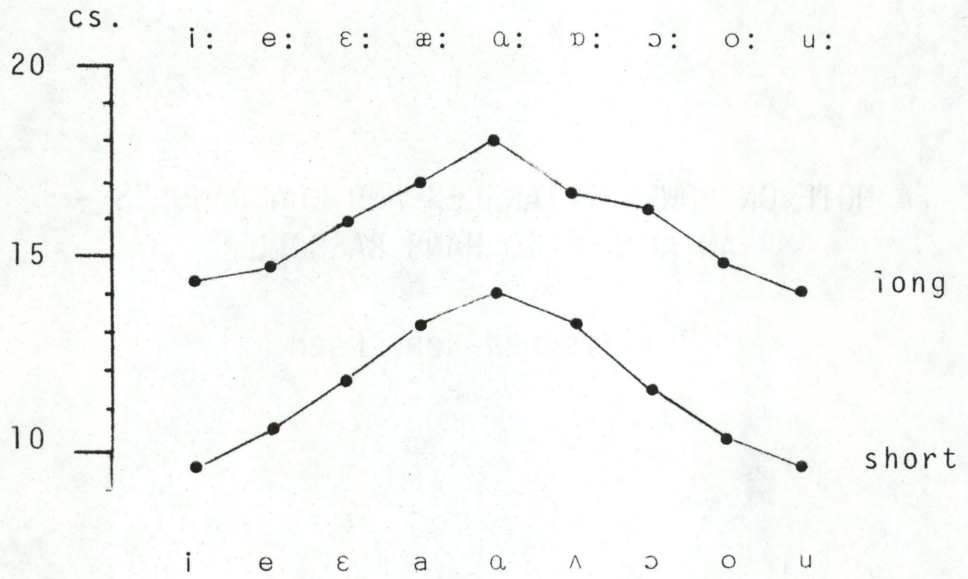


Figure 1

Intrinsic duration of Danish front unrounded vowels and back vowels (based on Bundgaard 1980, 5 subjects)

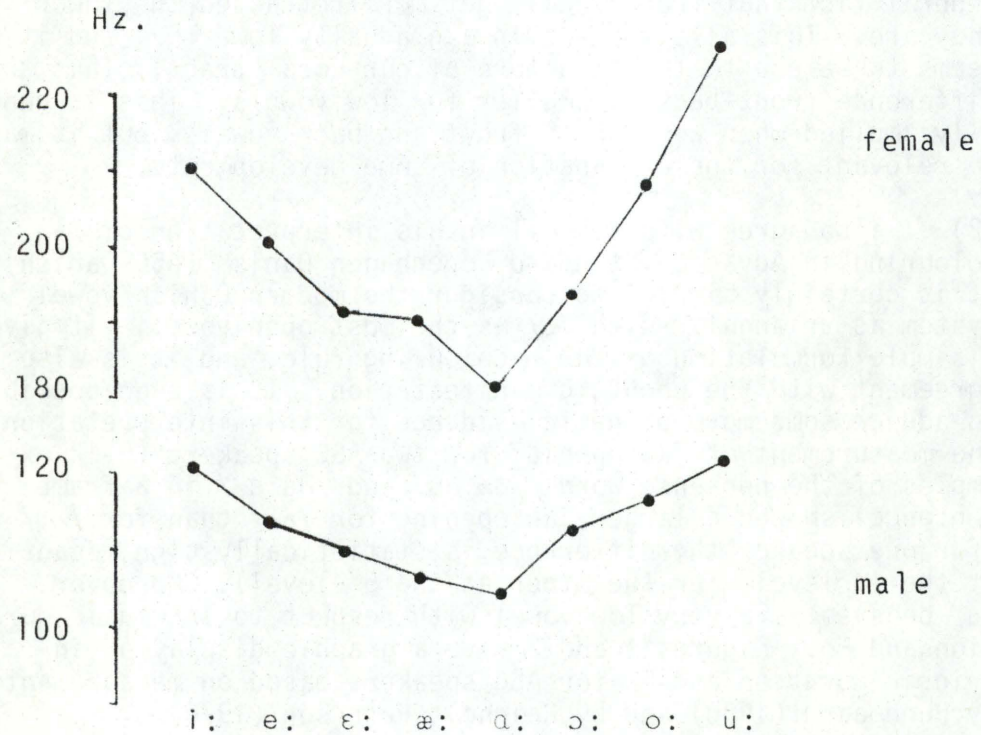


Figure 2

Intrinsic Fo of Danish front unrounded vowels and back vowels (based on Reinholt Petersen 1976, 5 subjects)

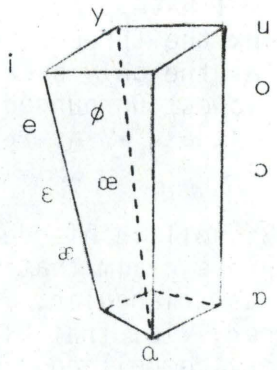


Figure 3

A three-dimensional vowel graph (| high-low, — front-back, / unrounded-rounded), with ASC Danish long vowels.

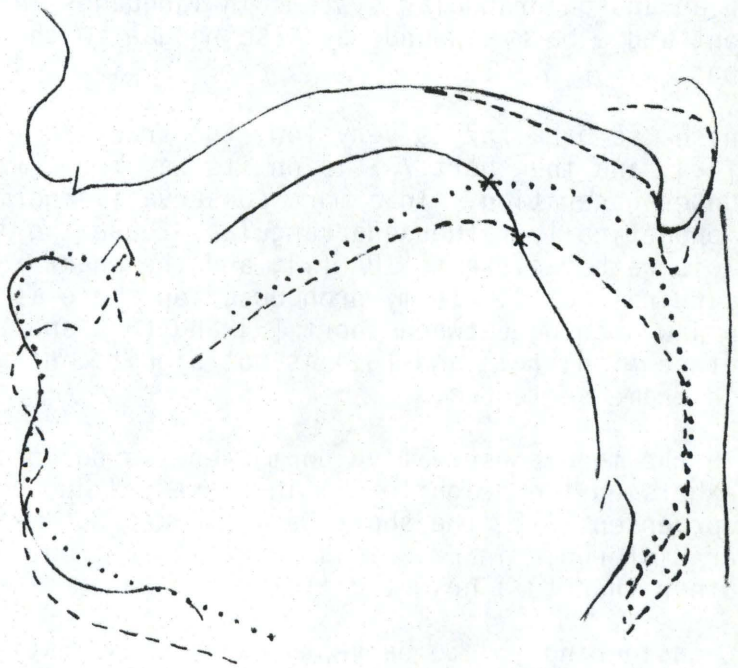


Figure 4

Tracings from X-ray photos of Danish [æ:] —— , [ɑ:] --- , *and* [ʁ] ··· *(subject EFJ); xx : approximate placement of pellets.*

tended? A three-dimensional version of Basbøll's graph might perhaps look like Figure 3. I have placed the long Danish vowels in the graph. I think the three-dimensional figure shows more clearly why /ɑ/ as the only back unrounded vowel can simply function as the lowest unrounded vowel and thus get in the same relation to /i e ɛ æ/ as /ɒ/ to the back rounded vowels.

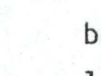
(3) Whereas I agree with Basbøll in his description of ASC Danish, I do not agree with his claim that the general feature system should be specified as triangular. I do not find the argument against the a-corner, viz. that [i] [u] [ɑ] but not [a] are close to consonants, very compelling. In the first place, [ɑ] is close to a consonant not because the tongue is low but because it is retracted, and in the second place [u] does not represent a clear corner either in this sense, since all vowels which are close to the palate or the uvula or the pharynx wall are close to consonants.

As for the relation between the sounds normally labelled [a] and [ɑ], Basbøll says himself that the main phonetic difference is front-back, and that an individual language may have an [ɑ] that is higher than cardinal No. 5 and may thus get a quadrangular system. I agree, but the differences between phonetically triangular and quadrangular systems in languages which have both a front and a back a-sound may also be due to the height of the front [a].

In ASC Danish the back /ɑ/ is very low; the front /a:/ is in fact IPA [ɛ:], and the short /a/ is on its way from [æ] to [ɛ] in the younger generation. In a more conservative norm the system is phonetically rather quadrangular. Here the long front /a:/ is rather close to IPA [æ], and the short /a/ is close to cardinal No. 4. In my pronunciation there is no difference in jaw opening between short [a] and [ɑ], and the difference between long [æ:] and [ɑ:] is not significant when said in words in frame sentences.

Further, in the more conservative norm there is no consistent difference between the height of F1 in short [a] and back [ɑ:] (Fischer-Jørgensen 1972, the short back [ɑ] was not measured), but a clear difference in F2. This points to a front back difference rather than to a height difference.

Moreover, considering [ɑ] to be kept low by preservative assimilation to [ʁ] is not a quite convincing formulation. In my pronunciation, at least, [ʁ] is higher than [ɑ]. Figure 4 shows tracings after X-ray photos of my vowels [æ:] and [ɑ:] pronounced in isolation, and of [ʁ] pronounced in slow speech in the sequence [ɑʁɑ]. It is obvious that when pronounced in isolation [ɑ] is both lower and more retracted than [æ], the lowering being due mainly to a wider jaw opening (there is not much difference in tongue height seen in relation to the jaw). The difference between [ʁ] and [ɑ] is almost exclusively one of height (cf. that [ʁ] has a lower F1 than [ɑ] (Heger 1975)), whereas the difference between [ʁ] and [æ] is mainly one of

fronting or backing of the whole tongue body. The jaw opening is the same and there is not much difference between the height of the raised parts of the tongue. Thus a preservative assimilation of [ɑ] to [ʁ] would keep [ɑ] back (with a concomitant narrowing of the pharynx), but there is no reason why it should be kept particularly low. In accordance with these X-ray pictures, a pellet tracking recording of the sequence [ɑʁɑ] undertaken in Tokyo 1976 showed the back of my tongue to go simply vertically up from [ɑ] to [ʁ] and down again from [ʁ] to [ɑ]. (The pellets were placed approximately at the place of the crosses in Figure 4 - unfortunately there was no pellet on the tongue root.) - Pellet tracking recordings of the sequences [eʁe, ɛʁɛ, øʁø, œʁœ] showed a movement obliquely back and up and down again with a slightly rolling movement, but opposite the one described by Delattre (1971). He found a movement down-back-up-front  , whereas in my case it was . The slight lowering before the fronting after /ʁ/ may perhaps contribute to explain the rising-falling movement in F1 found by Heger (1975) after [ʁ].

I suppose that in theory a front [a] may be just as low as the lowest back [ɑ] (you can in fact open the mouth more in the front). Even in Daniel Jones' own pronunciation of the cardinal vowels, where he tried to make No. 5 as low as possible but did not make any such effort for No. 4, the difference in tongue height is very small (Jones 1947) (F1 is even consistently lower in cardinal No. 5 than in No. 4 (Ladefoged 1967), but this may be due to a certain rounding).

In the phonological systems of individual languages the distinctive function and the behaviour in phonological rules must be taken into consideration. Trubetzkoy (1939) presented good arguments for considering some languages to have triangular and some to have quadrangular vowel systems. In my paper I mentioned that in Turkish and Finnish the facts of vowel harmony can be formulated most simply in terms of front-back, and here /ɑ/ functions as a back vowel.

Finnish				Turkish				
i	y	u		i	y	ɯ	u	
e	ø	o		ɛ	œ	ɑ	ɔ	
æ		ɑ						

In the conservative Danish norm [ʁ] may be pronounced after a back [ɑ] and [ɒ]. Thus, for instance, although my own pronunciation has been influenced by the ASC norm in various respects in later years, there may still be an [ʁ] in distinct speech medially before [i], e.g. *varig*, *Karry*, and I normally pronounce an [ʁ] in final voiceless clusters after short [ɒ] and [ɑ] (e.g. *kort*, *barsk*). In such a norm front and back a-sounds are variants of one phoneme, and phonologically the system is triangular with a very variable lowest phoneme, which is front in most positions, but back in some. In most triangular vowel systems, however, there is only one a-sound which may be more or less central.

It is possible that the Nordic *i*-Umlaut may support the assumption of a triangular system in Old Nordic. But in German the situation is not quite the same. In Old German /e/ may become /i/ both before /i/ and /u/, so that this should rather be regarded as raising before high vowels. The umlaut may then be described as a fronting of /u/ and /o/ and a fronting and raising of /ɑ/. This is not more complicated than Basbøll's formulation, and the common treatment of /u o ɑ/ speaks for the description of /ɑ/ as a back vowel.

I do not think there is much disagreement between Basbøll and me about all these different possibilities in different languages. But I deviate from him by drawing the conclusion that it would be very unpractical to set up a general system of distinctive features which was specified as either triangular or quadrangular. In my opinion the general system should be a system of independent dimensions, which do not meet, and a system of dimensions only, not of discrete steps. Thus neither the number nor the placement of the steps should be specified, although for purely practical reasons it is useful to agree on the transcription of some commonly used sounds. In this transcription [ɑ] is commonly used for a back vowel and [a] for a more fronted vowel. In some individual phonological systems [ɑ] may be the most open vowel, phonetically or phonologically or both, in other languages, with only one a-sound, front or central /a/ may be the most open vowel, and in so-called quadrangular systems there may be more than one vowel in the most open row. - But I consider these specifications as belonging entirely to the individual languages.

REFERENCES

Basbøll, H. 1984: "On the relation between vowel height and front-back: A comment on Eli Fischer-Jørgensen's paper: "Some basic vowel features, their articulatory correlates and their explanatory power in phonology" (this volume, p. 277-284)

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

Delattre, P. 1971: "Pharyngeal features in the consonants of Arabic, German, Spanish, French and American English", *Phonetica 23*, p. 129-155

Fischer-Jørgensen, E. 1972: "Formant frequencies of long and short Danish vowels", *Studies for Einar Haugen*, edited by Scherabon Firchow, Grimstad, Hasselmo and O'Neil (The Hague: Mouton), p. 189-213

Fischer-Jørgensen, E. 1984: "Some basic vowel features, their articulatory correlates and their explanatory power in phonology", (this volume, p. 255-276)

Heger, S. 1975: "Danish r and adjacent short stressed vowels", *Ann. Rep. Inst. Phon. Univ. Cph. 9*, p. 137-169

Jones, D. 1947: *An Outline of English Phonetics*, 6th edition (Cambridge: Heffer and Sons)

Ladefoged, P. 1967: *Three Areas of Experimental Phonetics*, (Oxford University Press)

Petersen, N. Reinholt 1976: "Intrinsic fundamental frequency of Danish vowels", *Ann. Rep. Inst. Phon. Univ. Cph. 10*, p. 1-27

Trubetzkoy, N. 1939: *Grundzüge der Phonologie*, *Travaux du Cercle Linguistique de Prague 7*.

ARIPUC 18, 1984, p. 293-385

VOCAL FOLD ADJUSTMENTS IN DANISH VOICELESS OBSTRUENT PRODUCTION[1]

BIRGIT HUTTERS

Vocal fold adjustments in the production of single intervocalic voiceless obstruents in Danish were investigated partly by photo-electric glottography combined with fiberoptic stills, and partly by electromyographic signals obtained from several intrinsic laryngeal muscles. The results show that the glottal opening-closing gesture, controlled by the laryngeal muscles, and its relation to the supraglottal articulation vary not only according to type of obstruent category but also within one and the same category. In the discussion I argue that in dealing with aspiration our understanding of the laryngeal mechanism involved may be hampered if we focus on the inter-articulatory timing rather than on the glottal gesture as such. With regard to the glottal gesture in the unaspirated stops I assume that the opening of the glottis is a consequence of the cessation of vocal fold vibrations. However, I tend towards the view that devoicing in these stops is not only a passive process, i.e. due to the changing aerodynamic conditions induced by the oral closure, but that some additional mechanism is directly involved in the devoicing process. I venture the controversial hypothesis that the PCA and INT activity actually found in these stops is a devoicing action rather than a means to open and close the glottis. Finally, some problems relating to the interpretation of the laryngeal EMG signals with reference to obstruent production are pointed out.

I. INTRODUCTION

In speech the vocal folds are sometimes abducted in order to satisfy the aerodynamic demands for producing sounds such as aspirated stops and fricatives, whereas other sounds such as unaspirated stops and sonorants are produced with adducted or nearly adducted vocal folds. The demand for different degrees of glottal opening necessarily results in movements of the vocal folds controlled by the intrinsic muscles of the larynx. Thus, during speech we observe an opening and closing of the glottis, which in fact may be considered transitions from one glottal "state" to another, alternating with periods of a more static appearance. This articulatory behaviour may under adequate conditions be overlaid by varying periods of vocal fold vibrations.

The present study deals with vocal fold adjustments in the production of Danish unvoiced obstruents occurring as single consonants in intervocalic position before a stressed vowel. Besides a short presentation of how the glottographic, fiberoptic, and electromyographic recordings were performed, section II contains a discussion of some more general aspects of the delimitation of acoustic signals and gives an account of the treatment of the recordings, including a fairly detailed description of the delimitation procedure applied to the mingographic material. In section III the results are presented as comparisons between obstruents belonging partly to different categories and partly to one and the same category. Differences in the glottal gesture itself and in its temporal relation to the supraglottal articulation are pointed out and related to the laryngeal muscular activity underlying the articulatory movements of the vocal folds, and the results are discussed in the light of other authors' findings. Section III closes with a presentation of the results concerning the period of vocal fold vibrations in the unvoiced obstruents including a discussion of the factors that may influence the offset and onset of voicing of these sounds. In section IV the production of (post-)aspiration is discussed from a more general aspect with special reference to the common view that aspiration is a matter of timing between the glottal and supraglottal articulation rather than the result of the glottal gesture per se. Then, the glottal gesture in Danish unaspirated stops is discussed from a devoicing point of view, and the section ends by pointing out some problems in the interpretation of the electromyographic signals with reference to vocal fold behaviour in the production of obstruents. The paper closes with some concluding remarks.

It will be relevant to discuss some results - found in the present Danish material as well as in the literature - from various points of view and in more contexts. This means that the same findings may appear several times here, which disrupts the logical development of the presentation. The main reason is that the results on Danish presented in section III and the relevant data and suggestions found in other studies may also be discussed in another and often in a more general context in the three sub-sections composing section IV. The

most obvious example relates to onset and offset of voicing in unvoiced obstruents: in section IIIC it is treated from the point of view of extent of voicing, in section IVB it appears in the context of devoicing in Danish *bdg*, and finally, in section IVC it is treated in relation to the discussion on interpretation of laryngeal EMG signals.

II. METHOD

The present study includes partly fiberoptic and glottographic recordings and partly electromyographic recordings (henceforth EMG). Unfortunately, the EMG signals were not obtained synchronously with the two other signals, which obviously reduces the information that can be extracted from the material. However, by way of consolation, several of the speakers served as subjects in both parts of the study.

A. FIBEROPTIC AND GLOTTOGRAPHIC RECORDINGS

Still-pictures of the vocal folds were taken through a fiberscope inserted through the nose and placed in the subject's pharynx. A still was taken during each test obstruent. The light guide of the fiberscope served as light source for the glottographic recordings, and the light was picked up by a photo-transducer placed on the frontal part of the neck - in a position approximately between the thyroid and cricoid cartilages. In order to synchronize the stills and the photo-electric glottograms a synchronizing pulse was generated, triggered by the synchronizing switch lever of the camera. After amplification the glottographic signal was fed into a professional DC tape recorder along with the synchronizing pulse and a microphone signal. The manual processing of the recorded material was based on mingographic registrations comprising the glottographic curve, the synchronizing pulse and various curves extracted from the microphone signal, intended for the acoustic delimitation. A detailed description of the fiberoptic and glottographic set-up is found in Andersen (1981).

B. ELECTROMYOGRAPHIC RECORDINGS

The EMG recordings were made using hooked-wire electrodes as described by Hirose (1979) and Hirose and Sawashima (1981).[2] EMG signals were obtained from the following four muscles, though only from two at a time: the posterior cricoarytenoid (PCA) and the interarytenoid (INT) muscles on one hand, and the vocalis (VOC) and the cricothyroid (CT) muscles on the other. The EMG signals were recorded and computer processed with a modified version of the system described by Holtse and Stellinger (1976).[3] It should be added, though, that each separate EMG signal had been highpass filtered before it was fed into the A/D converter in order to eliminate various disturbances, using a cut-off frequency well above the low range normally used, in accordance with the findings of Rischel and Hutters (1980).[4]

C. SUBJECTS

In the fiberoptic-glottographic study five adult speakers served as subjects - three female (HU, MF, FJ) and two male (LG, PA). HU (the author), LG, and PA are speakers of Standard Copenhagen, whereas FJ and MF have a more conservative pronunciation.

The EMG material presented in this paper includes seven speakers, three female (HU, FJ, BH) and four male (PM, LG, BM, JJ). All but FJ are speakers of Standard Copenhagen - except that PM is slightly influenced by his Jutland dialect. As it appears from the abbreviations, three of the speakers occur in both parts of the study.

Table I gives a survey of the distribution of subjects in the two parts of the study. It should be recalled that the EMG signals in no case were recorded simultaneously with the fiberoptic stills and the glottographic signal, and that the EMG activity was picked up from only two muscles at a time and always in the combination PCA + INT and VOC + CT.

Table I

Survey of subjects and muscles recorded in the two parts of the study. - means that the signal has been omitted due to its unreliable quality in some respects.

Subjects	PA	MF	FJ	LG	HU	PM	BM	BH	JJ
fiberoptic/ glottography	x	x	x	x	x				
laryngeal muscles									
PCA			x	x	x	x			
INT			-	x	x	x			
VOC					x		x	x	x
CT					x		x	-	-

D. LINGUISTIC MATERIAL

The Danish unvoiced obstruents include *ptk bdg fsh*[5], all contrasting in (absolute) syllable initial position before a full vowel. The main difference between the two stop categories is one of aspiration, *ptk* being aspirated and *bdg* unaspirated (see, e.g., Fischer-Jørgensen 1954, 1968b, 1980). The test material consisted of meaningful words as follows: *pile Tine/ tie kilde bile dine/die gilde file sile hige*. Only the *t*- and

d-words to the left of the slashes were included in the fiberoptic-glottographic material, whereas in the EMG material one or the other of the two words might be included.[6] In the fiberoptic-glottographic recordings the words were said in the frame sentence *"De ville sige..."* ('they would say...') pronounced [di vi si:], whereas in the EMG recordings they were placed in a variety of sentences of similar type. Thus, the test material includes single unvoiced obstruents occurring intervocalically in word initial position before a stressed vowel. The EMG material, however, is somewhat more comprehensive as it includes also sonorants and *v*, and the following vowel may be either *i* or *a*. In some relevant cases I will refer to this additional material.

Now, it is well-known that disturbances in the transmission between the light source (i.e. the fiberoptic cable) and the photo-transducer pose a problem for the glottographic method. The disturbances are caused partly by external factors such as coughing and swallowing, partly by the speech conditions themselves. These disturbances may cause variations in the level of the glottographic signal that are certainly not reflections of the variations in the size of the glottal aperture. So, in order to minimize the influence from the varying speech conditions, sounds produced with a more or less constricted pharynx should be avoided in the test material. Therefore, the test sentences for the glottographic recordings were designed so as to include only sounds articulated in the oral part of the vocal tract. Problems relating to the interpretation of the glottographic signal are discussed in Appendix C.

The test sentences were presented in a list containing two randomizations. In the fiberoptic-glottographic part each sentence was read about ten times, whereas in the EMG part the number of repetitions varied from six to about ten. In both materials, however, some tokens had to be eliminated for some reason or another. The exact number of tokens included in the statistical treatment of the data extracted from the mingographic material are found in the tables in Appendix A, whereas the number of tokens included in each averaged EMG signal appears from the figures.

E. TREATMENT OF THE RECORDINGS

1. FIBEROPTIC STILLS AND EMG RECORDINGS

The averaging of the EMG signals is a modified version of the processing described by Holtse and Stellinger (1976).[3] In the averaging procedure the EMG signals are aligned with reference to an acoustic line-up point, which for each averaging can be chosen freely among the time references specified in advance during the off-line preparation of the data. The line-up point actually used was the onset of the vowel preceding the test obstruent.

The averaged EMG curves will be described qualitatively only, but the description is supplemented by a fairly large number of figures which also illustrate the considerable variation that there may be between the subjects.

The treatment of the fiberoptic stills is likewise confined to a qualitative description serving to supplement the glottographic data.

2. DELIMITATION OF THE MINGOGRAPHIC MATERIAL AND EXTRACTION OF PARAMETERS

On the basis of the mingographic registration including both acoustic and glottographic curves a number of parameters have been extracted, which may give some additional insight into the glottal articulatory behaviour and its temporal relation to the supraglottal articulation. However, before going into the delimitation procedure actually applied, I want to point to some more general aspects of the whole topic.

It should be considered that statements and results relating to the temporal behaviour in speech are, to a certain extent, a function of the traces used and the criteria adopted for delimitation. As to the first point, delimitation is usually undertaken on the basis of acoustic curves. In many cases, however, we introduce a more or less implicit interpretation in terms of speech production and especially in terms of supraglottal articulation. It goes without saying that the temporal relations,found on the basis of acoustic curves interpreted in terms of production, may deviate from those found on the basis of curves directly representing the production. In this connection it should be recalled that since the acoustic signal reflects both glottal and supraglottal events, a comparison of the timing of acoustic and glottographic signals may lead to wrong statements about the temporal relation between glottal and supraglottal articulations.

Relating to the question of criteria applied for the delimitation of the acoustic signal I want to point out that if different criteria are applied according to the segments involved, this may, of course, also lead to false statements about differences and similarities between segment durations or durations of other sequences in which the acoustic delimitation is involved. If, for instance, the delimitation of the onset of the stops is taken to be the point in the acoustic signal where there is a clear reduction in intensity, whereas the onset of fricatives is taken to be the point where fricative noise appears, it is obvious that the duration of, say, the preceding vowel, or the period from the onset of the obstruent up to the maximum glottal aperture, may differ simply as a consequence of the different delimitation procedure.

It is rather common to use the offset and onset of voicing for delimitation of unvoiced obstruents. However, since the timing of these events are primarily dependent on the manner of pro-

duction, rather than on the timing of the articulatory movements as such, temporal findings of this kind are not relevant to the question of articulatory timing.

Particular attention should be given to the possibility that discrepancies between the temporal findings of different studies may be due to the delimitation procedures (choice of curves and criteria for delimitation) applied in the individual studies. This, in fact, may also result in erroneous statements about inter-language differences or similarities with regard to temporal phenomena.

The delimitation criteria applied to the present mingographic material are as follows (see figure 1):

a. delimitation of the acoustic curves

V symbolizes the onset of the vowel preceding the test obstruent.

C_1 symbolizes the start of the test obstruent. This is - in the case of labial and alveolar stops - taken to be the point in the intensity curve where it shows an abrupt fall. The oscillogram normally shows weak oscillations after this point. In the velar stops and in the fricatives, however, the decrease is often more gradual which introduces some uncertainty - especially in *h*. It should be underlined that the start of the obstruent is defined neither as the offset of periodic oscillations, nor - in the case of fricatives - as the onset of fricative noise.

C_2 stands for the offset of the test obstruent. This point is by definition identical with the onset of the following vowel. In most cases the start of the high *i* vowel after a voiceless obstruent does not pose any problems, since the energy starts almost simultaneously and abruptly over the whole spectrum, which is seen as an abrupt rise in the intensity curve. The delimitation of vowel start after aspirated stops is discussed in detail by Fischer-Jørgensen and Hutters (1981).

E is the moment of oral explosion of stops. It shows up as a clear rise in the intensity curve.

b. delimitation of the glottographic curves

G stands for the onset of the glottal gesture and is defined as the moment when the glottographic signal rises at the transition from vowel to obstruent. This increase may, though, be rather gradual, which introduces some uncertainty.

M is the moment of maximum glottal aperture defined as the moment of peak level of the glottogram.

Z symbolizes the offset of vocal fold vibrations and is taken to be the moment when the ripple on the rising glottographic signal is no longer visible.

c. Extraction of parameters On the basis of these delimitations the following durational parameters are extracted:

Acoustic parameters

(I) C_1C_2: total duration of the obstruent.

(II) C_1E: duration of the oral closure.

(III) EC_2: duration of the open interval. In the case of aspirated stops the open interval includes explosion noise, fricative noise in the case of *t* and *k*, and aspiration. The term 'aspiration' will be used in this wider sense as synonymous with 'open interval'.

(IV) VC_1: duration of the preceding vowel.

(V) VC_2: duration of the period including the whole obstruent and the preceding vowel.

(VI) VE: duration of the period including the oral closure and the preceding vowel.

Acoustic-glottographic parameters

(VII) VG: duration from the start of the preceding vowel to the onset of the glottal gesture.

(VIII) VM: duration from the start of the preceding vowel to the maximum glottal aperture.

(IX) GC_1: duration from the onset of the glottal gesture to the onset of the obstruent. This parameter gives a negative value if the onset of the obstruent precedes the gesture onset.

(X) ME: duration from the maximum glottal aperture to the explosion of the oral closure. This parameter gives a negative value if the explosion precedes the maximum glottal aperture.

(XI) C_1Z: duration of the time interval with sustained vocal fold vibrations after the onset of the obstruent, i.e. the extent of physiological voicing.

Glottographic parameters

(XII) GM: duration of the glottal abduction.

(XIII) A: In addition to the temporal parameters the peak level of the glottogram has been measured with the minimum level in the preceding vowel as reference.

