ARIPUC VOL. 8.

ANNUAL REPORT

of the
Institute of Phonetics
University of Copenhagen

COPENHAGEN 1974

ARIPUC VOL. 8.

ANNUAL REPORT

of the
Institute of Phonetics
University of Copenhagen

COPENHAGEN 1974

CONTENTS

PERSONNEL OF THE INSTITUTE OF PHONETICS 1973

Permanent Staff:

Eli Fischer-Jørgensen (professor, director of the Institute)

Lecturers:

Jørgen Rischel
Oluf M. Thorsen
Børge Frøkjær-Jensen
Hans Basbøll
Karen Landschultz (until August 1st)
Nina Thorsen
Hideo Mase (until April 1st)
Peter Holtse (permanently appointed since October 1st)
Birgit Hutters (until August 1st)

Technical Staff:

Carl Ludvigsen (M.Sc.) until March 1st
Mogens Møller (M.Sc.) from February 1st
Preben Dømler (B.Sc.) from April 1st
Svend-Erik Lystlund (technician)
Inger Østergaard (secretary) until April 1st
Aase Thiim (secretary) until April 1st
Else Parkmann (secretary) from April 1st

Part Time Teachers of General Phonetics:

Preben Andersen
Peter Molbæk Hansen
Ellen Pedersen
Niels Reinholt Petersen
Pia Riber Petersen

Temporarily appointed lecturers:

John Ohala (Berkeley)
Elizabeth Uldall (Edinburgh)

Furthermore several phoneticians from England, Sweden, Australia, Japan, India, and the United States of America have visited the laboratory.

PUBLICATIONS BY STAFF MEMBERS 1973:

Hans Basbøll: "A commentary on Hjelmslev's Outline of the Danish Expression System (II)", Acta Linguistica Hafniensia 14, p. 1-24.

Eli Fischer-Jørgensen: "Supplementary note to Hans Basbøll's commentary on Hjelmslev's Outline of the Danish Expression System", Acta Linguistica Hafniensia 14, p. 143-152.

KØBENHAVNS UNIVERSITET
OFFSET AFDELINGEN

LECTURES AND COURSES IN 1973

1. Elementary phonetics courses

One-semester courses (two hours a week) in elementary phonetics (intended for all students of foreign languages except French) were given by Preben Andersen, Peter Molbæk Hansen, Peter Holtse, Birgit Hutters, Ellen Pedersen, Niels Reinholt Petersen, Pia Riber Petersen, and Nina Thorsen.

There was one class in the spring semester, and 26 parallel classes in the autumn semester.

Courses in general and French phonetics for students of French (two hours a week in two semesters) were given through 1973 by Oluf M. Thorsen.

2. Practical training in sound perception and transcription

Courses for beginners as well as courses for more advanced students were given through 1973 by Jørgen Rischel, Nina Thorsen, Oluf M. Thorsen, and E. Uldall. (The courses which are based in part on tape recordings and in part on work with informants, form a cycle of three semesters with two hours a week.)

3. Instrumental phonetics

Courses for beginners as well as courses for more advanced students were given by Birgit Hutters and Peter Holtse in the spring semester (experimental acoustic phonetics and experimental physiological phonetics), and in the autumn semester by Nina Thorsen (registration of the intensity and fundamental frequency of speech).

4. Phonology

Jørgen Rischel, Hans Basbøll, and Eli Fischer-Jørgensen gave courses for beginners and advanced students. (The courses now form a cycle of two semesters with two hours a week. The contents being: problems in phonology and trends in phonological schools.)

5. Other courses

Eli Fischer-Jørgensen gave a course in German phonetics, held seminars on experimental phonetics, and gave a course in the interpretation of spectrograms and mingograph tracings.

Niels Reinholt Petersen gave a course in the physiology of the speech organs.

Oluf Thorsen gave a course in French phonetics.

Hans Basbøll gave a course in Danish phonology and phonetics.

Esther Dinsen (Institute of Applied and Mathematical Linguistics) gave a course in the theory and practice of the language laboratory.

John Ohala lectured on electromyography.

Carl Ludvigsen gave a course in mathematics, electronics, and statistics.

6. Seminars

The following seminars were held in 1973:

Dr. Annan (Leeds) lectured on the phonetics of some East African languages.

Steffen Heger presided at a discussion on the phonetic transcription of Danish.

Peter Holtse presented a study on categorial perception of vowels.

Björn Lindblom (Stockholm) gave an account of the activities at the phonetics institute in Stockholm.

Pia Riber Petersen presented a study on the Danish stød.

Eli Fischer-Jørgensen presided at a discussion on the relationship between generative phonology and other phonological schools.

We had a visit from the phonetics institute in Lund and exchanged reports on our research projects.

Professor Fujisaki (Tokyo) lectured on the measurement of jaw movements.

Eli Fischer-Jørgensen and John Ohala gave an account of their impressions from the symposium on auditory phonetics in Leningrad.

John Ohala reported on the congress on historical linguistics in Edinburgh.

John Ohala demonstrated and talked about the nasograph.

Anders Löfqvist (Lund) lectured on subglottal pressure during speech.

John Ohala gave a talk on a psycho-phonological test.

John Ohala presented a paper on an approach to sound change.

Steffen Heger presented his text-book on Danish phonetics.

7. Participation in congresses and lectures at other institutions by members of the staff

Eli Fischer-Jørgensen participated in the symposium on auditory analysis and perception of speech in Leningrad in August where she presented a paper on Perception of German and Danish vowels with special reference to the German lax vowels. Furthermore, Eli Fischer-Jørgensen was invited by the Alexander von

Humboldt Stiftung to visit the phonetics institute in Munich for one week in July.

Hans Basbøll visited the linguistics department of the University of Vincennes in Paris during the spring term.

Børge Frøkjær-Jensen visited the phoniatric clinics in Lund, Malmö and Brussels.

Mogens Møller and Jørgen Rischel visited the institute of linguistics in Edinburgh for one week with the purpose of working with the OLYMPUS fiberscope there.

Peter Holtse and Mogens Møller participated in the OS/8 seminar in Gothenburg.

Mogens Møller and Preben Dømler participated in the seminar on integrated semi-conductor circuits held by Texas Instruments.

Mogens Møller participated in a course in statistics given by the Danish Technical High School. He visited the Speech Transmission Laboratory at the Royal Technical High School in Stockholm. Mogens Møller furthermore took a course in PDP/8 hardware in Stockholm and participated in a seminar on software in Gothenburg.

Karen Landschultz, Mogens Møller, and Nina Thorsen participated in various courses at the pedagogical institute of the University of Copenhagen.

INSTRUMENTAL EQUIPMENT OF THE LABORATORY

The following is a list of the instruments that have been purchased or built since January 1st, 1973.[1]

1. Tape recorder

1 instrumentation recorder, Lyrec, type IR 86.

2. General-purpose electronic instrumentation

1 X-Y recorder, Hewlett & Packard, type 7044A.

3. Instrumentation for speech analysis

1 vocal chords fiberscope, Olympus, type VF.

4. Equipment for EDB

1 tape punch, GNT Automatic A/S, type 34.

1 real time clock, Digital, DK8-EP

1 a/d converter, Digital, AB8-EA.

---

1) For a complete list of the instrumental equipment of the laboratory, see the preceding volume (vol. 7) of ARIPUC.

ABBREVIATIONS EMPLOYED IN REFERENCES

AJPs. American Journal of Psychology

AL Acta Linguistica

ALH Acta Linguistica Hafniensia

ARIPUC Annual Report of the Institute of Phonetics, University of Copenhagen

Folia Ph. Folia Phoniatrica

FRJ For Roman Jakobson

F&S Form and Substance

Haskins SR Status Report on Speech Research, Haskins Laboratories

IJAL International Journal of American Linguistics

IPO IPO Annual Progress Report

JASA Journal of the Acoustical Society of America

JL Journal of Linguistics

JPh Journal of Phonetics

JSHD Journal of Speech and Hearing Disorders

JSHR Journal of Speech and Hearing Research

Lg. Language

Ling. Linguistics

LS Language and Speech

MIT QPR M.I.T. Quarterly Progress Report

NTTS Nordisk Tidsskrift for Tale og Stemme

Proc.Acoust. xx Proceedings of the xx International Congress on Acoustics

Proc.Ling. xx Proceedings of the xx International Congress of Linguistics

Proc.Phon. xx Proceedings of the xx International Congress of Phonetics Sciences

STL-QPSR Speech Transmission Laboratories, Quarterly Progress and Status Report

SL Studia Linguistica

SPE The Sound Pattern of English

TCLC Travaux du Cercle Linguistique de Copenhague

TCLP Travaux du Cercle Linguistique de Prague

UCLA Working Papers in Phonetics, University of California

Zs.f.Ph. Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung

THE INFLUENCE OF TONGUE HEIGHT ON THE PERCEPTION OF VOWEL DURATION IN DANISH

Niels Reinholt Petersen

1. Introduction

In numerous languages there is a tendency for high vowels to be shorter than low vowels, other things (quantitative category, environments, etc.) being equal. This tendency has been reported for Danish by Eli Fischer-Jørgensen (1955), for Swedish by C.-C. Elert (1964), for English by Peterson and Lehiste (1960) and by Holtse (1972), for Thai by Abramson (1962), and for Dutch by Nooteboom (1972).

Nooteboom also investigated the perception of vowel duration in a series of experiments in which the subjects adjusted the vowel duration of synthetic words according to a criterion of naturalness. In these experiments he found that the rules describing the effect on vowel duration of factors such as e.g. number of syllables in a word also described the effect on the durations preferred by his subjects in the perceptual tests. For instance, it was found in spoken words that the duration of a stressed first vowel in a word decreased as a function of the number of succeeding syllables in the word, and so did the preferred duration of that vowel in the perceptual tests. For the effect of tongue height on vowel duration, however, no such agreement between the production and perception of speech was found.

This is in disagreement with the results of a pilot experiment carried out by the present author. In that experiment listeners were asked to judge whether a vowel embedded in a word was phonemically short or long, the duration of the vowel being varied in lo ms steps from 80 to 200 ms. The vowels were Danish /i/ and /ε/. The identification functions showed a cross-over point from short to long /i/ at about 110 ms, and from short to long /ε/ at about 140 ms, i.e., the listeners wanted /ε/ to be longer than /i/ in order to judge it to be phonemically long. The listeners were also asked to compare /i/ and /ε/ (embedded in words) of varying duration. When /i/ and /ε/ were of equal physical duration, /i/ was perceived to be the longer, and when /ε/ was physically longer (about 20 ms) than /i/ (as is the case in natural speech), they were perceived to have the same duration. The results of this experiment were based on a very limited material, however, and cannot be considered entirely conclusive.

Thus the aim of the experiments reported below was to obtain data which could make it possible with more certainty to answer the question whether the perception of vowel duration is influenced by the tongue height[1] of the vowel perceived.

2. Stimuli and test procedure

2.1 Stimuli

The vowels chosen for the experiment were the Danish phonemes /i/, /i:/, /ε/, and /ε:/ (IPA [i] and [ε˔]). In order to avoid that the vowels were perceived as non-speech sounds they were embedded in the surroundings l-sə, giving meaningful Danish words for both long and short /i/ and /ε/, viz.: [lisə] lidse, 'lace', [li:sə] lise, 'relief' (or a personal name), ['lεsə] læsse, '(to) load', and [lε:sə] læse, '(to) read'.

---

1) By the term tongue height is meant the properties of the acoustic signal which convey information on that feature.

The words were synthesized on the parallel speech synthesizer of the Institute (Rischel 1969; Rischel and Lystlund 1972). The build-up of the synthetic words was based on spectrograms of the words spoken by four Danish speakers. Information about the formant frequencies and levels of the vowels [i] and [ɛ] (see table 1) was kindly provided by Mr. P. Holtse. These data corresponded to formant data that had given 100 per cent identification as /i/ and /ɛ/ in an identification test carried out by him (Holtse 1973). (There is only a slight difference between formant frequencies of short and long /i/ and /ɛ/ in normal spoken Danish.)

TABLE 1.

Formant frequencies and levels of the vowels [i] and [ɛ]. Formant frequencies in cps and levels in dB relative to the level of F1.

	i		ɛ	
	formant freq.	level	formant freq.	level
F1	226	0	416	0
F2	2326	-26	1970	-14
F3	3391	-19	2491	-18
F4	3800	-20	3750	-28
F5	4430	-26	4450	-26

The F_o contour was rising, beginning at 92 cps in the [l] and ending at 110 cps in the [ə]. In the vowel, however, F_o was kept constant at 100 cps; otherwise the variation of vowel duration would have caused a variation in the rate of F_o change during the vowel, which might influence the perception of vowel duration.

The vowels [i] and [ε] were built up by four 'acoustic segments'. (Dynamic control of connected speech on the synthesizer is obtained by means of a function generator producing varying or constant voltages in 20 successive steps - 'acoustic segments'. Each of these segments holds information on the parameter values and the transition time and can be varied in duration from normally 5 to 100 ms, see Rischel 1969; Rischel and Lystlund 1972.) The duration of one of the acoustic segments could be varied continuously by means of a remote control device which was to be used by the subjects for the adjustment of the vowel duration to meet certain criteria (see below). For this adjustment procedure it would have been most expedient if the subjects could cover the total range of expected durations in one sweep of the remote control. However, for the criteria employed in the tests, the expected adjustments varied from less than 100 ms to more than 250 ms. It turned out that this wide range could only be achieved at the expense of precision of adjustment and of linearity within the range. The widest range of continuous variation by means of the control device which gave an acceptable precision and linearity with our function generator was of the order of 90 ms (from 10 ms to 100 ms). Thus, in order to cover the total range of expected durations it did not suffice to vary one acoustic segment: it was necessary to select suitable durations of the other three acoustic segments of the vowel, depending on the type of stimulus and on the subject. The total duration of the parts of the vowel which were not affected by the remote control was thus set somewhere between 50 ms and 205 ms. If adjustments were to be made in the short vowel area, the three segments might, for instance, be given a joint duration of 50 ms by the experimenter, and the subject could then adjust vowels ranging from 60 to 150 ms. Similarly, if adjustments of longer durations were to be made, the three segments were lengthened, thereby shifting the range of the remote control device to cover adjustments of higher values of duration (e.g. from 175 to 265 ms with the duration of the three segments set at 165 ms). The highest range which could be obtained was 215 to 305 ms. The duration of the segment connected to the control device could be read by the experimenter from an electronic counter, and the total vowel duration was then the sum of that segment and the three segments set by the experimenter.[1]

2.2 Test procedure

The subjects' task was to adjust the vowel duration according to the following criteria by means of the remote control device:

1. Normal duration for phonemically short vowel.
2. Normal duration for phonemically long vowel.
3. Boundary between phonemically short and long vowel.

During the experimental sessions the subject was seated at the control device listening to the output from the synthesizer via ear phones (Sennheiser HD 414). The synthesizer was

---

1) The actual value of vowel duration may deviate from that specified by the function generator. It seems that there is an inherent source of error due to the fashion in which the amplitude is controlled: the point in time at which the sound is programmed to start, i.e. at which the voice source amplitude gate is activated to produce an increasing amplitude, is independent of the repetition rate of voice source pulses. Since the amplitude gate is placed before the formant filters, this means that the first pulse to produce an appreciable excitation of the formant filters (and hence a vowellike output) may occur with varying time-lags relative to the programmed onset, i.e. signals differing up to one period in duration (e.g. 10 ms at F_0 = 100 cps.) can be generated by exactly the same programming of the function generator. There is a similar possibility of error at the end of the synthesized vowel. - However, the results of perceptual experiments carried out at the present time suggest that the influence of these inaccuracies on the perception of duration is smaller than should be expected, but nothing conclusive can be said about the problem until investigations in progress have been completed.

set to repeat the programmed word once every two seconds, during which interval the subject could operate the control device to obtain a duration corresponding to the criterion given. When he found that he had adjusted the optimal duration the subject gave notice to the experimenter, who took down the reading of the electronic counter.

Because of the limitations of the apparatus mentioned above (section 2.1), every experimental session was initiated with a series of pilot adjustments to the criteria of that session. The purpose of these adjustments was to determine the best range of duration for the subject in question. This was done by presenting the subject with different ranges, e.g. 60 to 150 and 80 to 170 ms for the adjustment of normal short vowel durations, and choosing the one where his adjustments were well within the limits. This range was then kept for the succeeding adjustments to the criterion in question.

Adjustments to criteria 1 and 2 above were carried out in one experimental session, short and long vowels being adjusted alternately. Each subject adjusted 20 short and 20 long vowels of each quality.

In order to establish the perceptual phoneme boundary between short and long vowel, the subjects alternately started the adjustment from the lower limit of the range chosen, increasing the duration until he perceived a long vowel (i.e. perceived the word to be [li:sə] or [lɛ:sə], respectively), and from the higher limit of the range, decreasing the duration until he perceived a short vowel (i.e. perceived the word to be [lisə] or [lɛsə], respectively). Twenty 'short to long' and twenty 'long to short' adjustments were made by each subject for each vowel quality.

All [i]-adjustments were carried out first, and then, after a couple of weeks, the [ɛ]-adjustments were made.

Six subjects took part in the experiment. Five were university students of phonetics, and one (NT) was a member of the staff of the institute. All subjects made adjustments to criteria 1 and 2. Only four (BH, JJ, EBC, and NT) did criterion 3.

3. Results

Arithmetic means and standard deviations of the vowel durations adjusted are given in table 2, and the means are graphically displayed in fig. 1.

TABLE 2

Average adjustments ($\overline{X}$) and standard deviations (s) for each of the six subjects

		adjustments of normal durations				adjustments of boundaries			
		short vowel		long vowel		long to short		short to long	
		i	ε	i	ε	i	ε	i	ε
EBC	$\overline{X}$	75	116	163	196	97	141	99	143
	s	5.4	8.6	4.2	7.3	3.5	5.1	5.2	3.1
JJ	$\overline{X}$	84	108	148	208	95	126	106	156
	s	6.7	8.0	7.0	8.2	5.2	5.0	5.2	5.2
PA	$\overline{X}$	86	118	150	213				
	s	7.9	9.3	9.7	9.3				
BH	$\overline{X}$	86	109	151	216	98	138	96	138
	s	2.9	7.1	4.4	7.2	4.2	4.8	3.6	5.8
MJ	$\overline{X}$	89	110	182	226				
	s	6.0	11.0	10.1	11.8				
NT	$\overline{X}$	107	130	206	230	123	177	130	193
	s	8.4	5.5	8.7	10.9	5.1	7.9	5.6	4.7

grand mean diff. | 27 ms | 48 ms

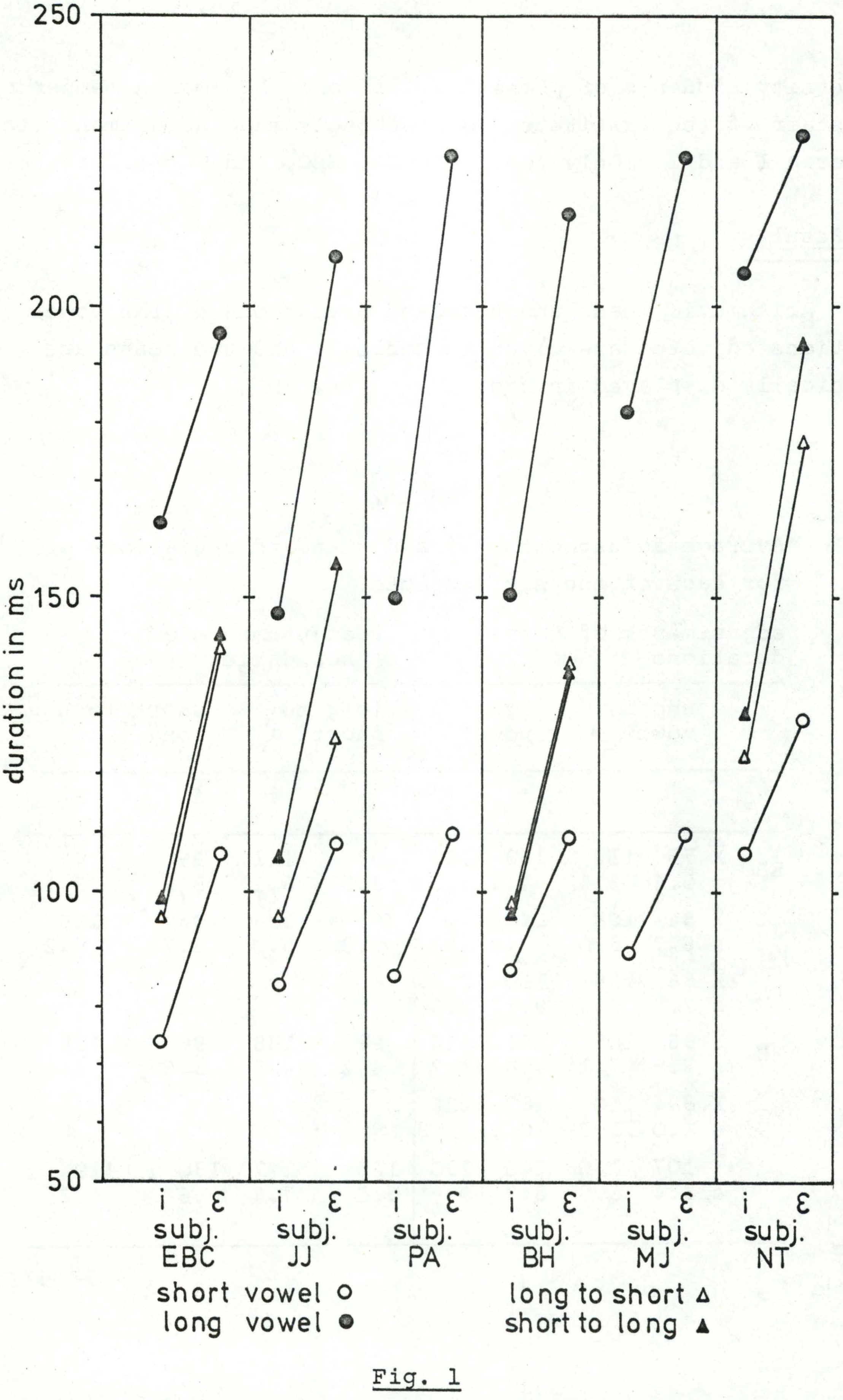


Fig. 1

Average adjustments of normal long and short vowels and phoneme boundaries for each of the six subjects.

From the data it appears to be quite evident that the subjects wanted [ɛ] to be longer than [i], when the vowels were adjusted in accordance with the same criterion. In all cases the difference between means was statistically significant at the 1 per cent level of significance.

Thus it seems reasonable to conclude that the perceptual duration of a vowel is influenced by the acoustic properties conveying information on the tongue height of that vowel, and, further, that the agreement between production and perception of vowel duration as found by Nooteboom for some influencing factors (see section 1 of this paper) may be extended to include tongue height as well.

References

Abramson, A.S. 1962: The vowels and tones of Standard Thai: Acoustical measurements and experiments (Bloomington Ind.: Indiana University)

Elert, C.-C. 1964: Phonologic studies of quantity in Swedish (Stockholm, Göteborg, Uppsala)

Fischer-Jørgensen, E. 1955: "Om vokallængde i dansk rigsmål", NTTS 15

Holtse, P. 1972: "Spectrographic analysis of English vowels", ARIPUC 6

Holtse, P. 1973: "Identification and discrimination of closely spaced synthetic vowels", ARIPUC 7

Nooteboom, S.G. 1972: Production and perception of vowel duration (dissertation)

Peterson, G.E. and I. Lehiste 1960: "Duration of syllable nuclei in English", JASA 32

Rischel, J. 1969: "Constructional work on a function generator for speech synthesis", ARIPUC 3

Rischel, J. and S.-E. Lystlund 1972: "A formant-coded speech synthesizer", ARIPUC 6

A MATHEMATICAL MODEL OF SPEECH AERODYNAMICS[1]

John J. Ohala

1. Introduction

A full understanding of the aerodynamic processes in speech would mean the ability to accurately predict the DC variations in air pressure and air flow in the vocal tract, including the subglottal cavities, given variations in the pulmonic force applied to the lungs and in the glottal and supraglottal air resistance. One of the problems in reaching this goal is that it is easier to sample and measure the dependent variables, air pressure and air flow, than it is the independent variables, pulmonic force and air resistance. In some cases indirect estimates of the air resistance can be obtained. Broad (1968), for example, derived effective mean glottal resistance, R_g, from simultaneous recordings of subglottal pressure, P_s, and transglottal air flow, U_g, via the relation, $R_g = P_s/U_g$.