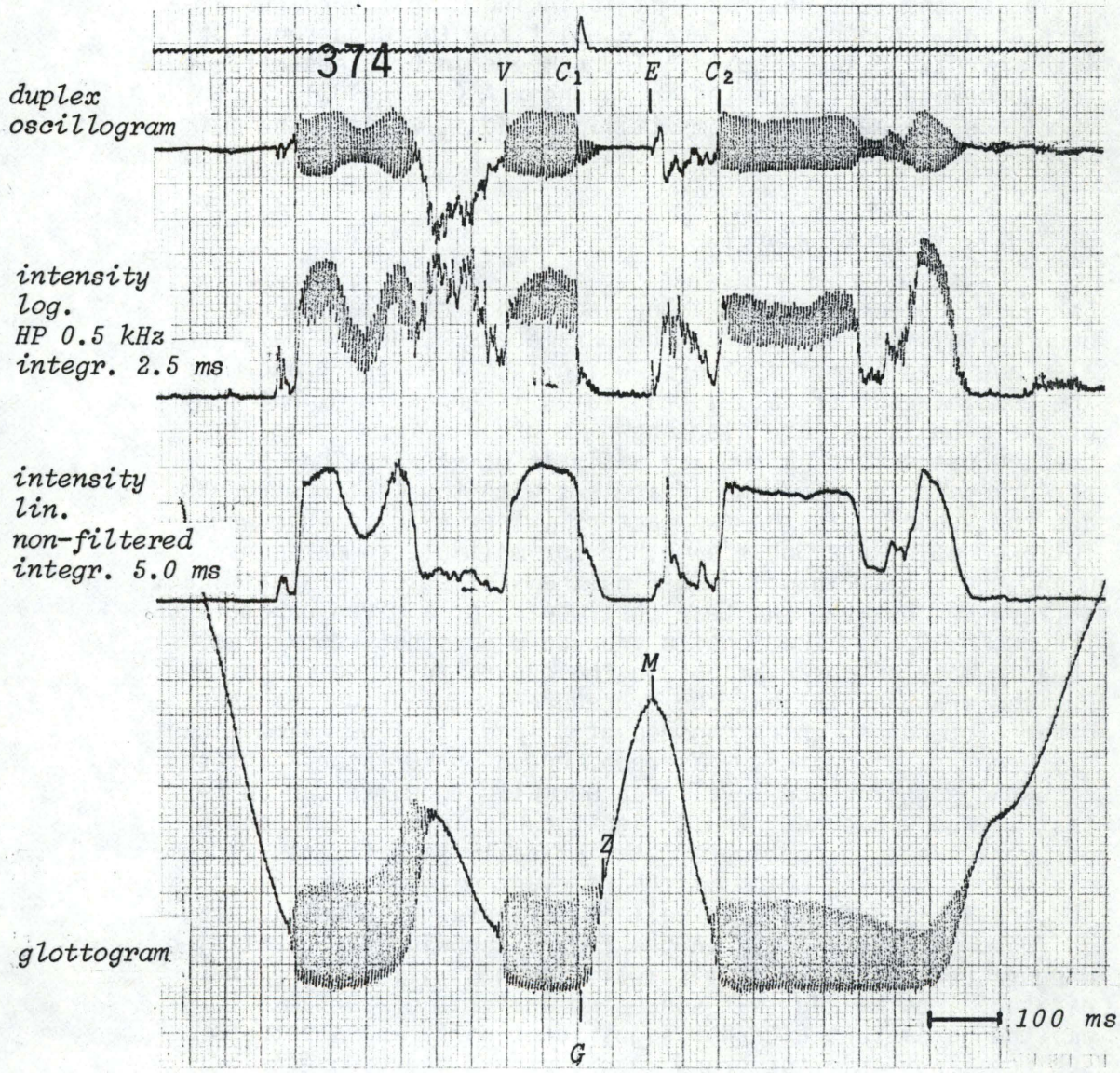


Figure 1

Delimitation of the mingographic material. V = onset of the vowel preceding the test obstruent; C_1 = onset of the test obstruent; E = the moment of oral explosion; C_2 = offset of the test obstruent (identical with the onset of the following vowel); G = onset of the glottal gesture; M = the moment of maximum glottal aperture; Z = offset of vocal fold vibrations.

The discussion in Appendix C about the interpretation of glottographic signals is relevant for parameter A and for the parameters including G (taken as the onset of the glottal gesture). It is also explained why no delimitation of gesture offset has been made. It should be added that since the onset of the preceding vowel V can be considered independent of properties of the test obstruent it is included in several parameters as an independent reference point for the occurrence of later events in the signals. The numbers in brackets refer to the numbering of the tables in Appendix A (see below). All durations are measured in ms (milliseconds), whereas the peak level of the glottogram is stated in mm.

3. STATISTICAL TREATMENT

For each subject the mean value and standard deviation has been calculated for each of the parameters. The raw data are also averaged across speakers, which is called the grand mean. The complete set of data is tabulated in Appendix A, but many of the mean values are also shown in graphic form in the figures. Furthermore, a two-way analysis of variance has been performed for each of the parameters with the test obstruents and the subjects as the two factors (Winer 1970). This was followed up by the Scheffé method for multiple comparisons using the F-test in order to find the statistically significant differences between the test obstruents for a given parameter. (Ferguson 1976). The Scheffé procedure does not require an equal number of observations in the groups to be compared. But the drawback is that the procedure is more rigorous than other procedures, which consequently leads to fewer significant differences. Therefore - as recommended by Ferguson - a less rigorous significance level than normally required for an F-test may be employed. Thus, significance levels of 10 and 5 per cent are also considered statistically significant. A complete list showing which of the relevant differences are statistically significant, as well as the level of significance, is found in Appendix B. In the figures showing the quantified mingographic data the statistical findings will also be indicated. Let me add that since the analysis of variance for all the parameters did show significant differences between subjects, not only the grand mean averaged over subjects but also the mean for each of the subjects are shown in the tables as well as in the figures.

III. RESULTS

The results will be presented as comparisons partly between obstruents belonging to different categories specified as aspirated stops versus unaspirated stops, and aspirated stops versus fricatives, and partly between obstruents belonging to the same category but differing as to place of articulation.

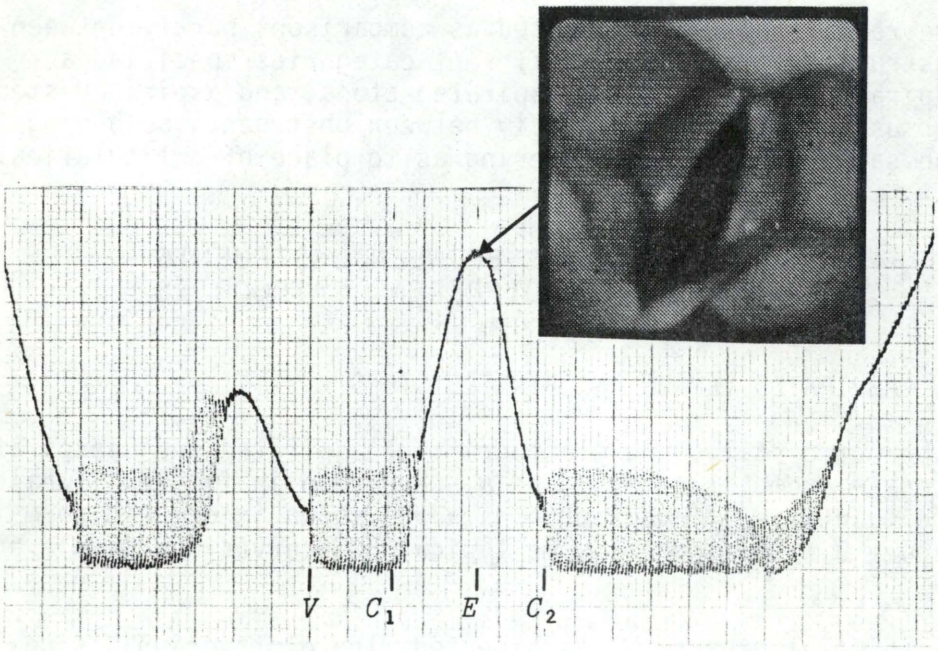
A. COMPARISONS BETWEEN OBSTRUENTS BELONGING TO DIFFERENT CATEGORIES

1. ASPIRATED VERSUS UNASPIRATED STOPS

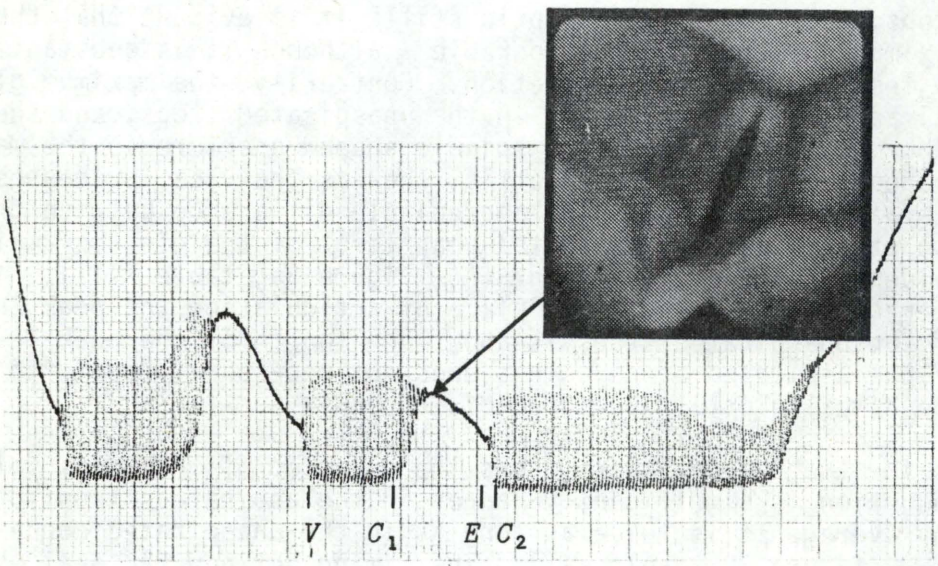
Fiberoptic stills and glottograms of aspirated and unaspirated stops - exemplified by *p* and *b* - are seen in figure 2, while the quantified mingographic data are shown in graphic form in figure 3. In figure 4 some typical EMG curves are displayed.

a. glottal gesture As expected, the glottographic signal shows a clear and almost symmetric opening-closing gesture in the aspirated stops, and from the fiberoptic stills it is evident that the maximum abduction is considerable - although it is substantially less than during respiration. Contrarily, the maximum glottal aperture is very small in the unaspirated stops, and the fiberoptic stills reveal a spindle shaped aperture in the membraneous portion of the glottis, whereas the vocal processes are almost adducted.[7] The appreciable difference between the two stop categories as to the peak level of the glottogram is evidently highly significant (A - figure 3a, table XIII). Averaged over subjects and place of articulation the peak level of the unaspirated stops is about one fourth of the peak level of the aspirated stops. This does not necessarily apply to the actual glottal aperture, cf. Appendix C.

As for the temporal parameters the duration of the vocal fold abduction is significantly longer in the aspirated stops; on the average it is twice as long as in the unaspirated cognates (GM - figure 3b, table XII). The timing between the oral explosion and the moment of the maximum abduction is also significantly different (ME - figure 3c, table X). In the aspirated stops the explosion occurs close to the moment of maximum aperture - on the average 20 ms before the maximum - whereas in the unaspirated stops the explosion always lags behind the moment of maximum abduction - by about 50 ms on the average. It appears from the glottograms that in the unaspirated cognates the explosion occurs close to the end of the glottal gesture, and shortly after the release - about 30 ms on the average - the pressure drop across the glottis is again sufficient for the resumption of vocal fold vibrations. In the aspirated stops this open interval - from the explosion and until the vocal folds have regained their adducted position -



de ville sige pile [b̥ʰi:lə]



de ville sige bile [b̥i:lə]

Figure 2

Fiberoptic stills and glottograms of aspirated and unaspirated stops exemplified by p and b, respectively. The arrows show where in the course of the glottal gesture the still originates. The acoustic events are shown below the glottogram (cf. figure 1).

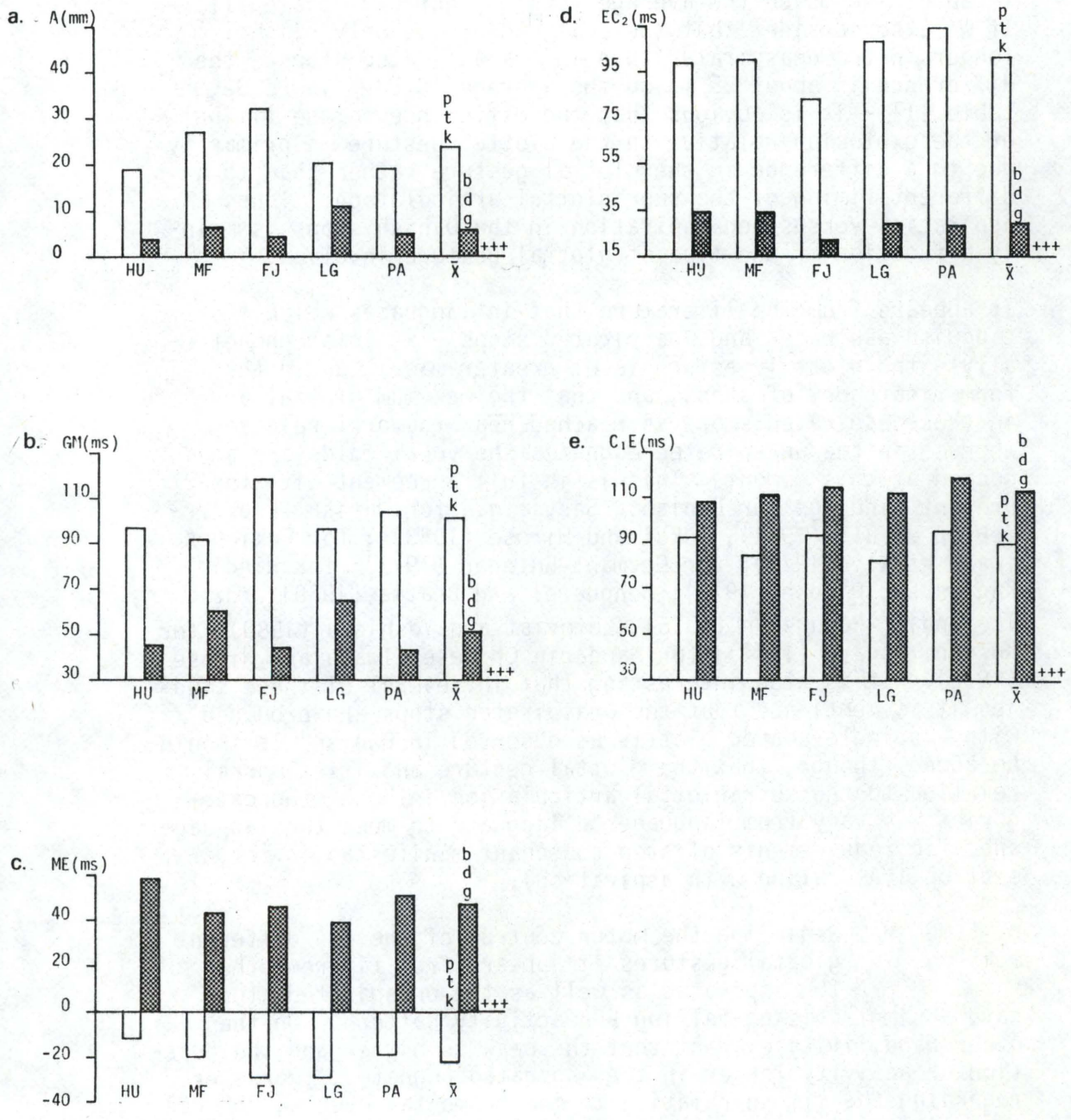


Figure 3

Mean values for aspirated and unaspirated stops - averaged over place of articulation - of the parameters A (peak level of the glottogram), GM (duration of the glottal abduction), ME (duration from the maximum glottal aperture to the explosion of the closure), EC_2 (duration of the open interval), and C_1E (duration of the oral closure). The mean values are shown for each speaker and averaged across speakers ($\bar{X}$ - computed from raw data). The left column represents the aspirated stop category, the right one the unaspirated category, as indicated to the right in each graph. The level of significance for the difference between the grand means ($\bar{X}$) is indicated as follows: xxx = $p<0.01$, xx = $p<0.05$, x = $p<0.1$, o = $p>0.1$.

is about 100 ms on the average (EC_2 - figure 3d, table III). If we also consider that the oral closure is only slightly longer in the unaspirated than in the aspirated stops - the difference is about 25 ms on the average (C_1E - figure 3e, table II) - it is obvious that the difference in the timing of the explosion relative to the glottal gesture is primarily due to a difference in the glottal gesture rather than to a different timing of the supraglottal articulation. Thus, aspiration versus non-aspiration in the Danish stops is mainly a function of the type of glottal gesture involved.

It appears from the literature that in languages which distinguish aspirated and unaspirated stops - at least phonetically - the glottal gesture is of greater magnitude in the former category of stops, and that the maximum glottal aperture in these aspirated stops is reached near the oral release, whereas in the unaspirated cognates the vocal folds are adducted at this moment. This is in full agreement with the present findings for Danish. See, e.g., for Danish: Frøkjær-Jensen et al. (1971), Fukui and Hirose (1983)[8]; for Fukienese: Iwata et al. (1979); for German: Butcher (1977); for Hindi: Kagaya and Hirose (1975), Benguerel and Bhatia (1980); for Icelandic: Pétursson (1976), Löfqvist and Yoshioka (1980); for Korean: Kagaya (1974); for Mandarin Chinese: Iwata and Hirose (1976). It is also interesting that in several of these studies it is mentioned that the unaspirated stops are produced with a spindle-shaped glottis as observed in Danish. It should be added, though, that the glottal gesture and its temporal relation to the supraglottal articulation in both stop categories may vary from language to language to meet the language specific requirements of stop consonant manifestation (cf. section IVA dealing with aspiration).

b. EMG Regarding the motor control of the two different glottal gestures it appears from figure 4 that the aspirated as well as the unaspirated stops have a simple rising-falling PCA activity pattern. On the other hand, it is evident that the peak is higher and the duration of activity longer in the aspirated cognates. Moreover, regarding the timing relative to the segmental events, the PCA peak activity occurs very close to the onset of the unaspirated stops, whereas in the aspirated stops it is attained only 20-30 ms after the implosion.[9]

Concerning the INT muscle the well-known reciprocity between PCA and INT is also seen in the present material: there is a dip in the INT activity, which in two out of three subjects descends to a slightly lower level in the aspirated than in the unaspirated stops. Furthermore, the clear increase in activity and the following maximum occur later in the aspirated stops, which means that in both stop categories the INT maximum almost coincides with the onset of the following vowel. Consequently, the period of reduced INT activity is longer in the aspirated stops than in their unaspirated cognates. It is also worthy of notice that two subjects show a somewhat higher INT maximum after aspirated than after unaspirated stops, even

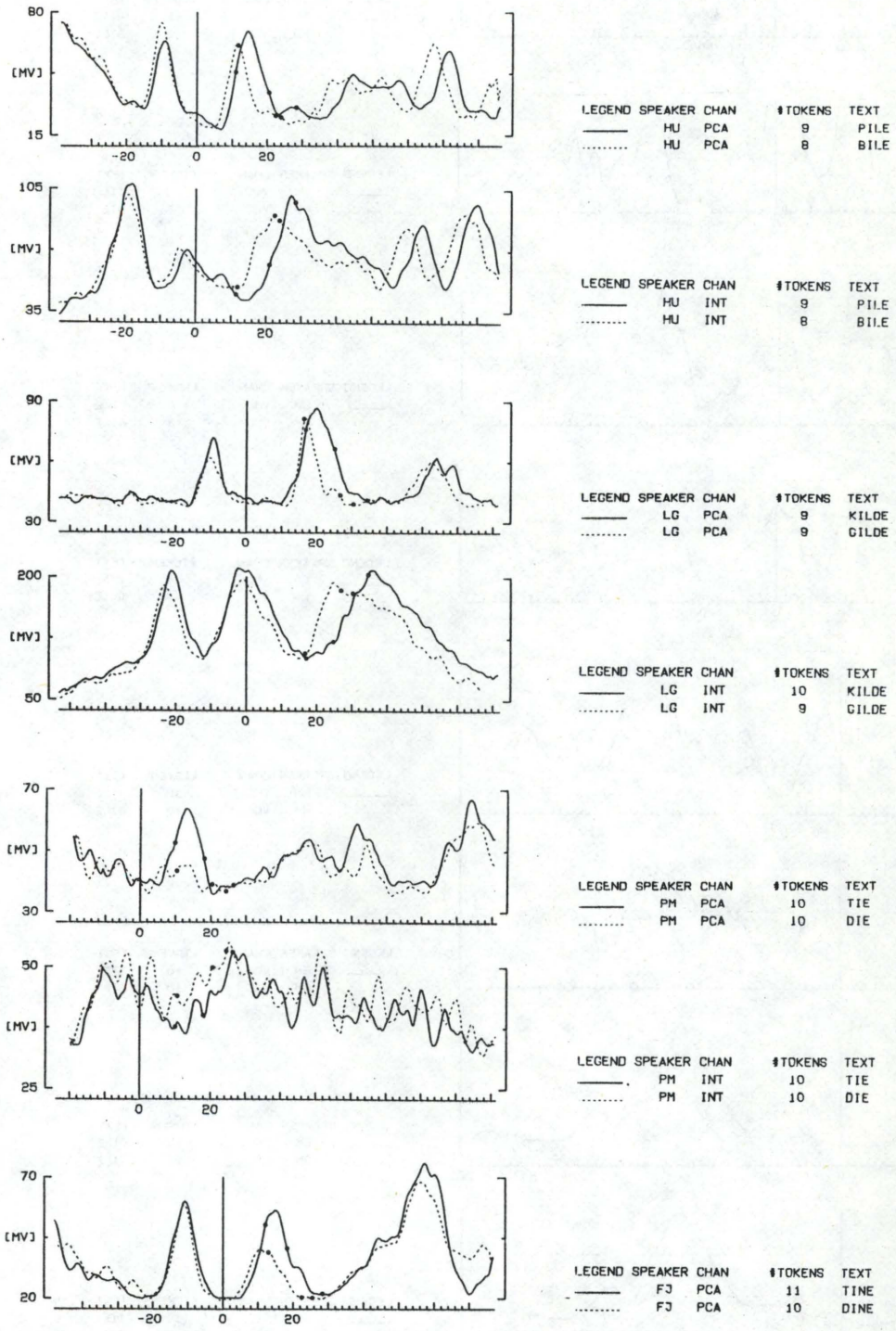


Figure 4

Averaged EMG curves of aspirated and unaspirated stops. The speaker, muscle, number of tokens, and place of articulation are indicated to the right of the curves. The line-up point indicated by the vertical line is the onset of the vowel pre-

(continued)

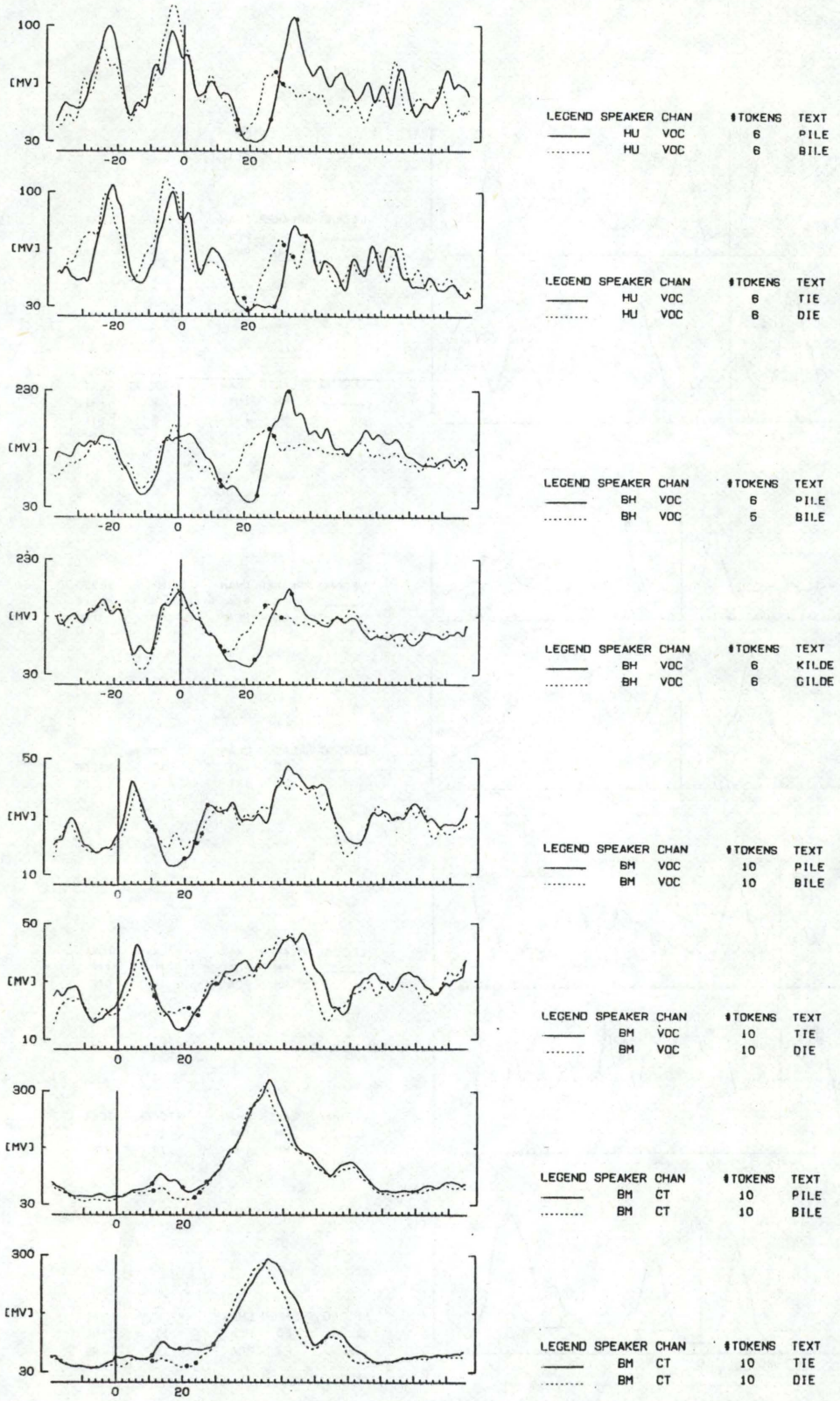


Figure 4 (cont.)

ceding the stop ("V" in figure 1). The small dots seen in each EMG curve indicate the acoustic events as follows: first dot = onset of the oral closure ("C_1" in figure 1), second dot = explosion of the oral closure ("E" in figure 1), third dot = the offset of the stop alias the onset of the following vowel ("C_2" in figure 1).

though the difference is not very stable in subject LG. Thus, generally it seems that the timing and amplitude differences in PCA and INT between the aspirated and the unaspirated stops reflect differences in the glottal gesture involved in the production of the two stop categories.

Turning now to the VOC muscle, there is a dip of activity in the aspirated as well as in the unaspirated stops, even though the dip is more pronounced in the aspirated ones. After the dip - almost at the explosion of the oral closure - the VOC activity starts to rise and reaches a peak at the transition from stop to the following vowel, nearly coinciding with the vowel onset in both stop categories, this pattern resulting in a longer period of reduced activity in the aspirated stops. The VOC peak is higher at the transition from aspirated stops, a difference that may continue into the vowel.

> *It appears from figure 4 that speaker BM deviates in several respects from this description. Furthermore, only JJ shows no dip in the unaspirated cognates, but of course there still remains a difference in minimum VOC level between his two stop categories.*

A differentiated pattern seems also to be present in the CT muscle which tends to have slightly higher activity during the aspirated stops, but it must be taken into account that the amount of data is very limited. Apart from this difference it is worthy of notice that the overall pattern of CT activity, with a clear rise in the following vowel correlating with a clear rise in the fundamental frequency, is almost identical in space and time, irrespective of stop category. This means that the level of CT activity in the following vowel - for any specific point in time reckoned from the vowel onset - is higher after aspirated stops due to the longer segment duration of these stops. This may be part of the explanation for the higher fundamental frequency found after aspirated than after unaspirated stops in Danish, and especially for the puzzling finding that the difference extends far into the vowel (Reinholt Petersen 1983). It should be added that also the tendency to a higher VOC activity in the vowel following aspirated stops may be relevant from the point of view of segmentally conditioned F_0 variation.

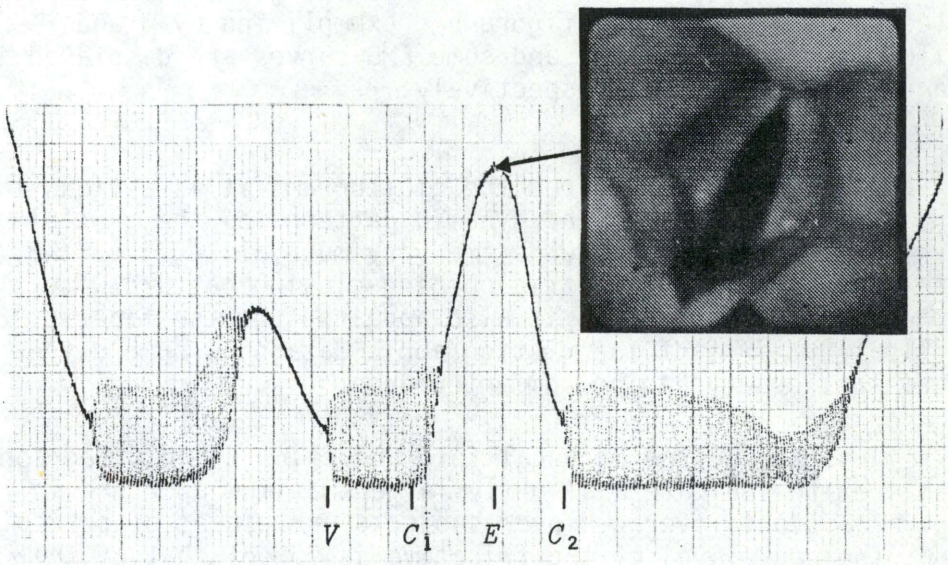
The EMG data presented above are in all essentials in good agreement with those found for Danish stops in Fischer-Jørgensen and Hirose (1974) and for stops in other languages. That is, the glottal gesture in aspirated and unaspirated stops seems to be primarily controlled by the reciprocal activity pattern of the PCA and INT muscles. A survey is found in, e.g. Hirose et al. (1978), Hirose and Sawashima (1981), and Sawashima and Hirose (1983).

In this context, I would like to take up one point in Fischer-Jørgensen and Hirose (1974). For one of two subjects they find a longer aspiration and consequently a later occurrence of the maximum INT level at the transition from stop to vowel. The authors wonder why this same subject also has an earlier start of his INT relaxation if it is for the purpose of aspiration. Indeed, this earlier relaxation does not reflect the aspiration, since the first part of the decreasing INT activity has to do with the preceding unvoiced obstruent in the frame sentence. This is obvious from the fact that in the test sentences with the sonorants l and m in the present material a similar INT reduction is seen during the preceding frame sentence. The difference in timing of the INT relaxation between Fischer-Jørgensen and Hirose's two subjects must therefore result from a difference in the duration of the vowel preceding the test obstruent. But since the first part of the INT relaxation in a sequence where the unvoiced obstruent under study is preceded by another unvoiced obstruent and a vowel, is in fact related to the preceding obstruent, we may conclude that the INT dip directly related to each unvoiced obstruent is much smaller than it appears at first. It should be added that, similar to the INT pattern, the first part of the VOC relaxation is a consequence of the last unvoiced obstruent in the preceding frame sentence. Thus, also the VOC dip relating to the production of obstruents may be smaller than it seems to be.

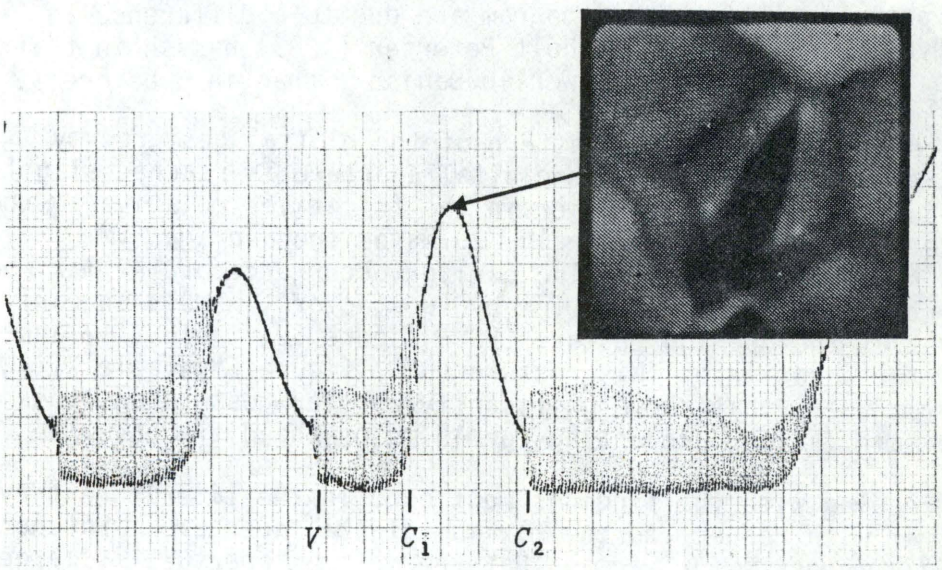
c. conclusion In conclusion, it is obvious that the production of aspiration versus non-aspiration in Danish stops is first of all a matter of the glottal gesture type that is produced, to satisfy the demand of an open versus a nearly closed glottis at the moment when the oral closure is released. Nor does there seem to be any doubt that the articulatory differences between the glottal gestures are reflected in the muscular activity pattern, primarily in the reciprocal pattern of the PCA and INT muscles. However, not even in the case of these two muscles does the relationship between muscle activity and articulatory behaviour appear to be quite simple, and the matter is still more complicated where the VOC and CT muscles are concerned, as will be discussed in section IVB and IVC dealing with the glottal gesture in Danish unaspirated stops and with the interpretation of laryngeal EMG signals, respectively.

2. ASPIRATED STOPS VERSUS FRICATIVES

The aspirated stops as well as the fricatives demand a high rate of air flow in order to generate the required aspiration and friction noise, which means that both categories must be



de ville sige pile[b̥ʰi:lə]



de ville sige file[fi:lə]

Figure 5

Fiberoptic stills and glottograms of aspirated stops and fricatives exemplified by p and f, respectively. The arrows indicate where in the course of the glottal gesture the still originates. The acoustic events are shown below the glottogram (cf. figure 1).

produced with a clear glottal opening-closing gesture. The point of interest is whether the gesture and/or its temporal relationship to the supraglottal articulation differ(s) between the two categories. Fiberoptic stills and glottograms of aspirated stops are shown in figure 5 - exemplified by *p* and *f* - while the quantified data and some EMG curves are displayed in figure 6 and figure 7, respectively.

a. glottal gesture — From the glottograms it is obvious that the dynamic patterns of the glottal gestures are very much alike, i.e. they both show an almost symmetric opening-closing gesture with a large maximum aperture as it also appears from the fiberoptic stills. But in fact the glottographic data show some differences that deserve to be commented upon.

As regards the degree of vocal fold abduction the glottographic data seem to indicate a slightly larger maximum aperture in the aspirated stops: averaged over subjects and places of articulation, the peak level of the fricative is 90% of that of the stop and the difference is significant (A - figure 6a, table XIII).[10] It may well be, however, that the small differences in peak level of the glottogram are due to a difference in larynx height, since Reinholt Petersen (1983) has shown that the larynx tends to be lower in Danish *f* than in *p* before *i*.

Concerning the temporal pattern of the glottal gesture, my data show that the maximum glottal aperture is reached slightly but significantly earlier in the fricatives relative to the onset of the preceding vowel - 20 ms on the average (VM - figure 6b, table VIII). The earlier occurrence of the maximum aperture may be accounted for by two factors, namely an earlier onset of the opening gesture (VG - figure 6c, table VII) and a shorter duration of this gesture (GM - figure 6d, table XII), but it seems that the significantly earlier gesture onset is the main factor, accounting for 15 ms of the difference.

Proceeding with the findings mentioned in the literature they differ with respect to the maximum glottal aperture. Butcher (1977) suggests for German that the maximum aperture is larger in the aspirated stops, whereas for Swedish (Lindqvist 1972; Löfqvist and Yoshioka 1979) and Icelandic (Löfqvist and Yoshioka 1980) it is stated that the aperture is larger in the fricatives. In American English, however, Sawashima (1970) does not observe any difference between the two types of obstruent. The same observation has been made in a preliminary study of British English undertaken at our institute.[11] If the discrepancy between these observations cannot be related to different experimental conditions, it may be due to language specific differences in the production of the sounds in question. It is, however, noteworthy that from the tables in Frøkjær-Jensen et al. (1971) dealing with Danish like the present study, it can be deduced that only one of the three subjects shows a higher peak level of the glottogram in *p* than in *f*. Also in the present material inter-subjective differences are seen:

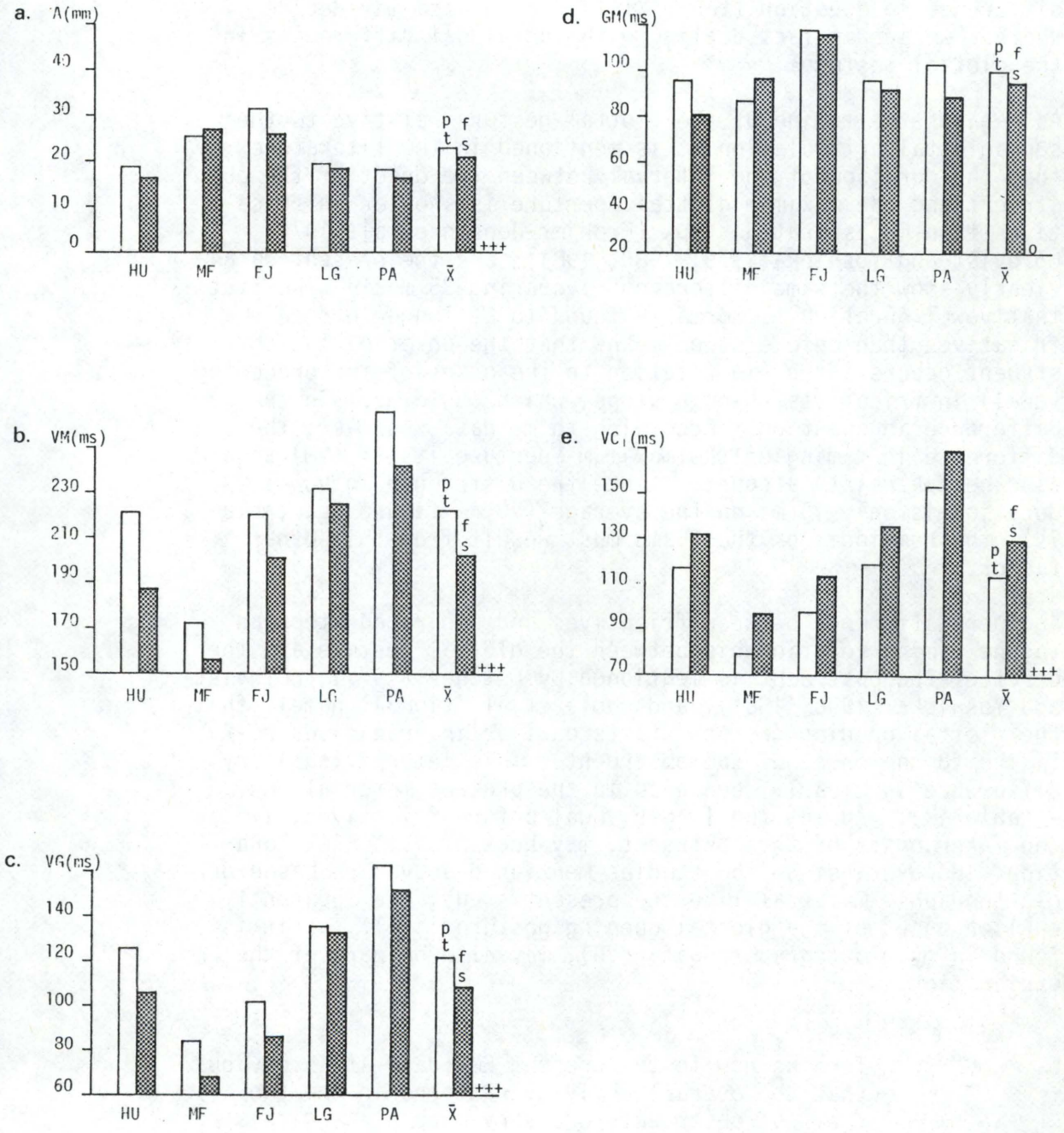


Figure 6

Mean values for aspirated stops and fricatives - averaged over place of articulation - of the parameters A (peak level of the glottogram), VM (duration from the start of the preceding vowel to the maximum glottal aperture), VG (duration from the start of the preceding vowel to the onset of the glottal gesture), GM (duration of the glottal abduction), and VC_1 (duration of the preceding vowel). (k and h are omitted, see note 9.) The left column represents the aspirated stop category, the right one the fricative category, as indicated above the rightmost columns. For further explanation, see the legend to figure 3.

for *p* and *f* only three out of five speakers show the peak level difference in question (table XIII). More data may decide whether we are in fact dealing with individual differences in the glottal gesture.

As regards the timing of the glottal gesture relative to the supraglottal articulation it is mentioned in the literature that the duration of the interval between the onset of the obstruent and the maximum glottal aperture is shorter in fricatives than in aspirated stops (Frøkjær-Jensen et al. 1971; Löfqvist and Yoshioka 1979, 1980, 1981), and the present data clearly show the same difference, averaging 35 ms.[12] The fact that vowel duration is normally found to be longer before fricatives than before stops means that the onset of the obstruent occurs later (in relation to the onset of the preceding vowel) in fricatives than in stops, which could explain the difference in question. According to my data, however, the difference in timing of the maximum aperture itself (VM) should also be taken into account: since the difference in vowel duration is only 15 ms on the average (VC_1 - figure 6e, table IV), the remainder of the 35 ms must result from this other factor.[13]

Another difference between fricatives and aspirated stops as to the timing relationship between the glottal gesture and the onset of the obstruent is mentioned by Butcher (1977), Löfqvist and Yoshioka (1980, 1981), and Hoole et al. (1983), namely that the glottal opening gesture starts earlier in fricatives relative to the onset of the obstruent. This interarticulatory difference is clearly seen also in the present material (GC_1 - table IX). Again, the longer vowel before fricatives, i.e. the later onset of the obstruent, may be the obvious explanation. However, since the studies mentioned above are based on glottographic material like the present study, the apparently earlier onset of the glottal opening gesture itself, as it is found in my glottographic data (VG), may also be part of the explanation.

b. EMG Turning now to the present EMG data it is obvious that the overall activity patterns in the two types of obstruents are very much alike with a clear rising and decline in the PCA activity and a reciprocal pattern of the INT muscles. Also the dip in the VOC activity and the small rising in CT resemble each other closely. There are indeed also some obvious differences, but only few of them show any noticeable degree of inter-speaker consistency.

Regarding the PCA muscle, the summit tends to be broader in the aspirated stops, whereas the peak activity may be higher or lower than in the fricatives. The INT dip also tends to be broader in the aspirated stops, whereas the minimum level may be higher or lower. It is not clear how these differences should be interpreted in terms of degree of glottal aperture, but it seems reasonable that not only the maximum and minimum levels but also the duration of increasing PCA activity and de-

creasing INT activity should be taken into account. If so, the data seem to indicate that the maximum aperture may be higher in the aspirated stops than in the fricatives. Moreover, the PCA maximum is attained slightly earlier in the fricatives in relation to the onset of the preceding vowel, due to an earlier onset of the rise and/or a shorter rise time. It seems very reasonable to suggest that this difference in timing of the PCA maximum is responsible for the earlier maximum glottal aperture in the fricatives as it was observed in the glottograms.

However, it is the VOC muscle that shows the most consistent temporal difference: the reduction of activity in the fricatives leads that of the aspirated stops - relative to the preceding vowel onset - due to a higher reduction rate and/or an earlier start of the reduction. Supposing that the reduction in the VOC activity during vocal fold vibrations results in a looser and shorter contact of the vibrating vocal folds, this will imply that the glottographic signal rises sooner in the case of fricatives. If this interpretation is true, then the different timing of the rising glottographic signal is a consequence of the vibratory pattern of the vocal folds and does not reflect a difference in the onset of the glottal opening gesture - provided that the glottal gesture is defined in terms of the movements of arytenoid cartilages. This interpretation should be considered in the light of the fact that the PCA and INT patterns do not seem to indicate an earlier onset of the abduction of the arytenoids - except with PM.

Only very few observations on laryngeal muscle activity comparing aspirated stops and fricatives are available in the literature. Löfqvist and Yoshioka (1979) find for their Swedish subject that the PCA peak is higher and the decrease in the INT activity is deeper and more rapid in *s* than in *p* resulting in a larger maximum glottal aperture and a higher abduction velocity in the fricative. It is worth of notice that these differences in the PCA and INT patterns may also be observed in the Danish subjects, whereas the glottographic data show the reverse relationship between aspirated stops and fricatives. This discrepancy may be due to methodological differences including comparisons between homorganic versus non-homorganic obstruents (see below). On the other hand, since the Swedish stops in general are produced with a shorter aspiration than the Danish ones, a slightly different glottal behaviour would not be surprising - a difference that may lead to deviating results when we compare aspirated stops and fricatives.

Regarding the VOC muscle Hirose and Gay (1972) find no difference in the degree of relaxation between English unvoiced fricatives and stops, as I did not in the Danish material either. Collier et al. (1979), however, consider Hirose and Gay's observation in contradiction with their own findings for Dutch. But taking into account that the stops in Dutch are unaspirated whereas the English ones are aspirated - in the position in question - and thus produced with a glottal gesture almost similar to the gesture in the fricatives, there may not be any discrepancy between the two findings.

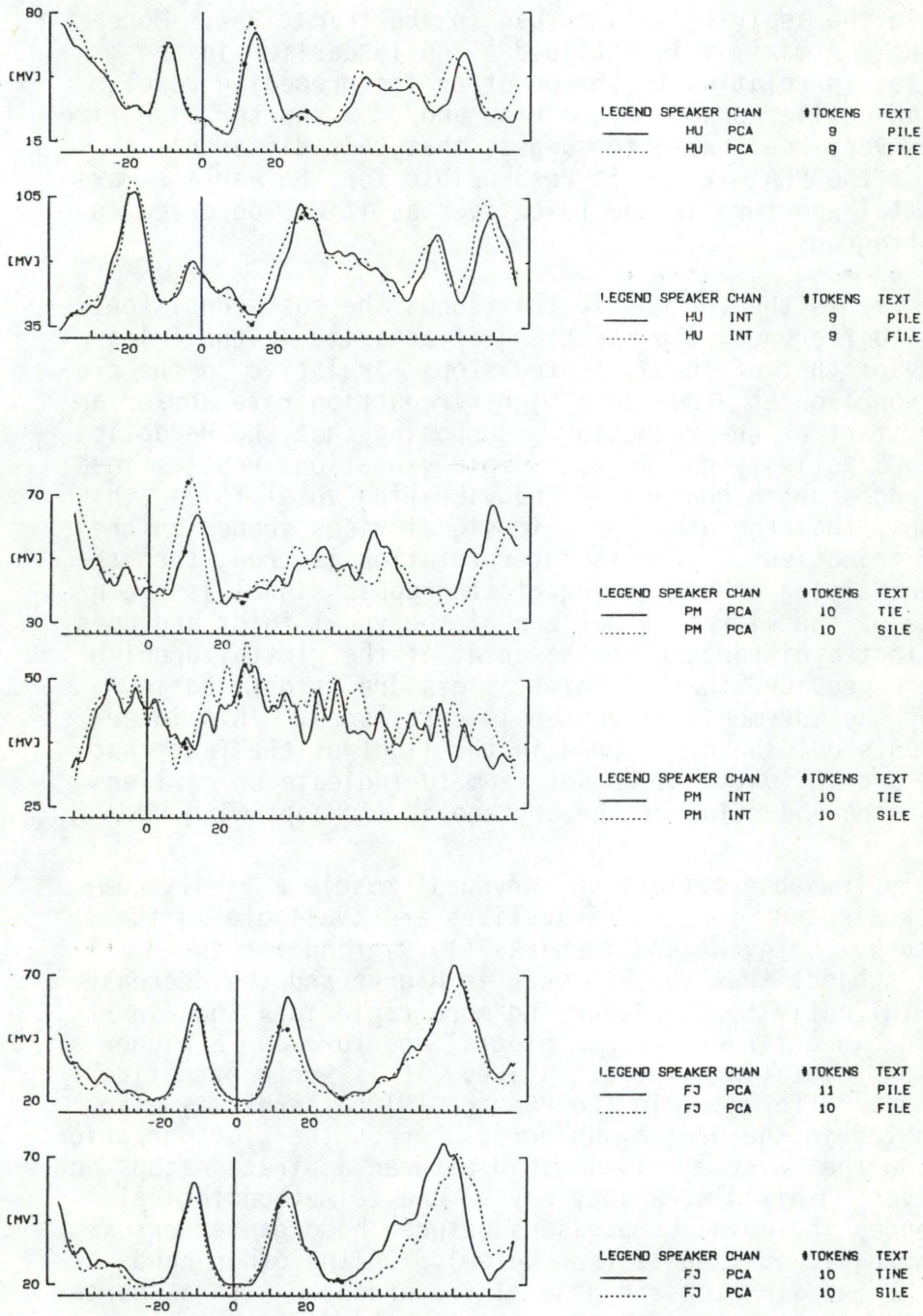
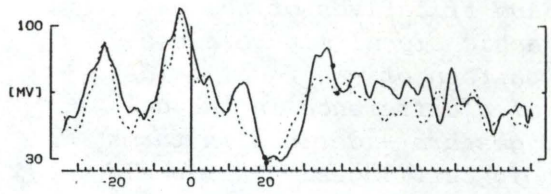
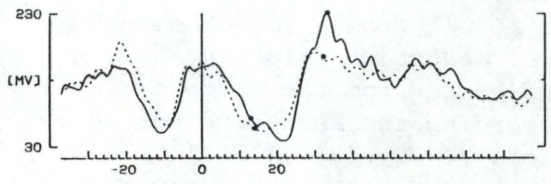


Figure 7

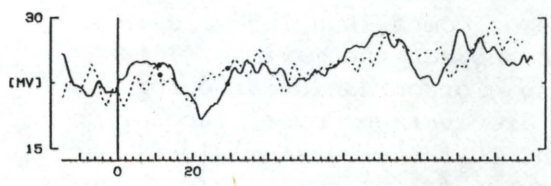
Averaged EMG curves of aspirated stops and fricatives. The small dots seen in each EMG curve indicate the acoustic events as follows: first dot = onset of the obstruent ("C_1" in figure 1), second dot = offset of the obstruent alias the onset of the following vowel ("C_2" in figure 1). For further explanation, see the legend to figure 4.



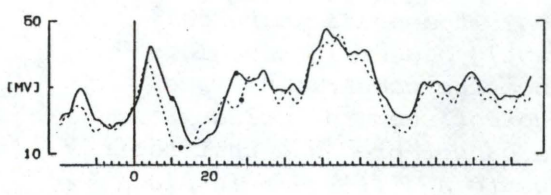
LEGEND	SPEAKER	CHAN	#TOKENS	TEXT
———	HU	VOC	8	TIE
.....	HU	VOC	8	SILE



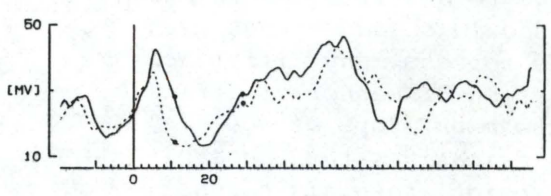
LEGEND	SPEAKER	CHAN	#TOKENS	TEXT
———	BH	VOC	8	PILE
.....	BH	VOC	8	FILE



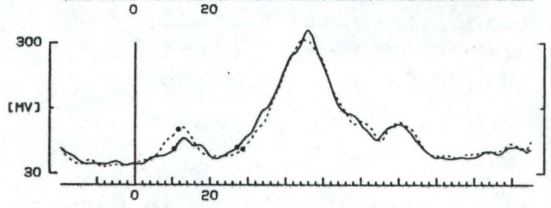
LEGEND	SPEAKER	CHAN	#TOKENS	TEXT
———	JJ	VOC	10	TIE
.....	JJ	VOC	10	SILE



LEGEND	SPEAKER	CHAN	#TOKENS	TEXT
———	BM	VOC	10	PILE
.....	BM	VOC	10	FILE



LEGEND	SPEAKER	CHAN	#TOKENS	TEXT
———	BM	VOC	10	TIE
.....	BM	VOC	10	SILE



LEGEND	SPEAKER	CHAN	REF	#TOKENS	TEXT
———	BM	CT		10	PILE
.....	BM	CT		10	FILE

Figure 7 (continued)

I have ventured the hypothesis that the different timing in aspirated stops and fricatives of the onset of the rising glottographic signal may be a consequence of the vibratory pattern of the vocal folds rather than a consequence of a difference in the onset of the glottal opening gesture - defined in terms of the movements of arytenoid cartilages. This would implicate that the true glottal opening gesture is longer in the aspirated stops since no clear difference was found in parameter GM (figure 6d - table XII). The longer abduction could result from a larger maximum aperture and/or a lower abduction velocity. As it appears above, an influence from the former factor seems possible. Regarding the abduction velocity, Löfqvist and Yoshioka (1979, 1980, 1981)[14] dealing with several languages find that the maximum abduction velocity is higher in fricatives than in aspirated stops as it appears from their glottographic material. For the sake of comparison I have calculated (manually) for one subject the maximum slope of the (LP-filtered) rising glottographic curve for pt and fs. When the data are averaged over place of articulation, it is evident that the maximum slope is certainly not higher in the fricatives. This discrepancy may be due to language specific differences. On the other hand, if the data are averaged over manner of articulation, it appears that the maximum slope is significantly higher in alveolars than in labials. Now, since the comparisons made by Löfqvist and Yoshioka do not seem to include homorganic obstruents, one cannot preclude that their observations are partly due to a difference in place of articulation. But it should be added that even though the present data do not show any difference in maximum slope as a function of manner of articulation, it does seem that the mean slope of the rising curve is higher in fricatives than in aspirated stops. The crucial point, then, is to what extent the slope of the glottographic curve actually reflects the abduction - and adduction - velocity of the vocal folds.

c. conclusion

In short, the glottal behaviour in Danish aspirated stops and fricatives is very much alike. On the other hand, there seems to be differences in the interarticulatory timing not only resulting from a different supraglottal behaviour but also from the timing of the glottal gesture itself. Thus, it is probably safe to state that the maximum aperture is reached slightly earlier in fricatives than in stops. It is hypothesized that the main factor accounting for this earlier maximum is a shorter abduction duration rather than an earlier gesture onset - on condition that the glottal gesture is defined in terms of arytenoid movements. The earlier onset of the rising glottographic signal in fricatives is tentatively explained as resulting from

a change in the vibratory pattern of the vocal folds due to the fact that the state of complete relaxation of the VOC muscle is reached sooner. A propos of the EMG data, the interpretation of the amplitude differences is somewhat doubtful, whereas the temporal course of the activity patterns seems in general to reflect the temporal differences in the glottal gesture.

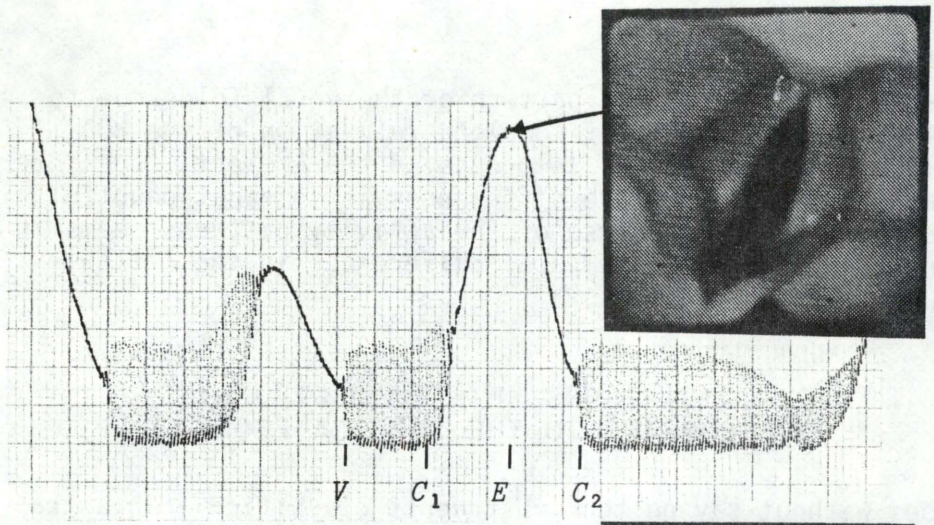
B. COMPARISONS BETWEEN OBSTRUENTS BELONGING TO THE SAME CATEGORY

It goes without saying that obstruents classified - in a traditional phonetic sense - as belonging to the same category, differing only as to place of articulation, are produced in the same way by definition. But it is a well-known fact that different places of articulation involve differences in how exactly the supraglottal articulation is performed, depending on the articulators involved and on the speech habits of a given speech community. The following sub-sections deal with how such differences in the supraglottal articulation may result in a slightly different timing between the glottal and supraglottal articulation. Moreover, it will be shown that minor differences in the glottal behaviour itself seem to be involved.

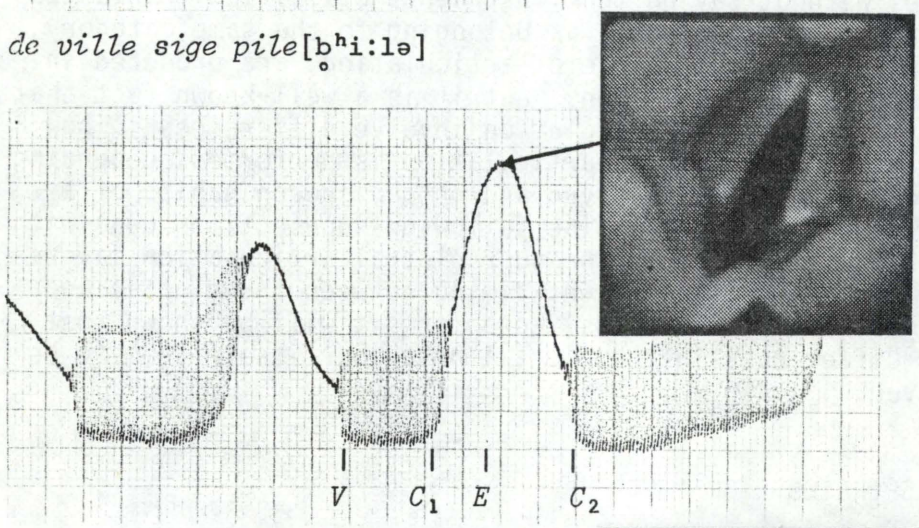
1. ASPIRATED STOPS

Fiberoptic stills and glottograms of the aspirated stops *ptk* are seen in figure 8, while the quantitative data and some EMG curves are displayed in figure 10 and figure 11, respectively.

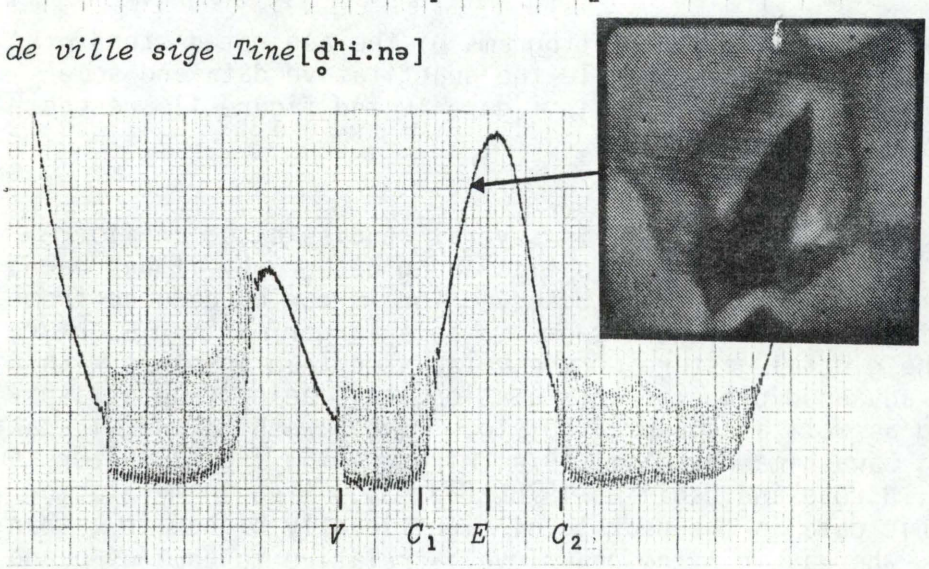
a. glottal gesture — In several studies based on fiberoptics and glottography it has been observed (or it can be deduced from the figures) that there seems to be a larger maximum aperture in the glottal gesture of the velar stop than in stops with a more advanced place of articulation, this being true of aspirated as well as unaspirated stops (see Sawashima and Miyazaki 1973; Sawashima and Niimi 1974; Hirose 1975, fig. 7; Pétursson 1976; Hirose and Ushijima 1978, fig. 3). Also in my glottographic data is the peak level significantly higher in *k* than in *p*, whereas in *t* the peak level - relative to that of *p* and *k* - varies considerably from subject to subject showing no significant difference from either *p* or *k* (A - figure 10a, table XIII). Contrarily, Löfqvist (1976) in his glottographic material fails to observe a higher peak in velar than in labial aspirated stops. Later, however, Löfqvist and Yoshioka (1979) report spuriously high levels in velar stops and explain that these are probably due to an artifactual influence from vertical movements of the larynx. They recommend that the transducer should be placed below the cricoid cartilage



de ville sige pile[b̥ʰi:lə]



de ville sige Tine[d̥ʰi:nə]



de ville sige kilde[g̊ʰilə]

Figure 8

Fiberoptic stills and glottograms of the aspirated stops ptk. The arrows indicate where in the course of the glottal gesture the still originates. The acoustic events are shown below the glottogram (see also figure 1).