Another approach to this problem is to construct a model of the aerodynamic system used in speech for which the time-varying values of pulmonic force and air resistance are guessed at and are used to derive the variations in air pressure and air flow (cf. Rothenberg 1968). We may have some confidence in the accuracy of our guesses if the derived pressure and flow values match those observed in real speech. I report here a preliminary attempt to devise such a model and to use it to explore certain controversial issues in phonetics.

---

1) Paper presented at Speech Communication Seminar, Stockholm, Aug. 1-3, 1974.

2. The issues

One of these issues is the relative contribution of the pulmonic and laryngeal systems in controlling fundamental frequency (F_o) of voice in speech. Recordings of P_s during speech frequently reveal it to be positively correlated with F_o (Ladefoged 1963, Lieberman 1967, Vanderslice 1967, Ohala 1970, Atkinson 1973). Since P_s can vary as a function of both the pulmonic force and glottal (and supraglottal) resistance, it is possible to attribute these P_s variations to either or both factors. Lieberman and Atkinson suggest that in certain circumstances the F_o variations are caused by the P_s variations which in turn are caused by variations in the pulmonic expiratory force. However, Isshiki (1969) and Ohala suggest that the subglottal pressure variation may be due in large part to variations in glottal resistance which would accompany the laryngeal muscles' action in varying F_o by changing the tension of the vocal cords.

Another issue surrounds the production of aspirated vs. unaspirated stops. Chomsky and Halle (1968), for reasons that are not entirely clear, suggest that aspirated stops, e.g., [p^h] and [b^h], are produced with heightened P_s in contrast to unaspirated stops such as [p] and [b] which would have normal P_s. It is implicit in their approach that they would regard this heightened P_s as a feature that is independent of (and thus not caused by) laryngeal features; therefore it could only be attributed to an increase in pulmonic force. One may guess that they thought the increased P_s necessary to account for the greater air flow accompanying aspirated stops. Ohala and Ohala (1972), however, sampled P_s during the speech of a Hindi speaker and found instances of heightened P_s during the closed portion of any stop, whether aspirated or not, thus showing that the heightened P_s was not a distinguishing characteristic of aspirated stops. They attributed these P_s peaks to the effects of increased oral resistance and a continued lung volume decrement during the closure. They also found markedly decreased P_s immediately after the release of the aspirated stops, but not after the unaspirated stops. They explained this lowered P_s as being due to lowered glottal resistance immediately after the release of the aspirated stops and this, in turn, would explain the high rate of air flow characteristic of these stops. (Halle and Stevens (1971) present a new analysis of stops and make no reference to heightened P_s as the distinguishing feature of aspirated stops, presumably indicating they have abandoned this feature. However, they cite no new evidence in support of this move.)

A final issue to consider is what special action, if any, is necessary to maintain voicing during voiced obstruents. Halle and Stevens (1967) suggest that a change in the vibratory pattern of the vocal cords equivalent to a decrease in glottal resistance plus the enlargement of the oral cavity are necessary for the maintenance of voicing during obstruents. The aerodynamic model to be reported here may be able to shed light on this and the preceding issues.

3. The model

The aerodynamic processes in speech were modeled mathematically with the model being implemented on a small general-purpose digital computer. The basic elements of the model are shown in figure 1.

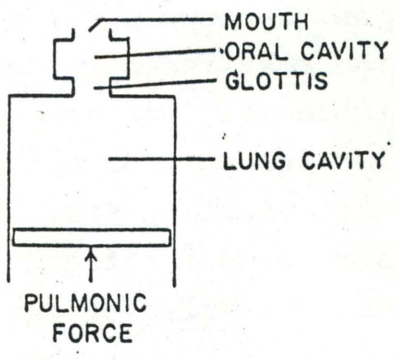
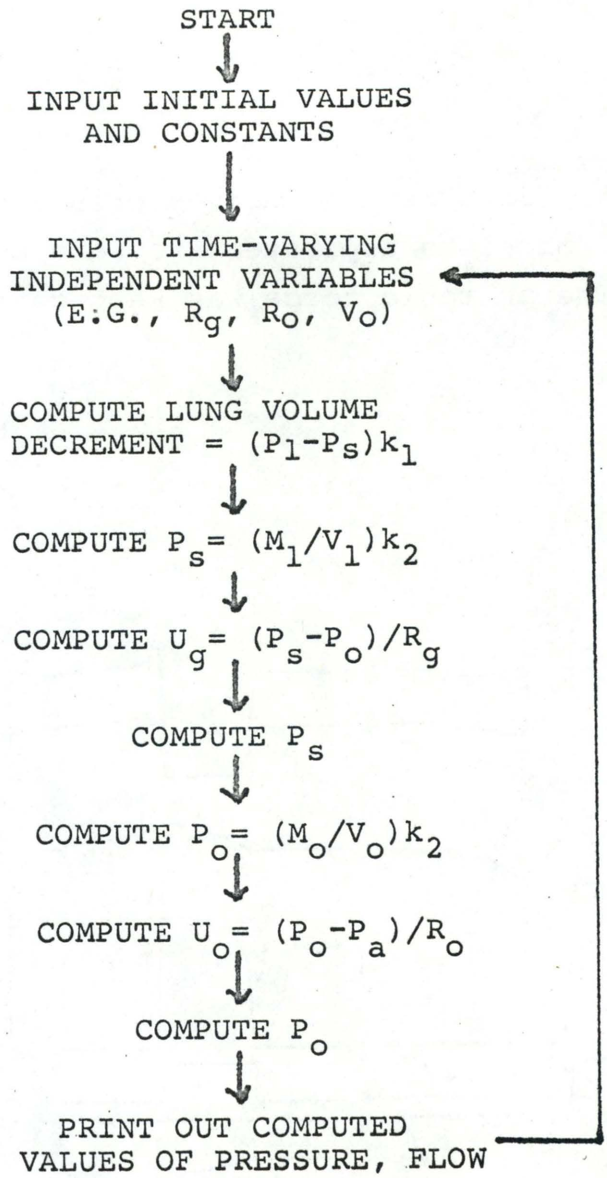


Figure 1.

Two connected air cavities, the lung cavity and the oral cavity, are defined by their respective volumes and air masses. Between the oral cavity and the "outside" there is an aperture, the mouth. Between the lung cavity and the oral cavity there is another aperture, the glottis. Both of these apertures are defined by their respective resistances. The volume of the lung cavity may decrease as the pulmonic force moves the chest wall and causes a lung volume decrement. The volume of the oral cavity is allowed to increase during voiced stop closures. The pressure inside a cavity is derived by the relation: pressure = air mass/volume. The mass of air inside a cavity varies as air flows in or out of it. The air flow through an aperture is a function of the pressure drop across the aperture and the resistance of the aperture: air flow = pressure drop/resistance.

To simulate the aerodynamic processes during a given sample of speech, the following are specified: the initial lung volume and air mass, the oral volume and initial air mass, pulmonic force, glottal and oral resistance, and various constants. The following are computed for each time increment: lung volume decrement, subglottal pressure, oral pressure, glottal air flow, and oral air flow. The program that performs these computations is given in flow chart form in figure 2. One pass through the program derives the relevant values for one short time increment. Then, on the next pass, the calculations are performed again with the most recently derived values serving as input for the computation of the values for the next time increment. These calculations must be performed for a sufficiently small time increment or the system may oscillate wildly. I found it necessary to use a time increment of .45 ms or less. Thus, for the 400 ms samples of speech to be discussed below, 880 passes through the program were required.



P_a = atmospheric pressure
P_l = pulmonic force
P_s = subglottal pressure
P_o = oral pressure
R_g = glottal resistance
R_o = oral resistance
U_g = glottal air flow
U_o = oral air flow
V_l = lung volume
V_o = oral volume
M_l = lung air mass
M_o = oral air mass
k_1, k_2 = constants

Figure 2

Flow chart of computer program simulating speech aerodynamics.

4. Results

Figures 3a - 3b show the derived pressure and air flow functions for a voiceless aspirated stop and a voiced stop, respectively. The pulmonic force was kept constant at 11 cm

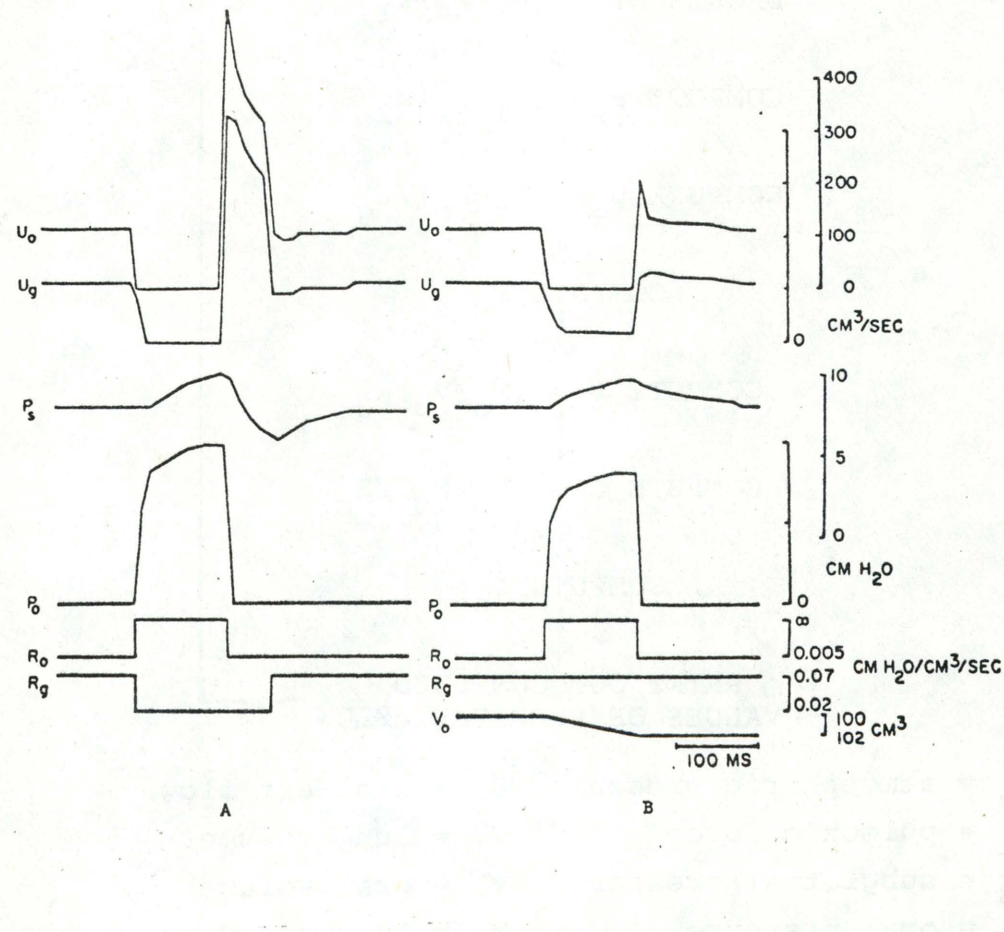


Figure 3
Output of aerodynamic model. A: intervocalic voiceless aspirated stop. B: intervocalic voiced stop. Parameters, from top: oral air flow, glottal air flow, subglottal pressure, oral pressure, oral resistance, glottal resistance, oral cavity volume.

H_2O (over atmospheric pressure) in both cases; only glottal resistance, oral resistance, and oral volume were allowed to vary as shown. (The step-function changes in resistance are unrealistic, of course, but these abrupt variations do not seem to give unusual results.) The P_s functions agree well with those obtained for real speech such as those in figure 4.

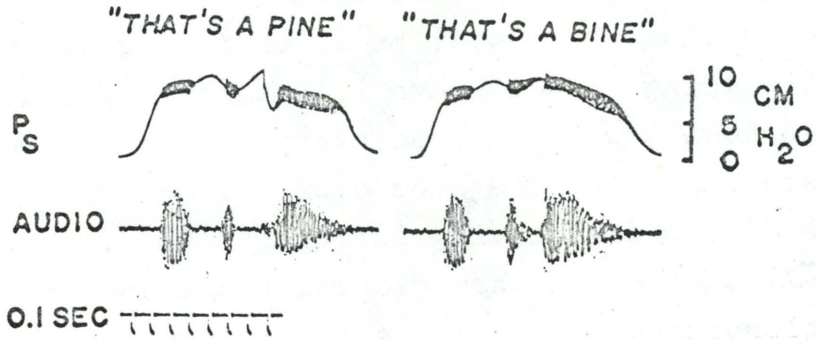


Figure 4

Subglottal pressure and microphone signal sampled during two utterances spoken by an adult male speaker of English.

(The P_s curves in figure 4 were sampled via a tracheal needle during the utterances "that's a pine" [ðætsə'pʰajn], on the left, and "that's a bine" [ðætsə'bajn], on the right, as spoken by a male adult speaker of American English. See Ohala 1970.) As was noted by Ohala and Ohala (1972) for a Hindi speaker, there are momentary increases in P_s during the stop closures - - in this case the rise is greater for the voiceless stop. These are a direct result of the increased oral resistance during the stop closure which causes an increase in oral pressure and a consequent decrease in the transglottal pressure drop which in turn causes diminished glottal flow. The P_s then approaches the pulmonic force asymptotically. For 50 ms after the release of the voiceless aspirated stop the glottal resistance remains low. Consequently the air flow out of the lung cavity is very high, with the result that the subglottal pressure is momentarily lowered. Again, this agrees well with the real speech data (cf. figure 4 and the findings of Ohala and Ohala 1972).

It is clear from many other studies that the oral pressure for voiced stops is significantly lower than that for voiceless stops (Fischer-Jørgensen 1972 and references therein). This is necessary in order that a positive transglottal pressure drop be maintained so that there will be a continuing glottal air flow and thus voicing. To achieve this with this model one or both of the following would be necessary: a) an increase in glottal resistance during the stop closure, or b) an increase in the volume of the oral cavity during the stop. Halle and Stevens' (1967) suggestion that glottal resistance be lowered during stop closures would make the problem worse: oral pressure would reach that of subglottal pressure even more rapidly and voicing would cease. As there is no evidence (that I know of) for (a), but there is evidence for (b) (Ewan and Krones 1972), I allowed the oral cavity to gradually increase by 2 cm^3 during the 100 ms stop closure. This allowed oral pressure to be less than subglottal pressure and thus yielded continued air flow and voicing throughout the stop closure.

Another interesting aspect of these curves is the fact that after "normal" glottal resistance is restored following the release of the voiceless aspirated stop, the subglottal pressure takes a considerable time to return to the normal "equilibrium" pressure proper to the given pulmonic force and glottal resistance. Likewise, after the release of the voiced stop, the subglottal pressure is maintained at a higher-than-normal level for some 90 ms into the following vowel. This pattern is also observed in the real speech samples in figure 4. Thus the average subglottal pressure is lower on vowels following voiceless aspirated stops and higher on vowels following voiced stops. Given the known causal correlation between subglottal pressure and the intensity of voice (Ladefoged and McKinney 1963), this accounts for the commonly observed higher intensity of vowels following voiced stops and the lower intensity of vowels following voiceless aspirated stops (House and Fairbanks 1953, Lehiste and Peterson 1959).

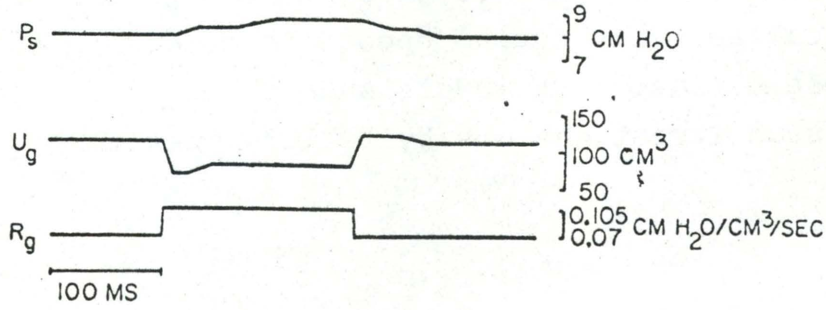


Figure 5.

Output of aerodynamic model showing effects of increased glottal resistance on subglottal pressure (top) and glottal air flow (second line).

Figure 5 presents the results of varying only glottal resistance and leaving the pulmonic force constant as before. As can be seen, when glottal resistance is increased by only 50%, subglottal pressure increases, although it takes a relatively long time to reach the equilibrium pressure. Air flow decreases in this case. A momentary increase in subglottal pressure could also be obtained by a momentary increase in the pulmonic force, leaving the glottal resistance unchanged. In this case, however, the air flow would also increase. The situation that actually prevails in speech during stressed or emphasized syllables (where brief increases of subglottal pressure have been observed) is probably that where there is primarily just a momentary increase in glottal resistance, since it is quite commonly the case that air flow on stressed syllables is less than that on unstressed syllables (Klatt, Stevens, and Mead 1968, Broad 1968). This, then, tends to support the notion that control of F_0 in speech is performed primarily by the larynx and not by the pulmonic system. The pulmonic system, in fact, can be assumed to be largely passive during speech except for providing a relatively constant force to the lungs.

Of course, more physiological investigation of pulmonic and laryngeal activity during speech is needed in order to verify these claims. But models such as the one reported here aid us in such investigations by telling what things to look for.

Acknowledgements

This research was conducted in the Institute of Phonetics, University of Copenhagen, and in the Phonology Laboratory, Department of Linguistics, University of California, Berkeley. I thank my colleagues in both laboratories, especially Robert Krones, for helpful comments. This work was supported in part by the National Science Foundation.

References

Atkinson, J.E. 1973: "Aspects of intonation in speech: implications from an experimental study of fundamental frequency", University of Connecticut dissertation.

Broad, D.J. 1968: "Some physiological parameters in prosodic description", Speech Communication Research Laboratories Monograph No. 3.(Santa Barbara).

Chomsky, N. and Halle, M. 1968: The sound pattern of English. (New York).

Ewan, W.G. and Krones, R. 1972: "A study of larynx height in speech using the thyro-umbrometer", JASA 53, p. 345.

Fischer-Jørgensen, E. 1972: "PTK et BDG français en position intervocalique accentuée" In: A. Valdman, (ed.) Papers in linguistics and phonetics to the memory of Pierre Delattre, p. 143-200.

Halle, M. and Stevens, K.N. 1967: "On the mechanism of glottal vibration for vowels and consonants", MIT QPR 85, p. 267-270.

Halle, M. and Stevens, K.N. 1971: "A note on laryngeal features", MIT QPR 101, p. 198-213

House, A.S. and Fairbanks, G. 1953: "The influence of consonant environment upon the secondary acoustical characteristics of vowels", JASA 25, p. 105-113.

Isshiki, N. 1969: "Remarks on mechanisms for vocal intensity variations", JSHR 12, p. 669-672.

Klatt, D.H., Stevens, K.N., and Mead, J. 1968: "Studies of articulatory activity and airflow during speech", In: A. Bouhuys (ed.) Sound Production in man. Annals of the New York Academy of Sciences 155, p. 42-54.

Ladefoged, P. 1963: "Some physiological parameters in speech", LS 6, p. 109-119.

Ladefoged, P. and N. McKinney 1963: "Loudness, sound pressure and sub-glottal pressure in speech", JASA 35, p. 453-460.

Lehiste, I. and G.E. Peterson 1959: "Vowel amplitude and phonemic stress in American English", JASA 31, p. 428-460.

Lieberman, P. 1967: Intonation, perception, and language, MIT Press

Ohala, J. 1970: "Aspects of the control and production of speech", UCLA 15.

Ohala, M. and J. Ohala 1972: "The problem of aspiration in Hindi phonetics", Annual Bulletin, Research Institute of Logopedics and Phoniatrics, University of Tokyo 6, p. 39-46, and Project on Linguistics Analysis Reports, Berkeley 16, p. 63-70.

Rothenberg, M. 1968: "The breath-stream dynamics of simple-released-plosive production", Bibliotheca Phonetica 6.

Vanderslice, R. 1967: "Larynx vs. lungs: cricothyrometer data refuting some recent claims concerning intonation and archetypality", UCLA 7, p. 69-79.

WORD-MEDIAL SP, ST, SK-CLUSTERS AND SYLLABIFICATION IN ENGLISH

Niels Davidsen-Nielsen

1. Introduction

In this paper[1] I wish to apply the delimitation method of "extrinsic allophones" (i.e. programmed as opposed to automatic, or "intrinsic", allophones) to English words with medial /sp, st, sk/ followed by a stressed vowel, either directly, as in despise, establish, confiscation, or with in intervening /r/, /l/, /w/, or /j/, as in estrange, disclaim, disqualify, dispute. It may be noticed that in words of this type syllabification cannot be accounted for phonologically by the principle of permitted initial and final clusters (cf. Hjelmslev 1936, p. 52), since the point of syllable separation according to this principle can be either between the sibilant and the stop (e.g. /iṣ-treindʒ, dis-paiz/), before the sibilant (/i-streindʒ, di-spaiz/), or, with a few exceptions, after the stop (/ist-reindʒ, disp-aiz/). Phonetically, however, it seems quite feasible to syllabify these words. It is well-known that /p, t, k/ in both British and American English are aspirated initially in monosyllabic words, and it is by now generally held that they are unaspirated post-initially, i.e. after initial /s/, in such words (cf. Davidsen-Nielsen 1969). Since /p, t, k/ are thus characteristically manifested as aspirated (voiceless) stops initially and as unaspirated (voiceless) stops post-initially, the interpretation naturally presents itself that the occurrence

1) This paper is an abbreviated version of Davidsen-Nielsen 1974. The phonetic experiments were carried out at the Institute of Phonetics, University of Copenhagen. I am grateful to Eli Fischer-Jørgensen and Jørgen Rischel for thorough and constructive criticism and to Svend-Erik Lystlund for technical assistance.

of an aspirated stop after s medially in a polysyllabic English word indicates syllable separation between the sibilant and the stop, and that the occurrence of an unaspirated stop in this environment indicates that the syllable border is not positioned between the sibilant and the stop but precedes s. It seems safe to exclude the possibility of syllable separation after the stop since a glottal catch never reinforces the stressed vowels in words of this type, i.e. pronunciations like [disp?aiz, rist?ɔ:] are not found.