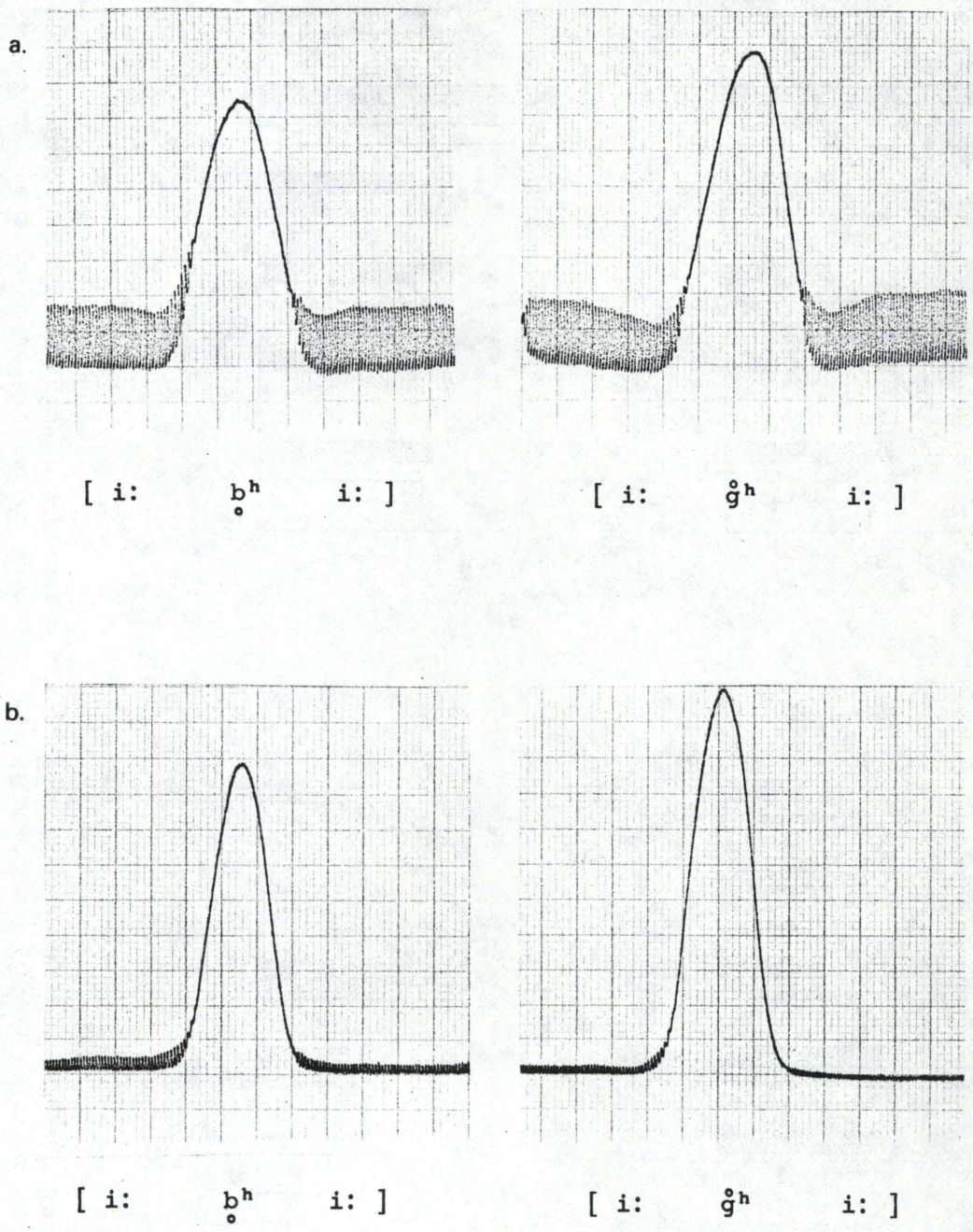


Figure 9

Glottograms of p and k recorded with the photo-transducer positioned just below the thyroid cartilage (a) and well below the cricoid cartilage (b).

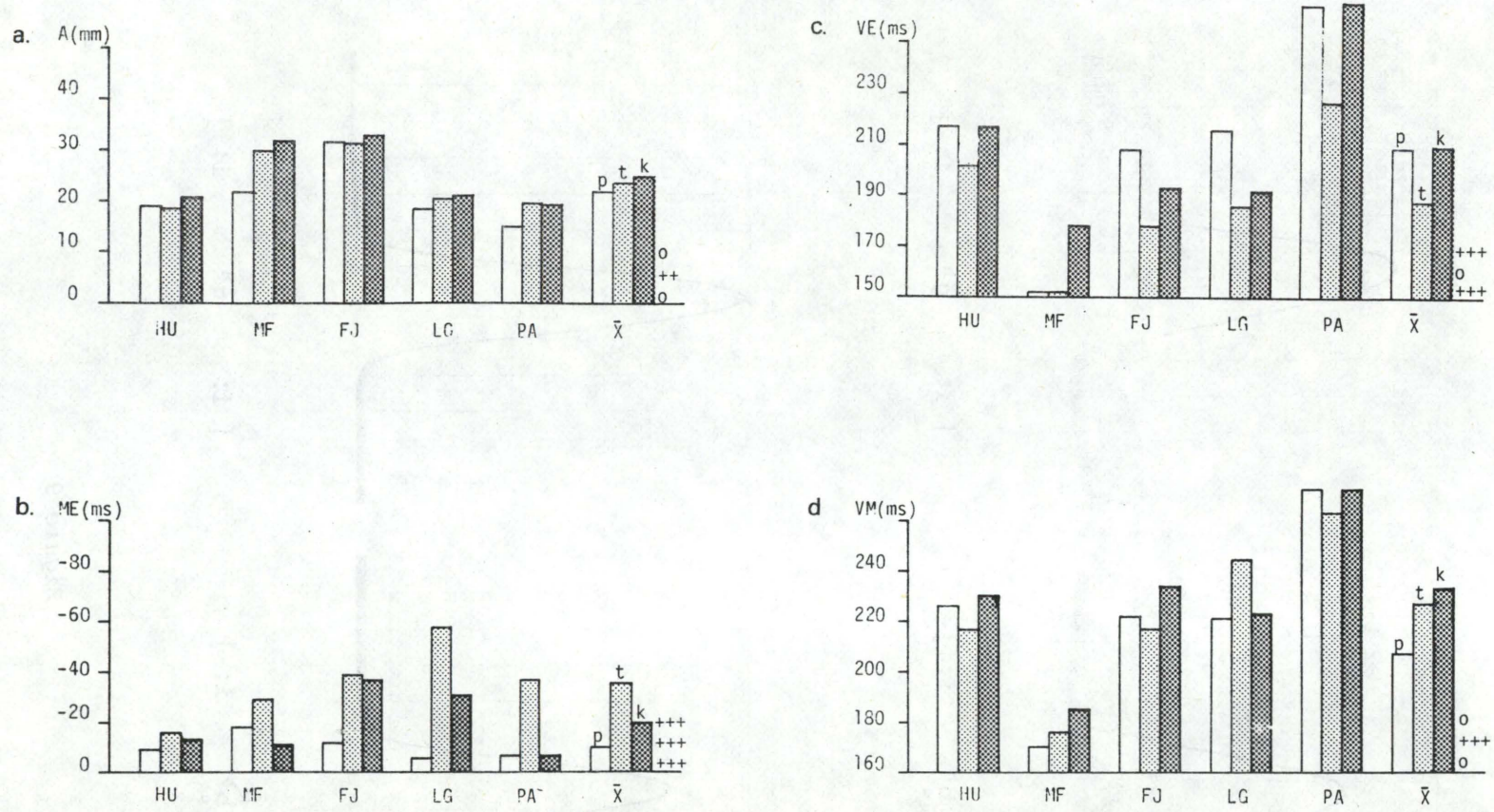


Figure 10

Mean values for the aspirated stops ptk of the parameters A (the peak level of the glottogram), ME (duration from the maximum glottal aperture to the explosion of the closure), VE (duration of the period including the oral closure and the preceding vowel), VM (duration from the start of the preceding vowel to the maximum glottal aperture), C_1E (duration of the oral closure), VC_1 (duration of the preceding vowel), and EC_2 (duration of the open interval). The left column represents p, the column in the middle t, and the right one k,

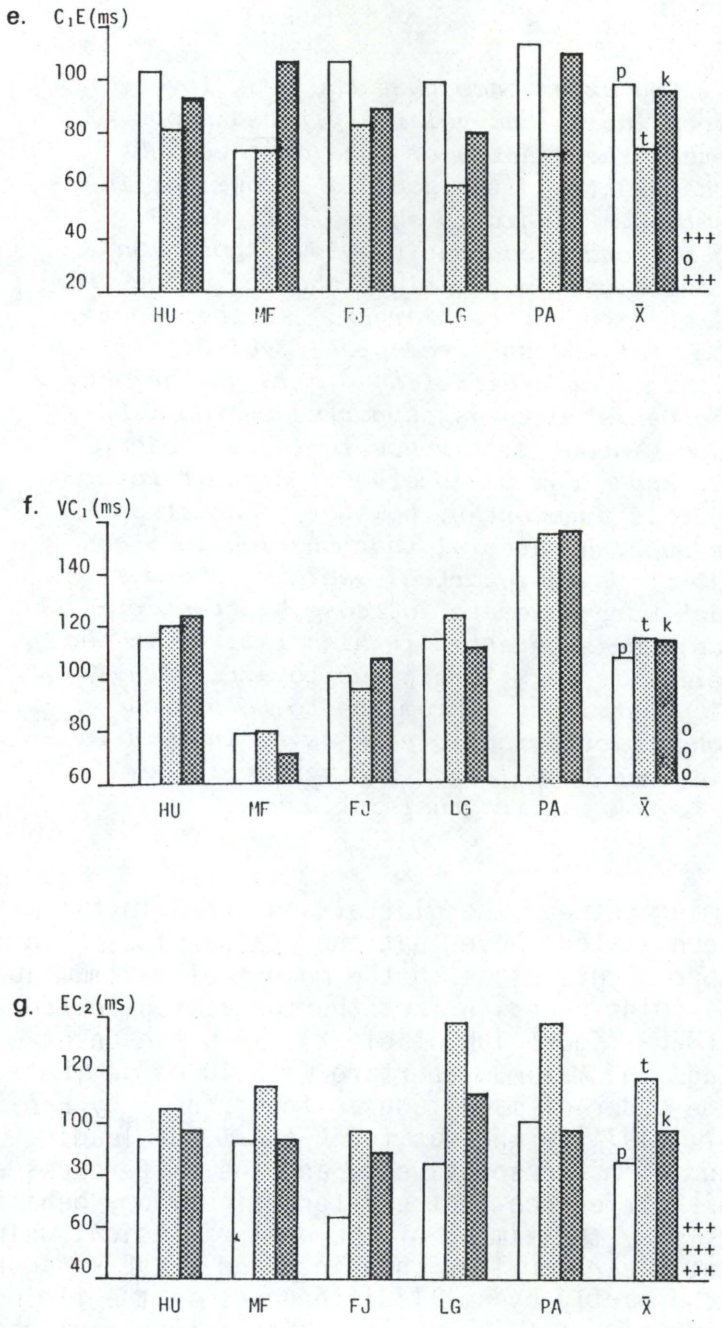


Figure 10 (cont.)

as indicated above the rightmost columns in each graph. The symbols for the level of significance should be interpreted as follows: the upper row shows the level of significance for p versus t, the row in the middle p versus k, and the lowest one t versus k. For further explanation, see the legend to figure 3.

rather than between the thyroid and cricoid cartilages - in order to avoid this undesirable effect.

> *However, it is my experience that the peak level difference between labial and velar aspirated stops shows up whether the transducer (and/or the light source for that matter) is placed in a lower or in a higher position.*[15] *This is shown in figure 9. Thus, it may be concluded that the lower position of the transducer need not reflect the influence from the vertical movement of the larynx. Another implication may be that the observed peak level difference does not reflect a difference in larynx height, which for the Danish data is supported by the finding that no consistent difference in larynx height between p, t, and k can be observed (Reinholt Petersen 1983). It is noteworthy, however, that it is only for the Swedish material that spuriously high peak in velar stops is reported, whereas in other studies a much more moderate increase has been observed. Thus, the substantially higher level in the Swedish material is very likely due to artifactual disturbances. But it still remains to be proved that the higher glottographic peak level in k is in general really due to a larger glottal aperture rather than to some artifactual influence.*

Regarding the timing between the glottal and supraglottal articulation it has been stated above that the oral explosion in the aspirated stops occurs close to the moment of maximum aperture of the vocal folds. But in fact the three stops differ in this respect (ME - figure 10b, table X): in *p* the onset of the explosion leads the maximum aperture with 10 ms on the average - the same order of magnitude as found for *p* by Frøkjær-Jensen et al. (1971) - whereas for *k* and *t* the leading averages 20 ms and 36 ms, respectively, and the differences are significant. This difference in the interarticulatory behaviour is explained partly by the timing of the oral explosion, which occurs significantly earlier in *t* than in *p* and *k* (VE - figure 10c, table VI), and partly by small differences in the timing of the maximum glottal aperture itself, which occurs earliest in *p* and latest in *k* , even though only *p* vs. *k* is significant (VM - figure 10d, table VIII). It should be added that the differences in timing of the glottal maximum aperture itself seem to be accounted for by a combined influence from the gesture onset (VG - table VII) and from the duration of the gesture opening (GM - table XII), even though substantial individual differences are seen regarding the degree of influence from each of the two factors.[16] But the fact remains that not only the timing of the oral explosion but also the timing of the maximum glottal aperture as such has to be taken into consideration in accounting for the time lag between these two articulatory events and their variation according to place of articulation.

Unfortunately, only a few studies include data on the glottal gesture and its temporal relation to the oral explosion according to place of articulation. One is Pétursson (1976) who states that in Icelandic (post-)aspirated stops the maximum aperture always leads the explosion, but he does not differentiate according to place of articulation. From his figures, however, it appears that the leading of the maximum aperture is shorter in *c* - due to a shorter oral closure - and in *k* - due to a later maximum - than in *p* and *t*.[17] Also in German it has been observed that the leading of the maximum aperture is longer in *p* than in *t*, which is suggested to be an effect mainly of the longer closure in *p* (Hoole et al. 1983). In Kagaya (1974) it is mentioned that in the Korean aspirated stops the explosion occurs around the moment of maximum glottal aperture, but it appears from his figures that the explosion in *k* may occur slightly earlier than in *p* and *t*. Furthermore, it can be deduced from his figures that in the aspirated affricate the explosion leads the maximum aperture considerably, which has also been observed in the Mandarin affricates (Iwata and Hirose 1976). These last findings seem very consistent with the clearly leading explosion relative to the glottal maximum aperture seen in the Danish affricated *t*.

Thus, it can be maintained that the temporal relationship between the oral explosion and the maximum aperture of the glottal gesture is definitely related to the place of articulation - whether it is the explosion that leads the maximum aperture or vice versa - although it seems that not only the timing of the oral explosion but also the timing of the maximum glottal aperture itself play a role.

> *The relations normally found for the duration of the oral closure in Danish aspirated stops is p>k>t (see, e.g., Fischer-Jørgensen 1954, 1980), which also appears in the present material, even though only t differs significantly from the two other stops due to the very deviating relations seen with one of the subjects (MF) (C_1E - figure 10f, table IV). The earlier explosion in Danish t resulting in a shorter closure is evidently connected with the fact that t is strongly affricated, which is in agreement with the general observation that affricated stops have a shorter closure than aspirated ones (see, e.g., Fischer-Jørgensen 1976). On the other hand, the tendency to an earlier implosion in p resulting in a slightly longer closure may be explained by the fact that the labial articulation is independent of the articulation of the preceding vowel (Fischer-Jørgensen 1980). This last explanation, however, is in fact somewhat dubious, at least if applied to the present data, taking into consideration the reservations that have to be made regarding the acoustic delimitation and its interpretation in terms of articulation.*

Proceding with the open interval, or aspiration, it is well documented that in Danish this interval is shortest in p, longer in k, and longest in t, i.e. the opposite relationships from those found for the closure (see again Fischer-Jørgensen 1954, 1980). Not surprisingly, these relations are also seen in the present data and the differences are highly significant (EC_2 - figure 10g, table III). It should be noted, however, that not only the moment of explosion relative to the glottal gesture but also the start of the following vowel may influence the duration of the open interval. Based on the observation that no clear systematic differences between the three stops as to the gesture onset (VG - table VII) seem to be present, it can be concluded that the longer open interval in t compared to k is primarily a function of the significantly earlier explosion, since no significant difference can be shown in the onset of the following vowel (VC_2 - table V). On the contrary, the shorter open interval in p compared to k is first of all the result of a slightly but significantly earlier onset of the following vowel (VC_2), since no significant difference could be proved for the timing of the explosion (VE). From these facts it can be deduced that the difference between p and t in the open interval is caused by both factors, as also appears from the relevant figures and tables.

The delay in the vowel onset after k and t compared to p seems to be most readily accounted for in terms of a difference in the voicing conditions: the slower release in k and especially in t leads to a slower decay in the intra-oral pressure which delays the attainment of the appropriate pressure drop across the glottis. But it may also be that the offset of the glottal gesture comes slightly earlier in p, taking into consideration the earlier timing of the maximum aperture, and if so, this factor should also be included in the voicing conditions.

b. EMG I shall turn now to the EMG curves. It is not surprising that the overall patterns are very much alike, irrespective of place of articulation. Particular attention should be given to the observation that only the muscular pattern for one subject (HU) supports the assumption that the glottal maximum aperture is slightly larger in *k* than in *p*, provided that a higher and broader PCA peak reflects such a difference in the gesture. Nor does the slightly later occurrence of the maximum aperture observed in *k* compared to *p* seem to be reflected in the EMG signals except in that same subject.

As regards the VOC muscle it seems that the differences in duration of the activity dip correspond to the differences in

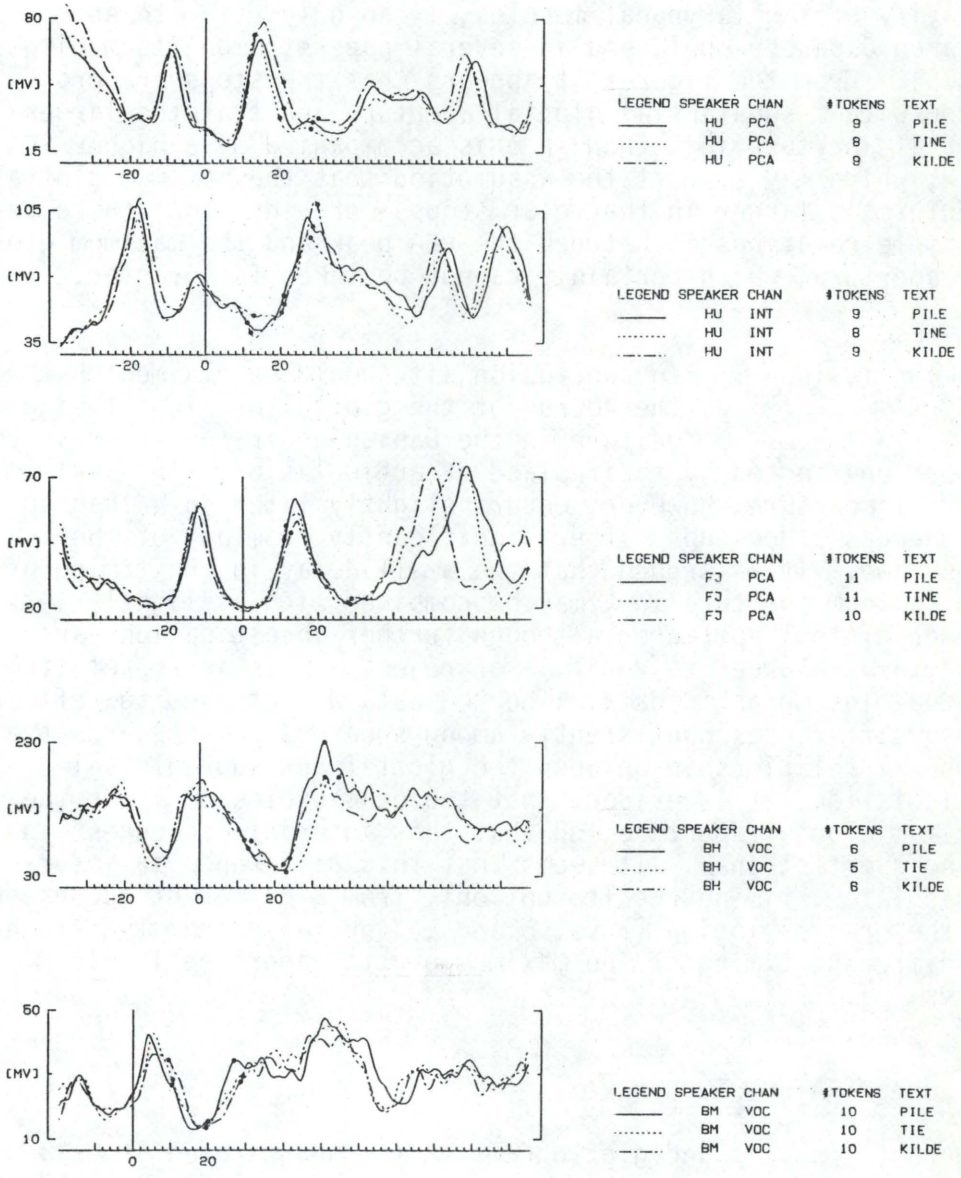


Figure 11

Averaged EMG curves of the aspirated stops ptk. For further explanation, see the legend to figure 4.

segment duration. Furthermore, the maximum level at the transition to the following vowel may be higher in *p* than in the two other stops.

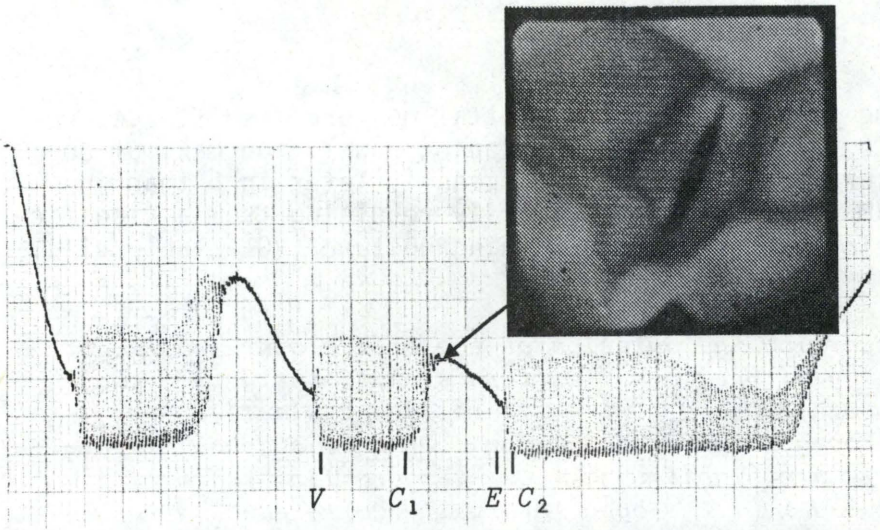
Finally, as far as the problem of the supposedly larger glottal aperture in velar stops is concerned and its reflection in the activity of the laryngeal muscles, I can only refer to some data on Japanese published in several papers, among them Hirose (1975). From the figures it appears that the stops are produced with a substantial glottal aperture and that the larger glottal aperture in *k* than in *t* is accompanied by a higher PCA peak, which may support the assumption that the maximum glottal aperture is larger in the velar stops - provided that there is a simple relationship between the PCA peak and the maximum glottal aperture, which certainly cannot be taken for granted.

c. conclusion — In conclusion, it should be claimed that the course of the glottal opening-closing gesture in the Danish aspirated stops is almost unaffected by their place of articulation. The maximum glottal aperture, however, occurs slightly later in *k* than in *p*, whereas *t* does not differ significantly from any of the other two. It is argued that the small delay in the timing of the maximum aperture in *k* may be combined with a slightly larger glottal aperture, although further investigations are necessary in order to verify - or reject - this interpretation of the glottographic data. The EMG data do not seem to reflect these differences consistently among speakers. As regards the temporal relationship between the glottal and supraglottal articulation, it is evident that the oral explosion leads the maximum glottal aperture and that this interval is longest in *t* and shortest in *p*. It seems that this difference in interarticulatory timing results not only from a different occurrence of the oral explosion (*t* vs. *p* and *k*), but also from the slightly different timing of the maximum glottal aperture itself (*k* vs. *p*).

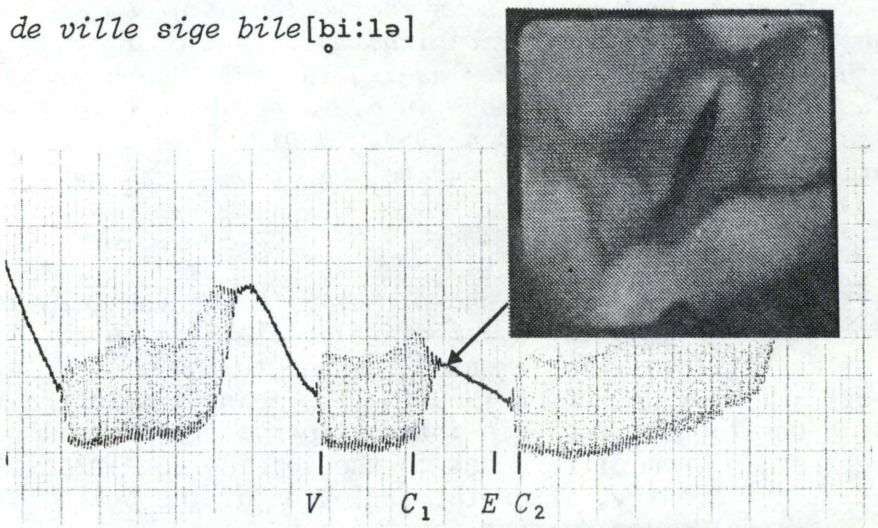
2. UNASPIRATED STOPS

Fiberoptic stills and glottograms of the unaspirated stops *bdg* are seen in figure 12, while the quantitative mingographic data are shown in figure 13. In figure 14 various EMG curves are displayed.

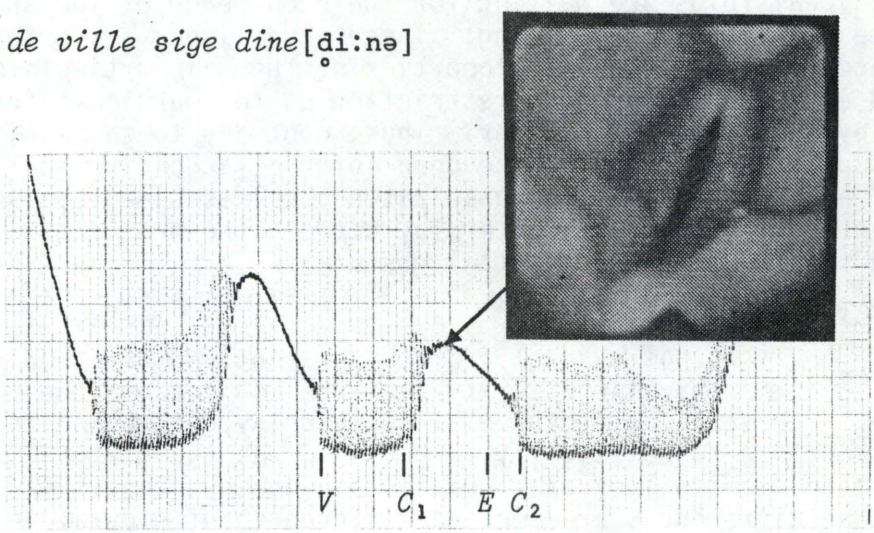
a. glottal gesture — As previously mentioned, the assumption that the maximum glottal aperture is larger in velar stops applies to aspirated as well as to unaspirated stops (Sawashima and Miyazaki 1973; Sawashima and Niimi 1974; Pētursson 1976). The present glottographic data also seem to indicate a slightly larger glottal aperture in Danish *g* than in the two other unaspirated stops, even though only *g* vs. *b* is significant (A - figure 13a, table XIII).



de ville sige bile[bi:lə]



de ville sige dine[di:nə]



de ville sige gilde[gilə]

Figure 12

Fiberoptic stills and glottograms of the unaspirated stops bdg. The arrows indicate where in the course of the glottal gesture the still originates. The acoustic events are shown below the glottogram (cf. figure 1).

Regarding the timing of the glottal gesture itself, i.e. relative to the onset of the preceding vowel, the maximum aperture occurs slightly but significantly later in *b* than in *d* and *g* (VM - table VIII), primarily resulting from a significantly delayed onset of the opening gesture (VG - table VII).

> *Concerning the duration of the oral closure the order is b>g>d, but only b differs significantly from the two other stops (C_1E - figure 13b, table II). For the open interval the order is b<d<g, and the differences are significant (EC_2 - figure 13c, table III). The order of the open interval - corresponding to the universal tendency - is on the whole in agreement with previous findings for Danish, whereas the order of the oral closure deviates as regards the tendency to a longer closure in g than in d, rather than vice versa (cf. Fischer-Jørgensen 1954, 1980). This deviation is again primarily due to the atypical relations seen with speaker MF.*

Many years ago already, it was shown that with the vocal folds in *bdg*-position voicing starts immediately after the introduction of an externally implemented leakage to the closed vocal tract (Fischer-Jørgensen 1963). This means that the differences in duration of the open interval normally found in Danish *bdg* result from differences in the voicing conditions. Or to put it differently, after the explosion of the oral closure the vocal folds are waiting for the occurrence of the appropriate conditions for voicing. These conditions vary primarily according to inherent properties of the main articulator involved and by the degree of restriction as to coarticulation imposed by this same articulator. Thus, contrary to the aspirated stops, the duration of the open interval according to place of articulation in the unaspirated cognates cannot be a consequence of the temporal interplay between the oral explosion and the course of the glottal gesture.

b. EMG

It is not evident from the present EMG data how the muscular activity patterns could cause the gesture onset to be slightly delayed in *b* and the glottal aperture to be slightly larger in *g*, as the glottographic data seem to indicate. Of course, one hypothesis may be that the differences are genuine differences of glottal gesture but caused by non-muscular forces, granted that there may be some such forces acting on the glottal gesture and influencing its size. On the other hand, regarding the slightly higher peak level of the glottogram for *g*, it is tempting to assume that it should be accounted for by a difference in larynx height, since Reinholt Petersen (1983), contrary to the aspirated stops, finds a tendency to a higher larynx in *g* before *i* than in the other two unaspirated stops.