2. The material investigated

English words with medial sp, st, sk-clusters followed by a stressed vowel may be classified in the following way:

(I) Words in which a prefix ending in -s immediately precedes the stressed syllable:

- (a) dis- (e.g. dispose, indiscriminate)
- (b) ex- (e.g. explain, inexpressible)
- (c) mis- (e.g. misprint)
- (d) sus- (e.g. suspend)
- (e) trans- (e.g. transplant)

(II) Words in which one of the following combinations or single graphemes (by far the majority of which are prefixes) immediately precedes the stressed syllable:
a-, ab-, anti-, be-, circum-, con-, de-, down-, e-, en-, fore-, hyper-, in-, inter-, intro-, non-, ob-, out-, over- per-, pre-, pro-, re-, retro-, sub-, super-, under-, un-, up-
(Examples: asparagus, abstain, antistrophic, bestir,

circumscribe, *conspicuous*, *despise*, *downstairs*, *establish*, *ensconce*, *forestall*, *hyper-space*, *inspect*, *intersperse*, *introspective*, *non-stop*, *obscure*, *outstanding*, *overstay*, *perspire*, *prescriptive*, *proscribe*, *restrain*, *retrospective*, *substantial*, *superstition*, *underscore*, *unspeakable*, *upstanding*)

(III) Proper names and words derived from proper names (e.g. *Aristophanes*, *Bostonian*)

(IV) Compounds (e.g. *cross-purpose*)

(V) Other words (e.g. *historian*, *trustee*)

The following words were now selected for recording (see Davidsen-Nielsen 1974 as regards the inclusion of the three words in parentheses):

(disburse)	miscalculate	perspire
disclaim	mistake	prescriptive
discomfit	mis-take ('take badly')	proscribe
discourteous	mistime	prospective
discover	mistook	prosperity
discrepancy	unmistakable	respectable
discussed		resplendent
(disdain)	suspense	responsible
(disgust)	sustain	restorative
disperse		retrospective
dispute	transcription	substantial
disqualify	transparent	unsteady
distain	transpire	
distemper		Australia
distinguish	ascribe	Shakespearean

distress	askance	
disturb	aspire	aristocracy
	astonish	austerity
exclaim	abstain	confiscation
excrescence	bespangle	dyspepsia
excursion	conspire	dexterity
expansion	constabulary	fastidious
expect	constituency	frustration
expediency	despise	gastronomy
expensive	despondence	gestation
experience	destruction	hysteria
extensive	especial	illustration
extenuate	establish	ministerial
exterior	estate	monstrosity
extinguish	estrange	
extrinsic	inspire	
	instalment	
	instinctive	
	interstellar	
	non-stop	
	overstep	

In order to provide a standard of comparison for the results obtained with word-medial stops, I furthermore selected the following words:

I	II	III
spear	pier	beer
spat	pat	bat
steam	team	deem
sty	tie	dye
scold	cold	gold
score	core	gore

I	II	III
spray	pray	bray
spume	pews	abuse
splint	plinth	blink
strain	train	drain
strip	trip	drip
stewed	tube	dude
scrape	crepe	grape
screw	crew	grew
skew	queue	gules
squad	quad	Gwen
	class	glass

3. Recording, instrumentation, and speakers

For various complicating reasons, which are explained in Davidsen-Nielsen 1974, the above words were recorded in two sessions. In the first recording session 22 native speakers of British and American English recorded the words with medial /sp, st, sk/ three times along with the words with initial and post-initial stops directly followed by a vowel (the former words were recorded in isolation after having been pseudo-randomized and the latter in rhythmically identical carrier sentences of the type A spear from the army). In the second session the words with initial and post-initial stops followed by a liquid or semivowel and (once again) the words with initial and post-initial stops directly followed by a vowel were recorded three times by 7 of the above 22 speakers (the former in rhythmically identical carrier sentences of the type A scrape on the elbow and the latter in isolation). At both sessions the informants were instructed to speak naturally and leisurely, and the tape speed was 19 cm/s. For each speaker one of the three recordings from the second session (two in the case of isolated words with (post-)initial stops immediately followed by a vowel and all three in the case of words with (post-)initial stops followed by a liquid or semivowel) were then analyzed by feeding the signal from the tape recorder into a pitch meter and an intensity meter and by using a mingograph as registering apparatus. On the mingograms there were two intensity curves (logarithmically and linearly registered respectively), one duplex oscillogram, and one pitch curve. On the three curves and on the oscillogram it was possible precisely to delimit the stops and their internal phases and thus to measure duration. The measurements comprised 2024 words with medial obstruent clusters, 714 words (462 from the first session and 252 from the second) with initial and post-initial stops immediately followed by a vowel, and 672 words with initial and post-initial stops followed by a liquid or semivowel.

4. Results

4.1 Initial and post-initial stops immediately followed by a vowel

On measuring the duration of the release stage in the 462 words with initial and post-initial stops recorded in the first session it appeared that the average duration of this element was 7·2 cs in the case of /p, t, k/, 2·4 cs in the case of /(s)p, (s)t, (s)k/, and 2·3 cs in the case of /b, d, g/. As was to be expected, furthermore, the place of articulation turned out to be relevant to the release stage, the duration of this phase increasing with growing retraction. The average figures obtained for labials, alveolars, and velars were as follows:

p-	6·4 cs	t-	7·6 cs	k-	8·1 cs
(s)p-	1·7 cs	(s)t-	2·6 cs	(s)k-	3·0 cs
b-	1·6 cs	d-	2·3 cs	g-	2·9 cs

On the basis of these findings I conclude that the stops after initial s are unaspirated along with /b-, d-, g-/, whereas /p-, t-, k-/ are aspirated. Considering the clarity of the results and the relatively large number of speakers I furthermore venture the generalization that unaspirated stops after initial s constitute the norm in both British and American English.

As for results obtained with words of this type in the second recording session, see Davidsen-Nielsen 1974, p. 24.

4.2 Medial stops immediately followed by a vowel

In the 1562 words where the medial stop is followed directly by a stressed vowel the duration of the release stage was measured. The results were then compared with those obtained for initial /p, t, k/ and /(s)p, (s)t, (s)k/ at the first recording session in order to determine whether the word-medial stop displayed initial manifestation (aspiration) or post-initial manifestation (non-aspiration).

In the following table of results (based on the first recording and only supplemented with figures from the second or third recording where borderline cases are involved) the figures which are considered to represent initial, aspirated manifestation have been underlined. See Davidsen-Nielsen 1974, (p. 26) as regards "disambiguation" of 84 borderline cases (i.e. 3 cs in the case of labials, 4 cs in the case of alveolars, and 5 cs in the case of velars); in nine remaining borderline cases a question mark has been added in the table.

TABLE 1

		American speakers											British speakers										
		JKD	JE	JOC	KOC	ACL	WW	PH	JH	KE	NW	JHP	IN	MCH	LC	MH	MC	DH	NS	GC	AC	JS	JB
1	(disburse)	2	1	2	1	2	1	2	2	2	2	2	2	1	2	3	2	2	2	2	1	1	2
2	discomfit	7	9	7	5?	6	7	6	3	6	8	6	5?	4	6	7	4	3	5?	8	6	3	2
3	discourteous	7	8	7	3	6	9	6	4	7	8	6	5?	3	7	7	4	3	6	6	4	4	4
4	discover	3	9	2	3	3	4	3	4	3	4	3	2	2	3	6	4	2	3	4	4	3	2
5	discussed	2	3	2	3	4	4	3	3	3	3	3	2	2	3	4	4	3	3	6	4	3	2
6	(disdain)	2	2	2	2	3	3	2	2	3	3	2	2	2	2	3	2	2	2	2	2	2	2
7	(disgust)	3	4	2	3	4	3	2	2	2	3	3	1	2	2	2	4	3	2	4	3	3	2
8	disperse	2	2	2	1	2	2	2	2	4	2	2	2	2	2	6	2	2	2	2	2	1	1
9	distain	2	2	2	2	2	2	2	2	2	2	2	7	3	4	5	4	7	2	2	9	2	2
10	distemper	7	8	6	6	5	5	3	5	5	6	5	5	2	6	6	4	3	3	3	3	3	3
11	distinguish	3	3	3	2	2	3	2	3	4	2	3	2	3	3	4	3	3	3	3	3	2	3
12	disturb	3	3	2	2	3	3	2	3	4	3	2	2	2	3	3	4	2	2	3	3	2	2
13	excursion	3	4	2	2	4	3	3	2	3	3	3	2	3	2	7	3	3	3	7	3	3	2
14	expansion	1	2	2	2	2	2	2	2	1	2	2	2	2	1	2	2	2	2	2	2	1	1
15	expect	2	1	2	1	2	2	2	2	1	1	2	2	2	1	2	2	2	2	2	2	1	1
16	expediency	1	2	2	1	2	2	1	2	2	2	2	2	2	1	5	2	1	2	2	2	1	2
17	expensive	1	1	2	1	2	2	2	1	2	1	2	1	2	1	2	2	1	2	2	2	2	1
18	experience	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2	2	1	2
19	extensive	3	1	2	3	3	2	3	3	3	2	2	2	3	2	5	3	3	2	5	3	3	2
20	extenuate	2	2	3	2	3	2	2	3	3	3	3	2	2	3	6	5	3	2	4?	3	3	2
21	exterior	3	3	2	2	5	2	6	3	3	7	3	5	2	4	8	3	3	3	5	3	2	3
22	extinguish	3	3	2	3	3	2	3	3	4	3	3	2	2	3	5	3	3	3	3	3	2	2
23	miscalculate	8	6	7	6	7	9	6	7	6	7	6	5?	4	7	7	8	7	6	6	6	6	7
24	mistake	3	3	2	3	2	5	3	2	4	2	2	2	2	4	4	6	3	2	3	3	2	3
25	mis-take	10	8	6	2	5	6	6	2	6	5	5	5	4?	10	7	6	6	7	6	7	5	6
26	mistime	7	7	7	6	6	9	8	6	7	6	5	6	5	9	11	7	7	7	6	8	6	8
27	mistook	7	9	6	5	3	9	3	5	7	7	5	2	2	4	5	4	5	7	3	5	2	2
28	unmistakable	3	3	3	2	3	2	2	2	3	2	2	2	3	3	3	5	2	2	3	3	3	3
29	suspense	2	1	2	1	2	2	1	1	2	1	1	2	2	2	2	2	1	2	2	2	1	1
30	sustain	3	2	2	3	3	2	3	3	3	3	2	2	3	2	3	4	2	1	3	3	2	2
31	transparent	1	4	7	4	4	10	4	4	5	1	4	4	2	2	5	2	6	5	2	2	2	2
32	transpire	2	1	1	1	2	2	2	2	5	1	2	2	2	1	2	4	2	2	2	2	1	2
33	askance	2	4	3	2	3	3	2	2	3	3	2	2	3	3	2	3	3	3	4	4	3	2
34	aspire	2	2	2	1	2	1	1	2	2	1	2	2	2	2	2	2	1	2	2	2	1	1
35	astonish	2	3	2	2	3	2	2	2	3	2	2	2	3	2	3	4	2	2	3	3	2	2
36	abstain	3	3	2	3	2	2	3	2	2	3	3	3	3	3	2	4	2	2	2	3	2	2

TABLE 1

(continued)

		American speakers											British speakers										
		JKD	JE	JOC	KOC	ACL	WW	PH	JH	KE	NW	JHP	IN	MCH	LC	MH	MC	DH	NS	GC	AC	JS	JB
37	bespangle	1	2	2	2	2	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	1	1
38	bestow	3	3	3	2	3	2	2	2	4	3	2	2	2	2	2	2	2	2	2	2	2	2
39	conspire	1	2	1	2	2	1	2	1	2	1	1	2	1	2	2	2	2	2	2	2	1	2
40	constabulary	2	2	3	2	3	2	3	2	3	2	2	3	2	2	3	4	2	2	3	3	2	2
41	constituency	2	3	2	2	2	2	2	3	4	3	3	3	3	2	3	4	3	2	5	3	2	3
42	despise	2	2	1	2	2	1	1	2	2	2	2	2	2	2	2	2	1	1	2	1	1	2
43	despondence	1	2	2	1	2	2	1	2	2	2	2	2	1	1	2	2	1	1	2	2	2	1
44	especial	2	2	1	1	2	2	2	2	2	1	1	1	2	1	1	2	1	2	2	2	1	1
45	establish	3	3	2	2	3	2	2	2	4	2	2	2	2	3	2	4	2	2	3	2	2	2
46	estate	2	3	2	2	3	2	2	2	3	2	3	2	2	3	3	3	2	3	3	2	2	2
47	inspire	2	2	1	1	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	1	1
48	instalment	3	3	2	2	2	2	2	2	3	3	3	2	3	2	3	4	2	2	3	2	3	2
49	instinctive	2	3	3	2	3	2	2	3	3	3	3	2	3	3	4	2	3	3	2	2	2	2
50	interstellar	2	3	2	2	2	3	2	3	3	3	3	2	2	3	3	4	2	2	3	3	3	3
51	non-stop	2	2	2	2	2	2	2	2	2	2	2	2	3	3	2	4	3	1	3	2	2	2
52	overstep	2	2	2	2	2	2	3	2	3	3	3	2	2	3	3	4	2	1	3	2	3	3
53	perspire	1	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2	1	2	2	2	1	1
54	prospective	1	1	1	1	2	1	2	1	1	1	1	2	2	1	2	2	1	2	2	2	2	2
55	prosperity	2	2	2	1	2	2	1	2	5	2	1	1	2	2	2	2	2	2	2	2	1	2
56	respectable	1	2	2	1	2	1	2	2	1	1	1	2	1	2	2	2	2	1	2	2	1	1
57	responsible	2	2	2	1	1	1	2	2	1	2	1	2	1	2	2	2	2	2	2	2	1	1
58	restorative	3	3	2	2	3	2	2	3	3	2	2	2	3	2	2	4	2	-	5	3	2	3
59	retrospective	2	1	2	1	2	2	2	2	2	1	1	2	2	2	2	2	1	1	2	1	1	1
60	substantial	2	3	3	2	3	2	2	2	2	3	3	2	2	3	3	4	2	2	3	3	2	2
61	unsteady	3	3	2	2	3	3	2	2	4	2	2	2	2	2	3	4	2	2	3	3	2	3
62	Shakespearean	1	2	2	1	2	2	1	2	2	2	2	2	2	2	5	4	1	2	2	2	1	2
63	aristocracy	2	3	3	2	3	2	2	2	4	3	2	2	2	3	3	4	2	3	4?	3	2	2
64	austerity	2	2	3	3	2	3	3	2	4	3	2	3	2	3	2	3	3	2	5	3	2	3
65	confiscation	3	4	3	3	4	3	3	4	3	6	3	3	3	3	4	4	3	3	4	3	3	2
66	dexterity	3	3	2	2	6	3	2	3	3	3	2	4?	2	4	9	4	2	2	3	3	3	1
67	dyspepsia	4	4	6	4	4	2	2	1	4	5	2	2	2	4	4	2	2	2	2	2	1	1
68	fastidious	3	3	2	2	2	3	3	3	4	3	3	2	3	1	6	4	3	3	5	5	2	3
69	gestation	3	5	2	3	5	3	3	3	3	2	3	2	2	4	3	4	3	2	3	3	2	3
70	hysteria	3	3	3	2	3	3	3	3	3	3	3	3	3	4	4	4	3	3	5	3	3	3
71	ministerial	2	6	3	2	2	2	3	3	3	3	3	3	3	2	4	6	3	2	5	3	2	3

4.3 Initial and post-initial stops followed by liquid or semivowel

On measuring the duration of the voiceless stretch of speech after the explosion in the words with initial and post-initial stops followed by a liquid or semivowel it turned out that the average duration of this phase was 9·0 cs in the case of /p, t, k/, 3·7 cs in the case of /(s)p, (s)t, (s)k/, and 3·2 cs in the case of /b, d, g/. The onset of vocal cord vibration was thus considerably delayed (20-35%) as compared with words in which the (post-)initial stop is directly followed by a vowel.

The difference between the initial and post-initial stops in pairs were as follows:

p(r)-	8.1 cs	(s)p(r)-	2·5 cs
t(r)-	9·4 cs	(s)t(r)-	4·4 cs
k(r)-	9·5 cs	(s)k(r)-	3·8 cs
p(j)-	8·8 cs	(s)p(j)-	3·0 cs
t(j)-	9·7 cs	(s)t(j)-	4·3 cs
k(j)-	10·0 cs	(s)k(j)-	4·2 cs
p(l)-	7·3 cs	(s)p(l)-	2·6 cs
k(l)-	8·2 cs		
k(w)-	8·8 cs	(s)k(w)-	4·0 cs

These results support the generally held assumption that /r, l, w, j/ in English are partially or completely devoiced after initial, stressed /p, t, k/ (aspiration manifested as devoicing of following liquid or semivowel). They furthermore show that /r, l, w, j/ are partially devoiced after initial, stressed /sp, st, sk/. However, partial devoicing also takes place after initial, stressed /b, d, g/, cf. the following average durations: br: 2·8, dr: 3·9, gr: 3·3, bj: 2·3, dj: 4·1, gj: 3·6, bl: 2·1, gl: 2·7, gw: 4·0.

4.4 Medial stops followed by liquid or semivowel

In the 462 recorded words in which a liquid or semivowel intervenes between the medial stop and the stressed vowel the duration of the release stage was measured. In order to determine whether the stops in these words displayed initial or post-initial manifestation, the results were then compared with those obtained at the second recording session for initial and post-initial stops followed by /r, l, w, j/.

In the following table of results (based on the first recording and only supplemented with figures from the two other recordings as far as borderline cases are concerned) the figures representing initial, aspirated manifestation are underlined. See Davidsen-Nielsen 1974, p. 31 for "disambiguation" of borderline cases (5 cs in the case of /spr, spl, spj, skl/, 6 cs in the case of /skr, skw/, 7 cs in the case of /str, stj, skj/); in two remaining borderline cases a question mark has been added in the table.

TABLE 2

		American speakers											British speakers										
		JKD	JE	JOC	KOC	ACL	WW	PH	JH	KE	NW	JHP	IN	MCH	LC	MH	MC	DH	NS	GC	AC	JS	JB
1	disclaim	6	12	8	7	7	11	7	3	6	9	7	6	2	7	9	9	6	8	7	10	3	4
2	discrepancy	4	4	4	3	4	2	3	3	3	7	3	5	3	5	5	3	2	4	5	7	4	3
3	dispute	1	4	3	3	3	7	4	2	4	7	7	2	2	3	3	2	6	4	6	7	3	2
4	disqualify	10	8	8	5	7	11	7	7	7	9	5	8	3	5	7	5	3	7	8	8	4	5
5	distress	3	5	5	5	4	5	3	4	4	6	5	3	3	4	8	6	4	5	8	4	4	4
6	exclaim	3	7	6	3	4	9	4	3	7	7	3	6	3	7	10	3	3	5?	9	6	4	2
7	excrescence	4	10	-	4	5	10	4	3	3	5	4	7	3	5	7	5	4	4	5	9	5	7
8	excusing	2	6	3	2	5	5	4	2	4	4	4	2	2	5	9	4	4	4	8	4	3	4
9	extrinsic	5	6	3	5	6	5	4	4	5	5	5	9	4	6	9	5	5	5	5	8	3	3
10	transcription	2	3	4	3	4	4	4	3	4	5	3	3	3	5	3	3	4	4	7	4	4	3
11	ascribe	3	4	3	3	3	4	4	4	3	5	3	3	2	3	3	4	3	5	8	4	3	3
12	destruction	4	5	3	3	3	4	2	3	4	5	5	3	3	4	3	4	5	3	5	4	4	5
13	estrange	3	4	3	2	4	4	5	4	3	5	4	5	4	4	5	5	4	4	6	6	5	4
14	prescriptive	3	4	3	4	4	2	3	4	3	5	2	3	3	4	3	3	5	4	5	4	4	3
15	proscribe	3	7	3	3	4	3	3	5	4	6?	3	3	3	5	3	4	4	5	7	5	3	3
16	resplendent	2	3	3	2	2	2	2	3	2	2	2	2	3	2	3	3	2	2	2	2	2	2
17	Australia	3	4	3	3	4	4	4	4	5	6	5	3	4	5	5	6	6	4	6	6	5	5
18	frustration	4	4	4	3	3	5	2	4	3	4	4	4	4	5	3	6	4	4	6	6	4	5
19	gastronomy	3	5	4	2	4	4	4	4	3	6	4	3	4	5	4	5	3	4	6	5	4	4
20	illustration	3	5	5	3	5	4	4	3	4	5	3	3	4	5	6	5	5	4	5	5	4	4
21	monstrosity	2	5.	4	3	4	4	4	3	5	4	4	4	5	4	5	4	4	3	5	5	4	4

5. Principles of syllabification in the words under investigation

The overall picture which emerges as regards syllabification in words with medial sp, st, sk followed by a stressed vowel, either directly or with an intervening liquid or semivowel, is that separation between the /s/ and the stop constitutes the exception rather than the rule. In only 10 out of the 89 words investigated did a majority of speakers aspirate the stop: mistime (22 speakers), miscalculate (20), mis-take (= "take badly", 19), disclaim (18), disqualify (14), discomfit (13), discourteous (13), distemper (13), transparent (13), mistook (13). Syllable division between the sibilant and the stop was furthermore characteristic of more than a few isolated speakers in the following words: exclaim (10 speakers), dyspepsia (9), exterior (6), excrescence (6), dispute (6).

Two striking features are shared by these words: they begin with a prefix ending in -s (or, in the case of dyspepsia, with a learned Greek borrowing ending in -s), and they contain an internal morpheme boundary which is so clear that it is presumably intuitively transparent to the speakers. Now in Basbøll 1972, p. 194 it is tentatively proposed that the syllable boundary in polysyllabic Danish words coincides with an intuitively transparent morpheme boundary in so far as the principle of permissible initial and final consonant clusters is not thereby violated, and my results seem to confirm this rule for English. The unfused character of the above words is most evident in miscalculate, mistime, mis-take, disclaim, disqualify and discourteous (the items with most consistently aspirated stops), where removal of the prefixes leaves behind words in their own right, and where the pejorative and reversative meanings of the prefixes seem obvious. In dispute, excrescence, transparent the existence of lexical items like putative, compute repute(d)/ crescent, crescendo, increase/ apparent as well as the relatively clear negative and locative ("out of", "through") semantic content of the prefixes may explain an awareness of internal boundary, although these cases are less easily accounted for. The composite nature of exclaim is supported by words like claim, disclaim, proclaim, reclaim, that of exterior by the antonym interior, and that of distemper by the relatively clear reversative sense of the prefix and the existence of the word temper. As regards mistook, it is claimed by Kenyon & Knott 1953, p. 282 that the syllable separation after the /s/ found in this form only, and not in mistake, mistaken, unmistakable, is due to the fact that the past form is less familiar. That a high degree of frequency should be conducive to the fusion of word elements seems quite probable, compare, for example, the reduction of the final vowel of milkman with the full vowel of chess-man in British English. Furthermore mistook is perhaps most commonly used in the sense "erroneously take A for B", e.g. I mistook him for the mayor, where the composite nature of the word is most obvious. Relatively low frequency of occurrence might explain the aspiration of the stops in dyspepsia and discomfit, together with a fairly clear negative/reversative semantic content of the prefix in the latter word and (possibly) with the existence of words like dyslalia and dyslexia.

Generalizing from my results and the principle that intuitively transparent morpheme boundaries coincide with syllable boundaries it can be predicted that syllable separation between the /s/ and the stop in English words of the type under investigation will be found in a small minority of cases, characteristic examples being disconsolate, displease, distaste (class Ia), expatriate (Ib), misprint, misquote (Ic), Fitzpatrick (III), cross-question (IV), and jurisprudence (V). Most words syllabified in this way begin with _dis-_, but there will also be a goodly number of words with _ex-_ and _mis-_ (actually all words of the last type except _mistake(n)_, _(un)mistakable_ can be expected to be syllabified in this way).

In a large majority of cases, then, syllable separation before the /s/ can be predicted, thus in the majority of words belonging to class I and III, in by far the greater number of words belonging to class V, and in practically all words belonging to class II.

If we examine the results obtained for each individual speaker, the homogeneity of the American material is conspicuous. One speaker, it is true, has aspiration of the stop in only six words, but he spoke fairly rapidly (JH). That speed of delivery could be conducive to the fusion of word elements does not seem implausible, cf. that NW, the American informant who spoke most slowly and deliberately, had 15 instances of aspiration. For British English the results are more heterogeneous. Three speakers have strikingly few cases of aspiration: MCH (only in _mistime_), JC (_miscalculate_, _mis-take_, _mistime_), and JB (miscalculate, mis-take, mistime, excrescence). On the other hand, two speakers, MH and GC, aspirated the stops in an exceptionally large number of words. Although the overall picture is the same in British as in American English, we are possibly on somewhat less sure ground in this type of English and therefore have to be more cautious in our generalizations.

References

Basbøll, H. 1972: "Some conditioning phonological factors for the pronunciation of short vowels in Danish with special reference to syllabification", _ARIPUC_ 6, p. 185-210

Davidsen-Nielsen, N. 1969: "English stops after initial /s/", _English Studies_ 50, 4, p. 321-339

Davidsen-Nielsen, N. 1974: "Syllabification in English words with medial _sp_, _st_, _sk_", _JPh_ 2, 1, p. 15-45

Hjelmslev, L. 1936: "On the principles of phonematics", _Proc. Phon._ 2, p. 49-54

Kenyon, J.S. and T.A. Knott 1953: _A pronouncing dictionary of American English_ (Springfield, Mass.)

THE PHONOLOGICAL SYLLABLE WITH SPECIAL REFERENCE TO DANISH

Hans Basbøll

1. Introduction

The present paper[1] is divided into three main parts. In the first of these (section 2) it will be argued that the syllable must be recognized as an important linguistic unit in any adequate phonology of Standard Danish. In section 3 I shall discuss the status of the syllable in a generative phonology, including such questions as the functions of the syllable in phonological rules, phonetic vs. phonological syllable, and the principles of syllabification. In section 4, finally, the nature of the syllable will be discussed in more general terms, and a distinction between what I call "hierarchical" and "cross-classificatory" distinctive features will be proposed. It will be shown that the hierarchy of phonological features can explain nearly all order relations among Danish consonants. These considerations have bearing on what the feature [syllabic] may actually mean.

Although the language material discussed in the present paper is taken almost exclusively from Standard Danish, the conclusions arrived at in sections 3 and 4 seem to be relevant to general phonological theory. The paper is thus not only meant to be a contribution to the phonological description of one language, viz. Standard Danish; it is intended to be a contribution to phonological theory as well.

---

1) This is an enlarged version of a paper entitled "Stavelsen i dansk fonologi" ('The syllable in Danish phonology') which I read at the Meeting of Nordic Linguists at Kungälv, March 30 - 31, 1974. I am indebted to Niels Davidsen-Nielsen for stylistic suggestions.

2. Arguments to the effect that any adequate phonology of Danish must include the syllable as a phonological unit

2.1. Allophonic variation of short /o/ and /a/

Since I have discussed the allophonic variation of the short /o/ and /a/ phonemes in Danish and their dependency upon syllabification elsewhere, I shall here only give a brief summary of what appear to be the main facts (see further Basbøll 1972a, p. 187-190).

The vowels [u:, u], [o:, ɔ] and [ɔ:, ʌ] have generally been taken to be manifestations of /u:, u/, /o:, o/ and /ɔ:, ɔ/ respectively (e.g. Ege 1965). However, the vowel [o] also occurs, partly as shortened /o:/ (in positions where the other long vowels are shortened too), partly in some foreign words like <u>foto</u> [fótˢo]. The vowel [ɔ] also appears in foreign words, e.g. <u>tundra</u>, <u>centrum</u> [tˢɔ́ndʁɑ, sɛ́ntʁ̥ɔm]. It has been pointed out (Basbøll 1969, p. 44) that the short vowels [o] and [ɔ] both occur posttonally in complementary distribution, [o] occurring in open syllables and [ɔ] in closed ones. This principle generalizes to all occurrences of [o] and [ɔ] not derived from long vowels. Such a principle of course presupposes a syllabification, and this syllabification turns out to be identical with the one which is presupposed by the other phonological rules examined here, and which is stated explicitly in section 3.4.1 below. (The distribution of short [o] and [ɔ] in French is in part due to similar principles, e.g. <u>sot</u>, <u>sotte</u> [so, sɔt].)

The short /a/ phoneme in Danish is pronounced as a back vowel in the environment of /r/, as a front vowel ([a]) before zero and dentals, and as a mid or back vowel ([ɑ̟] or [ɑ]) before velars and (in most language usages) labials. The fact that the first vowel of words like Amerika, akademi, tapir is pronounced [a] as opposed to the first vowel of e.g. amfiteater, Absalon, akkeleje suggests that this manifestation rule applies with the syllable rather than the word as its domain. The syllabification presupposed for this purpose is identical to that presupposed for the prediction of the manifestation of /o/ and of the consonant gradation phenomena mentioned below in section 2.2 (according to these principles a single intervocalic consonant goes to the preceding syllable if the following vowel is shwa, but to the following syllable if its vowel is a "full vowel", i.e. non-shwa, see section 3.4.1 below).

Particularly suggestive of an explanation in terms of syllables are alternative pronunciation patterns like the following. Amerikaner is either pronounced [ameʁikʰǽ:ʔnʌ, ameɒ̂ikʰǽ:ʔnʌ], where the first /a/ is followed by /m/, which again is followed by the full vowel /e/, i.e. /a/ occurs in an open syllable and is thus pronounced [a]; or it is pronounced [ɑmʁikʰǽ:ʔnʌ], where the second vowel is dropped (possibly via a reduction to shwa), and consequently the first /a/ occurs in a closed syllable ending in a labial, and it is thus pronounced [ɑ] (or [α]). Forms of the stem abbed like abbed, abbeder, abbedisse can be pronounced as [ábeð, ábe(:)ðʌ, abedísə], where the second vowel is the full vowel /e/ (which can even be long in abbeder), and /a/ thus occurs in an open syllable and is pronounced [a]; or it can be pronounced as [ɑ́bəð, ɑ́bəðʌ ɑbədísə], where the second vowel reduces to shwa, and the /b/ therefore belongs to the first syllable with the effect of retracting the vowel /a/.

2.2. Consonant gradation[1]

The term "Consonant gradation" is used here in accordance with Rischel 1970a. It covers a number of morphological alternations among which are the following (alternation (i) is mentioned by Uldall 1936, (ii) by Rischel 1970a, (iii) by Basbøll 1972a, (iv) by Hjelmslev 1951, and (v) by Rischel 1969):[2]

(i) Word-final unaspirated plosive alternates with aspirated plosive before a stressed vowel. Among the examples are a number of derived verbs in -ere, e.g. galopere, vattere, lakere [galop^{h}é:?ʌ, vat^{s}é:?ʌ, lak^{h}é:?ʌ], derived from galop, vat, lak [galʌ̂b, vad, lɑg]. Note particularly the last example where the /a/ is pronounced [ɑ] before the word-final and hence syllable-final /k/, whereas in lakere /k/ belongs to the following syllable, /a/ thus being pronounced [a].

---

1) For reasons of space, it is impossible to treat this very complicated subject in any detail here. I have discussed the "Consonant gradation"-phenomena extensively in my ditto'ed notes (in Danish) Konsonanter I-II (87 p., unfinished). (These notes, which will be continued, also include a chapter (46 p.) on the diphthongs.)

2) When I use formulations like "word-final X alternates with Y before Z" it does not imply that all instances of word-final X alternate with Y before Z, nor does it mean, of course, that all occurrences of Y before Z alternate with word-final X (notice that I do not use the word "alternates" in a "process" sense: the only implication here is that X before word boundary and Y before Z are in a relation of alternation, cf. Rischel 1974, p. 320ff). E.g. in (i) below it is said that word-final [b d g] alternate with [p^{h} t^{s} k^{h}] before stressed vowel, but alternations like klaustrofob-klaustrofobi [klɑu̯sdʁofó:?b, -fobí:?] show that some instances of word-final [b] alternate with [b] and not [p^{h}] before stressed vowels. And in (iv) below there are examples like dyr, dyrisk [dyɒ̯?/dy:?ɒ̯, dý:?ʁisg/dý:?ɒ̯isg/dýɒ̯?isg] which exhibit no alternation [ɒ̯] ∾ [ɒ̯d] (where the latter should occur before the derivative ending -isk). In most of the other cases, the alternations are general in the sense that the mentioned word-final sound types do not alternate with other sound types in the said context than those mentioned. It should be added that alternations (i)-(iii) and (v) are limited to "learned" derivations, cf. section 3.4.1 below.

(ii) Word-final voiced non-labial continuant consonant alternates with unaspirated plosive before stressed vowel. Examples are foreign words like _perfidi_, _pædagogik_ [p^{h}æɒ̯fidíːʔ, p^{h}ɛdagogíg], derived from _perfid_, _pædagog_ [p^{h}æɒ̯fíːʔð, p^{h}ɛdagóːʔɣ].

(iii) Word-final [ŋ] alternates with [ŋg] before a stressed vowel. There is at least one example of this alternation, viz. (_di_-, _mono_-, _poly_- etc.)-_ftongere_ [(di)ft^{s}ʌŋgéːʔʌ], derived from (_di_- etc.)-_ftong_ [(di)ft^{s}ʌ́ŋ].

(iv) Word-final [nʔ, lʔ, ɒ̯ʔ] alternate with [nd, ld, ɒ̯d] before derivative endings like -_ig_, _isk_. Examples are _mandig_, _skyldig_, _jordisk_ [mándi, sgýldi, jóɒ̯disg], derived from [manʔ, sgylʔ, joɒ̯ʔ/joːʔɒ̯] (in words with -_rd_, conservative usages have a long vowel in the underived word but the corresponding short one in the derived form).

(v) Word-final [ɒ̯] alternates with [ʁ] before stressed vowels. Examples are _kontorist_, _professorat_ [kɔnt^{s}oʁísd, pʁ̥ofɛsoʁɑ́ːʔd], derived from _kontor_, _professor_ [kɔnt^{s}óɒ̯ʔ, pʁ̥ofɛ́sʌ].

Now it is very interesting that all these alternations can be subsumed under a single principle, viz. the well-known one of _consonant weakening in syllable final position_. According to this principle the phonemes /p t k d g r/ are manifested as [p^h t^s k^h d g ʁ] in syllable initial position and as [b d g ð ɣ ɒ̯] in syllable final position (where [ɣ] is in younger standards substituted by [i̯], [u̯] or zero). It should be added that in utterance-final position, i.e. before pause, any final consonant can be followed by an [h]-sound, and that /t/ can be manifested [t^s], i.e. affricated (whereas final [b g] followed by [h] are not discernable from [p^h k^h]). It should also be mentioned that syllable-initial /p t k/ in word-internal position before an unstressed vowel can be pronounced [b d g], especially in non-distinct pronunciations, that [ɣ], but not [g], is dropped after nasals, and that [ð] is dropped after sonorants. (There are certain complications in accounting for the syllable-final pronunciations of the labial obstruents /b v/, which may also be weakened as compared with the syllable-initial manifestations [b v], but these cannot be treated here for reasons of space; the reader is referred to the notes referred to in footnote 1 on p. 42 for further information.)

The above-mentioned principle accounts not only for a great number of morphological alternations (as (i)-(v) above), but also explains (in a weak sense) quite a few distributional gaps like the following: [ð ɣ ŋ] occur only before #, consonants and shwa, never before full vowels; /h/ only occurs before full vowels, etc.

As already mentioned, it is, for reasons of space, impossible in this paper to fully discuss or even survey the very complicated phenomena connected with "Consonant gradation". However, it is clear that the value of the above-mentioned principle depends on whether there can be given explicit (and not unnatural) principles of the location of syllable boundaries that can account for all of the above-mentioned phenomena without giving rise to complications elsewhere in the phonology. This will be attempted in section 3.4.1 below.

By way of conclusion of this section it should be mentioned that there exists one explicit proposal for the delimitation of the consonant manifestations here called "syllable initial" and "syllable final" which does not include the syllable as a theoretical unit, viz. Rischel 1970a, p. 462. In order to show that the syllable is a <u>necessary prerequisite</u> for correctly accounting for the mentioned consonant phenomena, we must therefore demonstrate that Rischel's proposal is unsatisfactory. Rischel gives the following "working definition" for two positions, viz. "strong" and "weak", which "here will be taken as essentially synonymous with (syllable-)initial and (syllable-)final position":

"A consonant is in strong position if it fulfils the following two requirements: (1) it is preceded by juncture (morpheme border) or by a segment that is (phonemically) voiced; (2) it is followed by a full vowel (i.e., not shwa) with or without an intervening voiced consonant but without an intervening juncture. Examples are: [g] in [gli:ðə] 'glide', [b, g] in [lom'bæ:go] 'lumbago'.

Otherwise a consonant is in weak position. Examples are: [ɣ, ð] in [tɔ:ɣəðə] 'foggy' (plur.), [ð, g] in [jø:ðisg] 'jewish' (the latter word has a morpheme border between [ð] and [i]), and [ð] in [feðmə] 'fatness'."

It follows from Rischel's definition that a consonant is in weak position if it fulfils at least one of the following three requirements: (i) it is preceded by a (phonemically) voiceless segment belonging to the same morpheme; (ii) it is followed by a juncture; (iii) it is followed by a shwa, with a possible intervening consonant. (The alternations mentioned at the outset of this section show, in agreement with Rischel's note on p. 464, that "juncture" cannot simply mean "morpheme border" as the definition suggests, see section 3.4.1 below.)

In conservative varieties of Standard Danish there are certain counter-examples to implication (iii) of Rischel's definition. E.g. words like <u>kæntre</u>, <u>ændre</u> need not rhyme, and <u>verden</u>, <u>værten</u> can be distinguished: <u>kæntre</u> and <u>værten</u> can be pronounced with *unvoiced* /r/ ([kɛntʁ̥ə, væʁ̥tən]), which is never possible in *ændre*, *verden*, and in some usages, where the possibility of unvoiced /r/ in one or both of these examples does not exist, a distinction of length can be made in the non-syllabic sonorant so that /n, r/ etc. can be shorter before (written) *p*, *t*, *k* than before *b*, *d*, *g* (see Fischer-Jørgensen 1973, p. 146f).

Of course it can be argued that the proposed definition should only cover (younger) standards where these distinctions are not made, but I find it highly significant that *the only way* to account for these distinctions by means of "strong vs. weak position" is to claim that the obstruents in question are in *strong* and not weak position as they would be according to Rischel's definition; and this agrees well with the fact that [ð] is excluded in such positions, and that there is also a possible contrast between [p] and [b] (e.g. *jambe*, *lampe* [jɑm(.)bə, lɑmbə(lɑmpə)]. If Rischel's definition should be amended to cover these facts, it becomes even more complicated and unnatural. This brings me to my other point.

Rischel is, of course, well aware of the striking arbitrariness of his definition as shown e.g. by his term "working definition". It seems clear, to me at least, that if the [b] of [lom'bæ:go] is in strong position, it is not *because* it follows a phonemically voiced segment, viz. [m]; rather the voicedness (or better: sonority) of the preceding segment is a factor influencing the location of the syllable border, i.e. [b] is "strong" *because* it is syllable initial.

2.3. The stød

The stød has traditionally been considered to be a phonologically suprasegmental entity characterizing a syllable, although its place has generally been indicated as phonetically concurrent with a vowel or a consonant. Among the arguments that can be given for this phonological analysis are the following (cf. Basbøll 1972b, p. 6f): unlike most segments the stød can be lost under certain accentual and grammatical conditions, its place within the syllable is completely predictable (there is at most one stød which falls "on" the vowel if this is phonemically long, otherwise "on" the following consonant which in this situation is a sonorant), and the stød can be moved from a sonorant consonant to the following sonorant consonant without ever changing the identity of the word (Hjelmslev's (1951, p. 17) example jar'l = jarl').

However, the notational conventions used in what might be termed orthodox generative phonology (as codified in SPE) give no way to represent the stød as a characteristic of the syllable. Instead, the stød has been considered a feature characterizing the segment where it was normally said to occur phonetically (e.g. Austin 1971). Let us briefly consider some consequences of this latter stød-notation, contrasting it with the more traditional prosodic stød-description, which may be rephrased like this: The stød characterizes certain syllables (which syllables is, at least in part, predictable from grammatical and phonological information, but this is of no concern to us here). If the stød-syllable contains a long vowel, the stød is manifested on the vowel; if it contains a short vowel immediately followed by a sonorant consonant, then the stød is on this consonant. It might be added that if it contains a short vowel followed by an obstruent (such a combination is normally said to be unable to receive the stød), then the stød is not manifested (compare alternative pronunciations like pust [pu:?sd/pusd]) except by a "stød-like" intonation in certain Western varieties of Standard Danish; this amounts to considering an abstract stød ("Accent 1"), falling on syllables which never have stød phonetically but which occur in a morphological context normally implying stød (e.g. prefixed verbs, cf. beflitte, bestille [beflídə, besdél?ə]), further see Rischel 1970b, p. 128f, and Basbøll 1972b, p. 30.

Before the oral non-lateral non-syllabic sonorants ([ð, i̯, u̯, ɒ̯] manifesting /d, g, g/v, r/ syllable-finally) a stød-vowel can be shortened and the stød "moved" to the following sonorant, where it is normally placed (e.g. [fo:?ð/foð?, dæ:?i̯/dæi̯?, sgɛ:?u̯/sgɛu̯?, sbi:?ɒ̯/sbiɒ̯?]). Phonetically, it is quite possible that there is no "stød-movement" (i.e. change of the location of the stød) at all, the stød occurring at (roughly) the same place in the syllable (cf. Riber-Petersen 1973 and references cited there). If this is correct, a prosodic treatment of the stød would seem to present itself, but let us, for the sake of the argument, accept the segmental location of the stød (in the output) as traditionally given in phonetic notation and taken over by Austin. A non-syllabic treatment of the stød would then seem to need a phonological transformation like the following:

$$\text{(i)}\quad \underset{1}{\begin{bmatrix} \text{V} \\ \text{+long} \\ \text{+stød} \\ \end{bmatrix}} \underset{2}{\begin{bmatrix} \text{C} \\ \text{+son} \\ \text{-nas} \\ \text{-lat} \end{bmatrix}} \underset{\text{OPT}}{\Longrightarrow} \begin{bmatrix} 1 \\ \text{-long} \\ \text{-stød} \end{bmatrix} \begin{bmatrix} 2 \\ \text{+stød} \end{bmatrix}$$

However, this is clearly inadequate because part of the transformation states that the vowel is shortened before certain consonants, e.g. [ð], and a phonological rule with exactly this effect is needed since the shortening also occurs in words without stød (although less frequently), e.g. bide [bi:ðə/biðə]. We may then change the phonological transformation to the following, in order to avoid the duplication of the vowel-shortening rule (which it must follow in the ordering):

$$\text{(ii)}\quad \overset{V}{\begin{bmatrix}-\text{long}\\+\text{stød}\end{bmatrix}}\quad \overset{C}{\begin{bmatrix}\alpha\text{son}\end{bmatrix}} \underset{\text{OBL}}{\Longrightarrow} \overset{1}{\begin{bmatrix}-\text{stød}\end{bmatrix}}\quad \overset{2}{\begin{bmatrix}\alpha\text{stød}\end{bmatrix}}$$

The point to be emphasized is that this transformation (of a dubious formal status) is completely superfluous with a syllabic treatment of the stød, whereas it is, as far as I can see, indispensable in an orthodox generative framework.

Another type of stød-rule where a non-syllabic treatment seems hopelessly inadequate is constituted by rules ascribing stød to certain stressed syllables (e.g. ultimates and antepenultimates in foreign words). If an ad hoc transformation like (ii) is not recognized, it must in a non-syllabic framework be formulated as two processes, one concerning vowels and one concerning consonants, although it is clearly the same phenomenon which is involved. Such a rule will be explicitly formulated in section 3.1 below.

3. The syllable in a generative phonology

In the preceding section we have presented a number of facts taken from the phonology of Danish, the correct description of which seemed to presuppose, so we claimed, that the syllable should be included in phonological theory. But so far we have said next to nothing about the precise way in which the syllable should be included in phonology. In the following we shall discuss a number of different functions the syllable can have in phonological rules. We shall for this purpose accept the general framework of (non-orthodox) generative phonology (in section 4 the perspective will be widened). Since the most detailed and explicit discussion in print which I know of on the functions of the syllable within the framework of generative phonology is Hooper 1972, I shall mainly refer to her paper in section 3, and postpone the inclusion of some non-generative literature on the syllable until section 4.

3.1. The syllable as a unit in the structural description of phonological rules

Consider how a very simple stress rule (like the one in Polish) placing stress on the penultimate syllable of polysyllabic words and stressing monosyllables too would be formulated within an orthodox generative framework:

(i) $V \longrightarrow [+stress] \;/\; \_\_\_\; C_o\ (V)C_o\ \#$

or (ii) $V \longrightarrow [+stress] \;/\; \_\_\_\; C_o(VC_o)\#$

or (iii) $V \longrightarrow [+stress] \;/\; \_\_\_\; (C_oV)C_o\ \#$

The notational distinction between the three ways of parenthesizing the last syllable apparently does not correspond to anything empirically; this, of course, is a shortcoming of the notation. Another drawback is the fact that one has to state two instances of C_o in the structural description of a rule. Now, what does C_o mean? C_o means that it is _irrelevant for the application of the rule whether there are any consonants_ after the vowels mentioned in the structural description or not. Of course, the over-all principle of what ought to be mentioned in the structural description of a rule is that only what is relevant for the application should be included in the notation. But in the case under discussion here it is the notational convention that stress cannot be ascribed to syllables which forces one to state the irrelevant consonants. Compare how the rule can be formulated when stress is ascribed to the syllable (the S-symbol) (cf. McCawley 1968, p. 36):

(iv) $S \longrightarrow [+\text{stress}] \;/\; \_\_\_\_ \; (S) \; \#$

Rule (iv) avoids all the difficulties of rules (i)-(iii).

As mentioned in section 2.3 above, the basic principle governing the occurrence of the Danish stød in foreign words of a Latin-Greek type as well as the traditional school pronunciation of these languages (according to which the main stress is placed in agreement with the target language, whereas the other features of pronunciation have in general no connection with the classical languages) is the following: if stress is on the ultimate or the antepenultimate syllable, the stressed syllable has stød if it has "stød-basis" (i.e. contains a long vowel and/or a postvocalic sonorant), whereas a stressed penultimate syllable does not have stød (the relation between non-main stresses and stød is more complicated and can be ignored for the present purpose).

If the notation does not allow one to ascribe the stød to a syllable, then either one must rely on the ad hoc-transformation (ii) of section 2.3, or the rule in question must be split up into two parts, viz.:

(v) $\begin{bmatrix} V \\ +\text{long} \\ +\text{stress} \end{bmatrix} \longrightarrow [+\text{stød}] \;/\; \_\_\_\_ C_o (VC_oVC_o) \; \#$

or (vi) $\begin{bmatrix} V \\ +\text{long} \\ +\text{stress} \end{bmatrix} \longrightarrow [+\text{stød}] \;/\; \_\_\_\_ (C_oVC_oV) C_o \; \#$

$$\text{or (vii)} \quad \begin{bmatrix} V \\ +\text{long} \\ +\text{stress} \end{bmatrix} \longrightarrow [+\text{stød}] \;/\; \_\_\_ C_o(VC_oV)C_o \#$$

$$\underline{\text{and}} \text{ (viii)} \quad \begin{bmatrix} C \\ +\text{sonorant} \end{bmatrix} \longrightarrow [+\text{stød}] \;/\; \begin{bmatrix} V \\ -\text{long} \\ +\text{stress} \end{bmatrix} \_\_\_ C_o(VC_oVC_o)\#$$

$$\text{or (ix)} \quad \begin{bmatrix} C \\ +\text{sonorant} \end{bmatrix} \longrightarrow [+\text{stød}] \;/\; \begin{bmatrix} V \\ -\text{long} \\ +\text{stress} \end{bmatrix} \_\_\_ (C_oVC_oV)C_o \#$$

$$\text{or (x)} \quad \begin{bmatrix} C \\ +\text{sonorant} \end{bmatrix} \longrightarrow [+\text{stød}] \;/\; \begin{bmatrix} V \\ -\text{long} \\ +\text{stress} \end{bmatrix} \_\_\_ C_o(VC_oV)C_o \#$$

Notice that rules (v)-(vii) as compared to rules (viii)-(x) are notationally quite different: in (v)-(vii) the vowel is the affected segment (i.e. the segment which is changed by the rule), in (viii)-(x) it is the consonant; in (v)-(vii) the affected segment is stressed, which is not the case in (viii)-(x); and in (viii)-(x) the environment to the left is relevant which is not the case in (v)-(vii). I shall argue that all these differences have nothing to do with the rule itself. Furthermore, there is a notational ambiguity in the parenthesizing of the environment to the right which has no empirical content.

Quite independently of the phenomena under discussion here, there are good arguments for a general "stød-manifestation principle" according to which stød is manifested on the vowel if this is phonemically long, on the consonant immediately following the vowel if this consonant is a sonorant, and not manifested as a stød in the normal phonetic sense (but eventually by a "stød-like" intonation, at least in certain Western varieties of Standard Danish) if the short vowel is immediately followed by an obstruent (see section 2.3).

Granted this stød manifestation principle, the rule in question can be formulated quite simply in a syllabic framework:

(xi) $\begin{bmatrix} S \\ +stress \end{bmatrix} \longrightarrow [+stød] \; / \; \_\_\_ \; (SS) \#$

The phonological behaviour of the Danish stød thus strengthens the traditional concept of the syllable as bearer of prosodic features like stress and tone.