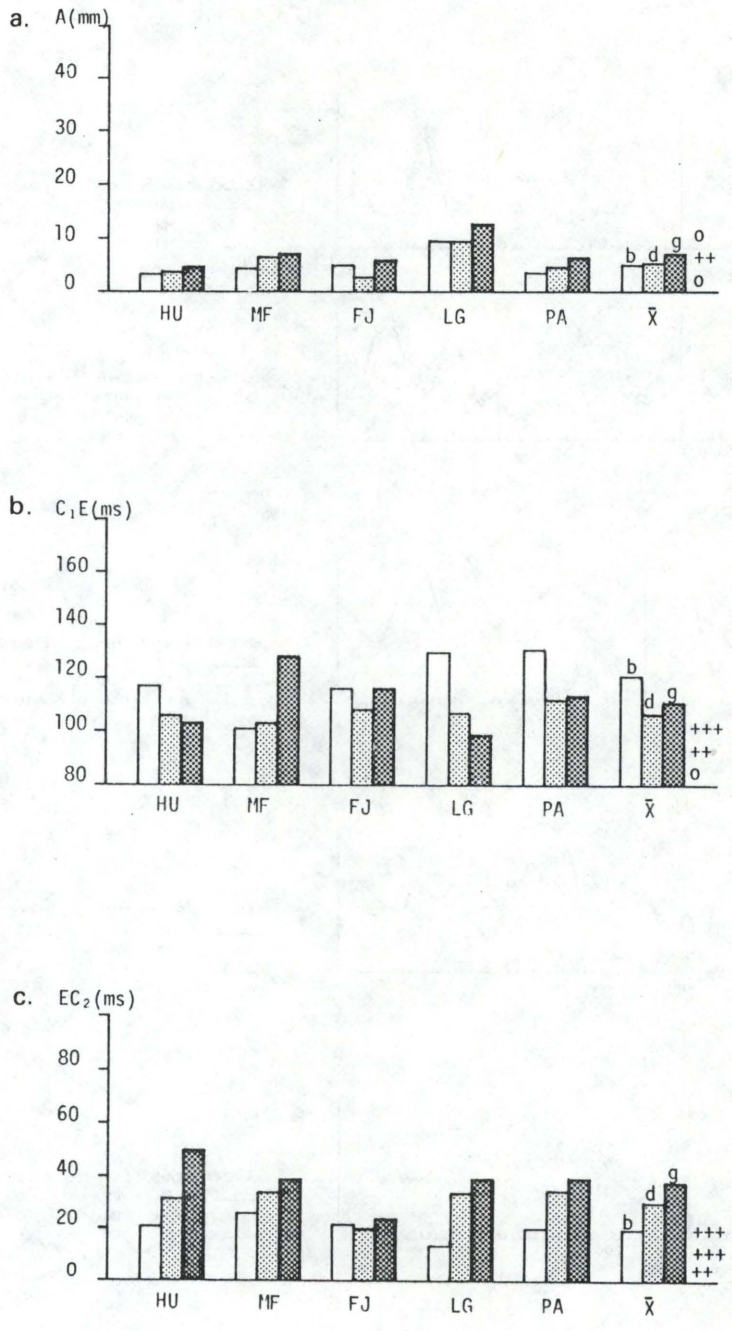


Figure 13

Mean values for the unaspirated stops bdg of the parameters A (peak level of the glottogram), C_1E (duration of the oral closure), and EC_2 (duration of the open interval). The left column represents b, the column in the middle d, and the right one g, as indicated above the rightmost columns in each graph. The symbols for the level of significance should be interpreted as follows: the upper row shows the level of significance for b versus d, the row in the middle b versus g, and the lowest one d versus g. For further explanation, see the legend to figure 3.

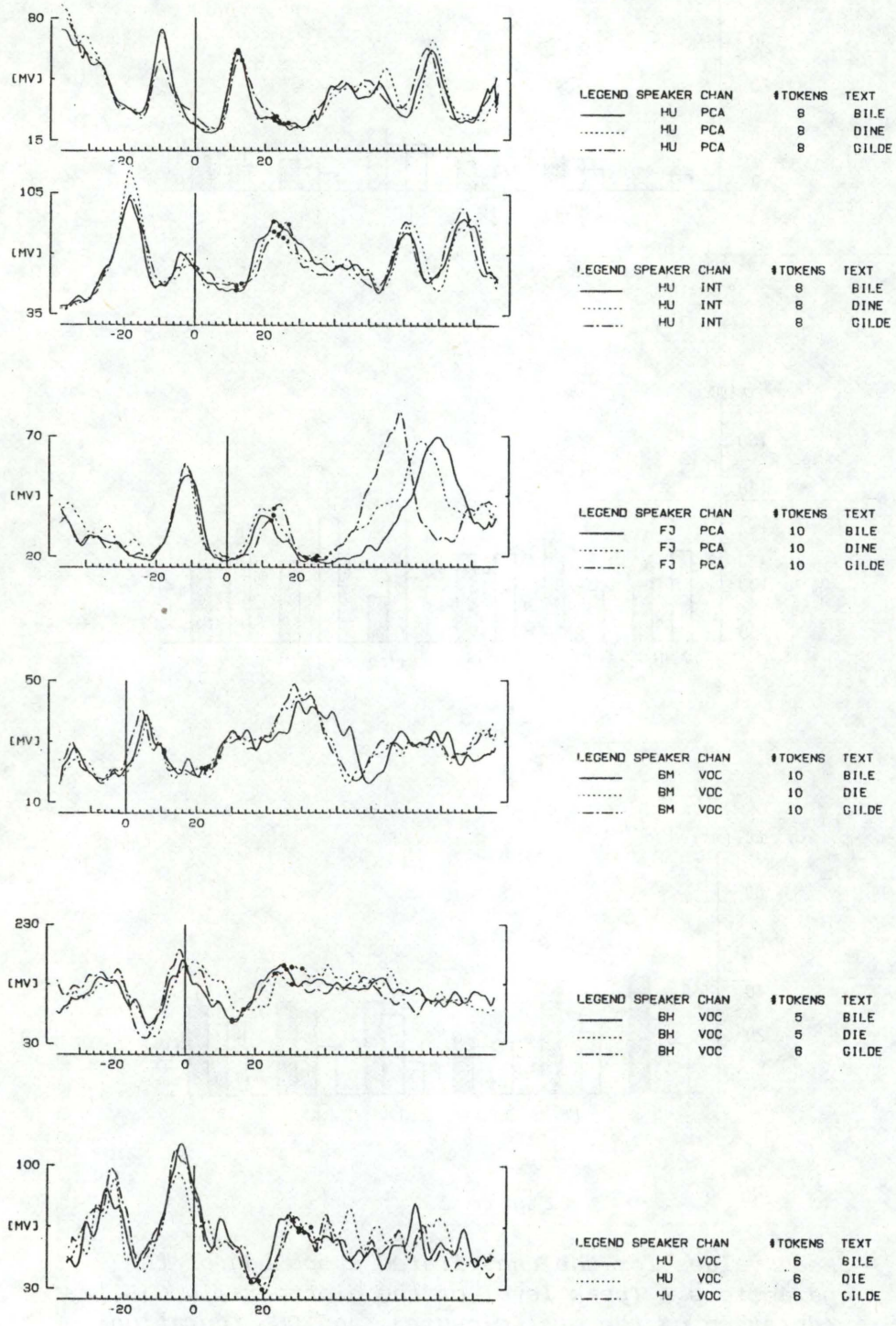
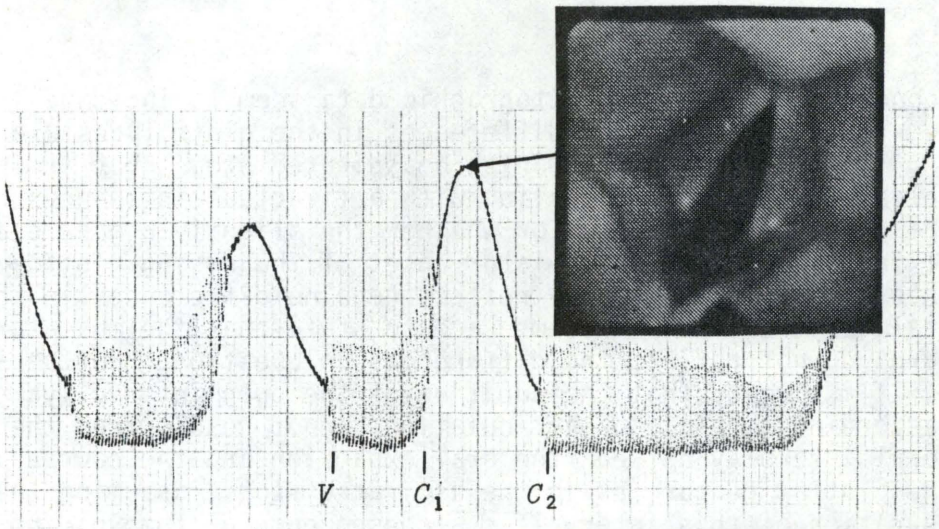
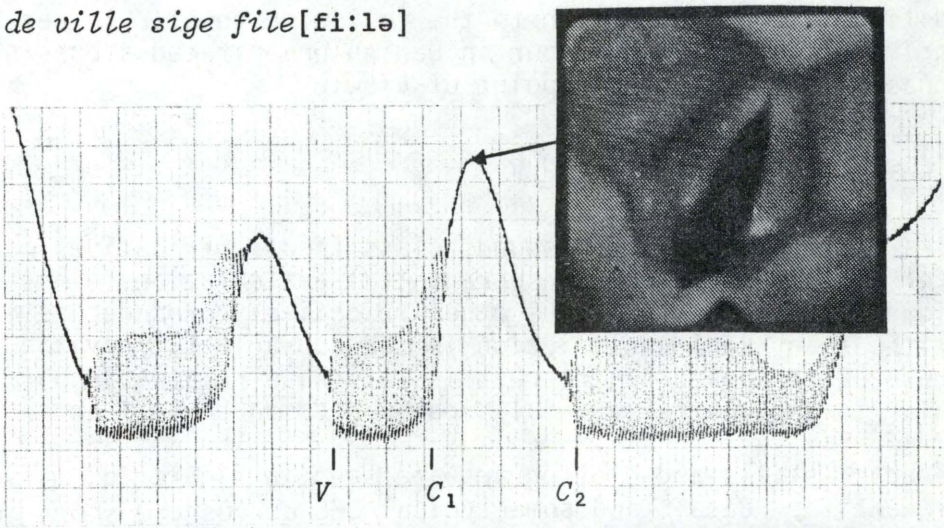


Figure 14

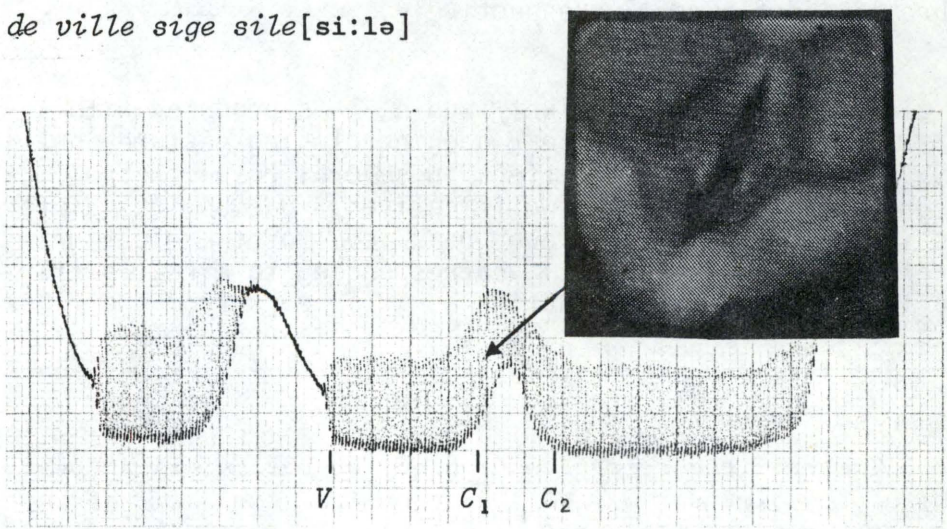
Averaged EMG curves of the unaspirated stops bdg.
For further explanation, see the legend to figure 11.



de ville sige file[fi:lə]



de ville sige sile[si:lə]



de ville sige hige[hi:ə]

Figure 15

Fiberoptic stills and glottograms of the fricatives fsh. The arrows indicate where in the course of the glottal gesture the still originates. The acoustic events are shown below the glottogram (cf. figure 1).

c. conclusion The glottographic data seem to indicate slight differences in the glottal gesture in the Danish unaspirated stops, but it is not clear whether we are dealing with artifactual influences on the glottographic signal or whether the slightly higher peak level in *g* and the slightly later onset of the rising glottographic signal in *b* actually reflect genuine differences in the glottal gesture. But it seems safe to state that the muscular patterns do not reflect the differences in question. Furthermore, it is argued that the duration of the open interval varying according to place of articulation cannot result from the temporal interplay between the oral explosion and the course of the glottal gesture as in the aspirated stops, assuming that the duration of this interval is a consequence of the voicing conditions at the transition to the following vowel. In section IVB the glottal behaviour in Danish unaspirated stops is discussed from a devoicing point of view.

3. FRICATIVES

Contrary to the stops, the sounds classified as fricatives may differ a good deal in their supraglottal articulation. This obviously applies to *h* versus other fricatives, produced as it is without any real constriction in the supraglottal cavities. Therefore, the point of interest in this case relates in particular to the glottal articulation in *h* compared to *f* and *s*.

Fiberoptic stills and glottograms are seen in figure 15, while the quantified data[18] and some typical EMG curves are shown in figure 16 and figure 17, respectively.

a. glottal gesture Like *f* and *s*, *h* is produced with a clear opening-closing gesture, but it differs in several respects from the gesture seen in the other two fricatives. One obvious difference is the substantially lower peak level of the glottogram (A - figure 16a, table XIII), corresponding to a real difference in the vocal fold abduction as it appears from the fiberoptic stills.

(Legend to figure 16)

Mean values for the fricatives fsh of the parameters A (peak level of the glottogram), VG (duration from the start of the preceding vowel to the onset of the glottal gesture), VC_1 (duration of the preceding vowel), and C_1C_2 (total duration of the obstruent). The left column represents f, the column in the middle s, and the right one h, as indicated above the rightmost columns in each graph. Notice that h is omitted for some of the parameters with speaker MF (see note [18]). The symbols for the level of significance should be interpreted as follows: the upper row shows the level of significance for f versus s, the row in the middle f versus h, and the lowest row s versus h. For further explanation, see the legend to figure 3.

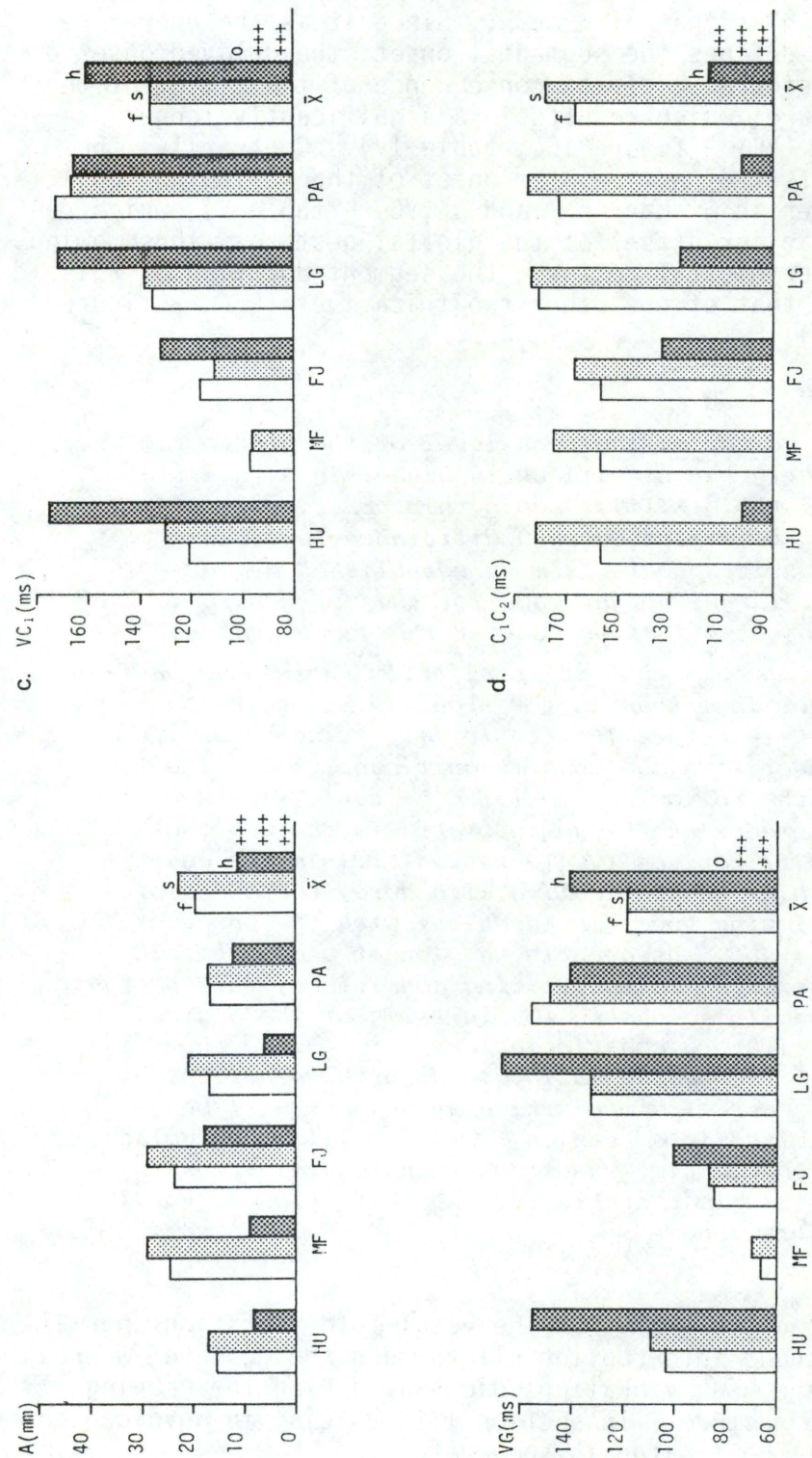


Figure 16

(legend on preceding page)

Regarding the temporal course of the gesture, the onset of the gesture in relation to the preceding vowel comes significantly later in *h* (VG - figure 16b, table VII). Now, in the case of *h*, the glottal aperture is the main factor accounting for the energy reduction seen in the intensity curves, this reduction occurring because of the absence of a narrow supraglottal constriction such as that of *f* or *s*. Since it is the energy reduction that defines the segmental onset, the delayed onset of *h* may be considered a direct consequence of the timing of the glottal gesture, which results in a significantly longer preceding vowel (VC_1 - figure 16c, table IV). Contrarily, the segmental offset of *h*, i.e. the onset of the following vowel, occurs earlier in *h* than in *f* and *s* (VC_2 - table V), which reflects the earlier offset of the glottal gesture. Thus, owing to the shorter glottal gesture, the segment duration of *h* is shorter than that of the other two fricatives (C_1C_2 - figure 16d, table I).

> *As to f versus s, the peak level of the glottogram is slightly but significantly higher in s (A - figure 16a, table XIII). Again this peak level difference may be explained by a difference in larynx height as it appears from my unpublished measurements of larynx height made for this very purpose (two subjects). If we look at the segment duration it appears that it is slightly but significantly longer in s than in f - about 15 ms on the average (C_1C_2 - figure 16d, table I). It seems as if the longer s is due to a (significantly) later onset of the following vowel (VC_2 - table V) since no difference can be demonstrated in the onset of the two fricatives (VC_1 - table IV). In the case of i, which we are dealing with here, the onset of the following vowel is identical with the onset of voicing, and consequently the longer s must be due to differences in the voicing conditions, this being most readily accounted for in terms of the degree of anticipatory coarticulation of i. Since the front of the tongue is involved in the s-articulation the anticipatory articulation of i will be more limited in s than in f, and it will thus delay the moment in time when the pressure drop across the glottis is suitable for the initiation of vocal fold vibrations.*

It should be noted that in *h* the vocal fold vibrations normally continue without interruption all through the gesture, whereas in *f* and *s* the weak vibrations die away during the opening phase of the gesture. In section IIIC voicing in unvoiced obstruents will be treated in some detail.

b. EMG It is obvious that the three Danish unvoiced fricatives are produced with the same type of overall muscular pattern: a pronounced PCA peak and a dip in INT and VOC with the PCA maximum and the INT and VOC minima in most cases positioned in the neighbourhood of the fricative onset, whereas the INT and VOC maxima after the dip occur in the vicinity of the offset of the fricatives. Thus, the timing of the muscular patterns differs more or less among the three fricatives in accordance with the differences in the timing of the segmental on- and offsets.

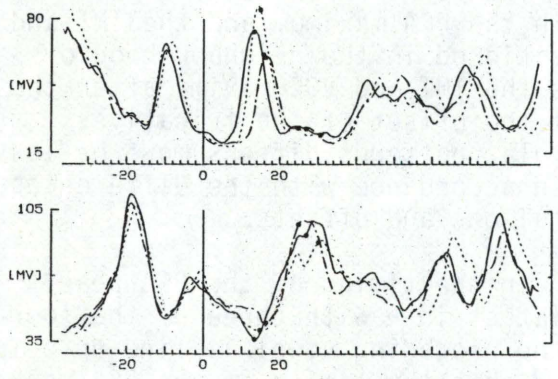
One clear difference is seen in the timing of the PCA peak which is earliest in *f* and latest in *h* with three of the four subjects with the onset of the preceding vowel as line-up point. Since it is normally supposed that the timing of the PCA peak corresponds to the timing of the maximum glottal aperture, it is somewhat surprising that no such difference can be observed in the averaged glottographic data (VM - table VIII). But if we look at the individual subjects included in both of the two materials, i.e. HU, FJ, LG, it appears that HU and LG do show the same timing differences in the glottographic data, *f*< *s*< *h*. FJ shows only a slight but significantly earlier maximum glottal aperture in *h* which corresponds fairly well with her PCA pattern.

Furthermore, it seems that the significantly later gesture onset found in *h* is reflected in the later rise in PCA, combined with a later reduction in INT. Also the general tendency to a later onset of the VOC reduction may reflect the later gesture onset in *h*, no matter how it actually influences the state of the vocal folds.

Regarding the EMG amplitude, two subjects, HU and PM, show a clearly lower PCA peak in *h* combined with a less pronounced dip in INT. It is tempting to assume that these muscular differences reflect the smaller and shorter glottal opening in *h*. On the other hand, FJ and LG do not show such a difference in the PCA peak. Provided, however, that it is the total amount of activity rather than the maximum activity of PCA that reflects the degree of the glottal opening, the somewhat more narrow peak in *h* may account for the smaller glottal aperture in *h*, probably combined with a less pronounced valley in INT. But also with regard to the INT valley individual deviations may be exemplified by LG, who shows a substantially more pronounced valley in *s* opposed to *h* as well as to *f*.

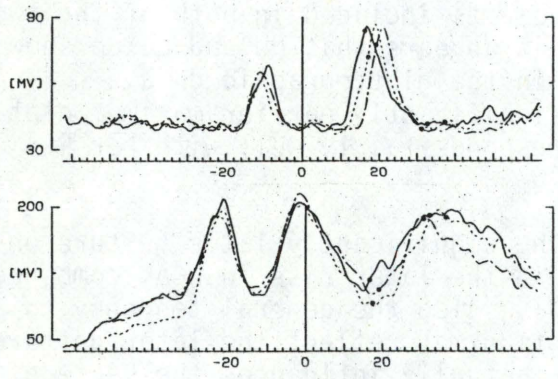
Other muscles may of course take part in the control of the difference in glottal aperture between *h* and *fs*. One such muscle could be VOC, even though the VOC differences also vary more or less from subject to subject, as it appears from the present material.

Only a few vocal fold studies have been reported in which *h* is compared with other unvoiced fricatives. In Frøkjær-Jensen et al. (1971) dealing with Danish, it is stated that the glottographic amplitude of *h* is similar in shape to that of *f*, but



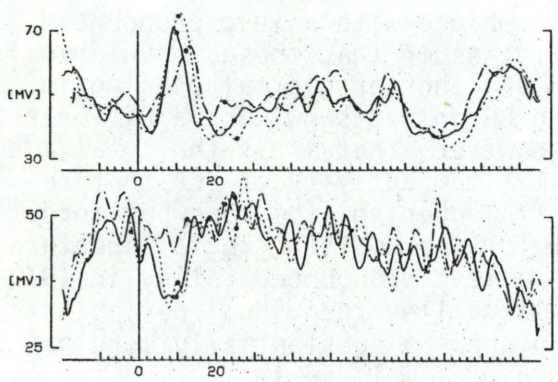
LEGEND	SPEAKER	CHAN	#TOKENS	TEXT
———	HU	PCA	9	FILE
.....	HU	PCA	8	SILE
-.-.-.	HU	PCA	8	HIGE

LEGEND	SPEAKER	CHAN	#TOKENS	TEXT
———	HU	INT	9	FILE
.....	HU	INT	8	SILE
-.-.-.	HU	INT	8	HIGE



LEGEND	SPEAKER	CHAN	#TOKENS	TEXT
———	LG	PCA	10	FILE
.....	LG	PCA	7	SILE
-.-.-.	LG	PCA	10	HIGE

LEGEND	SPEAKER	CHAN	REF	#TOKENS	TEXT
———	LG	INT		9	FILE
.....	LG	INT		7	SILE
-.-.-.	LG	INT		10	HIGE

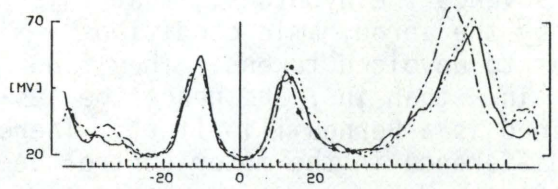


LEGEND	SPEAKER	CHAN	#TOKENS	TEXT
———	PM	PCA	10	FILE
.....	PM	PCA	10	SILE
-.-.-.	PM	PCA	10	HIGE

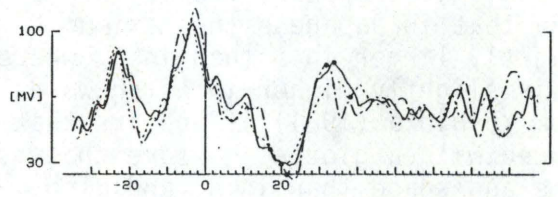
LEGEND	SPEAKER	CHAN	#TOKENS	TEXT
———	PM	INT	10	FILE
.....	PM	INT	10	SILE
-.-.-.	PM	INT	10	HIGE

Figure 17

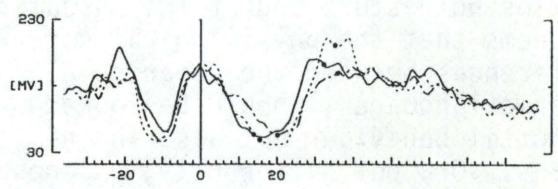
Averaged EMG curves of the fricatives fsh. For further explanation, see the legend to figure 7.



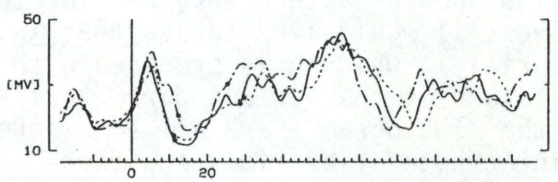
LEGEND	SPEAKER	CHAN	#TOKENS	TEXT
———	FJ	PCA	10	FILE
.....	FJ	PCA	10	SILE
-.-.-.	FJ	PCA	10	HIGE



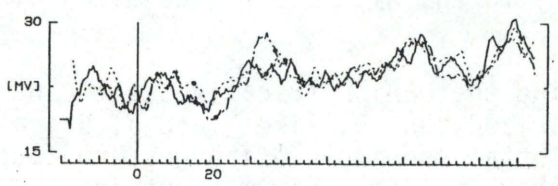
LEGEND	SPEAKER	CHAN	#TOKENS	TEXT
———	HU	VOC	6	FILE
.....	HU	VOC	6	SILE
-.-.-.	HU	VOC	6	HIGE



LEGEND	SPEAKER	CHAN	#TOKENS	TEXT
———	BH	VOC	6	FILE
.....	BH	VOC	6	SILE
-.-.-.	BH	VOC	6	HIGE



LEGEND	SPEAKER	CHAN	#TOKENS	TEXT
———	BM	VOC	10	FILE
.....	BM	VOC	10	SILE
-.-.-.	BM	VOC	10	HIGE



LEGEND	SPEAKER	CHAN	#TOKENS	TEXT
———	JJ	VOC	9	FILE
.....	JJ	VOC	10	SILE
-.-.-.	JJ	VOC	10	HIGE

Figure 17 (cont.)

that the aperture is generally smaller, and that this applies to intervocalic *h*, whether the vocal fold vibrations are interrupted or not. The authors advance the hypothesis that this difference may be explained by the aerodynamic conditions - applying to voiced as well as to unvoiced tokens. They find that the air flow is greater in *h* than in *f* and hence the lesser aperture could be explained as a Bernouilli effect. Therefore, they do not assume a difference in the motor control between *h* and the other fricatives. However, on the basis of the EMG material presented above, their hypothesis can hardly be maintained.

Opposed to the findings for Danish, it appears from figure 3 in Hirose and Ushijima (1978) that in Japanese the maximum glottal aperture is only slightly larger in *s* than in *h*, whereas the peak activity of PCA is slightly higher in *h* (in word initial position). Also from Yoshioka (1981) it appears that Japanese *h* tends to be produced with a glottal gesture that is only slightly smaller in time and space than in *s*, and also the EMG pattern is very similar. Therefore, the author supposes that the glottal adjustments may be almost identical in terms of the gross opening-closing gesture and of the muscular control as well. Thus, it seems that the physiological correlates to the well-known differences between the Japanese *h*, and the *h* found in the Germanic languages, should be looked for, not only in the supraglottal behaviour but also in the adjustments of the vocal folds. Or, put differently, Japanese *h* being produced with a substantial supraglottal constriction, i.e. produced as a supraglottal fricative, the glottal gesture and its motor control will also show a pattern very similar to the pattern of other fricatives. I would like to add that in the few cases seen in my material in which *h* is produced with interrupted vibrations, the values of the various parameters - including the peak level of the glottogram - are all somewhere between those for *h* with uninterrupted vibrations and those for *f* and *s*. Unfortunately, the present EMG material cannot tell us whether the muscle patterns also come closer to *f* and *s*, as it does not include any tokens with clearly unvoiced *h*.

c. conclusion

Concerning the Danish fricatives, it can be summarized that *h*, like *f* and *s*, has a clear glottal opening-closing gesture, but it is obviously produced with a smaller maximum aperture, a shorter duration, and a later onset of the opening gesture - all differences that seem to be reflected in the muscular activity pattern. One important question is why the gesture onset in *h* is delayed compared to the other two fricatives, if we are dealing with two independent systems with a high degree of synchronization in the control of glottal and supraglottal articulations. But maybe the question should rather be reversed, as follows: does the supraglottal constriction in fricatives like *f* and *s* somehow induce an earlier gesture onset? The question should be seen in the light of the finding in Löfqvist et al. (1981) that the control of the glottal opening is tightly coupled to activities in other parts of the speech apparatus.

One difference that should be kept in mind is that the vocal fold vibrations in intervocalic *h* normally continue throughout the glottal gesture, whereas in *f* and *s* they die away during the abduction of the vocal folds. This difference will be treated in detail in the following section which deals with the degree of voicing in unvoiced obstruents.

C. THE DEGREE OF VOICING IN THE UNVOICED OBSTRUENTS

It is obvious from the glottographic and fiberoptic material that the two obstruent categories with a large maximum aperture of the glottis, i.e. the aspirated stops and the fricatives, differ between them as to the degree of voicing, or more correctly stated: they differ as to the size of glottal aperture at the moment at which vocal fold vibrations can no longer be sustained and, conversely, are initiated. Such differences in voicing have also been observed by, among others, Frøkjær-Jensen et al. (1971).

In the intervocalic aspirated stops the vibrations cease shortly after the onset of the glottal opening gesture, whereas in the fricatives they continue into the opening phase of the gesture before they die away; in *h*, though, they are normally uninterrupted. These differences are obviously seen in the glottograms shown in figure 5, figure 8, and figure 15.

It is a well-known fact that the vocal folds only vibrate under adequate aerodynamic conditions combined with a suitable adjustment of the vocal fold tension and position. As it appears from figure 18 - showing glottograms and intraoral pressure of *p* and *f* - the glottal opening gesture in the fricative leads the glottal opening in the aspirated stop relative to the increasing intraoral pressure. Consequently, the glottis aperture is larger in the fricative when the pressure drop is no longer able to keep the vibrations going. Two factors seem to account for this difference. First, the rise of the intraoral pressure occurs later in the fricative than in the stop (in relation to a neutral line-up point such as the onset of the preceding vowel) due to the different manner of production. This explains the later acoustic onset of the fricatives resulting in the longer duration of the preceding vowel discussed in the section dealing with aspirated stops versus fricatives. The other factor is the difference in completion of the vocal fold abduction (also in relation to the onset of the preceding vowel), which occurs earlier in the fricative than in the stop as also discussed in the section dealing with aspirated stops compared with fricatives. In other words, the difference as to the size of the glottal aperture at the moment of offset of vocal fold vibrations can be accounted for in part by the different aerodynamic conditions induced by the supraglottal articulation - as suggested by Frøkjær-Jensen et al. (1971) - in part by a different timing of the glottal abduction itself. As a consequence, no difference in voicing duration relative to the acoustic onset of the obstruents should necessarily be expected

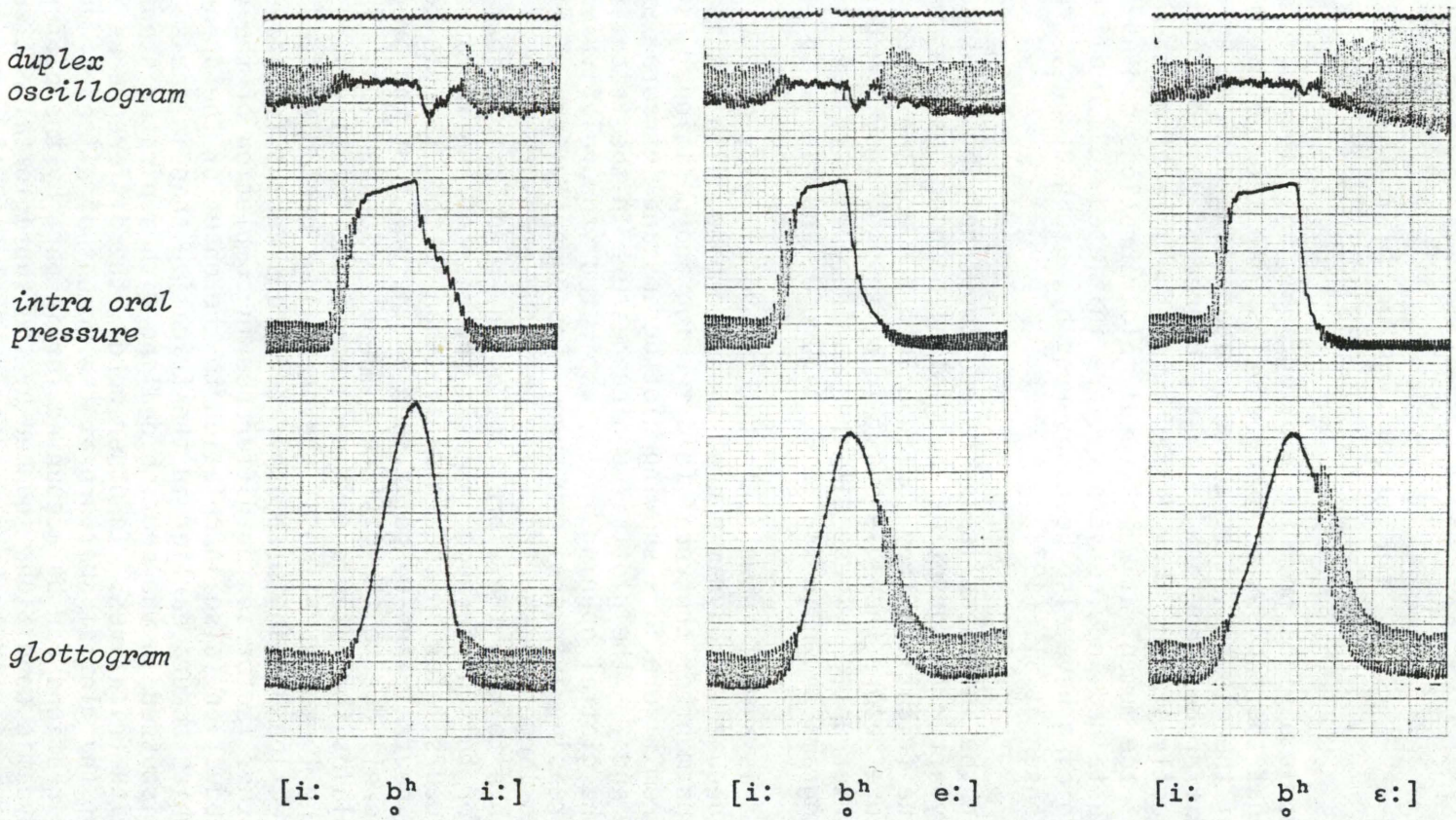


Figure 18

Intraoral pressure and glottograms of the aspirated stop p *and the fricative* f *followed by* [i:], [e:], *and* [ɛ:]

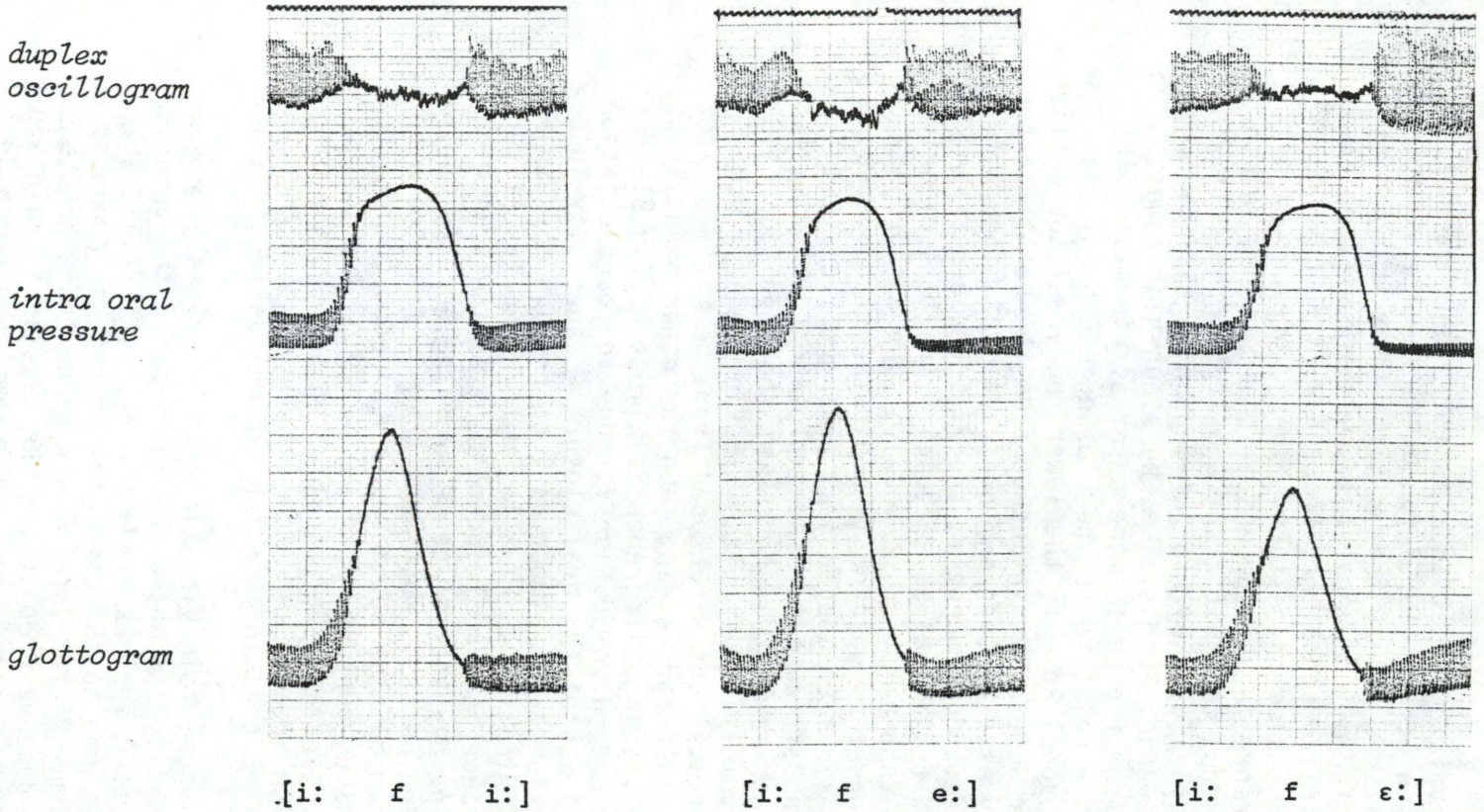


Figure 18 (cont.)

- other things being equal. And sure enough, the measurements do not show a longer voicing period in the fricatives than in the stops (C_1Z - figure 19a, table XI).[19] On the contrary, the tendency is toward a slightly longer voicing period in the stops, although no significant difference can be shown.

Turning now to the voicing onset in aspirated stops and fricatives, Frøkjær-Jensen et al. (1971) found the reverse relationship: in the fricatives (apart from *h*) vibrations did not resume until the closing of the vocal folds was almost completed, whereas in the aspirated stops they started well in advance of the completion of the closing gesture. In the present material, however, the difference in voicing onset is almost negligible. This discrepancy is due to a different quality of the following vowel: in the material used by Frøkjær-Jensen et al. the obstruents were followed by *e* or ɛ, i.e. by a non-high vowel. And as it appears from figure 18, the onset of vibrations in the aspirated stops - relative to the closing of the glottis - is correlated with vowel height: the lower the vowel the earlier do the vibrations resume. The influence from the height of the following vowel may also be seen in the fricatives, though to a much lesser degree. Consequently, the difference between aspirated stops and fricatives as to voicing onset is clearly greater before low than before high vowels. The authors explain the earlier onset of voicing in aspirated stops versus fricatives by the almost unimpeded oral passage, i.e. less resistance to the airstream after the explosion. This is also reflected in the intraoral pressure curves shown in figure 18: before a non-high vowel it takes less time in the aspirated stops than in the fricatives - reckoned from the maximum - to attain a level suitable for voicing. On the other hand, it is also clear why the voicing difference is almost negligible before the very high and narrow Danish *i*: due to coarticulation and to the slow release of the oral closure - especially in *t* - the pressure reduction takes almost as long time in the aspirated stops as in fricatives. Finally, within the fricatives the intraoral pressure is almost independent of the following vowel which explains why the influence from the quality of this vowel on the voicing onset is much less than in the aspirated stops: the airstream must be substantially impeded during most of the glottal closing gesture, irrespectively of the following vowel, in order to generate the fricative noise (cf. the discussion dealing with voicing condition accounting for the longer segment duration in *s* than in *f*, which closes section IIIB 3a).

The uninterrupted vocal fold vibrations in intervocalic *h* seem also to be most readily accounted for in terms of the aerodynamic conditions as supposed by Frøkjær-Jensen et al. (1971): the fact that *h*, as opposed to the other fricatives, is produced with an almost unconstricted vocal tract means that throughout this consonant a pressure drop across the glottis is maintained which is sufficiently great to keep the vibrations going in spite of a rather open glottis. Or, as stated by Sawashima and Hirose (1981): *"It is apparent that the crucial factor in the voicing distinction for these consonants [t s h] is the aerodynamic condition at the glottis rather than the extent of the*

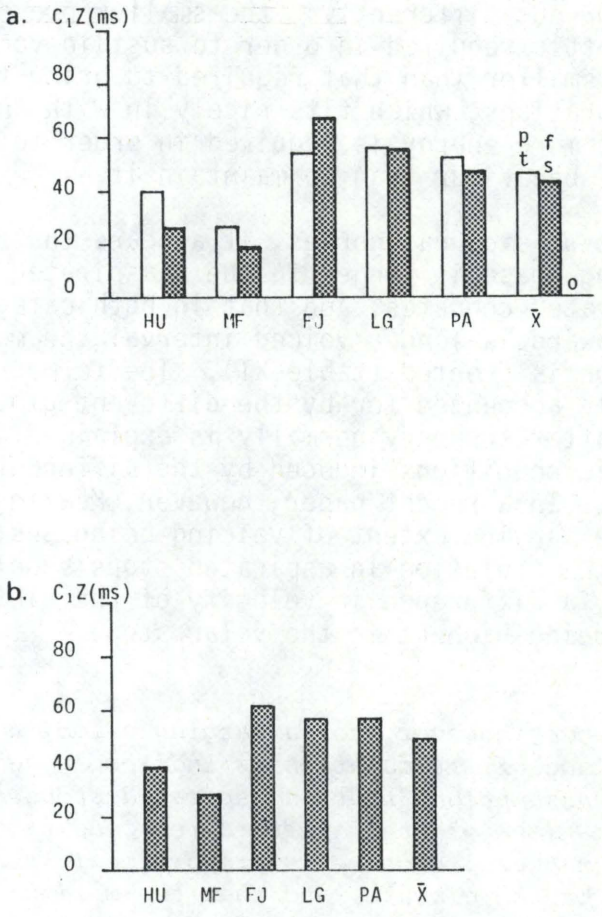


Figure 19

a. Mean values of voicing duration (C_1Z) in aspirated stops (pt) and fricatives (fs) - averaged over place of articulation. The mean values are shown for each speaker and averaged across speakers ($\bar{X}$ - computed from raw data). The left column represents the aspirated stop category, the right one the fricative category, as indicated above the rightmost columns. The difference is not significant (cf. the legend to figure 3).

b. Mean values of voicing duration for each single speaker, averaged over test obstruents. The grand mean averaged across speakers ($\bar{X}$ - computed from raw data) is shown in the rightmost column.

glottal abduction" (p. 339). But even if the crucial factor for the different voicing behaviour in these consonants is the aerodynamic conditions at the glottis, it seems plausible that the smaller glottal aperture in the Germanic *h*-type facilitates voicing. It should be added that the maximum intraoral pressure in *h* followed by *i* and the pressure level at which the vibrations die away in the other fricatives is much higher than the pressure level at which the voicing resumes in these other fricatives. Or put differently: the smaller pressure drop across the glottis required in order to sustain vocal fold vibrations is smaller than that required to bring the vocal folds into vibrations, which fits nicely in with the general principle that more energy is required in order to change a given state of behaviour than to maintain it.

Regarding the two stop categories, it appears that the physiological voicing phase is longer in the unaspirated stops than in their aspirated cognates, and that in both categories the tendency is towards a longer voiced interval the more the place of articulation is fronted (table XI). The former difference is most readily accounted for by the different glottal gesture, whereas the latter tendency normally is explained in terms of the aerodynamic conditions induced by the different supraglottal behaviour. In a recent paper, however, Keating (1984) assumes that the varying extent of voicing being sustained after the onset of the implosion in aspirated stops should be accounted for by a difference in velocity of the glottal opening, the velocity being highest in the velar stop.

> *Her evidence, however, for a varying velocity according to place of articulation is indirectly deduced from the assumptions that in aspirated stops the time to maximum glottal aperture reckoned from the onset of the oral closure is proportional to the duration of the closure, and that the maximum aperture does not vary across place of articulation (cf. the discussion in section IIIB1 above). Thus, the shorter the oral closure the higher the velocity of the glottal opening. But at least with regard to the first assumption, it is obviously not tenable for the present Danish data. This can be deduced from the fact that the time from explosion to maximum glottal aperture varies according to place of articulation, owing partly to the varying duration of the oral closure and partly to a small difference in the timing of the glottal maximum itself (cf. section IIIB1).*[20] *This, of course, does not imply that there may not be differences in the velocity of the glottal opening but only that we need other evidence (cf. the comments on abduction velocity at the end of section IIIA2).*

It should be added that for the Danish aspirated stops the data tend to show a trade-off between duration of the preceding vowel (VC_1 - table IV) and the voicing period (C_1Z - table XI). Thus, the very small and non-significant differences in voicing period may simply reflect the problems in delimitating C_1, i.e. the onset of the oral closure.

Finally, it should be observed that the duration of voicing - averaged over test sounds (except *h*) - is longer with the two male subjects than with two of the female subjects. In the third female subject (FJ), however, the voicing is even slightly longer than in the male subjects (figure 19b). Granted that substantial individual differences may occur, it nevertheless seems very likely that there is a clear tendency towards better voicing conditions in men's than in women's speech production, which has also been observed by others. Thorsen (1962), for instance, finds in his French material that the female subject has a smaller degree of voicing than the four male subjects, and that this is true of the voiced as well as the unvoiced obstruents. Smith (1977) also notices this tendency in phonologically voiced stops in American English. This effect of speaker's sex on the degree of voicing is normally explained by the sex conditioned difference in the volume of the supraglottal cavity which affects the aerodynamic conditions.

IV. DISCUSSION

The results presented above show that the glottal opening gesture and its motor control varies with obstruent category, and that even within the same category the glottal behaviour may vary more or less according to place of articulation. Many of the observations have already been discussed in relation to the findings of others. Thus, in this section I will take up some more general aspects concerning the glottal behaviour in obstruent production, primarily in relation to aspiration and voicing - or devoicing, rather - in stops. Finally, some problems relating to the interpretation of the electromyographic signals with reference to obstruents are pointed out.

A. ON ASPIRATION

The concept of aspiration associated with stop production has attracted the interest of phoneticians for many years. Today this interest is even intensified, partly due to the development of methods more suitable for examination of the laryngeal behaviour in normal speech, partly because of the focus in contemporary research on temporal phenomena in speech, including interarticulatory timing. There seems to be general agreement that aspiration is a matter of the glottal gesture and its temporal relationship to the supraglottal articulation. The discussion, however, of the phenomenon of aspiration and how it is produced can profitably be divided in two:
1. aspiration versus non-aspiration, and 2. varying duration of aspiration.

1. ASPIRATION VERSUS NON-ASPIRATION

As to the first point, it is well-established that aspirated and unaspirated stops are produced with different glottal gestures: the aspirated stops are produced with an almost symmetric opening-closing gesture with a relatively large glottal aperture at the moment of explosion - otherwise no aspiration would be produced - whereas the unaspirated stops are produced with a gesture which is of a smaller size in time as well as in space, being almost completed at the moment of explosion (references are given on p. 306). Thus, the difference between aspirated and unaspirated stops in the timing of the explosion relative to the glottal gesture is primarily due to a difference in the temporal course of the glottal gesture rather than to a different timing of the supraglottal articulation. In this connection it should be recalled that in Danish - as in several other languages - the oral closure has been found to be shorter in the aspirated than in the unaspirated stops (e.g. for Icelandic: Pétursson 1976; for Fukienese: Iwata et al. 1979; for Hindi: Kagaya and Hirose 1975; Benguerel and Bhatia 1980). In Danish the difference is very small, and I agree with Fischer-Jørgensen (1980) that this shorter closure *"is mainly due to weakness of articulation"* (p. 253). In any case it has obviously nothing to do with the control of the presence or absence of aspiration. It should be added that for Swedish it has been argued that the loss of aspiration in certain positions is due to an increase in the duration of the oral closure rather than to a change in the type of glottal gesture (Löfqvist 1976).

In papers dealing with aspiration it is very common to focus on the timing between the glottal and supraglottal articulations as the controlling factor of aspiration - rather than on the glottal gesture as such (see e.g. Löfqvist and Yoshioka 1980). But this emphasis on timing may hamper our understanding of the laryngeal mechanisms involved. In order to produce aspirated as opposed to unaspirated stops, we have to produce two basically different gestures to satisfy the demand of an open and a nearly closed glottis, respectively, at the moment when the oral closure is released. By emphasizing the inter-articulatory timing we may get the impression, rather, that the number of degrees of freedom existing between the glottal and supraglottal articulatory systems is less restricted than it actually is. The focusing on timing may even lead to formulations like: *"...we have shown evidence that the timing of peak glottal opening important as it is for distinguishing between aspirated and unaspirated stops..."* (Al-Bamerni and Bladon 1981, p. 8).

With regard to force of articulation, the unaspirated stops in Danish are evidently less fortis than those unaspirated stops of other languages which are normally labelled fortis, as it appears from the shorter oral closure in the Danish stops (Fischer-Jørgensen 1968a), even though they are probably more fortis than their aspirated cognates. However, not only the supraglottal but also the glottal articulation seems to

be different for the two types of unaspirated stops. Unfortunately, the glottal data available in the literature on fortis type stops are fairly scanty. Nevertheless, on the basis of the data on the glottal gesture in unaspirated fortis stops in Hindi (Kagaya and Hirose 1975; Dixit and MacNeilage 1980; Benguerel and Bhatia 1980) and in French (Benguerel et al. 1978), compared to the gesture observed in Danish unaspirated stops, I would speculate that the fortis type is produced with a more symmetric and somewhat larger glottal gesture than the lenis type. The more open glottis may be considered a sort of by-product from a general tensening of the speech organs, including the larynx. A difference as to fortisness in the glottal gesture of unaspirated stops is also proposed by Caisse (1982 - mentioned in Ohala 1982) as an explanation for the disagreement about effects of unaspirated stops on the fundamental frequency in the following vowel, the fortis type inducing a higher fundamental frequency than does the lenis type.

In this context it seems relevant to recall that Stevens and Halle (1971) in their paper dealing with laryngeal features suggest that voiceless, unaspirated stops normally are produced with stiff vocal folds opposed to the Danish type of unaspirated stops, which are supposed to be produced with neither stiff nor slack vocal folds. They assume that the different degrees of stiffness of the vocal folds are performed through adjustments of the VOC and the CT muscles. Data on Hindi compared with the present data on Danish seem to fit in nicely with their hypothesis inasmuch as a pronounced CT activity has been observed in Hindi unvoiced, unaspirated stops, whereas in the Danish cognates no such CT activity is present (cf. section IVC dealing with interpretation of laryngeal EMG signals).

A third type of unvoiced, unaspirated stop is the Korean forced stops which show a small glottal opening during the first part of the oral closure, whereas no opening is normally seen in the last part (Kagaya 1974). This constriction of the glottis results from an increased activity in the lateral cricoarytenoid muscle and in particular in the vocalis muscle, a muscular pattern that is specific to this type of stop (Hirose et al. 1974).

2. VARYING DURATION OF ASPIRATION

Let us now turn to the control of the varying duration of aspiration which may be induced by various linguistic and para-linguistic factors. It should be noticed that the term duration is used here instead of degree of aspiration, since the latter may in fact refer to time as well as to intensity. Very often, however, "strong" and "weak" are used synonymously with "long" and "short". This can normally be done without any confusion, because - as is stated by Löfqvist (1976): *"The role of the respiratory system in the control of aspiration seems to be limited or none whatsoever, perhaps with the exception of Korean stops"* (p. 17).

One of the factors that may result in varying duration of aspiration is the language specific requirements for the manifestation of the aspirated stops. Even though it may be difficult to compare measurements derived from different languages if they are not based on the same methods and procedures, there is no doubt that aspiration is longer in some languages than in others. This is exemplified in table III, which includes data from aspirated *p* in comparable environments. It is seen that the aspiration is longer in Danish and Hindi than it is in Swedish, English, and Icelandic, and vice versa for the closure. The closure is about 55% of the total segment duration in the former group and about 75% in the latter.

Table III

Data on aspirated p in comparable environments derived from several languages.

language	closure ms	aspir. ms	total ms	closure %	from max. glot. aperture to oral explosion ms
Danish (Hutters, Appendix A)	98	86	184	53	-10
Hindi (Kagaya and Hirose 1975)	90	70	160	56	0
Swedish (Löfqvist 1976)	115	32	147	78	14
English[11] (Quinlan 1978)	128	50	178	72	18
Icelandic (Pētursson 1976)	129	36	165	78	49

Furthermore, the data in table III indicate that in languages with the longer aspiration the explosion slightly leads or coincides with the maximum glottal aperture, whereas in the others it comes after this maximum. Provided that these examples are representative of what is going on in languages with aspirated stops in general, it would appear that one important factor responsible for the inter-language differences is the duration of the oral closure. But the varying duration of aspiration may also be influenced by different voicing conditions, causing a difference in voicing onset, at the transition from the stops to the following vowel. Nevertheless, if in

general a shorter closure in aspirated stops is due to a weaker articulation, as it seems to be the case in Danish, then it may be concluded that in aspirated stops a more lenis manifestation is followed by a longer aspiration compared to aspirated stops with a more fortis manifestation. This would be an argument against including aspiration in the fortis feature as most recently proposed by Kohler (1983). This topic is discussed at greater length by Fischer-Jørgensen (1968b). From a comparison of table III with the data for Danish *bdg* (cf. figure 13c - table III) it also appears that in slightly aspirated stops the duration of the open interval may be of the same order of magnitude as the duration of this interval in unaspirated, voiceless stops. This supports the assumption that the gesture in unaspirated stops differs in nature from the gesture in aspirated stops. Moreover, it implies the well known fact that a temporal parameter like the 'open interval' or VOT may not necessarily reveal a difference in aspiration versus non-aspiration. The fact that two unvoiced stops with the same delay in voice onset time can be categorized as aspirated and unaspirated, respectively, presupposes that aspiration is not simply conceived of as a delay in voice onset time.

In this context I want to point to the fact that the disagreement regarding aspiration as it is discussed in the literature is due to some extent to a conceptual confusion of defining and explanatory descriptions. Aspiration has been defined as, for instance, "a puff of air" (e.g. in Jespersen 1897-99), as "a large delay in voice onset time" (e.g. in Lisker and Abramson 1964), as "a period of voicelessness" (in Ladefoged 1975), and as "a glottal friction" (Dixit 1979) - following the oral explosion. But since defining aspiration is a matter of purpose or of point of view - within certain limits - it may not be very relevant to discuss which of the definitions are right or wrong, and they certainly should not be compared to explanatory descriptions. On the other hand, it is very relevant to discuss various views on the physiological explanations of aspiration.

In relation to the language specific requirements for the production of aspirated stops it should be recalled that the duration of aspiration varies according to place of articulation, the normal ordering being *k>t>p* for the three most common places of articulation. As mentioned above, these differences can be accounted for by the inherent characteristics of the main articulator involved in performing the oral closure and by the degree of restriction as to coarticulation imposed by this same articulator. Both are factors which may influence the closure duration and the voicing onset. Other explanations are discussed in Fischer-Jørgensen (1980). However, as mentioned above, the relations may deviate from the more usual ones due to language specific manifestation requirements like in Danish.

Other factors that may influence the duration of aspiration are those which influence segment duration in general, namely factors such as speaking rate and linguistic stress. In

Andersen (1981) it is shown that in Danish aspirated stops an increase in speaking rate results in a decreased duration of both oral closure and aspiration, accompanied by a reduction - in time and space - of the glottal gesture. Furthermore, it was found that the temporal relationship between the glottal opening and closing was almost preserved. Therefore, it seems very likely that the timing of the oral explosion relative to the glottal gesture is preserved, even though no information on this point is given by Andersen. A similar change in the glottal and supraglottal articulatory events - with preservation of the interarticulatory timing - probably takes place in case there is a stress conditioned variation in the duration of aspiration. This assumption is based on a glottographic pilot study including test material with aspirated stops placed under main stress, reduced stress, and weak stress.

Thus, it seems that the physiological behaviour resulting in varying duration of aspiration - including no aspiration - cannot always be described as a matter of interarticulatory timing. And in cases where it may be appropriate to describe the variation as a matter of timing, it does not necessarily explain the actual physiological processes involved. A very well-known spatial parameter supposed to explain the variation in duration of aspiration has been suggested by Kim in his widely quoted article on aspiration claiming that *"it seems safe to assume that aspiration is nothing but a function of the glottal opening at the time of release of the oral closure"* (Kim 1970, p. 109). To put it differently, he states that it is only the degree of glottal opening at the oral release that determines the onset of the vocal fold vibrations - alias the onset of the following segment. Now, as pointed out by Kim himself and later by others (e.g. by Lisker and Abramson 1971) this implies that the rate of the glottal closure is constant, which probably cannot be taken for granted. Furthermore, it seems appropriate also to take into consideration not only the glottal aperture as such, but also the point at which, in the temporal course of the glottal gesture, the explosion takes place, inasmuch as the oral release may precede the maximum glottal aperture. But there is a more important point to be made, viz. that it is not only the time it takes after the oral release to close the glottis but also the aerodynamic conditions at this point in time that determine the onset of vocal fold vibrations, and thus the duration of aspiration as it is defined by Kim - and as it is normally defined. In the following section the glottal gesture in Danish unaspirated stops will be discussed in more detail.

B. ON THE GLOTTAL GESTURE IN DANISH UNASPIRATED STOPS

According to the myoelastic-aerodynamic theory of phonation, the vocal folds will vibrate when they are properly adducted and tensed and when a sufficient airflow passes between them. Based on an electrical model it was shown by Rothenberg (1968) that if the vocal folds are vibrating, these vibrations will

cease shortly after a total blocking of the flow of air is performed somewhere in the vocal tract. This is due to a decrease in the pressure drop across the glottis and thus in the airflow through it. This seems to agree with the fact that in the Danish unaspirated stops the vibrations - as they appear from the acoustic signal[19] - die away shortly after the onset of the oral closure. It is generally accepted that in order to keep the vibrations going in spite of the oral closure the vocal tract must be expanded - actively or passively. In other words, some mechanism(s) must be added in order to produce a voiced stop.