It should be added that the recognition of the syllable as a possible unit (S) in the structural description of phonological rules as recently used e.g. by Hooper (1972) in no way implies that one is forced to accept Hooper's definition of this symbol as "a sequence of segments between two syllable boundaries" (p. 537), see section 4.1 below.

3.2. Syllable boundaries in the structural description of phonological rules

There is no doubt that phonological theory should allow the notation of a syllable boundary ($) in the structural description of phonological rules. This is typically the case when $ constitutes the left-hand side or right-hand side environment of the segment affected by the rule, i.e. when the rule expresses a process taking place in absolute syllable-initial or absolute syllable-final position. (I use the qualification "absolute" to resolve the ambiguity of e.g. "syllable-final position" as meaning either the final segment (perhaps the final consonant) of the syllable (i.e. "absolute syllable-final position"), or a segment (perhaps a consonant) in the final part of the syllable, i.e. occurring after the syllabic peak.) Examples of such processes, which should be stated by means of the syllable boundary symbol $ in their structural description, are strengthening of glides (in Spanish) in absolute syllable-initial position (cf. Hooper p. 528, where the environment is given as / $___) and tensening of non-low stressed vowels (in Castilian) in open syllables, i.e. in absolute syllable-final position (cf. Hooper p. 530, where the environment is given as / ___$).

However, Hooper mentions the $-symbol in a much wider range of phonological rules than those referred to above, and in most of these other cases I find that her use of $ in the structural description of phonological rules is unjustified and in fact obscures the nature of the processes. This is the case with several assimilation rules which, according to her statements, apply only across $. We shall now devote some attention to a typical example of this sort, viz. the place-of-articulation assimilation of nasals before consonants in Spanish (in "Allegretto" speech style, p. 525f).

The problem to account for is that nasals assimilate to obstruents both within words (e.g. campo, ganga [-mp-, -ŋg-]) and across word boundaries (e.g. un beso, un gato [-mb-, -ŋg-]). Before glides ([j, w]), on the other hand, the assimilation occurs only across word boundaries (e.g. un hielo, un huevo [-ɲj-, -ŋw-]), not within words (e.g. miel, nieto, muevo, nuevo [mj-, nj-, mw-, nw-]). Hooper writes:

"With the syllable boundaries included in the strings, the solution to the problem becomes obvious. Nasal assimilation occurs before a consonant or a glide only if a $-boundary intervenes. Thus one rule generates the desired output for all cases:

(1) $$[+\text{nasal}] \rightarrow \begin{bmatrix} \alpha\text{coronal} \\ \beta\text{anterior} \\ \gamma\text{back} \\ \delta\text{distributed} \end{bmatrix} \quad / \; \underline{\quad} \; \$ \; \begin{bmatrix} \alpha\text{coronal} \\ \beta\text{anterior} \\ \gamma\text{back} \\ \delta\text{distributed} \end{bmatrix}$$

Rule 1 accomplishes all that is required of a phonological rule: it generates all instances of nasal assimilation; it states the constraints on nasals followed by non-vowels; and it provides an explanation of the facts. Using $-boundaries, the difference between the glides of [uñ$yelo] and [nyeto] can be stated formally: the first [y] begins a syllable and thus is functioning as a non-syllabic segment, while the second is part of the syllable nucleus. Rule 1 states the appropriate generalization, that nasals assimilate only before segments that begin syllables, i.e. segments in a non-syllabic function." (p. 526; my italics.)

The first thing to be noted in the passage quoted is that Hooper's "explanation" (which even "can be stated formally") does not follow from her rule formulation, but that it presupposes an additional axiom like the following: a glide which begins a syllable is a non-syllabic segment as opposed to a glide which does not begin a syllable. If this is not in plain contradiction to other passages of hers, e.g. "two contiguous syllabic segments form separate syllables. This rule applies to Spanish ..." (p. 534), at best her terminology is confusing (the word "syllabic" apparently meaning "part of the syllable nucleus" while the feature name "[syllabic]" seems to mean "constitutes the peak of a syllable"; see section 4.5 below). Anyhow, the "explanation" is based upon a dubious distinction between those prevocalic glides which do and those which do not form part of the syllable nucleus, and I find this far from convincing (it should be noted that in the speech style discussed here, the glides in absolute syllable-initial position are not strengthened to obstruents, p. 528).

But if her explanation in the text is not connected with her rule formulation and in itself unconvincing, let us then take a look at her notation. The rule in fact states that the assimilation takes place only if the segment to be assimilated is separated from the segment causing the assimilation by a $-boundary. But to me it is unbelievable that it is the occurrence of a boundary which conditions the assimilation. I suggest the universal principle that if a certain assimilation under certain conditions takes place across a certain boundary, then the same assimilation under the same conditions will also take place in the absence of this boundary (and in the presence as well as absence of a weaker boundary).

This is, of course, not to deny the observational adequacy of Hooper's assimilation rule, but I think the explanation should be sought in a different direction, although within the same basic framework, viz. that the syllable plays a crucial role for the assimilation. Even though the function of the glides in un hielo and nieto is the same, as I suggest is the case, the function of the nasals in the two examples is quite different. In the final part of the syllable there is no contrast of place-of-articulation among nasals, and - with the Prague School terminology - there is thus only one nasal archiphoneme syllable-finally, whereas syllable-initially there is a three-way contrast between labial, dental, and palatal nasals (similarly the contrast between dental and palatal laterals is neutralized in syllable-final position). These neutralizations in syllable-final position are only one manifestation of a more wide-spread phenomenon: that syllable-final consonants are "weaker" than their initial counterparts in a number of respects: they are shorter, have a laxer articulation, and are much more susceptible to assimilatory phenomena (cf. Grammont 1933, p. 184ff). Whatever the correct formulation of the rule may be, I think this is the heart of the matter. Notice that Hooper must claim that her rule "states the constraints on nasals followed by non-vowels" in such a way that if a nasal should be followed by a word-final obstruent, the former would not necessarily be homorganic with the latter. Such a word-final cluster does not, of course, occur in normal Spanish words, but it seems improbable that a nasal preceding e.g. a word-final /p/ could be anything but [m].

According to Hooper, "these assimilation rules are two examples of generalizations that cannot be stated without the $-boundary" (p. 526; my italics). This assertion can only be due to lack of imagination, and I shall just sketch two alternative ways of accounting for the nasal assimilation in what seems to me a more insightful way. I have at present no strong arguments in favour of one or the other of these two proposals, and I do not believe that they are completely independent of each other.

One possibility is to re-include the notion of the archiphoneme in phonology, in the sense that a rule can refer to a segment which is not fully specified, i.e. which has zero as its coefficient for one or more features. Furthermore, the place-of-articulation assimilation expressed by the rule is obviously not a simultaneous but independent assimilation of four distinct features: [coronal, anterior, back, distributed], as the formulation suggests. If we therefore use one multi-valued feature [place-of-articulation] (abbreviated [art]) (cf. Ladefoged 1971, p. 42ff) [1], the rule can

---

1) By the term "multi-valued" I mean that the feature can take a finite number larger than two (but in practice a small number) of coefficients. Schane (1973, p. 96) apparently considers "multi-valued" equivalent to "scalar" in the sense that multi-valued features (i.e. features on a late, "phonetic" level of representation within his framework) can specify fine phonetic distinctions, e.g. between different languages, within a certain "scale". I should emphasize that this implication

be formulated as follows:

$$\begin{bmatrix} +\text{nasal} \\ \emptyset\ \text{art} \end{bmatrix} \longrightarrow [\alpha\ \text{art}]\ /\ \_\_\_\ \begin{bmatrix} \text{C} \\ \alpha\ \text{art} \end{bmatrix}$$

This rule expresses the generalization that the nasal archiphoneme is assimilated as regards place-of-articulation to a following consonant, whereas a fully specified nasal is unchanged by the rule; this explanation also holds true for the lateral assimilation. (This statement in terms of archiphonemes may be interesting in view of evidence from speech error analysis discovered by Niels Davidsen-Nielsen (forthcoming) which points to an encoding of *sp*-, *st*-, *sk*- as /s/ followed by a stop *archiphoneme* unspecified as regards the feature(s) distinguishing /ptk/ from /bdg/.)

Another possible solution would be that [-syllabic] segments are subcategorized into [+final] and [-final]. Thus the underlying forms could be - but need not be - sequences of syllables, each syllable consisting of an unordered set of segments which would then have to become a sequence by a process of *linearization* in accordance with the syllabic hierarchy, see section 4 below. This idea of linearization

---

1) (cont.) (due to the orthodox generative (Jakobsonian) unwillingness to recognize non-binary features on the *phonological* level) does not apply to my use of "multi-valued" (which is in agreement with Ladefoged). Ladefoged, however, uses "scalar" synonymously with "linearly ordered" (as opposed to "independent"), i.e. a ternary feature can be scalar according to him. I think this terminology should be refined so that "scalar" means "*continuously* varying within certain limits" (this is in fact a normal property of a "scale"), in agreement with Schane's use of the word, whereas "a multi-valued linearly ordered feature" can take only a finite number of values on a certain "scale". Whatever terminology is chosen, the *distinctions* mentioned above all seem necessary, i.e. both Schane's and Ladefoged's terminologies are too crude.

has evident parallels in higher parts of the grammar (cf. Chafe 1970, p. 250ff). Thus the assimilation rule could be stated like this:

$$\begin{bmatrix} +\text{nasal} \\ +\text{final} \end{bmatrix} \longrightarrow [\alpha\text{art}] \; / \_\_\_ \begin{bmatrix} \text{C} \\ \alpha\text{art} \end{bmatrix}$$

In favour of a solution along these lines would be the fact that for a large number of assimilation processes the segment affected could be defined as being [+final], and this is true also in cases where there is no neutralization in final position, as in the voicing assimilation rule which Hooper states (p. 530) in the following way:

$$[\quad] \longrightarrow [+\text{voice}] \; / \_\_\_ \$ \quad [+\text{voice}].$$

3.3. The syllable as a domain for phonological rules

At the outset of section 3.2 we mentioned certain types of phonological rules where it seemed to be the best solution to mention the $-boundary in the structural description, viz. mainly processes applying to segments in absolute syllable-initial and absolute syllable-final position (there are maybe some others as well, cf. Hooper p. 531). We claimed, however, that in a number of cases this was not the correct solution. In this section we shall discuss some further phonological rules where the syllable seems to play a crucial role although not the one ascribed to it by Hooper.

The $-boundary is, of course, not the only boundary in phonology which can constitute the left or the right-hand side limit of a rule environment. For example, special processes can occur phrase-finally (aspiration (and affrication in the case of /t/) in Danish is such an example), and such rules must contain the boundary symbol in their structural description (devoicing before # or ## in many languages is another well-known example).

It must be emphasized that such cases by no means illustrate the only function of boundaries in phonology; they do not even constitute the most important one. This is not the place for a general discussion of grammatical boundaries in phonology (cf. Basbøll (forthcoming) for an extensive discussion of (particularly word-internal) boundaries in French phonology). However, in order to place syllable boundaries in their proper perspective, I find it useful very briefly to discuss the function of the boundaries in SPE and to compare it with the proposals of McCawley (1968, p. 55ff).

Let us begin with the conceptually simpler proposal, viz. that of McCawley. He suggests that the grammatical boundaries in phonology (junctures) in a given language can be arranged in a unique rank order B_1, B_2, ... B_n, in the sense that B_1 is the weakest (of lowest rank), then comes B_2, etc. (B_1 is the morpheme boundary, B_n the pause.) The main function of the boundaries is to define the domain of phonological rules in the following way: Each phonological rule has a certain rank expressed in terms of a specific boundary, e.g. [B_3]; when an input string to a rule of the rank [B_3] is scanned to see whether it matches the structural description of the rule, the string is divided into chunks in such a way that each chunk is limited on both sides by a boundary of rank 3 or a higher rank (i.e. B_3^n), and that the only boundaries that occur within each chunk are of a rank lower than 3 (i.e. B_1^2); each such chunk is thus separately matched to the structural description of the rule, ignoring those internal lower rank boundaries (i.e. B_1^2) in the input string which are not mentioned in the structural description of the rule.

(According to McCawley, a rule of rank B_n is higher valued than a similar rule of a lower rank.) Although the proposal is in need of precision as to the restrictions on the inventory of possible boundaries, it is conceptually very satisfying.

The same can scarcely be said about the use of boundaries in SPE. First of all, boundaries (which are [-segment]) are cross-classified by means of distinctive features, just like vowels and consonants ([+segment]). This is a very powerful principle (the inclusion of which in the phonology of boundaries seems by and large unmotivated). Chomsky and Halle use three different intra-phrase boundary symbols, viz. + ("morpheme boundary", i.e. [-segment, +formative boundary, -word boundary]), # ("word boundary", i.e. [-segment, -formative boundary, +word boundary]), and = ([-segment, -formative boundary, -word boundary]). This latter boundary, forming separate natural classes together with + ([-segment, -word boundary] defining = and +) and # ([-segment, -formative boundary] defining = and #), destroys the hierarchical principle inherent in McCawley's proposal, but since it is limited to certain quite specific non-native formations (SPE p. 94f), we shall ignore it here. Boundaries stronger than # are in SPE represented by sequences of # , which in fact approaches the idea of a hierarchy of boundaries; e.g. a rule applying in the environment /___# will also necessarily apply in the environment /___## , but not inversely.

Now, what is the function of the boundaries + and # in rules? According to SPE, the distribution of the boundary + is irrelevant to the application of phonological rules unless their structural description explicitly mentions a

+-boundary.[1] A # , on the other hand, prevents the application of a phonological rule which does not mention a # -boundary at the indicated place in its structural description. This is empirically equivalent to saying that a phonological rule which does not mention any boundaries in its structural description only applies if the string of segments matching the structural description does not contain any instances of # . For non-cyclical rules, this amounts to saying that <u>their domain of application is the maximal string between two successive #-boundaries</u> unless boundaries occur in the structural description. And, in fact, also <u>the maximal string between two successive +-boundaries</u> containing no internal boundaries <u>defines a rule domain</u>, viz. the domain for morpheme structure rules or morpheme structure conditions.

The point I wish to make by this digression is the following: despite the great differences between the boundary theories of McCawley and <u>SPE</u>, they in fact agree that <u>a main function of boundaries in phonology is to define the domain of rules</u>. This follows from the fact that <u>boundaries can block the application of rules</u> where the sequences of segments

---

1) The condition expressed in the last clause does not seem very well motivated. Jørgen Rischel has suggested that the distinction between the boundaries # and + should be, simply, that # can have phonological effects, whereas the distribution of the boundary + in phonological strings is irrelevant to the application of phonological rules, i.e. <u>neither the presence nor the absence of +</u> can be presupposed for the application of a phonological rule. This hypothesis is a way of formulating the well-known insight that certain grammatical boundaries, but not all, have phonological consequences, cf. Rischel 1972, p. 226 (notice that the argument above does not apply to morpheme structure rules or -conditions). However, compare Kiparsky 1973, p. 57ff.

match the structural description (the occurrence of + blocks a morpheme structure rule, and the occurrence of other junctures can block phonological rules).

Now, if this is true of grammatical boundaries let us then examine the possibility that also syllable boundaries can block certain phonological rules, or that the syllable could be a domain of phonological rules. Apparently this possibility is not considered by Hooper (who seems implicitly to assign the same junctural status to the \$-boundary as to the +-boundary, i.e., that if a \$ occurs in the structural description, then the rule applies only to input strings containing \$ at the indicated place, whereas rules not mentioning \$ apply regardless of the distribution of \$; in an *SPE*-framework, this seems in fact to be the only possibility). A case in point is the following example:

"Many P-rules in Akan depend upon syllable shape and thus syllable boundaries[...] For example, consider the nasalization of high vowels which occurs before a nasal consonant *in the same syllable*. (Note that this environment is extremely common for vowel nasalization – it occurs in French for all vowels *in just such an environment*.) It is possible to avoid the mention of 'syllable' in formally stating the rule, by recognizing two environments: before a nasal before another consonant, and before a morpheme-final nasal:

(15) $$\begin{bmatrix} \mathrm{V} \\ \mathrm{+high} \end{bmatrix} \longrightarrow [\mathrm{+nasal}] \; / \; \_\_\_ \; [\mathrm{+nasal}] \begin{Bmatrix} \mathrm{C} \\ \# \end{Bmatrix}$$

It is simpler, however, to state the environment in terms of \$-boundaries:

(16) $$\begin{bmatrix} \mathrm{V} \\ \mathrm{+high} \end{bmatrix} \longrightarrow [\mathrm{+nasal}] \; / \; \_\_\_ \; [\mathrm{+nasal}] \quad \$$$

Note that Rule 16 is NOT an abbreviation of 15. Rule 16 is the real statement of the environment, while Rule 15 is merely an ad-hoc contrivance that produces the same results" (Hooper, p. 533; my italics).

The point to be emphasized is that the two formulations do not in all cases produce the same results, but that the "ad-hoc contrivance" gives the right output, whereas her rule 16 often gives the wrong output. For example, French words like cent, sans [sɑ̃] (I use examples from French where I know the data) in their underlying forms end in obstruents (cf. centaine, centenaire [sɑ̃tɛn, sɑ̃t(ə)nɛ:ʁ] and sans amour [sɑ̃ z amu:ʁ]). The latter example suggests that the phonological $-boundary should occur after the /z/ at the point of the derivation where the vowel is nasalized, since this is the place of the word boundary (according to Hooper, one should not expect a readjustment of the $-boundary to the place before /z/ until /z/ is intervocalic, i.e. after the nasal has dropped, see p. 527, 537). This conclusion seems strengthened by the observation that a word like sens [sɑ̃:s] has a nasalized vowel before a word-final pronounced obstruent (in certain varieties of French, at least, there is no underlying shwa in such words). A further argument for the position that nasalization of vowels in French can occur also before nasals that are not in absolute syllable-final position comes from words like pente [pɑ̃:t] in which the vowel is nasalized before a pronounced obstruent followed by pause. Such words end in an underlying shwa, and it might be claimed that the /t/ is syllable-initial at the point of the derivation where the vowel is nasalized. However, this does not agree with the fact that in a word like étiquette [etikɛt] (which is underlying /#etikətə#/, cf. étiqueter [etikte]), the phonological syllable border must go after the final /t/

since the /ə/ is converted to ε, see Dell 1973, p. 198ff and Basbøll, forthcoming. Although each of these arguments taken alone is not irrefutable, they nevertheless give a certain implausibility to Hooper's formulation. And at any rate, it is certain that the two formulations are not equivalent in all cases; and I find it hard to accept a notation in which every form /VN$/ is nasalized, but no form /VNC$/.

We therefore claim that the nasal consonant in these examples does not occur immediately before the $ which Hooper's formulation demands. In order to make her rule 16 work, it should be changed to (for the French material):

(i) $V \rightarrow [+nasal] / \_\_\_ [+nasal]\, C_o\$$

But this is only one more example where one is forced to state in the rule what is irrelevant for the application of it, viz. C_o.

How should this rule, then, be formulated? In the text Hooper in fact gives the correct condition, viz. "before a nasal consonant in the same syllable". Thus it is irrelevant whether the $-boundary occurs immediately after the nasal or not, what counts is that it does not occur between the vowel and the nasal consonant. If we recognize the possibility that a rule can have the syllable as its domain, or can have the $-boundary as its rank, or can be blocked by the occurrence of a $-boundary within its structural description (all these formulations here being equivalent), then the rule can be formulated like this:

(ii) $V \rightarrow [+nasal] / \_\_\_ [+nasal]$

where it should be a property of the rule, but not of its structural description, that it has the syllable as its domain.

(If it cannot be deduced from the process itself that the syllable is the domain of the rule, this must be given as an index of the rule, as McCawley must state the rank of rules with smaller domains than the string occurring between two pauses; however, as far as I know, it is the normal case for rules assigning consonant features to vowels, as well as for rules coalescing a vowel and a following consonant, that they do not apply across syllable boundaries.)

Let us now turn to a rule from Danish in which there is absolutely no way out of mentioning C_o in the structural description if we are to accept Hooper's framework. I have in mind the rule for the manifestation of short /a/ (see section 2.1 above). The colouring of /a/ by the following homosyllabic consonant is quite independent of any additional final consonants, e.g. the /a/ in samt [sɑm?d] is pronounced exactly like the /a/ in Sam [sɑm?]. Consequently Hooper would have to formulate the rule like this:

(iii) $a \longrightarrow$ [+back] /___ [-cor] C_o\$

(Notice that in the SPE-framework of distinctive features used by Hooper, there is no way to capture the assimilatory nature of the process in question, because what is alike in the vowel and the consonant is not articulatorily defined, see below.)

However, I argue that the domain of the rule is the syllable, and that its correct formulation is the following:

(iv) $a \longrightarrow$ [+grave] /___ [+grave] (domain: syllable)

This rule too is a case where a feature of a consonant (viz. the acoustically-auditorily defined feature [+grave] characterizing labials and velars, cf. Ladefoged 1971, p. 44) is ascribed to a preceding homosyllabic vowel.

Of course, I do not claim that the syllable is the <u>only</u> domain for phonological rules, not even that it is the most important one. Some other domains of phonological rules in Danish will be briefly mentioned below.

It is clear from the discussion of grammatical boundaries in phonology at the outset of the present section that according to the notational conventions of <u>SPE</u>, non-cyclical rules which do not include any boundary symbol in their structural description have the maximal string between two consecutive #-boundaries as their domain. The rule [-son] ⟶ [-voi]/___ [-voi] in Danish is such an example. This rule turns an underlying voiced obstruent, i.e. <u>b</u>, <u>d</u>, <u>g</u>, <u>v</u>, <u>ʁ</u>, into its voiceless counterpart, i.e. <u>p</u>, <u>t</u>, <u>k</u>, <u>f</u>, <u>ʁ̥</u>, in the position before a voiceless obstruent not separated from the preceding segment by any #-boundary (on this rule, see the notes referred to in footnote 1 on p. 42).

The compound stress rule in Danish seems to have the word as its domain: it is characteristic that the stress pattern which is typical of ordinary Danish compounds never applies to more than one word (cf. Rischel 1972). Finally, certain assimilation rules have larger domains, in part depending on stylistic factors, and so on.

It is an urgent need in phonology to find the relations which undoubtedly exist between the nature of a phonological process and its domain. I suggested above that rules which assign consonant features to a preceding vowel normally have the syllable as their domain. However, this relation is probably not a universal law, since certain rules seem able to change their domain. Such conditions of rules changing their domain ought to be investigated; for example, rule (iii) above seems to be in the process of enlargening its domain by younger speakers, so that words like <u>papir</u>, <u>akademiker</u> are now often pronounced [p^hɑp^híɒ̥ʔ, ɑk^hadé:ʔmigʌ] (the pronunciation [p^h, k^h] suggests that it is not the syllable boundary which has been moved).

3.4. Syllabification

By "syllabification" I understand the division of a phonological string (consisting of phonological segments and boundaries) into consecutive syllables. "Syllabification" is thus roughly equivalent to "placement of syllable boundaries", and it therefore concerns the border between contiguous syllables, and not their internal structure in terms of peak, nucleus, margin, and the like (see further section 4).

As to the cases where the syllable functions as a unit in phonological rules (i.e. typically in rules concerning prosodic features like stress, tone, and stød, cf. sections 2.3 and 3.1 above), syllabification is not required for the correct application of the rules: what is necessary is only that the number of syllables be known, and this information can possibly be given with an identification of the syllabic peaks. The consequences of this fact for the formal definition of the syllable will be taken up in section 4.1 below.

Obviously, if a syllable boundary occurs in the structural description of a phonological rule (section 3.2 above), the input string to this rule must be syllabified. Similarly, regarding the rules having the syllable as their domain (section 3.3 above), their input strings also appear to presuppose a previous syllabification.

However, on this point I would like to claim that phonological syllable boundaries can be partly indeterminate, i.e. they need not always be fully determined. In a great number of cases where the syllable plays a role in determining phonological processes, the only relevant distinction is one between open and closed syllables, whereas a distinction between syllables ending in one, two, or three consonants does not matter. This is the case e.g. with the stress rule in classical Latin and with the rule predicting the manifestation of Danish short /o/ (section 2.1 above).