It is therefore quite understandable that Frøkjær-Jensen et al. (1971) propose the interesting hypothesis that the glottal opening gesture in Danish unaspirated stops may be a consequence of the offset of the vocal fold vibrations, i.e. of the change in the aerodynamic conditions induced by the oral closure, rather than the result of muscular activity such as normally controls the abduction and adduction of the vocal folds. This seems even more evident from the fact that with the arytenoids in adducted position the muscular part of the vocal folds will form a spindle-shaped glottis (see, e.g., Zemlin 1982; Sawashima and Hirose 1983) - very similar to the glottis shape in Danish *bdg*. Thus, the new idea proposed by Frøkjær-Jensen et al. (1971) is that the glottal gesture in the unaspirated stops is a passive process in the sense that it is a consequence of the aerodynamic conditions introduced by the oral closure, whereas the glottal gesture in the aspirated stops is considered an active process, i.e. controlled by the abductor and adductor muscles of the vocal folds. Later, this idea has been accepted by others as explaining the slight glottal slit seen in devoiced and partially devoiced stops as found in English and German (Kohler 1977). However, the passive gesture hypothesis was later rejected as a consequence of EMG recordings showing activity in the abductor and adductor muscles in the Danish aspirated as well as in the unaspirated stops (Fischer-Jørgensen and Hirose 1974), a finding that is fully confirmed in the present EMG material.[21]

Some years ago I carried out a small experiment in which a total blocking of the airflow was externally implemented. While the subject was phonating a sustained *i* into an airtight mask with a small aperture, the phonation was momentarily interrupted, at unexpected points in time, by closing the external orifice of the mask. Fiberoptic stills and glottograms were recorded, and it appeared that the state of the glottis and the surrounding structures - after the cessation of voicing - were very similar in the externally implemented stop and in the natural unaspirated stop. But it also appeared that in the mask-condition the period of voicing was much longer, implying that it took much more time after the onset of the closure before the spindle-shaped glottis appeared. This longer voicing period can only to some extent be accounted for by the extension of the vocal tract induced by the mask.

Later, EMG recordings of the PCA muscle have been made during the production of externally implemented closures, in order to throw light on the question whether the laryngeal muscle activity found in the natural Danish unaspirated stops might result from a change in the aerodynamic conditions, by means of some reflex mechanisms sensitive to such changes. The idea was not pure imagination, considering that such reflex mechanisms seemed to be present in the VOC muscle (Wyke 1976). But already from the start of the experiment it seemed rather unlikely that the hypothesis would be confirmed, taking into account that in the unaspirated stops the onset of the rising PCA activity precedes the oral implosion, which initiates the changes in the aerodynamic conditions. And sure enough, the PCA muscle did not show a pattern that could have any relation to the opening of the glottis in the externally implemented stops.

Consequently, on the basis of the finding that the much longer period of voicing in the externally implemented closure implies that it takes much more time before the spindle-shaped glottis appears after the blocking of the vocal tract, it is tempting to assume that the abduction of the vocal folds in the natural unaspirated stops is a consequence of the cessation of vocal fold vibrations, and that this cessation of vibrations is not exclusively a passive process but that some additional mechanism is directly involved in the devoicing process. Thus, considering that the glottal opening was performed without any PCA activity when the closure of the vocal tract was externally implemented, I venture the controversial hypothesis that the activity in the PCA and INT muscles is a devoicing mechanism rather than a means to open and close the glottis. This implies that the slight abduction of the arytenoid cartilages that may be seen in the Danish *bdg* is considered a by-product of the vocal fold adjustment that causes the vocal fold vibrations to die, rather than an end in itself.

> *In this connection, it seems relevant to mention that in speech produced by children suffering from a pronounced velopharyngeal insufficiency, the Danish unaspirated stops may be produced with a nasal puff of air. This indicates another state of the glottis than in normal production, since with the vocal folds in "bdg-position", the voicing starts immediately after the introduction of an externally implemented leakage to the closed vocal tract (cf. Fischer-Jørgensen 1963 and the mask experiment presented above). Thus, it appears that it is not possible to produce the bdg-gesture without a complete blockage of the airflow. This may of course suggest that the devoicing in normal Danish bdg is after all due to the change in the aerodynamic conditions induced by the oral closure as originally proposed. But it may also be concluded that the PCA and INT muscles cannot work as they do in Danish unaspirated stops with a leakage in the vocal tract - no matter whether their*

function is to produce the spindle-shaped glottis or to terminate the vocal fold vibrations.

The general problem as to whether some devoicing mechanism is involved in the production of unvoiced obstruents - and in particular of unvoiced stops - is not unknown. Lisker (1977) claims in his paper dealing with factors in the maintenance and cessation of voicing that *"it has not been generally agreed that voiceless closure intervals require no devoicing maneuvre other than the articulatory closure itself"* (p. 306). In support of the hypothesis that some additional active devoicing mechanism may be needed in order to produce unvoiced unaspirated stops, like those found in Danish, I refer to Westbury (1983). He suggests, on the basis of his electrical model, that voicing in intervocalic stops may continue considerably longer after the oral implosion than supposed by Rothenberg (1968), due to the compliance of tissues surrounding the supraglottal cavity, a factor that has not been taken into account in Rothenberg's model. On the basis of the same electrical model as used by Westbury, Keating (1984) has shown that the more lax the cheeks the longer the voicing continues after the onset of the oral closure. The present data on the total (physiological) voicing interval (table XI) with which the simulations shown in Keating's paper can be compared, suggest that Danish *bdg* should be produced with moderately tensed cheeks. This seems plausible taking into account that Danish *bdg* are considered more tense than voiced stops and less tense than fortis stops. On the other hand, it appears from her paper that the simulated data should rather be compared with the shorter voicing intervals deduced from acoustic curves, and in that case (cf. note[19]) her data suggest that Danish *bdg* should be produced with very tensed cheeks, which is not very likely, indeed. Another devoicing action could be a decrease in the supraglottal volume, but according to Keating *"the contribution of this parameter is quite small compared to the contribution of the surface area of the cavity walls"* (p. 30). Thus, I maintain the more controversial hypothesis that the PCA and INT activity seen in Danish *bdg* is a devoicing action.

The devoicing mechanism originally proposed by Halle and Stevens (1971) should be mentioned in this context. They claimed that *"an increased stiffness of the vocal folds tends to narrow the range of the transglottal pressure and glottal apertures over which the vocal fold vibration occurs"* (p. 202). The empiric foundation, however, is not very solid except that the very consistent relaxation of the VOC muscle in unvoiced obstruents may be considered a devoicing mechanism. This will be discussed in the following section dealing with interpretation of laryngeal EMG signals. It should be added that in Stevens (1977) it is suggested that the theory on the "horizontal" stiffness of the vocal folds should be changed into one of "vertical" stiffness of the folds - a theory which is even more lacking on empirical verification.

I also want to refer to the general observation that phonologically unvoiced unaspirated stops tend to be weakly voiced in non-strong positions, as it is the case for Danish intervocalic *bdg*. This may support the active devoicing theory, supposing that a general reduction in articulatory effort takes place in non-strong positions (Kohler 1983): when the articulatory effort is reduced the devoicing mechanism - whatever its nature may be - is consequently less effective or not effective at all.

Finally, I want to add a comment on the widespread assumption that the glottal opening in unvoiced fricatives and aspirated stops is a devoicing mechanism rather than a mechanism permitting the production of the required airflow. In Weismar (1980), for instance, the glottal gesture in English unvoiced obstruents is called "a devoicing gesture". The same view also appears - directly or indirectly - from descriptions of the articulatory behaviour of the glottis as it is found in various textbooks (e.g. Daniloff et al. 1982, Ladefoged 1971). But as the vocal folds may vibrate in spite of a considerable degree of glottal opening, *"...the extent of the glottal opening itself is not necessarily a crucial condition for the cessation of vocal fold vibration"* (Sawashima and Hirose 1983, p. 17). On the face of it the concept of a devoicing gesture in unvoiced fricatives and aspirated stops may lead to the impression that voicelessness and aspiration - even if it may only be a matter of terminology - should also be avoided in the feature description, provided that features are intended to reflect the physiological behaviour in speech production.

In summary, there are indications in favour of the assumption that the opening of the vocal folds seen in the Danish unaspirated tops is a consequence of the cessation of the vocal fold vibration as originally proposed by Frøkjær-Jensen et al. (1971). Accepting that the offset of the gesture is equivalent to the resumption of the vocal fold vibrations, the glottal gesture as such may be considered a consequence of the voicing conditions. The keystone is to find out whether the cessation of vibrations is simply a passive process, in the sense that it is only due to the changing aerodynamic conditions induced by the oral closure, or an active process performed for instance by a change in the vocal fold adjustment or by a change in the supraglottal conditions such as in the compliance of tissues - or both. I have ventured the controversial hypothesis that the PCA and INT activity seen in the Danish unaspirated stops is a devoicing action rather than a means to open and close the glottis. *"What is most certain in all this is that stop voicing will continue to provide problems to exercise us..."* - to quote the final remarks in Lisker (1977, p. 306).

C. ON THE INTERPRETATION OF LARYNGEAL ELECTROMYOGRAPHIC SIGNALS

The last discussion leads to the final topic which deals with some problems relating to the interpretation of laryngeal EMG signals in terms of vocal fold behaviour in the production of obstruents. Some of these problems have already been touched upon in previous sections, but in the following a more coherent account will be given.

The articulatory behaviour of the vocal folds is normally considered as being controlled by the intrinsic laryngeal muscles (for a recent survey I refer to Sawashima and Hirose 1983). But it is obviously not possible to deduce the influence of one particular pattern of muscle activity on vocal fold behaviour, due primarily to the fact that the influence of too many factors are still not known.

It is generally recognized that the reciprocal pattern between the PCA and INT activity is the muscular behaviour primarily responsible for the abduction and adduction of the vocal folds in speech production, and that the PCA peak and the INT suppression tend to be more marked with a larger glottal aperture. From a visual impression and based on the observation that there is a positive correlation between the maximum glottal aperture and the PCA peak, a very direct relationship has been suggested between the glottal opening and the PCA muscle, in time as well as in space (Hirose 1975; Hirose and Ushijima 1978). However, as pointed out by Löfqvist and Yoshioka (1979) such a simple positive relation applies only to pooled data, whereas many exceptions can be observed if only single data points are taken into consideration. Löfqvist and Yoshioka find that *"within one and the same utterance type the temporal changes of glottal opening area and PCA activity levels are monotonically related"* (p. 119), and from their material including obstruent clusters they conclude that *"PCA activity thus seems more directly related to changes in glottal area than to glottal area per se"* (p. 121). The authors also suggest that the clear reciprocal relation between the PCA and INT muscle activity applies only to single voiceless obstruents, whereas in consonant clusters the relation is not that simple.

Concerning the INT muscle, I question whether it can be taken for granted that the increasing activity that follows the suppression is more pronounced when the obstruent has a larger glottal aperture, as suggested by Fischer-Jørgensen and Hirose (1974): *"...the activity of the closing muscle [INT] is more pronounced when the preceding consonant has a larger glottal opening"* (p. 250). My reservation is due to the fact that in the present material this relation is not very consistent across speakers and in particular to the finding that the relation seems to be highly influenced by the following vowel: the difference is diminished or may even be absent before *a*, due to a reduction of the INT peak level in the aspirated stops. In Fischer-Jørgensen and Hirose's material including also *i* and *a* (and *u*) no such influence from vowel type is observed.

Furthermore, results from other languages are apparently somewhat inconsistent. I am referring to Hirose et al. (1974) who observe a higher INT peak level in the Korean aspirated stops than in the unaspirated cognates, whereas Kagaya and Hirose (1975) find no such difference.

If these - and maybe other - uncertain factors are taken into account, though, it seems safe to state that for single obstruents there is a strong tendency toward a reciprocal relationship between the activity in the PCA and INT muscles. Furthermore, it appears that their activity patterns are somewhat differentiated according to the degree of glottal opening - if the spatial as well as the temporal dimensions of their activity patterns are taken into account. In this context I refer to the very hypothetical suggestion presented above, namely that the PCA and INT activity in Danish unaspirated stops has to do with the cessation of vocal fold vibrations rather than serving to open and close the glottis. If this idea is proved to be correct, then it must be realized that the function of these two muscles may not only be related directly to the abduction and adduction of the vocal folds but also to some other kind of adjustment of the vocal folds. In support of this assumption it should be mentioned that according to the present EMG material also *l* and *v* may be produced with increased PCA activity.

As regards the VOC muscle it is normally reported that the activity is somewhat reduced in unvoiced as well as in voiced obstruents, but the findings are somewhat fluctuating inasmuch as the reduced activity may not always differ in the degree of reduction (Kagaya and Hirose 1975; Hirose and Ushijima 1978; Collier et al. 1979). Furthermore, it generally appears - whether explicitly or implicitly - that in unvoiced obstruents the glottis tends to be more open the more the VOC activity is suppressed. From the present material, however, it appears that the difference in suppression may be very small in spite of a considerable difference in glottal aperture. On the other hand, unvoiced obstruents always differ in the duration of the period of suppression according to differences in segment duration. It should be kept in mind that the suppression of VOC - and INT - directly related to unvoiced obstruents may be smaller than it appears at first, when the obstruent is preceded by another obstruent followed by a vowel.

The point of interest is the influence that this reduction in VOC activity may have on vocal fold behaviour. It has been supposed to reflect the cessation of voicing (Hirose et al. 1974, 1981). This may seem reasonable according to the cover-body theory originally proposed by Hirano (1977), implying that the lower VOC activity results in a reduced slackness of the cover which should hamper the vocal fold vibrations - ceteris paribus. Consequently, the reduction in voiced obstruents may serve other purposes - if it is not secondary to some primary articulatory behaviour. Furthermore, if the reduced activity is crucial for the devoicing process in all unvoiced obstruents, it has to be admitted that the mechanism

is, paradoxically, more pronounced in obstruents produced with a large glottal aperture than in stops produced with the vocal folds in voicing position like in Danish unaspirated stops. Conversely, a contraction of VOC may result in a slackening of the cover, facilitating vocal fold vibrations (Fujimura 1977), provided that the CT activity is not increased.

In relation to the theory of reduced VOC activity as a devoicing factor it is worthy of notice that when the obstruent is followed by *a* instead of *i* the present Danish material indicates no difference in the VOC activity pattern between the unaspirated stops and the sonorants. This means that the small dip seen in the unaspirated stops before *i* is no longer present before *a*. If this observation turns out to be a general phenomenon, it seems reasonable to conclude that the devoicing interpretation is less probable than otherwise assumed.

But it has also been suggested that the differentiated degree of VOC suppression in unvoiced obstruents corresponds to differences in the glottal opening-closing gesture (Hirose et al. 1974, 1981). This interpretation is also assumed in Collier et al. (1979), who claim that the unaspirated *"stops [in Dutch] show less relaxation than fricatives suggesting that the vocal folds are slacker in the latter case. Probably the slackening of the vocal folds in the fricatives also contributes to their abduction"* (p. 364). The last interpretation apparently sticks to the older theory about the influence of VOC on the stiffness of the vocal folds.

Concerning the interpretation that the reduced VOC activity somehow reflects the opening-closing gesture of the glottis, Sonesson (1982) claims that in adducted position the VOC muscle has but little influence on the cricoarytenoid joint, whereas with the vocal folds in abducted position the VOC muscle may assist in moving the vocal folds from abducted to adducted position. This might explain the higher VOC activity that can be observed after fricatives and aspirated stops than after unaspirated stops (see Hirose et al. 1974 - figure 4 and figure 5; Collier et al. 1979). But again, the situation seems more complex inasmuch as in the present material the increased VOC level in obstruents with a large glottal aperture is in fact substantially reduced when followed by *a* versus *i*. Fujimura (1977) has suggested that this "momentary activity" in VOC after the 'forced' Korean stops (Hirose et al. 1974) and after aspirated stops *"function as a relatively fast-response voicing trigger mechanism which may be available for vocal fold vibration under otherwise unfavourable conditions"* (p. 286). It may be tempting to see the higher VOC level before *i* than before *a* in the light of this interpretation, since the aerodynamic conditions are in fact less favourable for voicing onset before the very narrow Danish *i* than before the low vowels. This interpretation, however, is obviously weakened by the fact that in aspirated stops followed by *i* the vocal fold vibrations are not resumed until the completion of the glottal gesture, whereas if followed by *a* they may be resumed very early in the closing gesture. In this context it should

be recalled that contrary to the present findings no influence from the following vowel on the INT and VOC patterns in obstruents has been observed in other studies including more than one vowel (Fischer-Jørgensen and Hirose 1974; Sawashima and Hirose 1981; Sawashima and Hirose 1983).

Finally, even if the INT and VOC activity patterns in obstruents may look fairly much alike, their functions are indeed different not only with reference to prosody but also to the production of segments. This is clear from the fact that in sonorants the INT muscle never shows any increasing activity at the transition to the following vowel, whereas the VOC muscle always show such increasing activity. The VOC peak in Danish 'stød', in glottal stops, and in the 'forced' Korean stops also indicate that the two muscles serve different purposes, since no such peak is seen in the INT muscle. It is tempting to take this to mean that INT is basically an adductor muscle, whereas VOC primarily serves other purposes.

Even though the lateral cricoarytenoid muscle (LCA) traditionally is classified as an adductor muscle it is evident from EMG studies that its function with reference to segmental events is more complex. Studies which include LCA report that this muscle behaves almost like the VOC muscle, i.e. it may be involved in the control of the adduction of the vocal folds as well as of the voicing distinction (Hirose et al. 1974; Kagaya and Hirose 1975; Collier et al. 1979). On the other hand, it can hardly be stated from the data available in the literature whether VOC and LCA can be functionally differentiated in the production of obstruents.

It appears from studies dealing with the segmental aspect of the CT muscle that there may be differences partly between voiced versus unvoiced stops, and partly between aspirated versus unaspirated stops. As regards the CT activity related to voiced opposed to unvoiced stops the findings are rather contradictory (Kagaya and Hirose 1975; Hirose 1977; Collier et al. 1979; Sawashima 1979; Dixit and MacNeilage 1980). But if there is a difference the level is higher in the unvoiced cognates - a difference that has been supposed to facilitate voicelessness.

> *It should be added that the difference may appear as a difference within an overall decrease or increase in activity probably due to non-segmental differences in the test utterances.*

As regards the CT activity related to aspirated versus unaspirated stops a higher level has been found in Hindi unaspirated stops, a difference that seems to be more pronounced at the onset of the stops than later in the closure (Kagaya and Hirose 1975; Dixit and MacNeilage 1980). It seems reasonable, as suggested by Dixit and MacNeilage (1980), that it is the unaspirated cognates with the smaller aperture that need a de-

voicing mechanism - in the form of a higher CT activity - in addition to the elimination of the pressure drop across the glottis due to the oral closure. However, since the tendency in the present very limited amount of data is rather toward a slightly increased CT level in the Danish aspirated stops (and not especially marked at the onset of the closure) without any change in the unaspirated ones, increased CT activity can hardly serve the same purpose in Danish as in Hindi - if it serves a purpose at all. But purposeful or not, the increasing CT activity may of course influence the state of the vocal folds.

Sawashima and Hirose (1983) observe - in their Danish material - that also in physiologically voiced intervocalic *h*, a relatively high CT activity is present, serving as an argument against as being very important for the devoicing process. In the present Danish *h*-material, however, no such clear CT activity is present, which is another case of contradictory EMG findings.

As it appears, it is not very clear how the CT and VOC activity patterns actually observed may function as devoicing mechanism, but more data are called for, of course. Or in more general terms, as Lisker (1977) puts it: "*...if the voicelessness of particular consonants is said to involve, necessarily or even optionally, some action to stiffen the folds, we are so far without observational data to support it*" (p. 305).

Finally, I want to point to the considerable inter-speaker variation that can be observed in the present EMG material, and which emphasizes that it may be hazardous to generalize on the basis of very few subjects - not to mention statements on the basis of one single speaker. Very often our claims and statements about laryngeal muscle activity in speech are based on very few speakers due to the fact that subjects do not exactly queue up for laryngeal EMG recordings. One obvious risk is, however, that differences which are in fact speaker specific are taken to be language specific - and vice versa - when we compare findings relating to different languages. Another problem that arises is that we cannot always be sure whether differences regarded as inter-subject are due to some specific experimental conditions such as placement of the electrodes within a given muscle, or whether they in fact reflect individual variation in the laryngeal control in speech. The first interpretation can of course be ruled out if the findings with a given speaker can be reproduced with reinserted electrodes. In case the inter-speaker differences actually reflect individual variation the problem is whether such a variation in muscular activity results in the same or in a slightly different overt articulatory behaviour. Last, it should be pointed out that for obvious reasons it is normally only some of the laryngeal muscles that are successfully registered synchronously, which is an evident draw-back in light of the fact that the muscular system is supposed to work as a whole.

V. FINAL CONCLUDING REMARKS

Considering the very inaccessible placement of the larynx it is somewhat surprising that it is one of the more well described parts of the speech production apparatus. However, as regards the unvoiced obstruents, most studies deal with single sounds in intervocalic, stressed position at normal speech rate, even though some studies on obstruents under various other linguistic and para-linguistic conditions have been carried out. It is evident that the glottal gesture in single unvoiced obstruents differs according to the aerodynamic demands, i.e. primarily setting of aspirated stops and fricatives as against unaspirated stops, and that the different gesture types are reflected in the motor control. But it seems reasonable to put the question whether the EMG activity patterns actually found in the laryngeal muscles in unaspirated stops such as Danish *bdg* should in fact be considered an action causing devoicing rather than serving to perform a slight opening gesture. The clear opening-closing gesture produced in aspirated stops and fricatives is often considered ballistic in nature. If, however, it were a true ballistic movement, i.e. not controlled for duration and extent, the gesture and its motor control should be almost similar across the two types of obstruents. The tendency, however, is rather toward a more or less differentiated course of the gesture, specific not only to each category but also, to a certain extent, to each particular speech sound. Likewise, it has become very clear that each muscle seems to show a very delicate pattern of activity almost specific to each particular condition, including the phonetic context in which the obstruent is produced. On the other hand, it is so far impossible to give a detailed interpretation of the muscular activity patterns in terms of vocal fold behaviour as it appears from glottographic data. This can mainly be accounted for by two factors, viz. by the problems related to the interpretation of the glottographic signal and by the fact that the motor patterns may not always be interpreted in terms of movements. It is crucial for our progress in this field that we still increase the amount of data in order to determine to what extent our results are speaker specific and to what extent they are specific to a given language or rather to the speech habits of a given speech community. In this context I want to point out that most studies are based on averaged data which highlight the invariance in speech production. However, even though invariance in the production of speech is very important from a linguistic point of view, it is mandatory to focus also on the variability that occurs, i.e. to consider speech production from a more biological point of view, in order to better understand the basic behaviour not only of the larynx but of the speech production system in general.

VI. NOTES

1. This project is part of a larger framework titled: "The glottal behaviour in Danish consonants, stress, and stød".

2. The insertation of electrodes was performed by Dr. Hajime Hirose, Institute of Logopedics and Phoniatrics, University of Tokyo. With subject FJ, however, the insertation was performed by Seiji Niimi, also of the Institute of Logopedics and Phoniatrics, University of Tokyo.

3. The modification has been performed by Preben Dømler and Peter Holtse.

4. The planning of the EMG project and the preparation of the material for computer processing were carried out primarily by Eli Fischer-Jørgensen.

5. /*sj*/, pronounced [ʂ] or [ʃ], has not been included in the test material.

6. With PA all the test words were of the structure [Ci:lə] including some nonsense words.

7. It should be pointed out that the speaker shown on the fiberoptic stills has a small leakage between the vocal processes in voiced sounds.

8. In Fukui and Hirose (1983), which also includes subject FJ (= their EFJ), it is shown that her *p*-explosion occurs after the maximum abduction of the vocal folds, whereas in the present study it slightly leads the maximum as also observed in other subjects. Furthermore, the difference between *p* and *b* as regards the maximum glottal aperture is apparently less in Fukui and Hirose's fiberoptic material than in mine. These discrepancies may be due to methodological differences.

9. PM's *bdg*, however, deviate from the general pattern by having a very low-level activity without a clear rising-falling pattern. It is tempting to relate this deviating pattern to the fact that voicing in his unaspirated stops continues for a longer period of time after the implosion than is normally seen, i.e. these stops are partially - though weakly - voiced, probably due to his dialect background. But it cannot be ruled out that the reason is the fairly bad quality of his EMG signals.

10. *k* and *h* are omitted in the comparisons between aspirated stops and fricatives, since they have no counterpart with identical place of articulation in the other category. From section IIIB it will appear that place of articulation is a factor that has to be taken into account.

11. Unpublished essay by Glenn Quinlan (1978).

12. The difference averaging 35 ms can be deduced from VM - VC_1 (figure 6b - table VIII and figure 6e - table IV).

13. The delay in the onset of the fricatives - resulting in longer vowel before fricatives than before stops - has been explained in terms of a slower rate of movements of the speech organs "related to the precision required for fricative production" (MacNeilage 1972, p. 27) or because, as Kohler (1983) puts it, "the fricative requires a greater muscular coordination than stops" (p. 276). The slower rate of movements of the speech organs in fricative production seems to be reflected in the delayed rise of the intraoral pressure (see section IIIBc dealing with voicing), but I also refer to the discussion in section II and particularly to the fact that our statements and results about temporal relations may be influenced by the curves and criteria used for delimitation.

14. It is not explicitly stated in Löfqvist and Yoshioka (1981) that the difference concerns aspirated stops and fricatives. But it appears that the stops are produced with a considerable opening of the glottis, and apart from Japanese the languages included are all considered as containing aspirated stops.

15. By increasing the distance between the glottis and the transducer the signal is considerably modified in sounds produced with a small glottal aperture primarily positioned in the muscular part of the vocal folds, and the increasing-decreasing appearance of the signal may even disappear.

16. The substantial inter-subject variation may be explained by uncertainty in the delimitation of G.

17. In the Icelandic material published by Löfqvist and Yoshioka (1980) the explosion in *p* coincides with maximum aperture of the glottis.

18. C_1, C_2, G, and M in *h* could not be delimitated with subject MF.

19. It should be noted that the parameter C_1Z shows higher values than normally given for the period of voicing after the onset of the oral closure in stops. However, in the present material we are dealing with physiological rather than acoustic voicing. Furthermore, C_1 is probably leading the moment of total blockage of the vocal tract.

20. It also appears directly that in the present data a shorter closure is not followed by a corresponding reduction in duration from the onset of the oral closure to the maximum glottal aperture. The oral closure duration (C_1E) is seen in figure 10e - table II, while the time from the maximum glottal aperture reckoned from the onset of the oral closure (C_1M) can be deduced from VM - VC_1 (figure 6b - table VIII and figure 6e - table IV).

21. In the reprint from 1973 these later findings are mentioned by Frøkjær-Jensen, Carl Ludvigsen, and Jørgen Rischel.

ACKNOWLEDGEMENTS

My sincere thanks are due to Jørgen Rischel for his valuable suggestions for improvements in the manuscript.
The project has been supported by the Danish Research Council for Humanities.

REFERENCES

Al-Bamerni, A.H. and Bladon, R.A.W. 1981: "On the role of peak glottal opening", *Papers in Experimental Phonetics*, (University of Oxford)

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. 15*, p. 103-146

Benguerel, A.-P., Hirose, H., Sawashima, M., and Ushijima, T. 1978: "Laryngeal control in French stop production: a fiberscopic, acoustic and electromyographic study", *Folia Phoniatrica 30*, p. 175-198

Benguerel, A.-P. and Bhatia, Tej K. 1980: "Hindi stop consonants: an acoustic and fiberscopic study", *Phonetica 37*, p. 149-158

Butcher, A. 1977: "Coarticulation in intervocalic voiceless plosives and fricatives in connected speech", *Arbeitsberichte 8*, p. 154-213 (Kiel)

Caisse, M. 1982: *Cross-linguistic differences in fundamental frequency perturbation induced by voiceless unaspirated stops*, Master's thesis, University of California, Berkeley, (quoted in Ohala 1982)

Collier, R., Lisker, L., Hirose, H., and Ushijima, T. 1979: "Voicing in intervocalic stops and fricatives in Dutch", *J. Phonetics 7*, p. 357-373

Daniloff, R., Schuckers, G., and Feth, L. 1980: *The physiology of speech and hearing*, (Prentice-Hall)

Dixit, R.P. 1979: "Aspiration: what is it and how is it produced?", *ASA-50 Speech Communication Papers* (edited by J.J. Wolf and D.H. Klatt), p. 83-88

Dixit, R.P. and MacNeilage, P.F. 1980: "Cricothyroid activity and control of voicing in Hindi stops and affricates", *Phonetica 37*, p. 397-406

Ferguson, G.A. 1976: *Statistical Analysis in Psychology Education*, (McGraw-Hill Book Company)

Fischer-Jørgensen, E. 1954: "Acoustic analysis of stop consonants", *Miscellanea Phonetica II*, p. 42-59

Fischer-Jørgensen, E. 1963: "Beobachtungen über den Zusammenhang zwischen Stimmhaftigkeit und intraoralem Luftdruck", *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung 16, 1-3*, p. 19-36

Fischer-Jørgensen, E. 1968a: "Les occlusives françaises et danoises d'un sujet bilingue", *Word 24, 1-3*, p. 112-153

Fischer-Jørgensen, E. 1968b: "Voicing, tenseness, and aspiration in stop consonants, with special reference to French and Danish", *Ann. Rep. Inst. Phon. Univ. Cph. 3*, p. 63-114

Fischer-Jørgensen, E. and Hirose, H. 1974: "A preliminary electromyographic study of labial and laryngeal muscles in Danish stop consonant production", *Status Rep. Speech Res., Haskins Labs. 39/40*, p. 231-254

Fischer-Jørgensen, E. 1976: "Some data on North German stops and affricates", *Ann. Rep. Inst. Phon. Univ. Cph. 10*, p. 149-200

Fischer-Jørgensen, E. 1980: "Temporal relations in Danish tautosyllabic CV sequences with stop consonants", *Ann. Rep. Inst. Phon. Univ. Cph. 14*, p. 207-261

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. 15*, p. 77-102

Frøkjær-Jensen, B., Ludvigsen, C., and Rischel, J. 1971: "A glottographic study of some Danish consonants", *In: Form and Substance*, p. 123-140; also *Ann. Rep. Inst. Phon. Univ. Cph. 7, 1973*, p. 269-295

Fujimura, O. 1977: "Control of the larynx in speech", *Phonetica 34*, p. 280-288

Fukui, N. and Hirose, H. 1983: "Laryngeal adjustments in Danish voiceless obstruent production", *Ann. Bull. Res. Inst. Logopedics and Phoniatrics 17*, p. 61-71, (Tokyo)

Halle, M. and Stevens, K.N. 1971: "A note on laryngeal features", *MIT Quart. Prog. Rep. 101*, p. 198-213

Hirano, M. 1977: "Structure and vibratory behaviour of the vocal folds", *In: Dynamic Aspects of Speech Production* (edited by M. Sawashima and F.S. Cooper), p. 13-27 (Tokyo)

Hirose, H. and Gay, T. 1972: "The activity of the intrinsic laryngeal muscles in voicing control: An electromyographic study", *Phonetica 25*, p. 140-164

Hirose, H., Lee, C.Y., and Ushijima, T. 1974: "Laryngeal control in Korean stop production", *J. Phonetics 2*, p. 145-152

Hirose, H. 1975: "The posterior cricoarytenoid as a speech muscle", *Ann. Bull. Res. Inst. Logopedics and Phoniatrics 9*, p. 47-66

Hirose, H. and Ushijima, T. 1978: "Laryngeal control for voicing distinction in Japanese consonant production", *Phonetica 35*, p. 1-10