Let me briefly discuss a third example of this sort, viz. the rule E-Adjustment in French which in certain contexts (which can tentatively be defined as "in closed syllables") converts an /e/ or /ə/ into /ɛ/ (see Basbøll, forthcoming, for a discussion of this rule in connection with syllabificational problems). Examples like *genevois*, *jetterons* [ʒənvwa,ʒɛtʁ̥ɔ̃], derived from /# ʒənəv+uaz# , #ʒət+ə+r+ɔN# z #/, show that an intervocalic consonant between two unstressed shwas belongs to the following vowel if no morpheme boundary intervenes between it and the consonant, otherwise to the preceding vowel. But where does the syllable boundary occur in an example like *sévrerons* [sɛvʁəʁɔ̃], derived from /#səvr+ə+r+ɔN#z#/ (cf. *sevrer* [səvʁe] /#səvr+ə+r#/)? If we say that the syllable boundary coincides with the morpheme boundary, then the first syllable ends in the otherwise unknown final consonant combination /vr/ (see below); but if we say that the syllable boundary occurs between /v/ and /r/, then we must recognize that the factor determining the syllable boundary, viz. the morpheme boundary, occurs at a different place, i.e. that it can only retract the syllable boundary one place, and this sounds somewhat mysterious to me. In fact I think the choice between the two mentioned locations of the syllable boundary is a pseudo-problem which only arises in a theory forcing one to state one unique inter-segmental location of every syllable boundary. A more realistic solution is to define the notions *open* and *closed* syllable formally in such a way that the first shwa in *sévrerons* will occur in a closed syllable, but without commitment to the choice between several locations which will *in any context under any circumstances give the same result*.

Before we turn to the discussion of several factors which can influence the syllabification of phonological strings, a word should be added about *the level at which syllabification applies*.

First of all, I do not think it makes much sense to speak of syllables and syllabification at very abstract levels of representation (cf. Fudge 1969). Of course, one can claim (Hooper 1972, p. 538) that Spanish estable is derived from /stable$/ via (by an e-epenthesis rule followed by a resyllabification) /estable/, but I fail to see why there must be $-boundaries (before /s/) in the abstract string /stable$/. Similarly, she notes that Spanish pan is derived from /pane$/ via (by an e-deletion rule followed by a resyllabification) /pan/. In fact, I think that the only arguments for abstract syllable boundaries different from the phonetic ones are the impact of such boundaries on the application of phonological rules. There is thus no reason to postulate an "abstract syllable boundary" between /a/ and /n/ in pan if this boundary can have no phonological effect whatever and, of course, no phonetic existence at all (this is in line with our proposal above that phonological syllable boundaries are in some cases partly indeterminate). Notice that the present discussion concerns the level of description, i.e., I do not intend to exclude an abstract syllabification /pane$/ if this word-internal syllable boundary could be phonologically justified in other forms with e-deletion; but a phonological syllabification applying before a certain rule P should be excluded in principle if the syllabification in question can have no effects on the application of P or of any phonological rule ordered before P.[1]

At the other end of the abstractness scale there is ample evidence for the existence of phonetic syllables and phonetic syllable boundaries. At the phonetic level the postulation of a syllable boundary should, of course, be justified on purely phonetic (i.e. non-phonological) grounds, including the distinction between syllable initial- and syllable-

---

1) The rule order involved here is descriptive order, i.e. priority is concerned, not time (e.g. in a psychological sense), see Rischel 1974, p. 311ff.

final extrinsic allophones, etc. The postulation of phonetic syllable boundaries can thus be confirmed or disconfirmed by experimental-phonetic evidence, as opposed to phonological syllable boundaries which are merely "descriptive devices" in cases where they differ from the phonetic ones. Since the subject of the present paper is the phonological and not the phonetic syllable, I shall leave the purely phonetic questions here (but cf. section 4), and turn to the relation between phonological and phonetic syllable boundaries.

"The claim made here is that Rule 22 [which inserts syllable boundaries in certain contexts defined in terms of segments; HB] represents the universal definition of the phonological syllable, and that languages may or may not have additional late rules that define a language-specific, phonetic syllable" (Hooper 1972, p. 536; my italics). Cf. the following quotation: "We have been assuming that it is always possible to assign a non-arbitrary syllable division in any word; this assumption is justified on the grounds that we are referring to the phonological syllable, not to the phonetic syllable" (ibid.). Hooper refers to Hoard 1971 who shows that stress influences syllabification (which is not predicted by her Rule 22) in English. Apparently Hooper here confuses two issues; what is specific to English is not, as Hooper suggests it is, that "a stressed syllable attracts a maximum number of segments to it" (ibid.), but rather the fact that English is a language which has heavily stressed syllables contiguous to unstressed ones, as well as (naturally) the regularities and irregularities of the location of the stresses.

In fact, I think that the relation between phonetic and phonological syllable boundaries is just the opposite of what Hooper believes (cf. Basbøll 1972a, p. 193[1]). It is highly

---

1) I have permitted myself to restate some of the points I made in Basbøll 1972a, since Hooper's article had not appeared at the time I published my paper, and I could therefore not take into consideration her partly similar, partly deviant proposals on that occasion.

significant that the universals concerning syllable types in different languages, possible consonant clusters and their division by syllable boundaries, etc. which have been signalled in the literature (cf. e.g. Jakobson 1941, Greenberg 1965, and Cairns 1969) all apply at a phonetic or a "classical" phonemic level, but not at an abstract phonological (morpho-phonemic) level. This suggests that an abstract phonological syllable may be more or less language-specific (although there is also here a high degree of convergence between the syllabificational criteria used by different languages, e.g. as to the specific role of the unstressed shwa in both French and Danish); phonetically, on the other hand, languages seem to agree, and I suggest that the universally unmarked way to syllabify a given sound chain is the phonetic syllabification (but it depends, of course, on a number of language-specific factors, such as the distribution of stresses (and maybe junctures, see below), a dependence which is probably not language-specific). As a hypothesis, I think it has much more inherent plausibility than Hooper's, and it seems to be supported by observations like the following:

In Danish, phonological criteria clearly point to a syllable division after [ð] in words like bade [bæ:ðə] (e.g., [ð] never occurs word-initially, and it alternates with [d] before stressed vowels as in abbed, abbedisse [ábeð, abedísə]), but both phonetically and psychologically (see below) it seems to occur before [ð], which is in agreement with the tendency towards CV-syllables, especially when the vowel is long (note that in this example there is no question about segments being inserted or deleted, or changing their coefficient for the feature [syllabic], which could motivate a resyllabification within Hooper's framework).

As for the psychological reality of syllable boundaries, the only thing that can be said here is that the problem of determining whether such psychologically real syllable boundaries exist, and if so, where they are, is empirically quite distinct from determining both the phonetic and the phonological syllable boundaries. This question should be enlightened by means of psychological test methods, but its interwovenness with orthography will probably make the issue extremely difficult to settle.

Let us now return to the phonological syllable boundaries. Below I shall discuss briefly a number of factors which can influence the location of the phonological syllable boundaries, viz. the following: grammatical boundaries; the existence of initial and final segment combinations in the utterance and/or the word and/or the morpheme; stress; the surrounding vowels; and the sequences of consonants. It should be emphasized that these do not represent a number of alternative principles of syllabification of which every language chooses one; on the contrary, those principles interact in different ways in different languages. In section 3.4.1 below I shall state briefly how Danish uses these different criteria in an ordered fashion to form a set of principles of phonological syllabification.

Sensitivity of syllable boundaries to grammatical boundaries. It has always been recognized that grammatical boundaries can determine the place of syllable boundaries in such a way that the latter must coincide with the former (cf. Hooper 1972, p. 527, 537). For example, it seems to be true for a great number of languages that boundaries between words are also phonological syllable boundaries, if the latter are relevant at all.

The relation between intra-word grammatical boundaries and syllable boundaries is more complex. If the hypothesis sketched in the footnote on p. 62 can be upheld, viz. that the distinction between the boundaries # and + is that the former but not the latter can play a role for the application of phonological rules, then the hypothesis suggests itself that the boundary # always coincides with the syllable boundary (this hypothesis seems by and large valid for the Danish and French material I have investigated with regard to syllabification); however, the general examination of it is an enormous task because the junctural structure of the language in question must be established independently of syllabificational phenomena if the reasoning shall avoid circularity (see Basbøll, forthcoming).

The last problem I shall mention concerning the relation between grammatical and syllabic boundaries is the well-known fact that, in general, Germanic languages as opposed to Romance languages exhibit a phonetic syllabification which is sensitive to word boundaries (compare the fact that in German word-initial vowels have a glottal attack, e.g. ein Esel [ʔɑin ʔé:zl̩], with the liaison- and enchaînement-phenomena in French, e.g. en avril [ɑ̃ n avʁíl]). In my view, this casts doubt on the existence of a phonological rule of universal applicability that erases all grammatical boundaries at the end of the phonological component. An alternative would be that (intra-phrase?) grammatical boundaries are erased at an earlier point of the derivation in French than in German, or maybe that they are not erased in German at all but manifested phonetically in various ways. This issue has scarcely been investigated at all.

Sensitivity of syllable boundaries to the existence of initial and final segment combinations in the utterance and/or the word and/or the morpheme. It is a traditional insight that there is generally a high correlation between the possible initial and final consonants in a syllable on one hand and in grammatical units like the utterance, the word or the morpheme on the other, in the sense that an intervocalic consonant cluster can be partitioned on different syllables so that the syllable final cluster is also found finally in the grammatical unit in question, and that the syllable initial cluster is also found initially in the grammatical unit.

As for the grammatical unit used for the comparison, I think most can be said in favour of the word, which is in fact also the traditional choice. In general, there seem to be no phonological restrictions determining which words can be combined in utterances, can occur utterance initially or -finally, etc. (of the type "words with three initial consonants never start utterances"), this of course being related to the fact that the word is the "minimal free form". Thus the word will give the same results as the utterance for this purpose, and is of a more manageable size (there may be methodological reasons for preferring one or the other frame of reference within different linguistic schools, but this can be ignored here). On the other hand, there can be phonological restrictions on the combination of morphemes into words (for a Danish example, cf. Basbøll 1973, p. 129f) and, furthermore, one morpheme often has several alternants differing with respect to consonant combinations (i.e. with different phonological rules being applied in different contexts), and this makes the consonant combinations less well defined. A Danish example showing the difficulty in using the morpheme as frame of reference is mentioned in section 3.4.1 below.

Within generative phonology, a similarity between initial and final consonant combinations in the syllable and in the morpheme has often been pointed to (e.g. Hooper 1972, p. 535f). The interest of this comparison within the orthodox generative framework is no doubt due to the fact that the concept "possible morpheme initial (final) cluster" can be given a well defined theoretical status by means of the morpheme structure rules or -conditions applying in the lexicon (cf. Stanley 1967). But it should be emphasized that these morphemes are generally taken to be very abstract entities, and in cases where the initial or final clusters of these morphemes are changed by phonological rules, they do not play any role for syllabificational phenomena, as far as I know. The dangers in using abstract morphemes as frame of reference for syllabificational phenomena can be seen in the following quotation from Hooper:

"In Spanish the division is /s$l/, as in is$la, but in English it is /$sl/. Again the constraints for syllable-initial position correspond to the constraint for word-initial position.

To account for such exceptions [viz. to the general principle that the syllable boundary occurs before the obstruent in an obstruent-liquid sequence; HB] formally, I propose that, in addition to the general rule for inserting $-boundaries, the metatheory provide a list of possible exceptions [...] The choice of exceptions applicable in a given language is determined by the morpheme structure conditions of that language [...] The fact that such a relation obtains may bear on the question of the existence of $-boundaries in the lexicon." (p. 535f; my italics).

Hooper jumps easily from the similarity between the syllable-initial constraint and the word-initial constraint of /sl/ in Spanish (which is real), to an imaginary similarity between the syllable-initial constraint and the morpheme-structure constraint which is in flagrant contradiction to her underlying forms on p. 538 like /slabo/ which has an initial /sl-/. The correct conclusion is, of course, that there is no relation between clusters in abstract morphemes and syllabification, unless these abstract clusters are also found at more concrete levels of representation.

This conclusion agrees well with one of the claims made at the outset of section 3.4, namely that phonological syllable boundaries are merely descriptive devices in cases where they differ from the phonetic syllable boundaries, and that it is unjustified to speak of phonological syllable boundaries on more abstract levels than those on which they are used by phonological rules. It follows, then, that the comparison between word-initial (or -final) and syllable-initial (or -final) clusters cannot be made on such abstract levels.

However, the comparison in question cannot always be made on the phonetic level either, since the clusters to be compared may differ in ways which have no bearing on the validity of the comparison. E.g., if final voiceless stops in a given language are aspirated in word-final position but not in syllable-final position within the word, then it seems reasonable to disregard this difference in aspiration when comparing a certain postulated syllable-final cluster to the set of word-final clusters to see whether one of these matches it. A level appropriate for this comparison would be something like a phonemic level in the pre-generative sense, and this has in fact been used for distributional descriptions by many authors, e.g. Sigurd 1965 and Vestergaard 1968. However, within orthodox generative phonology no such intermediate level between the systematic phonemic and the systematic phonetic level has been given any theoretical status (cf. Chomsky 1964). If the generative model is not expanded to include one or more theoretical levels of this sort (which it probably should, cf. Rischel 1974, p. 361ff), then the distributional description - including the comparison between word-initial (or -final) and syllable-initial (or -final) clusters - can, at any rate, be given at an intermediate level defined by the rules which have, respectively have not, applied at that level (cf. Basbøll 1973 for a distributional description of Danish consonants at such a level, and Basbøll 1974 for a partly similar description of Italian).

It should also be said that the relation discussed in these lines obtains between syllable-initial and -final clusters on the one hand, and possible word-initial and -final clusters on the other, not necessarily registered clusters in a given corpus (i.e. clusters which are accidentally missing should be included in the material). It is of course a moot question how to define the distinction between structurally missing and accidentally missing clusters (cf. Spang-Hanssen 1959 and chapter VI of Fischer-Jørgensen 1952), and it has both statistical, psychological, and, above all, purely phonological aspects of the sort discussed in section 4.3 below. E.g., a Danish word like lingvist [leŋgvísd] should be syllabified before /gv/ despite the non-occurrence of word-initial /gv/, [g] being the syllable-initial allophone of /g/. That the non-occurrence of word-initial /gv/ may be regarded as accidental is seen from the existence of word-initial [sgv] as found in skvat [sgvad], etc. (the interpretation of [sg-] as /sk-/, suggested by Uldall 1936 and taken over by Vestergaard 1968, is not only far-fetched phonetically, but also at variance with the unanimously accepted interpretation of [sv-] as /sv-/ and not /sf-/).

There are at least two principal difficulties which arise when we try to evaluate the importance of the syllabificational criterion under discussion here. First: we argue for a certain intra-word syllable boundary by saying that alternative locations of it would lead to a syllable-initial (or -final) cluster which is structurally impossible as a word-initial (or -final) cluster. However, it will in general be possible to get (roughly) the same results by defining the location of the syllable boundary in terms of the sequence of intervocalic consonants instead, without reference to the set of structurally possible clusters in the word. This difficulty arises in Danish, and it will therefore be discussed in section 3.4.1 below.

The other principal difficulty is the following: When a certain constraint is valid both for syllable-initial (or -final) clusters within the word and for word-initial (or -final) clusters, then it is no simple matter to decide which constraint is the basic one and which is the derived one, if they are not both consequences of a more general principle. It is rather evident that the set of word-initial and word-final clusters are primary data in a way in which intra-word (phonological) syllable boundaries are not, and in the language analysis the word-initial and -final clusters have therefore generally been taken as primary, and the intra-word syllable-initial and -final clusters as secondary. But in the final language description model (in the synthesis, if you like), it is not so clear what to do. Within the descriptive framework advocated in section 4 below, according to which most phonotactic information can be derived from a general model of "maximal syllabic structure", the constraints for word-initial and word-final consonant clusters are a consequence of the fact that these clusters are by necessity also initial, respectively final, in the syllable.

(An approach to syllabificational phenomena relying heavily on the relation to initial and final clusters is found in Anderson and Jones 1974.)

Sensitivity of syllable boundaries to stress. As already mentioned, Hoard 1971 contains a detailed discussion of the influence of stress in English on the occurrence of syllable boundaries (see the beginning of this section). The general principle seems to be that a stressed syllable attracts a maximum number of segments to it.

In Danish, this principle is not very important for the *phonological* syllabification (but the unstressed vowel shwa plays a great role, see section 3.4.1 below). I shall, however, mention one type of example where its influence can be clearly seen. A word like *eskorte* is nearly always pronounced with an unaspirated [g]: [ɛsgɒ́:də]. The derived verb *eskortere*, on the other hand, is often pronounced with an aspirated [k^h], i.e. with the /k/ in syllable-initial position not occurring after a homosyllabic /s/: [ɛsk^{h}ɒt^{s}é:?ʌ/ɛsgɒt^{s}é:?ʌ]. The only relevant difference between the two occurrences of /s/ + velar stop is that in *eskortere*, both surrounding vowels are unstressed, whereas in *eskorte*, the following vowel is stressed and it thus obligatorily attracts both preceding obstruents to it.

Sensitivity of syllable boundaries to the surrounding vowels. It has been pointed out in Basbøll 1972a that the distinction between the shwa-vowel and all other vowels (i.e. "full vowels") plays a major role for phonological syllabification in Danish; e.g. a single intervocalic consonant, as well as a /g/ preceded by a sonorant, belongs to the preceding syllable if the following vowel is shwa, but to the following syllable if its vowel is a "full vowel". It is of course not accidental that it is exactly the vowel shwa which occupies such a unique place with regard to syllabification: The distinction between shwa and full vowels is also essential to the description of stress (cf. Rischel 1970b), of the Danish "word types" (Poul Andersen 1955 and later works), and of the creation of syllabic consonants and other assimilatory phenomena in unstressed syllables (see section 4.6 below).

The importance of the vowel shwa for syllabificational phenomena is also suggested by the fact that the "consonant gradation"-phenomena (section 2.2 above), i.e. the consonant weakening in syllable-final position, are peculiar to Danish as opposed to e.g. Swedish, in which there is a contrast /a/:/e/ in unstressed position, corresponding to the one Danish vowel shwa (cf. Danish gade, Swedish gata [gæ:ðə, gɑ:ta] 'street').

Finally, it is interesting to notice that the unstressed vowel shwa in French plays a role for phonological syllabification very much as in Danish. For example, an intervocalic consonant as well as an obstruent-liquid cluster belongs to the following syllable if its vowel is a "full vowel" (this term denoting all other vowels than shwa, as well as the underlying shwas which receive the word stress, e.g. appel [apɛl] /#apəl#/, cf. appeler [apl̥e] /#apəl+ə+r#/), but to the preceding syllable if its vowel is a full vowel and the following vowel is an unstressed shwa. If both surrounding vowels are unstressed shwas, then the preceding syllable is open unless there is a morpheme boundary between the post-vocalic consonant and the following shwa (cf. genevois, appellerons [ʒənvwa, apɛlʁɔ̃] /#ʒənəv+uaz#, #apəl+ə+r+ɔN#z#/) as mentioned above in the present section. Also in French there is, of course, ample evidence for the special phonological role played by the unstressed shwa, both with regard to the word stress-rule (a word-final shwa in a polysyllable being the only final vowel which does not bear the word stress) and to the vowel deletion rules (where unstressed shwas regularly drop in most contexts, as opposed to all other vowels with a couple of quite isolated exceptions like the /a/ of /la/).

Sensitivity of syllable boundaries to the sequences of consonants. It is a well-known fact that different types of consonant clusters can have different effects on syllabification. E.g., in many languages, like Latin, Italian, and French, and to some degree also in Germanic languages, a sequence obstruent-liquid "counts as" one single consonant with regard to the principles of syllabification (it should be added, however, that in the Romance languages mentioned above it is a condition that the obstruent be non-sibilant). Also the clusters of /s/ plus a stop can in certain cases in Germanic languages, but generally not in Romance, act as single consonants with respect to syllabification (see Davidsen-Nielsen 1974 for a phonetic analysis of some English material in this respect). Since this type of conditioning on syllabification has been discussed extensively by others, e.g. Hooper 1972, I shall limit myself to the above remarks on this matter.

3.4.1 Some principles of phonological syllabification in Danish

As an illustrative appendix to the more general discussion in the preceding section, I shall here briefly survey some principles which seem, by and large, to account for the phonological syllabification in Danish (I have discussed these principles in greater detail in the notes mentioned in footnote 1 on page 42).

It is probably superfluous to emphasize that the principles to be presented below constitute only one out of several possible ways to account for the phonological syllabification in Danish, and that the formulations are highly sketchy. As pointed out above, it is very difficult to determine the interaction between two of the principles to be mentioned below, namely (ii) (on the relation between clusters in the syllable and in the word), and

(iii) (on the sensitivity of syllable boundaries to the surrounding segments), and it is not even certain that principle (ii) will be indispensable in the final analysis. The following section should thus be taken as exemplifying the preceding discussion rather than explaining the location of every phonological syllable boundary in Danish.

The practical difficulties in effectively testing such principles (as well as phonological rules in general) for observational adequacy should be evident. I would therefore like to mention that Kjeld Kristensen and the author of these lines are at present engaged in computer testing of a part of a Danish phonology which includes a phonological syllabification of Danish words, as well as a number of phonological rules (both categorial and variable). This work in progress is planned to be reported in the forthcoming volume of ARIPUC.[1]

I shall now discuss one model for determining the place of the phonological syllable boundaries in Danish, containing the following three types of factors (it should be borne in mind that (ii) below may turn out to be dispensable in the final analysis, presupposing a certain elaboration of (iii)): (i) grammatical boundaries, (ii) the relation between initial and final consonant combinations in the syllable and in the word, and (iii) the sequences of segments in question (both concerning vowels and consonants). The principles apply in a hierarchical fashion so that (ii) only applies if (i) has not decided the location of the syllable boundary, and (iii) only if (ii) has not (i.e., the principles (i)-(ii)-(iii) are disjunctively ordered). It is presupposed that the chain of segments to be syllabified is fully specified as to the feature [syllabic], i.e. it is known which segments form a syllabic peak and which do not ("vowels" and "consonants", respectively, in a somewhat loose terminology, see below), cf. sections 4.1 and 4.5 below.

---

1) A preliminary account is given in the present volume of ARIPUC.

The function of the principles can be illustrated by means of the following (somewhat metaphorical) description: The first time a phonological string is input to a rule which either has the syllable as its domain or which mentions a syllable boundary in its structural description, it is syllabified. Every boundary of a certain "rank" occurring between two [+syllabic]-segments is concurrent with a syllable boundary (see below). In the cases where there is no such boundary between two [+syllabic]-segments (possibly with intervening [-syllabic]-segments), all the places a syllable boundary can occur without giving rise to a syllable-final cluster which is impossible word-finally, or a syllable-initial cluster which is impossible word-initially, are marked off (ii). If there is only one such place, then the syllable boundary occurs there (there will always be at least one). If there are several such places, then the syllable boundary is placed at one of these, according to the principles (iii) which are sensitive to sequences of segments.

(i) The syllable boundary coincides with certain grammatical boundaries. This seems true of the boundaries between words, the intra-word boundaries before stems, and the boundaries before stressed native suffixes like -*inde* and -*agtig*. It thus seems possible to define the occurrence of these grammatical boundaries only in terms of what follows; viz. if what follows is a word, a stem (i.e., roughly, a major lexical category: N, V or A), or a stressed native suffix, or if nothing follows, then the grammatical boundary is obligatorily a syllable boundary too. Notice that both words, stems, and native stressed suffixes contain at least one [+syllabic]-segment, i.e. the principle suggested above can give rise to at most one syllable boundary between two [+syllabic]-segments (which is, of course, necessary in order for the definition "to work").

I leave the question open whether the syllable boundaries in these cases could or should be defined by means of junctures, or whether they have to refer directly to notions like "stem", "suffix", etc.

Concerning the grammatical boundary before stems, it will place the correct syllable boundaries between prefixes or "small words" (prepositions etc.) and stems, e.g. *medgørlig* [mɛðgŒ́ɒ̯?li], as well as between stems, i.e. between the two parts of a compound, e.g. *mados* [mā́ð(?)ò:?s] (in this connection it may be interesting to notice that in French as well, the grammatical boundaries *before* stems are always syllable boundaries too, which is not the case for all grammatical boundaries *after* stems, cf. Basbøll, forthcoming).