Hirose, H., Yoshioka, H., and Niimi, S. 1978: "A cross-language study of laryngeal adjustment in consonant production", *Ann. Bull. Res. Inst. Logopedics and Phoniatrics 12*, p. 61-71

Hirose, H. 1979: "Laryngeal EMG", *Ann. Bull. Res. Inst. Logopedics and Phoniatrics 13*, p. 13-22

Hirose, H. and Sawashima, M. 1981: "Functions of the laryngeal muscles in speech", *In: Vocal Fold Physiology* (edited by K.N. Stevens and M. Hirano), p. 137-154 (University of Tokyo Press)

Hirose, H., Park, H.S., Yoshioka, H., Sawashima, M., and Umeda, H. 1981: "An electromyographic study of laryngeal adjustments for the Korean stops", *Ann. Bull. Res. Inst. Logopedics and Phoniatrics 15*, p. 31-43

Holtse, P. and Stellinger, J. 1976: "A system for computer aided processing of phonetic measurements", *Ann. Rep. Inst. Phon. Univ. Cph. 10*, p. 201-219

Hoole, P., Pompino-Marschall, B., and Danes, M. 1983: "Glottal timing in German voiceless occlusives", *Abstracts 10th Int. Congr. Phon. Sc.*

Iwata, R. and Hirose, H. 1976: "Fiberoptic acoustic studies of Mandarin stops and affricates", *Ann. Bull. Res. Inst. Logopedics and Phoniatrics 10*, p. 47-60

Iwata, R., Sawashima, M., Hirose, H., and Niimi, S. 1979: "Laryngeal adjustments of Fukienese stops. Initial plosives and final applosives", *Ann. Bull. Res. Inst. Logopedics and Phoniatrics 13*, p. 61-81

Jespersen, O. 1897-99: *Fonetik*, (Copenhagen)

Kagaya, R. 1974: "A fiberscopic and acoustic study of the Korean stops, affricates and fricatives", *J. Phonetics 2*, p. 161-180

Kagaya, R. and Hirose, H. 1975: "Fiberoptic, electromyographic and acoustic analyses of Hindi stop consonants", *Ann. Bull. Res. Inst. Logopedics and Phoniatrics 9*, p. 27-46

Keating, P. 1984: "Physiological effects on stop consonant voicing", *Working Papers in Phonetics, UCLA 59*, p. 29-34

Kohler, K. 1977: "The production of plosives", *Arbeitsberichte 8*, p. 30-49 and 98-110 (Kiel)

Kohler, K. 1983: "Phonetic timing as explanation in phonology", *Abstracts 10th Int. Congr. Phon. Sc.*, p. 275-283

Ladefoged, P. 1971: *Preliminaries to Linguistic Phonetics*, (University of Chicago Press)

Ladefoged, P. 1975: *A Course in Phonetics*, (New York)

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

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

Lisker, L. and Abramson, A.S. 1971: "Distinctive features and laryngeal control", *Language 47*, p. 767-785

Lisker, L. 1977: "Factors in the maintenance and cessation of voicing", *Phonetica 34*, p. 304-306

Löfqvist, A. 1976: "Closure duration and aspiration for Swedish stops", *Working Papers 13*, p. 1-39 (Lund)

Löfqvist, A. and Yoshioka, H. 1979: "Laryngeal activity in Swedish voiceless obstruent clusters", *Status Rep. Speech Res., Haskins Labs. 59/60*, p. 103-126

Löfqvist, A. and Yoshioka, H. 1980: "Laryngeal activity in Icelandic obstruent production", *Status Rep. Speech Res., Haskins Labs. 63/64*, p. 275-292

Löfqvist, A. and Yoshioka, H. 1981: "Interarticulator programming in obstruent production", *Phonetica 38*, p. 21-34

Löfqvist, A., Baer, T., and Yoshioka, H. 1981: "Scaling of glottal opening", *Phonetica 38*, p. 265-276

Löfqvist, A. 1982: "Laryngeal articulation in speech", *Folia Fennistica & Linguistica*, p. 219-230 (Tampere)

MacNeilage, P.F. 1972: "Speech physiology", *In: Speech and Cortical Functioning* (edited by J.H. Gilbert), p. 1-72 (Academic Press)

Ohala, J.J. 1982: "Physiological mechanisms underlying tone and intonation", *Preprints: Working Group on Intonation*, p. 1-12 (Tokyo)

Pétursson, M. 1976: "Aspiration et activité glottale. Examen expérimental à partir de consonnes islandaises", *Phonetica 33*, p. 169-198

Reinholt Petersen, N. 1983: "The effect of consonant type on fundamental frequency and larynx height in Danish", *Ann. Rep. Inst. Phon. Univ. Cph. 17*, p. 55-86

Rischel, J. and Hutters, B. 1980: "Filtering of EMG signals", *Ann. Rep. Inst. Phon. Univ. Cph. 14*, p. 285-316

Rothenberg, M. 1968: "The breath-stream dynamics of simple-released-plosive production", *Bibliotheca Phonetica No. 6*, (Basel)

Sawashima, M. 1970: "Glottal adjustments for English obstruents", *Status Rep. Speech Res., Haskins Labs. 21/22*, p. 187-200

Sawashima, M. and Miyazaki, S. 1973: "Glottal opening for Japanese voiceless consonants", *Ann. Bull. Res. Inst. Logopedics and Phoniatrics 7*, p. 1-9

Sawashima, M. and Niimi, S. 1974: "Laryngeal conditions in articulations of Japanese voiceless consonants", *Ann. Bull. Res. Inst. Logopedics and Phoniatrics 8*, p. 13-17

Sawashima, M., Hirose, H., Ushijima, T., and Niimi, S. 1975: "Laryngeal control in Japanese consonants, with special reference to those in utterance-initial position", *Ann. Bull. Res. Inst. Logopedics and Phoniatrics 9*, p. 21-26

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 12*, p. 53-60

Sawashima, M. 1979: "Laryngeal control for voicing distinctions: a review of recent works", *Ann. Bull. Res. Inst. Logopedics and Phoniatrics 13*, p. 23-26

Sawashima, M. and Hirose, H. 1981: "Abduction-adduction of the glottis", *In: Vocal Fold Physiology* (edited by K.N. Stevens and M. Hirano), p. 329-344, (University of Tokyo Press)

Sawashima, M. and Hirose, H. 1983: "Laryngeal gestures in speech production", *In: The Production of Speech* (edited by P.F. MacNeilage), p. 11-38 (Springer-Verlag)

Smith, B.L. 1977: "Effects of vocalic context, place of articulation, and speakers' sex on "voiced" stop consonant production", *West Coast Phonetics Symposium, Santa Barbara*

Sonesson, B. 1982: "Vocal fold kinesiology", *In: Speech Motor Control* (edited by Grillner et al.), p. 113-117 (Pergamon Press)

Stevens, K.N. 1977: "Physics of laryngeal behaviour and larynx modes", *Phonetica 34*, p. 264-279

Thorsen, O. 1962: *Voice assimilation of stop consonants and fricatives in French and its relation to sound duration and intra-oral air-pressure*, unpublished thesis

Weismer, G. 1980: "Control of the voicing distinction for intervocalic stops and fricatives: some data and theoretical considerations", *J. Phonetics 8*, p. 427-438

Westbury, J.R. 1983: "Enlargement of the supraglottal cavity and its relation to stop consonant voicing", *J. Acoust. Soc. Am. 73*, p. 1322-1336

Winer, B.-J. 1970: *Statistical Principles in Experimental Design*, (McGraw-Hill)

Wyke, B.D. 1976: "Laryngeal reflex mechanisms in phonation", *Folia Phoniatrica 28, Suppl. 1*, p. 528-537

Yoshioka, H. 1981: "Laryngeal adjustments in the production of the fricative consonants and devoiced vowels in Japanese", *Phonetica 38*, p. 236-251

Zemlin, W.R. 1981 (1968): *Speech and Hearing Science*, (Prentice-Hall Inc.).

Table I

Total duration of the obstruent C_1C_2 (ms)

	HU			MF			FJ			LG			PA			Grand mean		
	$\bar{X}$	sd	N	$\bar{X}$	sd	N	$\bar{X}$	sd	N	$\bar{X}$	sd	N	$\bar{X}$	sd	N	$\bar{X}$	sd	N
pi	197	11	10	164	16	10	171	11	10	185	11	13	214	7	5	183	19	48
ti	186	11	10	185	27	11	179	9	10	199	16	12	209	11	10	192	19	53
ki	190	6	10	203	12	10	180	9	10	192	13	10	208	11	10	194	14	50
bi	138	7	10	123	18	8	138	6	10	144	7	10	152	7	11	140	13	49
di	137	10	11	137	20	10	128	11	10	141	11	10	141	8	10	137	13	50
gi	152	7	10	165	18	10	140	16	10	138	11	11	153	8	10	149	16	48
fi	157	9	10	157	20	10	157	8	11	181	10	10	173	7	10	165*)	15	51
si	182	11	11	175	17	10	167	7	10	184	16	10	185	8	10	178*)	14	51
hi	102	15	10	-	-	-	133	19	10	126	15	10	102	12	10	115	21	40

*) If MF is not included, $\bar{X}$ = 167 ms for fi and 179 ms for si.

Table II

Duration of the oral closure C_1E (ms)

	HU			MF			FJ			LG			PA			Grand mean		
	$\bar{X}$	sd	N	$\bar{X}$	sd	N	$\bar{X}$	sd	N	$\bar{X}$	sd	N	$\bar{X}$	sd	N	$\bar{X}$	sd	N
pi	103	5	10	73	15	9	107	13	10	99	8	13	113	14	5	98	17	47
ti	81	8	10	73	17	10	83	9	10	60	8	9	72	9	10	74	13	49
ki	93	5	10	107	17	9	89	13	9	80	4	10	110	8	10	96	15	48
bi	117	8	10	101	19	6	116	8	10	130	7	10	131	8	10	120	14	46
di	105	11	11	103	23	10	108	11	9	107	9	10	113	7	9	107	14	49
gi	103	9	8	128	23	8	116	9	10	99	15	11	114	7	9	111	16	46

Table III

Duration of the open interval EC_2 (ms)

	HU			MF			FJ			LG			PA			Grand mean		
	$\bar{X}$	sd	N	$\bar{X}$	sd	N	$\bar{X}$	sd	N	$\bar{X}$	sd	N	$\bar{X}$	sd	N	$\bar{X}$	sd	N
pi	94	11	10	93	11	9	64	15	10	85	8	13	101	13	5	86	16	47
ti	105	9	10	114	19	10	97	4	10	138	17	9	138	9	10	118	21	49
ki	97	5	10	94	16	9	89	12	9	112	12	10	98	11	10	98	13	48
bi	21	8	10	26	2	5	22	4	10	14	4	10	21	5	10	20	6	45
di	32	9	11	34	5	10	20	4	9	34	5	10	27	3	9	30	8	49
gi	50	5	10	39	12	8	24	9	10	39	6	11	39	5	9	38	11	48

Table IV

Duration of the preceding vowel VC_1 (ms)

	HU			MF			FJ			LG			PA			Grand mean		
	$\bar{X}$	sd	N	$\bar{X}$	sd	N	$\bar{X}$	sd	N	$\bar{X}$	sd	N	$\bar{X}$	sd	N	$\bar{X}$	sd	N
pi	115	11	10	79	6	10	101	14	10	115	14	13	152	27	5	108	24	48
ti	120	12	10	80	11	11	96	13	10	124	10	12	155	22	10	115	29	53
ki	124	15	10	71	12	10	107	14	10	112	14	10	156	14	10	114	31	50
bi	117	9	10	93	11	9	123	11	10	135	6	10	158	15	11	126	24	50
di	117	8	11	87	14	10	113	14	10	139	10	10	160	11	10	123	27	51
gi	118	9	10	82	14	10	105	11	10	118	8	11	156	19	10	116	27	51
fi	121	8	10	97	8	10	117	11	11	135	14	10	174	10	10	128*)	27	51
si	130	10	11	96	11	10	111	10	10	138	9	10	168	14	10	128*)	26	51
hi	175	16	10	-	-	-	132	24	10	172	11	10	166	24	10	161	26	40

*) If MF is not included, $\bar{X}$ = 136 ms for fi and si.

Table V

Duration of the period including the obstruent and the preceding vowel VC_2 (ms)

	HU			MF			FJ			LG			PA			Grand mean		
	$\bar{X}$	sd	N	$\bar{X}$	sd	N	$\bar{X}$	sd	N	$\bar{X}$	sd	N	$\bar{X}$	sd	N	$\bar{X}$	sd	N
pi	311	7	10	240	13	10	272	18	10	299	14	13	366	28	5	290	39	48
ti	305	17	10	264	23	11	275	18	10	324	15	12	364	23	10	306	40	53
ki	314	17	10	274	16	10	287	22	10	304	16	10	364	17	10	308	36	50
bi	254	12	10	214	21	8	261	12	10	279	6	10	310	16	11	267	34	49
di	255	15	11	224	17	10	241	15	10	279	13	10	301	13	10	260	31	51
gi	270	8	10	247	19	10	245	17	10	256	13	11	309	19	10	265	28	51
fi	278	11	10	254	19	10	274	17	11	316	19	10	347	10	10	293*)	37	51
si	311	11	11	271	16	10	277	10	10	322	16	10	352	17	10	307*)	33	51
hi	277	14	10	-	-	-	265	15	10	298	21	10	267	23	10	276	22	40

*) If MF is not included, $\bar{X}$ = 303 ms for fi and 315 ms for si

Table VI

Duration of the period including the oral closure and the preceding vowel VE (ms)

	HU			MF			FJ			LG			PA			Grand mean		
	$\bar{X}$	sd	N	$\bar{X}$	sd	N	$\bar{X}$	sd	N	$\bar{X}$	sd	N	$\bar{X}$	sd	N	$\bar{X}$	sd	N
pi	217	11	10	152	16	9	208	15	10	213	8	13	265	38	5	207	35	47
ti	200	19	10	152	13	10	178	17	10	186	9	10	227	19	10	188	30	50
ki	217	15	10	178	14	9	193	18	9	192	11	10	266	18	10	210	35	48
bi	234	13	10	199	20	6	239	13	10	265	6	10	289	14	10	249	31	46
di	223	15	11	190	20	10	223	18	9	245	10	10	274	11	9	230	32	49
gi	220	8	10	198	32	8	221	12	10	217	14	11	260	21	9	223	26	48

Table VII

Duration from the start of the preceding vowel to the onset of the glottal gesture VG (ms)

	HU			MF			FJ			LG			PA			Grand mean		
	$\bar{X}$	sd	N	$\bar{X}$	sd	N	$\bar{X}$	sd	N	$\bar{X}$	sd	N	$\bar{X}$	sd	N	$\bar{X}$	sd	N
pi	128	9	10	86	9	10	100	14	10	129	15	13	162	26	5	117	27	48
ti	123	12	10	83	8	11	104	14	10	143	14	12	164	22	10	123	32	53
ki	130	14	10	84	9	10	116	14	10	128	7	10	166	12	10	125	29	50
bi	130	9	10	94	15	7	143	19	7	146	6	10	188	18	11	144	33	45
di	120	13	11	85	14	9	142	16	9	147	13	10	187	13	10	137	36	49
gi	118	8	10	89	15	10	125	12	10	123	9	11	175	18	10	126	30	49
fi	103	7	10	66	8	10	85	12	10	133	20	10	156	12	10	108*)	35	50
si	109	10	11	70	9	10	87	8	10	133	11	10	149	12	10	109*)	31	52
hi	155	15	10	-	-	-	101	17	10	168	14	10	141	22	9	141	31	39

*) If MF is not included, $\bar{X}$ = 119 ms for fi and si.

Table VIII

Duration from the start of the preceding vowel to the maximum glottal opening VM (ms)

	HU			MF			FJ			LG			PA			Grand mean		
	$\bar{X}$	sd	N	$\bar{X}$	sd	N	$\bar{X}$	sd	N	$\bar{X}$	sd	N	$\bar{X}$	sd	N	$\bar{X}$	sd	N
pi	226	9	10	170	12	10	222	14	10	220	10	13	272	31	5	217	32	45
ti	217	13	10	176	8	7	217	15	10	244	9	12	263	24	10	227	31	46
ki	230	14	10	185	17	5	234	20	10	223	12	10	272	15	10	233	29	45
bi	170	10	10	159	25	7	185	10	6	212	6	10	227	18	11	194	30	44
di	168	11	11	141	12	9	206	12	9	212	11	10	231	12	10	192	34	49
gi	164	7	10	146	20	8	174	7	10	188	13	11	223	20	10	181	29	49
fi	183	7	10	161	10	10	204	15	10	221	13	10	243	11	10	202*)	31	50
si	190	11	11	168	10	10	199	8	10	229	11	10	240	15	10	205*)	28	51
hi	230	11	10	-	-	-	188	28	10	247	11	10	212	21	9	219	29	39

*) If MF is not included, $\bar{X}$ = 213 ms for fi and 214 ms for si.

Table IX

Duration from the onset of the glottal gesture to the onset of the obstruent GC_1 (ms)

	HU			MF			FJ			LG			PA			Grand mean		
	$\bar{X}$	sd	N	$\bar{X}$	sd	N	$\bar{X}$	sd	N	$\bar{X}$	sd	N	$\bar{X}$	sd	N	$\bar{X}$	sd	N
pi	-14	6	10	-7	6	10	1	6	10	-14	5	13	-10	7	5	-9	8	48
ti	-1	6	10	-3	10	11	-9	9	10	-19	8	12	-9	7	10	-8	10	53
ki	-7	5	10	-14	5	10	-9	7	10	-16	10	10	-10	6	10	-11	8	50
bi	-14	6	10	-5	8	7	-16	21	7	-11	6	10	-30	7	11	-16	13	45
di	-3	8	11	-2	7	9	-31	6	9	-8	7	10	-28	5	10	-14	14	49
gi	-1	5	10	-7	10	10	-20	6	10	-5	7	11	-19	3	10	-10	10	51
fi	18	9	10	32	3	10	33	5	10	2	11	10	18	8	10	20*)	14	50
si	20	7	11	27	5	10	24	9	10	6	6	10	18	6	10	19*)	10	51
hi	21	9	10	-	-	-	31	12	10	5	11	10	23	7	9	20	14	39

*) If MF is not included, $\bar{X}$ = 18 ms for fi and 17 ms for si.

Table X

Duration from the maximum glottal opening to explosion of the closure ME (ms)

	HU			MF			FJ			LG			PA			Grand mean		
	$\bar{X}$	sd	N	$\bar{X}$	sd	N	$\bar{X}$	sd	N	$\bar{X}$	sd	N	$\bar{X}$	sd	N	$\bar{X}$	sd	N
pi	-9	8	10	-18	5	10	-12	9	10	-6	6	13	-7	14	5	-10	7	48
ti	-16	6	10	-29	11	7	-39	7	10	-60	6	9	-37	7	10	-36	16	46
ki	-13	5	10	-11	14	5	-37	13	9	-31	8	10	-7	11	10	-20	16	44
bi	63	7	10	18	6	4	54	8	9	53	4	9	62	10	10	54	14	42
di	55	11	11	47	15	10	37	10	8	34	9	10	42	6	9	43	13	48
gi	56	7	10	53	20	6	47	8	10	29	11	11	46	6	9	45	14	46

Table XI

Duration of the time interval with vocal fold vibrations in relation to the onset of the obstruent C_1Z (ms)

	HU			MF			FJ			LG			PA			Grand mean		
	$\bar{X}$	sd	N	$\bar{X}$	sd	N	$\bar{X}$	sd	N	$\bar{X}$	sd	N	$\bar{X}$	sd	N	$\bar{X}$	sd	N
pi	42	6	10	29	4	10	55	5	10	55	5	13	52	4	5	47	12	48
ti	38	6	10	26	7	11	54	11	10	57	10	12	52	7	10	45	15	53
ki	34	6	10	29	4	10	47	6	10	53	7	10	50	4	10	42	11	50
bi	52	7	10	46	10	7	67	11	10	62	5	10	74	5	11	61	13	48
di	51	5	11	37	8	10	72	9	9	62	10	7	69	8	10	58	15	47
gi	39	4	10	35	5	10	65	6	10	59	9	11	60	4	10	51	13	51
fi	28	5	10	18	8	10	76	8	11	55	8	10	46	6	10	45	22	51
si	27	5	11	21	3	10	58	10	10	55	6	10	48	8	10	41	17	51

Table XII

Duration of the glottal abduction GM (ms)

	HU			MF			FJ			LG			PA			Grand mean		
	$\bar{X}$	sd	N	$\bar{X}$	sd	N	$\bar{X}$	sd	N	$\bar{X}$	sd	N	$\bar{X}$	sd	N	$\bar{X}$	sd	N
pi	98	5	10	85	12	10	122	11	10	92	11	13	110	12	5	100	17	46
ti	94	6	10	90	10	7	113	14	10	100	11	12	100	8	10	100	12	49
ki	99	5	10	104	7	5	118	11	10	95	8	10	106	7	10	104	11	45
bi	41	5	10	64	5	7	38	12	6	66	6	10	39	2	11	49	15	44
di	48	5	11	56	11	9	42	10	9	65	9	10	44	5	10	51	12	49
gi	46	9	10	61	16	8	49	5	10	65	7	11	48	6	10	54	12	49
fi	80	5	10	95	9	10	119	6	10	88	12	10	88	8	10	94	16	50
si	81	6	11	99	8	10	113	10	10	96	13	10	91	8	10	96	14	51
hi	75	8	10	81	13	8	88	17	10	79	10	10	71	7	9	79	13	47

Table XIII

Maximum amplitude A (mm)

	HU			MF			FJ			LG			PA			Grand mean		
	$\bar{X}$	sd	N	$\bar{X}$	sd	N	$\bar{X}$	sd	N	$\bar{X}$	sd	N	$\bar{X}$	sd	N	$\bar{X}$	sd	N
pi	19.0	1.5	10	21.8	5.8	10	31.6	3.4	10	18.7	2.6	7	15.2	1.0	5	22.2	6.6	42
ti	18.3	3.1	10	30.1	3.5	7	31.5	4.2	10	20.5	3.6	9	19.8	3.7	10	23.7	6.6	46
ki	20.8	3.4	10	32.2	4.9	5	33.0	2.8	10	21.2	2.6	7	19.7	2.8	10	24.9	6.7	42
bi	3.2	0.5	10	4.3	1.4	7	4.6	1.9	9	9.6	1.3	7	3.6	1.3	11	4.8	2.5	44
di	3.8	0.7	11	6.5	2.0	10	2.7	1.6	9	9.6	2.4	7	4.7	1.1	10	5.2	2.7	47
gi	4.8	0.7	10	7.2	2.2	9	5.7	2.4	10	12.8	1.9	8	6.3	1.3	10	7.1	3.2	47
fi	15.5	2.2	10	24.3	4.2	9	23.8	3.2	11	16.6	3.0	7	16.5	2.6	10	19.5	4.9	47
si	17.1	1.9	11	29.0	3.9	10	29.0	3.8	10	20.6	1.5	7	16.8	1.3	10	22.5	6.3	48
hi	8.3	2.2	10	8.9	1.4	9	17.7	2.3	10	6.1	1.8	7	12.3	1.3	10	11.0	4.4	46

Results of a multiple comparison procedure (the Scheffé method). The level of significance is indicated as follows: xxx = p<0.01, xx = p<0.05, x = p<0.1, o = p>0.1. The compared obstruents and obstruent categories are seen in the leftmost column. > means that the mean value averaged across speakers is higher in the obstruent or obstruent category to the left of the hyphen, < means that it is lower.

	acoustic parameters						acoustic/glottal parameters					glottal param.	
	I	II	III	IV	V	VI	VII	VIII	IX[1]	X[1]	XI	XII	XIII
	C_1C_2	C_1E	EC_2	VC_1	VC_2	VE	VG	VM	GC_1	ME	C_1Z	GM	A
p-b	>+++	<+++	>+++	<+++	>+++	<+++	<+++	>+++	>o	<+++	<+++	>+++	>+++
t-d	>+++	<+++	>+++	<o	>+++	<+++	<++	>+++	>o	<+++	<+++	>+++	>+++
k-g	>+++	<+++	>+++	<o	>+++	<o	<o	>+++	<o	<+++	<+++	>+++	>+++
ptk-bdg	>+++	<+++	>+++	<+++	>+++	<+++	<+++	>+++	>o	<+++	<+++	>+++	>+++
p-f	>+++			<+++	<o		>o	>+	<+++		>o	>o	>+++
t-s	>+++			<+++	>o		>+++	>+++	<+++		>o	>o	>o
pt-fs	>+++			<+++	<o		>+++	>+++	<+++		>o	>o	>+++
p-t	<o	>+++	<+++	<o	<+++	>+++	<o	<o	<o	>+++	>o	=o	<o
p-k	<++	>o	<+++	<o	<+++	<o	<o	<+++	>o	>+++	>o	<o	<+++
t-k	<o	<+++	>+++	<o	<o	<+++	<o	<o	>o	<+++	>o	<o	<o
b-d	>o	>+++	<+++	>o	>o	>++	>o	>o	<o	>+++	>o	<o	<o
b-g	<o	>++	<+++	>o	>o	>+++	>+++	>+++	<o	>+++	>++	<o	<++
d-g	<+++	<o	<++	>o	<o	>o	>o	>o	<o	<o	>++	<o	<o
f-s[2]	<+++			=o	<+++		<o	<o	<o		>o	<o	<+++
f-h	>+++			<+++	>+++		<+++	<o	=o			>+++	>+++
s-h	>+++			<+++	>+++		<+++	<o	>o			>+++	>+++

1) x-y $\left\{\begin{matrix}>\\<\end{matrix}\right\}$ means that C_1/E occurs $\left\{\begin{matrix}\text{later}\\\text{earlier}\end{matrix}\right\}$ in x than in y in relation to G/M.

2) With subject MF no data are available for h in parameters including C_1, C_2, G and M, and therefore she has been omitted from the significance tests for fricatives.

APPENDIX C

It is widely recognized that interpretation of the level of the glottogram in terms of glottal aperture is a crucial point in the glottographic method (see also Frøkjær-Jensen et al. 1971). If the glottographic signal was simply a function of the glottal aperture, the relation between these two variables should be positively linear, since the input-output characteristics of our glottographic set-up are linear except in the very lowest end of the voltage range (Hutters 1976). But the signal is influenced by external and internal disturbances in the transmission between the light source and the photo-transducer, which induce variation in the level of the glottogram that do not reflect variations in the size of the glottal aperture.

The influence from external factors such as coughing and swallowing can be reduced by attending to the subject's well-being - physically and psychologically - and by an appropriate external fixation of the fiberoptic cable. As regards the influence from the internal factors, i.e. from the speech conditions involved, it is normally recommended to have the test material comprise sounds that are produced in the oral part of the vocal tract only, since the signal is very sensitive to even slight retractions of the tongue root. But other articulatory events have to be taken into account such as vertical movements of the larynx and movements of the velum.

When fiberoptic stills and films are recorded synchronously with the glottographic signal the correlation coefficient is very often calculated in order to control the degree of linear positive correlation between the level of the glottogram and the size of the glottal aperture. Since data normally includes measurements in the whole range from small to large glottal apertures, the two variables will in general be highly correlated for pooled data. If, however, we look at single data points many exceptions from the overall trend can be observed, and the present material is no exception to this. Therefore, the correlation coefficient is not a very appropriate means to expose influences from sources of error. The slope of the regression line may serve better to reveal these influences, as mentioned by Andersen (1981). With regard to the present material, for instance, it seems that for some subjects the different obstruent categories are better described by their different regression lines.

It has to be considered, however, that the control procedure itself may also be influenced by sources of error. In most cases the distance between the vocal processes, which in general are clearly discernible, can be taken to represent the glottis aperture, but in fact the relative distance between the vocal processes is only approximately proportional to the area. Since the level of the glottogram ideally reflects the glottal area (provided that the photo-transducer does not selectively pick up light from some dominant part of the glottis, which was not the case in the present study), some exceptions to a simple, positively linear relationship should be

expected. A fluctuating distance between the glottis and the light source/fiberoptic object, which influences the glottographic signal and the size of the glottal aperture as it appears from the stills, should also be mentioned as a potential artifact.

In this context I want to suggest that the light should rather be picked up from the cartilagenous part of the glottis if we are interested in the glottal aperture as a function of the movements of the arytenoid cartilages, since this kind of glottographic signal probably is a better representation of the arytenoid movements. If so, another implication would be that the onset and offset of the glottal gesture in terms of these arytenoid movements can be unambiguously defined from the glottographic curves. Contrarily, with the light and the phototransducer positioned as in the present study - and in many others - these glottal events cannot be identified in the glottographic signal, since it indiscriminately reflects variations in glottal aperture whether they result from movements of the arytenoids or not. Therefore, the onset and offset of the gesture as identified from the gross movements of the glottographic trace do not necessarily reflect the physiologically well-defined gesture interval. Furthermore, voicing onset at the transition from an unvoiced obstruent to the following vowel may occur not only before but also after the offset of the falling slope of the glottogram. Since this different behaviour influences the appearance of the signal, it is difficult to provide a consistent delimitation of the offset even if we define the glottal gesture period in terms of the rise of the signal from the minimum level and the reattainment of this level. The problem relates to aspirated stops compared with fricatives and to comparisons within the two categories when the voicing conditions differ considerably at the transition to the following vowel. Also in case of aspirated stops versus unaspirated stops like the Danish ones, a comparison of their gesture offsets, identified in terms of the glottographic signal, does not seem very meaningful from a physiological point of view, since in the latter category the "gesture offset" simply equals onset of voicing. Therefore, no such offset point has been defined for the obstruents in the present study. Contrarily, the offset of voicing in single (intervocalic) obstruents normally never leads the onset of the rising amplitude reflecting the beginning of the opening-closing gesture, and thus a more consistent delimitation of the latter can be performed here, whatever its physiological interpretation might be.

Finally, a problem relating to the zero-line should be mentioned. The photo-glottographic signal has no physiologically well-defined zero-line due to the translucent vocal folds, which means that even with a closed glottis the light will be picked up whose intensity varies with the degree of compression and with the area of contact between the vocal folds. Incidentally, this not only influences the DC-level of the signal but also makes the interpretation of the signal in terms of glottal area somewhat dubious at small signal levels such as

those encountered during vocal fold vibration. In the present study the minimum level attained in the vowel preceding the test obstruent has been used as reference line. An arbitrary reference line might also have been used, but it seems inappropriate to add a constant representing DC-offset to the peak level measurements. The DC-level of the glottographic signal depends on the relation between the light source, the glottis, and the photo-transducer. Therefore, the degree of fluctuation of this level may - as a rule of thumb - indicate the influence from external sources of error during a given recording session. As regards the speech internal factors due to the articulatory behaviour in consonant production, these are naturally much less reflected in the DC-levels of the surrounding vowels. Needless to say, artifactual influences from the speech condition itself will decrease the more the phonetic content of the test material is restricted, but at the same time the number of phenomena that can be studied will also be reduced. Thus, interpretation of the glottogram amplitude in terms of glottal aperture is still a crucial problem in the application of the photo-glottographic method.