Concerning the syllable boundary before native stressed suffixes like -*inde*, -*agtig*, -*dom*, -*hed*, it should be emphasized that "stressed" covers main stress as well as the reduced main stress typically found in second parts of compounds (e.g. *violinistinde*, *barnagtig*, *sølvagtig*, *guddom*, *guddommelig*, *godhed* [violinisdénə, bɑ:nɑ̃́gdi, søl̀ɑ̆gdi, gū́ðdʌ̀m?, guðdʌ̃́m?əli, gṓðhè̆ð?]). (To account for the distribution of main vs. reduced stresses is in many cases difficult, and non-pertinent for our purpose.)

The qualification "native" of the stressed suffixes in question is necessary in view of the many foreign (Greek-Latin type) stressed suffixes like -*at*, -*ist*, -*i*, -*isse* which are preceded by *syllable-initial* consonant allophones (e.g. *doktorat*, *kontorist*, *perfidi*, *abbedisse* [dʌgt^{s}oʁɑ̃́:?d, k^{h}ɔnt^{s}oʁĩ́sd, p^{h}æɒ̯fidĩ́:?, abedĩ́sə], cf. *doktor*, *kontor*, *perfid*, *abbed* [dʌ̃́gdʌ, k^{h}ɔnt^{s}ṓɒ̯?, p^{h}æɒ̯fĩ́ð?, ā́beð]). Another fact which points to the foreignness of these suffixes, as opposed to -*agtig*, -*dom*, etc., is that they have never *reduced* main stress which is in general characteristic of typically native formations.

The special role of the boundary before stressed native suffixes agrees well with the vague hypothesis advanced in Basbøll 1972a (p. 194) that the <u>intuitively transparent</u> morpheme boundaries are also syllable boundaries. This amounts to postulating that native derivative formations like <u>professorinde</u> are more easily analysable for the native speaker than learned forms like <u>professorat</u>.

Finally it should be said that morphemically complex words can be lexicalized in the sense that their internal (phonologically relevant) grammatical boundaries are erased, with the consequence that they are treated <u>phonologically</u> as if they were <u>monomorphemic</u>. E.g. when the word <u>rødspætte</u> [ʁœ́sbɛ̀də] is pronounced without [ð], it is in accordance with the general constriction on <u>monomorphemic</u> words that [ð] does not occur before [s] (which it does in compounds etc.). This agrees well with the semantics of the word in question, which means 'plaice', whereas its second "constituent" in isolation can only mean 'woodpecker' (cf. Basbøll 1973, p. 122ff). Similarly, several names of Danish islands ending in -<u>ø</u> ('island') are pronounced as if they were not compounded: <u>Ærø</u>, <u>Thurø</u> [ɛ́:ʁœ̀:ʔ, t^sú:ʁœ̀:ʔ] (cf. <u>Færøerne</u> [fɛ́ɒø̀:ʔʌnə]), while others have alternating pronunciations corresponding to forms with and without the grammatical boundary (e.g. <u>Rømø</u>, <u>Femø</u> [ʁŒ́mø̀:ʔ, fɛ́mø̀:ʔ] and [ʁœ́:mø̀:ʔ, fɛ́:mø̀:ʔ]).

<u>(ii) The location of syllable boundaries as decided by the possibility of word-final and word-initial clusters</u>. As already mentioned, phonological syllable boundaries in Danish, as well as in many other languages (not in all, e.g. not in Classical Greek), can always be - and, in fact, also are - placed in such a way that the syllable-final cluster does not violate any general restrictions holding for word-final clusters, and so that the syllable-initial cluster does not violate any general restrictions holding for word-initial clusters

(I deliberately avoid the formulation that "syllable-final clusters should also occur word-finally" or the like, in order to stress the fact that the issue concerns structurally possible clusters, see above).

Examples which clearly demonstrate the necessity of having recourse to such a principle for the location of phonological syllable boundaries in Danish are not easy to find. One could mention words like angre(r), buldre(r) [ɑŋʁʌ, bulʁʌ], where the manifestation of /r/ is the one elsewhere found in syllable-initial position, but what these examples in fact show is only that the morpheme boundary and the syllable boundary need not coincide (in both words the morpheme boundary occurs after /r/). Whether the explanation for this syllabification is that neither /-ŋr/ nor /-lr/ are structurally possible word-final consonant clusters, or that the syllable boundary in an intervocalic sequence of two sonorant consonants always separates these consonants, is an open question.

The use of this criterion in the present model for phonological syllabification in Danish depends on a previous phonotactic description allowing one to distinguish between (structurally) excluded and accidentally non-occurring clusters in the beginning and end of the word (cf. Basbøll 1973). This criterion has been used here in order to permit as simple and general a formulation of principle (iii) below as possible.

(iii) The location of syllable boundaries as a function of the sequence of segments. When the location of a syllable boundary between two syllabic peaks with intervening [-syllabic]-segments must be determined by principle (iii), the primary distinction is whether the following syllabic peak is a "full vowel" or a "weak syllabic peak". The basic generalization is, then, that the syllable border is situated in the leftmost position indicated by principle (ii) if the following syllabic peak is a full vowel, otherwise in the rightmost position with one reservation (see below). In other (and vaguer) words: the "stronger" of the two syllabic peaks attracts the consonants (cf. the influence of stress briefly mentioned in the preceding section), but when the vowels are of equal "strength" (i.e. with respect to the distinction full vowel:weak syllabic peak), then the following syllabic peak is the decisive one (thus the syllable boundary goes "to the left" between two full vowels, but "to the right" between two weak syllabic peaks).

First, let us define the distinction between full vowels and weak syllabic peaks on the phonetic surface (but it should be remembered that it is used at an earlier point of the derivation, cf. below). The latter comprise all occurrences of the vowel [ə] and of all syllabic consonants[1] (i.e. [ð̩, l̩, m̩, n̩, ŋ̩], of which [m̩, ŋ̩] only occur after labials and velars, respectively, where they are in free variation with [n̩], see Basbøll 1969, p. 44f). Certain occurrences of the unstressed vowels [i, e, ʌ] are weak syllabic peaks, too. This is always true for the vowels of the (unstressed native) derivational suffixes ig, ing (and the rare ik), e.g. dydig, madding(,maddik) [dý̠:ði, mā́ðeŋ (,mā́ð?ig)]. Also all unstressed [ʌ]s which are derived from one of the sequences /ər, rər, rə/ or from /ə/ preceded by /r/ (e.g. kuer=kurer=kure, angre [ku:ʌ, ɑŋʁʌ], cf. Rischel 1969, p. 196ff) count as weak syllabic peaks. The vowel of the derivational suffix isk in some words counts as a weak syllabic peak (e.g. metodisk, pædagogisk [met^só:?ðisg, pɛdagó:?(ɣ)isg]), in others as a full vowel (erotisk, parodisk [eʁó:?t^sisg/eʁó:?disg, pɑʁó:?disg]), cf. Rischel 1970b, p. 133f. All other syllabic peaks are full vowels (i.e. all stressed vowels, all unstressed vowels different from [i, e, ʌ, ə], and all instances of unstressed [i, e, ʌ] which do not satisfy the (mainly non-phonological) conditions stated above).

---

1) I here follow the traditional use of the term "consonant", but in section 4.2 below [ð] will be defined as a vocoid.

When viewed on the phonetic level, the distinction between full vowels and weak syllabic peaks may look cumbersome and unnatural. But it should be remembered that the phonological syllabification under discussion takes place at an intermediate level of the phonological component, and on that level the distinction turns out to be much more natural. First of all, the phonological processes deriving unstressed [ʌ][1] from a number of different inputs all containing shwa (discussed by Rischel 1969, p. 196ff) have not yet applied, and all syllables with a weak syllabic peak and which phonetically contain [ʌ] thus have shwa as their vowel when the phonological syllabification applies. Furthermore, syllabic consonants are generally created by a process of shwa-assimilation (see section 4.6 below) which is a late one and thus applies after phonological syllabification (it will be considered in section 4.5 below whether underlying syllabic consonants should be postulated, but at any event there is a general relation of equivalence between pronunciations with [ə] and a non-syllabic

---

1) My transcription differs from most of the earlier IPA-transcriptions of Standard Danish used in ARIPUC in that I identify the unstressed vowel derived from /ə/ in certain /r/-contexts with the stressed vowel in kom! [k^hʌm], and not with (the quality of) the stressed vowel in hård [hɒ:ʔ]. From a purely phonetic (in the sense of 'non-phonemic') point of view the former identification is clearly superior to the latter for the majority of Standard Danish usages; but if each phonetic symbol is taken to represent a certain range of different pronunciations within the norm, then a new symbol (i.e. a symbol not identical with any stressed Danish vowel symbol) is probably needed for this unstressed vowel: the unstressed vowel in damers can be identical to that of Amos, but also less rounded; the unstressed vowel of Amors, on the other hand, can be identical to that of Amos (and thus also to that of one of the possible pronunciations of damers), but also more rounded and back/low. All these pronunciations can even occur in the speech of one individual (e.g. my own).

consonant (in more formal styles etc.), and pronunciations with syllabic consonants). Finally, an assimilation rule has been proposed with the effect of raising a /ə/ in /əng, əg, ək (əsk)/ to *i* (Basbøll 1972a, p. 201):

$$ə \longrightarrow [+\text{high}] / \_\_\_\_ \ ([+\text{cor}]) \begin{bmatrix} \text{C} \\ +\text{high} \end{bmatrix}$$

thus accounting for the fact that the vowels of the derivational suffixes *ing* (where *i* is regularly lowered before a nasal), *ig*, *ik* (and possibly, for purely descriptive reasons, certain occurrences of *isk*, namely those which seem to have a weak syllabic peak) count as weak syllabic peaks, the reason being that their vowel is an underlying shwa. With the possible exception of the derivational suffix *isk*, the set of syllables with weak syllabic peaks is thus identical to the set of syllables whose vowel is shwa at the point of the derivation where phonological syllabification applies.

There is one exception to the general principle that the syllable boundary goes to the left before a full vowel but to the right before a shwa, viz. when a consonant cluster which contains a stop other than /g/ preceded by either a (underlyingly) voiced continuant or a nasal occurs before shwa. In that case the syllable boundary goes *before* the stop (see section 2.2 above):

kæn$tre, *æn$dre* [(k^{h}ɛntʁ̥ʌ/)k^{h}ɛndʁʌ, ɛn(.)dʁʌ];
lam$pe, *jam$be* [(lamp^{h}ə/)lambə, jam(.)bə];
vær$ten, *ver$den* [(væʁ̥t^{s}ən/væʁ̥dən/)væɒ̯dən, væɒ̯(.)dən];
mal$ke, *alg$e* [(malk^{h}ə/)malgə, al(.)ɣ̥ə];
vær$ke, *værg$e*, *ærg$re* [(væʁ̥k^{h}ə/væʁ̥gə/)væɒ̯gə, væɒ̯(.)ɣ̥ə, æɒ̯(.)ɣ̥ʁʌ];

van$te (sb.), ban$de (sb.), van$dig [(vant^{s}ə/)vandə, ban(.)də, van(.)di].[1]

There is overwhelming evidence that a single intervocalic consonant before a shwa belongs to the preceding syllable (Basbøll 1972a), e.g. bade, koge [bæ:ðə, kɔ:ɣə]. Examples showing the location of the syllable boundary before full vowels are ek$stra, a$ttrap, O$ta [ɛ̃gsdʁɑ, at^{s}ʁ̥ɑ́b, ó:t^{s}a] (that /s/ belongs to the preceding syllable in eksport [ɛgsp^{h}ɒ́:d] is in accordance with principle (i) above since its stem is port, cf. import [emp^{h}ɒ́:d]). As already mentioned, the location of the syllable boundary can be dependent on stress as shown by certain clusters of /s/ plus stop.

---

1) I thus no longer believe (as in Basbøll 1972a, p. 199) that the syllable boundary in /CVndə, CVldə/ regularly goes to the right of the /d/, cf. bande (sb.) and see section 2.2 above. (The fact that /d/ is deleted (possibly via a lenition to ð) in vande, skylde, jorde, etc., is thus not a consequence of the location of the syllable boundary, but of an earlier rule which seems to obey the proposed universal condition (Kiparsky 1973, p. 65ff) that (non-automatic?) neutralization processes only apply to derived forms (i.e. not to the underlying form of one morpheme); the underlying logic is, not surprisingly, that if they applied to the non-derived forms, too, a word like bande (sb.) could then have no possible use for its underlying /d/ since it would be deleted in all contexts.) This change in the earlier syllabification principles has the desirable consequence of turning the class of "exceptions" (1972a, p. 201f) vandig, skyldig, etc., into regular consequences of the principles. The main argument for the location of the syllable boundaries proposed here is that they permit the most general statement possible on the manifestation of /p t k/ vs. /b d g/: the distinction is one of aspiration when they occur initially in a syllable with a full vowel, the distinction is most often neutralized, in favour of the unaspirated stops, when they occur initially in a syllable with a weak syllabic peak (whereas Conservative usages in that position can make a distinction in the length of the preceding consonant), and /d g/ are regularly manifested as voiced continuants when they occur syllable finally (for details, see my notes referred to above).

4. The structure of the syllable and the hierarchy of distinctive features

The theoretical framework used in the preceding sections was that of (non-orthodox) generative phonology.[1] In this section we shall take a more general view of the phonological syllable. It is, of course, impossible here to survey all the different theories of the syllable which have been proposed in the literature (Kloster Jensen 1963 gives a good overview of the subject; cf. also Malmberg 1965). Our main concern will be to discuss the sonority structure of the syllable in terms of currently used phonological distinctive features and to define a syllabic hierarchy of these features (sections 4.2 - 4.4).

Before this central part of section 4 (on the structure of the syllable), a word will be said on how to define the phonological syllable (i.e. on the nature of the syllable) (section 4.1), and we will conclude by discussing the feature [syllabic] (section 4.5) and a rule which changes the coefficient for this feature (section 4.6).

---

1) It should be emphasized that I consider this generative framework a purely descriptive model (which, in my view, is a useful complementary tool to structuralist methods of analysis emphasizing the notions of contrast etc. (cf. also Rischel 1974)), and I thus do not share the commonly held belief among generativists that their phonologies are (ideally) models of the psychological language mechanism of the (ideal) speaker-hearer. See Linell 1974 and (particularly) Derwing 1973 for recent critical discussions of the psychological implications of generative phonology and syntax.

4.1 On the nature of the phonological syllable

Hooper (1972, p. 537) defines the syllable (S) in the following way:

$$X \longrightarrow S \; / \; \$ \_\_\_\_ \$$$

Condition: X contains no $.

I.e., according to Hooper, a syllable is defined as the maximal string between two consecutive syllable boundaries ($). This seems to me a poor definition. It is well known that the number of syllables as well as the location of their peaks can very often be indicated - by the native speaker as well as by the linguist - even in cases where the precise location of their borders is indeterminate. Hooper's definition makes it impossible to speak about e.g. bisyllabic words without being able to indicate the precise location of the border between the syllables. This is to me suggestive of a weakness in her theory (recall that I do not agree with Hooper's claim that phonological syllable boundaries (as opposed to phonetic ones) are always, by necessity, uniquely defined, cf. section 3.4 above).

It is, of course, true that if the location of all syllable boundaries in a given phonological string is known, then the extension of each syllable of that string is, ipso facto, also known; but the peaks of each syllable are not determined by the boundaries. If the location of all syllabic peaks in a given phonological string are known, on the other hand, then the number of syllables is also known, but not the location of their boundaries. Thus the syllabic peaks give some information on the syllable that cannot be derived from the location of the syllable boundaries, and vice versa. This means that the syllable is crucially characterized by both of these concepts, and a satisfactory general definition of the syllable ought thus to include both types of information, cf. Kloster Jensen's conclusion (1963, p. 34): "Es ist dringend notwendig, dass man sich klar macht, was für eine Hybride die Silbe ist. Sie hat einen phonetischen Kern und phonemisch und distributionell bestimmte Grenzen". In the following we shall briefly survey a couple of definitions which in different ways make reference to the "phonetic kernel" of the syllable.

One such approach to defining the phonological syllable is that exemplified by Haugen 1956 (cf. also O'Connor and Trim 1953), who defines the syllable as "the smallest unit of recurrent phonemic sequences. We will then have to include not only the segmental phonemes, but also the prosodic ones like stress, tone, length, and juncture. Any or all of these occur in sequence with each other, and the syllable is that stretch of phonemes which makes it possible to state their relative distribution most economically" (p. 216). Haugen thus indirectly refers to both the peak and the boundaries of the syllable: the peak is characterized by means of "prosodic phonemes" like stress and tone, the boundaries by the segmental phonemes. This certainly makes sense as an operational definition within phonemic theory, but from our point of view the definition is not entirely satisfactory. For one thing, it is not quite clear that the meaning of the phrase "to state their relative distribution [i.e. of the phonemes] most economically" can be made explicit, and if so, there is no reason to believe that the syllable is in all languages the most satisfactory frame of reference for the phonotactic description (for some evidence bearing on this question, see Pike 1947, e.g. p. 144ff, 174ff; Basbøll 1974 argues that the word is the preferable frame of reference for the distributional description of Italian consonants). It is also problematic in Haugen's definition what "recurrent phonemic sequences" means in cases where there are great differences between the phonemic inventory and/or structure of e.g. "stressed and unstressed syllables" (cf. Fischer-Jørgensen 1952, p. 17).

I thus see Haugen's definition as a practical way of trying to solve the problem of the (phonemic) syllable within an American structuralist framework, but with results which are not quite satisfactory from other points of view.

The difficulty met with above, i.e. that two types of syllables (e.g. "strong" and "weak", of which only the "strong" ones occur as one-syllable utterances) may differ structurally, also shows up in Eli Fischer-Jørgensen's definition (1952, p. 16ff) as she herself notes. She defines the "syllabic base" as "the class of the smallest units, of which each (in connection with stress, tone, and intonation, if such units are distinctive in the language in question) is capable of constituting an utterance by itself" (my italics). The difficulty in question cannot, of course, be avoided by changing "each" to "some" in the definition, since segments (i.e. vowels and consonants) will then fulfil this definition in a language in which a vowel alone can form a "strong" syllable and thus also an utterance.

To Hjelmslev (1939), the essential feature of the syllable is its being a unit consisting both of a syllabic theme and a distinctive accent. This leads to the well known terminological absurdities that French does not have syllables, and that Finnish vocoids and contoids are not vowels and consonants, respectively, in the structural sense, cf. Fischer-Jørgensen ibid. (Hjelmslev therefore defines the notion pseudo-syllable, which is a unit consisting of a pseudo-vowel (capable of constituting an utterance by itself) and pseudo-consonants, to be applied in languages without distinctive accent, e.g. French; the pseudo-syllable is close to Fischer-Jørgensen's definition of the syllabic theme.) According to Hjelmslev, the vowels (in the structural sense) are presupposed by the consonants (this idea has recently been developed within the framework of dependency grammar, see Anderson and Jones 1974). The difficulty of working with Hjelmslev's definitions is that they are made without regard to the phonetic substance, and what appears to be basically the same unit in different languages, viz. the syllable, will therefore be given a number of quite different structural interpretations within Hjelmslev's framework (cf. Fischer-Jørgensen: "The most suitable method will probably be to choose the structural unit presenting the closest affinity to the phonetic syllable", but this unit "will hardly be structurally the same in different languages" (1952, p. 16; my italics)). This is in fact to deny the very existence of a universally defined phonological syllable.

But does such a universally defined phonological syllable in fact exist? What I mean by "phonological" in this context is simply that it is more abstract than "phonetic". But it is a consequence of the claims made in section 3.4 above that we should expect a universal definition of the phonological syllable to apply at a level of representation not very distant from the phonetic surface. And the characterization of the phonological syllable to be given in the following sections in fact applies at such a level (see section 4.2 below). However, it should be emphasized that within generative phonology the distinction between phonological (as used here) and phonetic (e.g. syllable) is a matter of degree, not of principle.

It is a common feature of most definitions of the phonological syllable that they only make reference to one aspect of the problem of the syllable, either to its boundaries, or to its peak, or to the fact that it is the bearer of prosodic entities, or to its capability of occurring as an utterance. These definitions all have a certain value within given methodological frameworks, but they are far from giving an adequate picture of the nature of the syllable. A more insightful characterization of the syllable can be given, I think, if we go back to Jespersen's (1897-99) theory of the syllable as a top of sonority. The idea that each syllable consists of exactly one relative sonority maximum (the peak), that the least sonorous segments occur farthest off the peak, etc., is a way of condensing information concerning several aspects of the nature of the syllable: the syllabic peak is defined by the syllabic sonority structure, the syllable boundaries can be defined as the relative sonority minima in the cases where they are not determined by grammatical boundaries (see section 3.4 above), the voiced part of the syllable (which is also defined by the syllabic structure) is phonetically the bearer of tone, the non-consonantal part of the syllable is "the nucleus", etc. In the following sections we shall try to develop Jespersen's general theory of the syllable within the framework of modern phonological theory, i.e. to make the content of the sonority hierarchy precise, defined with currently used phonological distinctive features. We shall also state some limitations of this general theory (it goes without saying that not every part of Jespersen's syllabic theory can - or ought to - be redefined within a quite different conceptual framework from his of the late nineteenth century).

4.2 The distribution of distinctive features around the syllabic peak

In Basbøll 1973 (p. 106ff) I argued that it is an empirical problem at what level phonotactic restrictions in a given language are best described. I also claimed that the most general distributional statements of Danish vowels and consonants can be given at an intermediate level of the phonological component, defined by the fact that the diphthongization of short vowels plus homosyllabic voiced continuants has been applied, but neither the rule which devoices continuants after aspirates, nor the rule which deletes [ɒ̯] after low back vowels. Furthermore, the variable rule which shortens long vowels before homosyllabic voiced (non-lateral) continuants ought not to be applied.

At this intermediate level the inventory of non-syllabic segments (i.e. the inventory of segments which can occur outside the syllabic peak; ̯ under a vowel symbol indicates that the vocoid in question does not form the syllabic peak, see below) is the following:

p t k b d g s f h v ɣ ð m n ŋ l ʁ u̯ ɒ̯ i̯.

Words like skærv, lav (adj.), ud, bor, var, hagl, elv, talg, plaske, sjæl at this level consist of the following sequences of segments: [sgæɒ̯?v, læ:?v, u:?ð, bo:?ɒ̯, vɑɒ̯, hɑu̯?l, ɛl?v, tal?ɣ, plasgə, sjɛ:?l].

At this level of description I postulate that syllables containing a full vowel (see section 3.4.1 above) have the following hierarchical structure (cf. Basbøll 1973, p. 130ff):

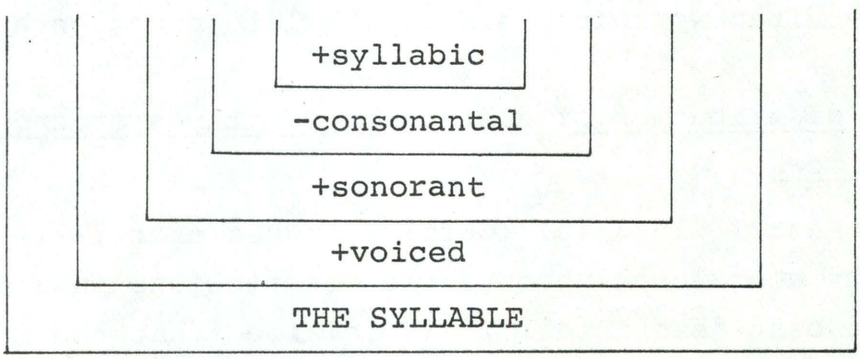


Fig. 1

The distinctive features are used in accordance with the principles of Ladefoged 1971. I shall briefly comment on each of the hierarchical features.

syllabic means "constitutes the peak of a syllable".[1] When a Danish syllable contains several vocoids, it is clear from an auditory point of view that exactly one of these constitutes the syllabic peak. E.g. the first syllable of ivrig and juridisk both contain the sequence of segments [iu], but in the former word [i] constitutes the syllabic peak, whereas it is [u] in the latter. I follow the traditional transcription in writing a [‸] under a vowel symbol which does not constitute the syllabic peak: ivrig, juridisk [iu̯ʁi, i̯uʁidisg] (more examples of this sort are given in section 4.5 below). However, since the articulatory and acoustic correlates of the feature [syllabic] are largely unknown, there are evidently great difficulties in defining it (see Ladefoged 1971, p. 81f). Nevertheless, I do not think it is logically circular to include this feature in the syllabic hierarchy, since its placement at the top of that hierarchy reflects the observation that when the syllable contains one or more vocoids (i.e. non-consonantal segments, see below), then its peak falls in a vocoid. But it is undeniable that the feature [syllabic] has another status in the syllabic hierarchy than the other features since it is not defined by any inherent quality of the segment in question (thus each of the pairs [i] and [i̯], [l̩] and [l], etc., are in a sense the same segment type), but only in a given context within the syllable. On the special status of this feature within the present model, see section 4.5 below.

consonantal is, according to Ladefoged, a cover feature "in that it can be defined only in terms of the intersection of classes already defined by other features. Thus non-consonantal sounds are non-lateral and sonorant. They correspond largely to what Pike (1943) called vocoids, which he defined as central, resonant orals" (1971, p. 91). I here use [-consonantal] as

---

1) In Basbøll 1973 I used a terminology which I now think should be abandoned: I then used [+syllabic] roughly for what is here called [-consonantal], and for the syllabic peak I employed the label vowel (which then, in fact, had to be taken as a distinctive feature).

equivalent to [+sonorant, +continuant, -lateral] (it must be a lapse (which is, however, repeated on p. 108) in Ladefoged's formulation that he does not include "continuant" (in the sense of SPE) or "non-stop" (in his own terminology) in his characterization of non-consonantal sounds). According to this definition, the class of consonants is the union of the class of obstruents (i.e. [-sonorant]), the class of nasals (i.e. [+sonorant, -continuant]), and the class of laterals (i.e. [+lateral]). According to the syllabic hierarchy, all syllabic segments found at the present level are non-consonantal. Of the non-syllabic segments found at this level, only [i̯ u̯ ɒ̯ ð] are non-consonantal.[1]

sonorants have "a comparatively large amount of acoustic energy within a clearly defined formant structure" (Ladefoged 1971, p. 58). It is a consequence of this definition that voiceless sounds are always obstruents. All non-consonantal segments are sonorants. Of the consonantal segments found at the present level, [p t k b d g s f h v ɣ ʁ] are obstruents, [m n ŋ l] sonorants.[2]

---

1) It is a consequence of the definition of [-consonantal] adopted here that [ð] must be classified as a vocoid (in disagreement with Basbøll 1973), since it is (at the level used throughout the present discussion) a sonorant and neither a nasal nor a lateral. This is in accordance with Heger's classification (forthcoming), which is purely phonetically based.

2) It is also a consequence of the definitions that [ʁ] must be classified as an obstruent (and not as a consonantal sonorant as in Basbøll 1973), just like [v] ([ʁ] is neither a vocoid nor a nasal nor a lateral). Also this is in agreement with Heger's phonetic classification (as well as with the phonetic classification of Eli Fischer-Jørgensen, personal communication), and since I have argued elsewhere that the Danish /r/ is an underlying obstruent, the so-called "consonantal r" must be [-sonorant] throughout the derivation.

voiced is used here as a binary feature defined by the vibration of the vocal cords during the articulation. All sonorants are voiced. Of the obstruents found at the present level, [p t k b d g s f h] are voiceless, [v ɣ ʁ] voiced.

Those are the features which define the syllabic hierarchy of Danish. There are, of course, other distinctive features too (see section 4.4 below).

Fig. 1 should be read as follows: the syllabic peak together with possible adjacent glides (i.e. [-syllabic, -consonantal]) constitute the non-consonantal part of the syllable, often called the "nucleus". The non-consonantal part together with possible sonorant consonants constitute the sonorant part of the syllable. The sonorant part together with possible adjacent voiced obstruents constitute the voiced part of the syllable. The voiced part together with possible adjacent voiceless segments constitute the syllable (or syllabic theme, in Hjelmslev's terminology).

Thus the following "implication chains" hold true without exception:

[+syllabic] ⊇ [-consonantal] ⊇ [+sonorant] ⊇ [+voiced]
and: [-voiced] ⊇ [-sonorant] ⊇ [+consonantal] ⊇ [-syllabic].

(The two implication chains are notational variants in a system where all the features are binary.)

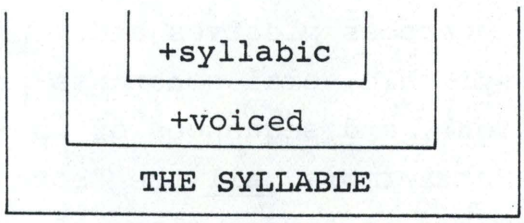
These implication chains predict the systematic nonoccurrence of certain combinations of distinctive features, viz. [+syllabic, +consonantal], [-consonantal, -sonorant], [+sonorant, -voiced], [+syllabic, -sonorant], [+syllabic, -voiced], [-consonantal, -voiced]. That only vocoids can be syllabic is a trivial consequence of the fact that the present model only concerns syllables with a full vowel as peak (in syllables with a weak syllabic peak, also consonantal sonorants can be syllabic, at least at the phonetic level, see below). That only consonantal sounds can be obstruents is a consequence of our <u>definition</u> of [-consonantal] as equivalent to [+sonorant, +continuant, -lateral]. That only obstruents can be voiceless is an empirical consequence of the definition of [sonorant] (by "empirical (consequence)" I mean that it does <u>not</u> follow from the <u>formulation</u> of the definition itself). That the last three combinations mentioned are excluded follows from the systematic non-occurrence of the first three, mentioned in the beginning of the paragraph.

However, the important point in this connection is not only to what degree the implication chains are based on observable facts on (Danish) syllable structure (see the following section), on the nature of the human speech and hearing apparatus, or on the formal definition of the distinctive features used. But it is in itself significant that exactly these features can be arranged into a syllabic hierarchy. It should be emphasized that all the features used here are well motivated in phonological theory, and that their definitions have not been given with the syllabic hierarchy in mind (as a matter of fact, such considerations are not included in Ladefoged's treatment of the distinctive features at all). Furthermore, all the hierarchical features except [syllabic] have well-defined articulatory and/or acoustic correlates. As mentioned above, the acoustic and articulatory correlates of the auditive feature [syllabic] are largely unknown, cf. sections 4.5 and 4.6 below.

One possible interpretation of the model is that the hierarchical features do not "belong to" one segment at a time, but that they characterize the "central part" of the syllable as a whole (where this part has a different extension for different features).

It can be shown that each of the hierarchical features has its function in the syllabic hierarchy, i.e. the removal of any of these features will cause a decrease in the explanatory value of the model (see the following section). But it is clear that in many concrete cases, the extent of several hierarchical features will be identical: e.g. in a monosyllable like <u>tak</u>, the syllabic part, the non-consonantal part, the sonorant part, and the voiced part of the syllable all equal the vowel <u>a</u>.

The feature hierarchy for syllables with a weak syllabic peak is a strict subset of the feature hierarchy for syllables with a full vowel as peak, viz. that given in fig. 2.



<u>Fig. 2</u>

The "implication chains" which can be read from fig. 2 are the following:

$$[+\text{syllabic}] \supseteq [+\text{voiced}]$$

and $$[-\text{voiced}] \supseteq [-\text{syllabic}]$$

(i.e., [-voi]-segments are never [+syl]; this has the status of an empirical finding).

Within this weak syllabic hierarchy there is thus no motivation for a distinction between the non-consonantal and the sonorant part of the syllable, nor for a distinction between the sonorant and the voiced part of the syllable (see below).

4.3 Ordering relations in Danish as explained by the proposed feature hierarchy

Fig. 3, which in all relevant respects is a notational variant of fig. 1, expresses the order relations holding between segments within the same syllable more directly (in fig. 3 I only write the features which cannot be deduced from the implication chains).

Fig. 3 correctly predicts the order of any permitted (unordered) set of <u>either</u> initial <u>or</u> final segments, together with the following additional restrictions: In the initial part of the syllable, <u>s</u> precedes plosives and <u>v</u> precedes <u>ʁ</u>; in the final part of the syllable, oral sonorants precede nasals (see the following section), and sequences of non-sibilant obstruents end in a dental. Furthermore, <u>all</u> the "boxes" of fig. 3 can be descriptively motivated as "order classes" (cf. Sigurd 1965, Vestergaard 1968) for the phonotactic description of Danish (see Basbøll 1973). This highly general model of the syllabic hierarchy, defined in terms of phonologically well motivated distinctive features, thus explains a very large part of the phonotactic structure in Danish.

Fig. 4 is the similar "linearization" of fig. 2, i.e. it should predict the combinatory possibilities and order in syllables with a weak syllabic peak. (Included in fig. 4 are only segments which can occur in such weak syllables, cf. Basbøll 1973, p. 134ff. The syllable boundaries are supposed to be in accordance with the principles of section 3.4.1 above. p t k are only found initially in weak syllables in certain pronunciations, cf. section 3.4.1 above.)

Examples showing the structure of weak syllables in this respect are <u>ændre</u>, <u>ellers</u>, <u>himlens</u> [ɛndʁʌ, ɛl?ʌs, heml̩ns], which are, at the level used througout this discussion, \$ɛn\$dʁə\$, \$ɛl?\$əʁ̭s\$, \$hem\$ləns\$.

-voi	-son +voi	+cons +son	-syl -cons	+syl	-syl -cons	+cons +son	-son +voi	-voi
sfh ptk bdg	vʁ	l mn	i̯	vowels	u̯ɒ̯i̯ð	l mnŋ	vɣ	sf bdg

Fig. 3

-voi	-syl +voi	+syl	-syl +voi	-voi
bdg (ptk)	lʁn	ə	lnɒ̯ð	sd

Fig. 4

A reasonable question to ask at this point is the following: Since phonotactics plays a crucial role as evidence for the syllabic hierarchy according to our model, and since the phonotactics of different languages can be highly diverse, can this model in any sense be a candidate for a phonological universal?

First of all, the fact that some of the "boxes" (i.e. of the features in the hierarchy) have no descriptive justification in a given language, does in itself not prove that the model is not universal. As long as a language (with a less rich range of possible consonant combinations than e.g. Danish) does not offer counter-examples to the orderings predicted by the general model, then it is in accordance with the model in the sense that a strict subset of the hierarchical features in the same order will be a relevant model for its syllabic structure (just as the model in fig. 2 is no counter-example of the validity of the model in fig. 1 for Danish).

But cases like initial [ŋk-, mb-] in many African languages and [mgl-, lg-] in Russian are more serious to the hypothesis. However, [ŋk-, mb-] can be considered counter-examples only if [ŋ, m] in these clusters are voiced and the "clusters" in fact function as such, i.e. only if [ŋk-, mb-] do not function as unit phonemes (as "prenasalized stops" or the like). And in the cases where /ŋk-/ etc. are true combinations and thus apparent counter-examples to the hypothesis, it is a question whether the nasal does not constitute a syllabic peak (or a "mora") of its own (as could be shown in a tone language if /ŋkɑ/ may have different tones on /ŋ/ and /ɑ/). In all probability, however, there will turn out to be some real counter-examples to the hierarchy (cf. the description of the Mazateco syllable by Pike and Pike 1947). In that case, the syllabic hierarchy might be considered the maximally "natural" or "unmarked" arrangement of distinctive features in the syllable (a universal tendency, if you like), but not a universal law without exceptions.

As for the Russian examples, the violation of the model in fig. 1 is clear. But notice that fig. 1 will apply to Russian if the feature [sonorant] is removed from the model. And notice, furthermore, that it is not possible to reverse any of the hierarchical features if the model is still to be consistent with the data. I.e., it is not true in Russian that sonorant consonants cannot be nearer to the syllabic peak than voiced obstruents (e.g. [gl-, br-]). This may lead to the following hypothesis concerning the universality of fig. 1: The hierarchical features are [+syllabic, -consonantal, +sonorant, +voiced]. Each language takes all of these features or any subset of them in the order given to form its syllabic hierarchy. Viewed from this angle the hypothesis predicts that the syllabic hierarchy for a given language will always be a strict (ordered) subset (in the sense in which fig. 2 is a strict subset of fig. 1) of the universal hierarchy. I.e., if a certain universally hierarchical feature plays a role for the syllabic hierarchy of a given language, then its place in that language-specific hierarchy is predictable from the universal hierarchy.

It is clear that the explanatory force of the model for a given language increases when the proportion of mirror-image ordering relations which are accounted for by the model (see the following section) increases. If this proportion is small, then the model is unsuccessful with respect to the language in question. And since a hierarchy must contain at least two units, it will be nonsense to speak about [syllabic] as the only feature involved in the syllabic hierarchy in a given language (phonologically speaking, this would only amount to saying that the syllabic peak occurs within the syllable). In languages with no hierarchical features the predictive value of the model is, of course, nil. And languages which use two or more of the hierarchical features in their syllabic hierarchy, but in the reverse order of that predicted here, will be counter-examples to the hypothesis suggested above.

The feature hierarchy proposed here agrees well with the psycho-linguistic test results reported by Pertz and Bever (1973). These authors tested a number of the markedness relations among different consonant clusters postulated, mainly on the basis of typological evidence, by Greenberg 1965 and Cairns 1969 (of the type 'initial cluster AB is more marked than initial cluster CD'). A number of monolingual English-speaking persons, both adolescents and children, were asked which cluster was likely to be found in the largest number of languages, and their responses agreed (with one exception, see below) with the postulated markedness relations (so that the less marked cluster was supposed by the test subjects to occur in more languages). The exception was that the sequence liquid-nasal seemed to be less marked than liquid-voiced obstruent, in disagreement with Greenberg and Cairns (who did, in fact, not distinguish between voiced and voiceless liquids; this is, of course, an error, cf. Pertz and Bever 1973, p. 72).

The feature hierarchy proposed here in fact <u>explains all</u> the test results of Pertz and Bever, in connection with the following two very natural axioms: (i) a cluster which violates the hierarchy is more marked than one which does not, and (ii) if two clusters both violate the hierarchy, and if they both begin with a certain hierarchical feature, then the cluster which violates the hierarchy "with more steps" is the more marked one (e.g. liquid-nasal is less marked than liquid-voiced obstruent, and nasal-voiced obstruent less marked than nasal-voiceless obstruent).

Notice that this explanation of their results does <u>not</u> depend on the <u>universality</u> of the syllabic hierarchy proposed here: it only presupposes that the syllabic hierarchy is valid for the language of the test subjects, i.e. English (which it clearly is). As a matter of fact, I am anything but convinced by Pertz and Bever's general line of reasoning that if English-speaking test subjects can react differently to clusters which do not occur in English, and which have the same "Distance from English" in the sense of Greenberg and Jenkins 1964, then they must have internalized a universal hierarchy of clusters. They do not at all consider the possibility that the internalized hierarchy may be relevant only to English and more or less "English-like" languages.

4.4 Hierarchical, semi-hierarchical, and cross-classificatory features

According to the hypothesis sketched here, the following features in the order given constitute the "syllabic hierarchy": [+syllabic, -consonantal, +sonorant, +voiced][1]. Of these features, [syllabic] differs from the others in that its phonetic correlates are undefined. The content of this feature label will be discussed in the following section.

Other distinctive features are not hierarchical, but cross-classificatory (cf. the distinction between a hierarchical

---

1) Or [αsyllabic, -αconsonantal, αsonorant, αvoiced] (cf. the implication chains of section 4.2 above). The reason why we need to write a coefficient together with the feature dimension is, of course, that [consonantal] always has the opposite coefficient of the other hierarchical features. This is due to the fact that we have chosen (like Ladefoged 1971, p. 108) not to deviate unnecessarily from the usual feature labels. But it would have been more satisfactory to call the feature [vocoid] (in the sense of Pike 1943), not only because it is [+vocoid] (but [-consonantal]) which is defined by a configuration (as opposed to a disjunction) of distinctive features (cf. Ladefoged ibid.), but also because it is [+vocoid] which plays a part in the syllabic hierarchy like the positive values of the other hierarchical features.

(McCawley) and a cross-classificatory (SPE) conception of grammatical boundaries, discussed in section 3.3 above). This is true of the feature or features accounting for different places of articulation (cf. section 3.2 above), the features accounting for differences of vowel quality, and for further features characterizing different "articulation modes" of consonants (like the features characterizing sibilance, stridency, aspiration, etc.; I shall not commit myself as to what the exact inventory of such features is).

It is an important characteristic of most cross-classificatory features that they are not involved in ordering relations, whereas they are sometimes essential for the statement of other distributional restrictions like "aspirates do not combine", "labials do not combine". However, some cross-classificatory features can, in fact, be involved in ordering relations, but then these ordering relations are not mirror-image-like. E.g. s and the plosives in Danish are not distinguished by any hierarchical feature. Nevertheless, the ordering relation obtains in the initial part of the syllable that s precedes a plosive, but it is well known that the inverse relation does not hold in the final part of the syllable (e.g. aks, ask, cf. Vestergaard 1968, p. 64f). Similarly, it is a general ordering restriction in the final part of the syllable that sequences of non-sibilant obstruents end in a dental (Basbøll 1973, p. 127). But the inverse relation does not hold in the initial part of the syllable, etc.[1]

---

1) Jørgen Rischel suggests (personal communication) that the ordering relations which are here called "mirror-image-like" should be termed "distance relations" instead, since they refer to the (relative) distance (of non-syllabic segments) from the syllabic peak. According to this terminology, cross-classificatory features are not involved in distance relations (whereas they can be involved in other ordering relations). I entirely agree with this suggestion.

There is at least one phonological feature in Danish which is neither hierarchical nor cross-classificatory in the above-mentioned sense, viz. [continuant].[1] Within the sonorant part of the syllable this feature seems to be hierarchical (see fig. 5):

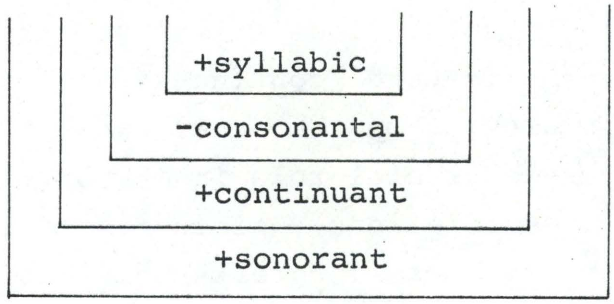


Fig. 5

---

1) By different linguists the nasal consonants have been defined either as [+continuant], viz. if the definitions are acoustically/auditorily based (e.g. Jakobson, Fant and Halle 1952), or as [-continuant], viz. if the definitions are articulatorily based (e.g. SPE and Ladefoged 1971, who uses [α stop] as equivalent to [-αcontinuant] in SPE). I follow the latter definition which seems to me to fit well into the general framework of distinctive features: The nasals in a given language regularly follow the same pattern as the stops (as opposed to the fricatives) with regard to place of articulation. Furthermore, the natural class consisting of vocoids, laterals, and nasals is already defined by the feature [+sonorant] (it is, of course, due to the nature of the human speech apparatus that a sonorant with no passage of air through the mouth must be a nasal). The most important argument for the classification of nasal consonants as [-continuant] within the present framework is, however, the fact that it permits the class of vocoids (i.e. the [-consonantal] segments) to be defined as the segments which are [+sonorant, +continuant, -lateral]. And there seems to be no way of defining this natural class (which includes, of course, the nasalized vocoids) as equivalent to a configuration of already established distinctive features if the nasal consonants are [+continuant], i.e. [-consonantal] will not be a cover feature in the sense of Ladefoged 1971. This seems to me a serious drawback of this alternative definition of [continuant].

The following implication chains hold in Danish syllables with a full vowel as peak:

[+syllabic] ⊇ [-consonantal] ⊇ [+continuant]

and: [-continuant] ⊇ [+consonantal] ⊇ [-syllabic]. But notice that it is not true that [+continuant] ⊇ [+sonorant], or [-sonorant] ⊇ [-continuant], since fricatives are [+continuant, -sonorant]. I.e., the feature [continuant] cannot be included in the syllabic hierarchy. On the other hand, the feature [continuant] can play a crucial role in mirror-image ordering relations (or distance relations), since nasals (i.e. [+sonorant, -continuant]) generally occur in greater distance from the syllabic peak than laterals (which are [+consonantal, +sonorant, +continuant] within the present framework). I therefore use the label "semi-hierarchical" for a feature like [continuant].

I have not used the feature [nasal] in the discussion for two reasons. First, this feature is not hierarchical within the sonorant part of the syllable in languages with a contrast between oral and nasalized vowels (like French), whereas it follows from our definitions that [continuant] will always be "semi-hierarchical" (since [-consonantal] always equals [+sonorant, +continuant, -lateral], and since the relation between (sonorant) laterals and (sonorant) nasals seems universally to be that indicated here, cf. Jespersen 1897-99, p. 523ff). Second, in languages like Danish which do not have phonologically relevant nasalized vowels, [+nasal] can, in the phonology, be used as a cover feature which is always equal to [+sonorant, -continuant]. It is thus completely redundant in the phonology of e.g. Danish (just as [consonantal] is completely redundant since [-consonantal] always equals [+sonorant, +continuant, -lateral]; but it may very well, of course, be economical to include redundant features in the phonological description).

I have not included the feature [lateral] in the discussion for a similar reason. It follows from our definitions that the class of sonorant laterals (that is, all the laterals which exist in Danish at the level under discussion here, viz. l) can be defined as [+consonantal, +sonorant, +continuant]. This means that the class of (sonorant) laterals equals the class of consonantal sonorants minus the nasals. I.e., the place of the feature [-lateral] in that part of the syllabic hierarchy which only comprises the continuant sonorants would necessarily be equal to the place of [-consonantal] and thus completely redundant. All this is, of course, a consequence of the fact that we do not recognize the existence of any consonantal sonorants which are not laterals or nasals. (In languages with a contrast between sonorant and obstruent laterals, the feature [lateral] will be cross-classificatory.)

4.5 The feature [syllabic]

In section 4.2 above we said that the feature [syllabic] indicates that the segment in question constitutes the peak of its syllable. It was also emphasized that the articulatory and acoustic correlates of this (auditorily defined) feature are largely unknown.

In this section we shall briefly examine two questions: (i) is the feature [syllabic] redundant at the systematic phonetic level, i.e. can the occurrence of the syllabic peak in a syllable be predicted from the occurrence of other distinctive features in the chain of segments which constitutes the syllable? and (ii) which role does the feature [syllabic] play in a generative phonology of Danish, i.e. must it be specified in the phonological representations in the lexicon, can the coefficients of this feature change in phonological derivations, etc.?

The features involved in the syllabic hierarchy in Danish except [+syllabic], i.e. [-consonantal, +sonorant, +voiced], together with the semi-hierarchical feature [+continuant] define the following syllabic hierarchy of segment types:

vocoids (i.e. [-consonantal] ≡ [+sonorant, +continuant, -lateral])

laterals (i.e. [+lateral] or [+consonantal, +sonorant, +continuant]

nasals (i.e. [+sonorant, -continuant])

voiced obstruents (i.e. [-sonorant, +voiced]; they are in Danish always [+continuant] too)

unvoiced segments (i.e. [-voiced]).

In the spirit of Jespersen (1897-99, p. 525, also 1926, p. 112)[1], we shall (in fig. 6) give a graphical illustration of the "sonority relations" in a number of segment sequences on the systematic phonetic level:

---

1) I differ from Jespersen in that I do not distinguish between voiceless stops and fricatives (his groups 1a and 1b, 1897-99, p. 523f, and 1926, p. 191), and between sonorant r-sounds (i.e. the "syllable final /r/") and vocoids, nor between vocoids of different height (i.e. his groups 4-8). However, in order not to get a high number of counter-examples to his scheme, viz. all the initial clusters of s plus a stop, Jespersen in the graph considers 1a, 1b as one group. Within his framework, this is nothing but a trick; within our model, however, this is due to the fact that no feature(s) distinguishing between s and the stops can be arranged into the syllabic hierarchy, quite apart from any ordering considerations (e.g. it is the very existence of the class of fricatives that makes it impossible to include [continuant] into the syllabic hierarchy, see the preceding section).

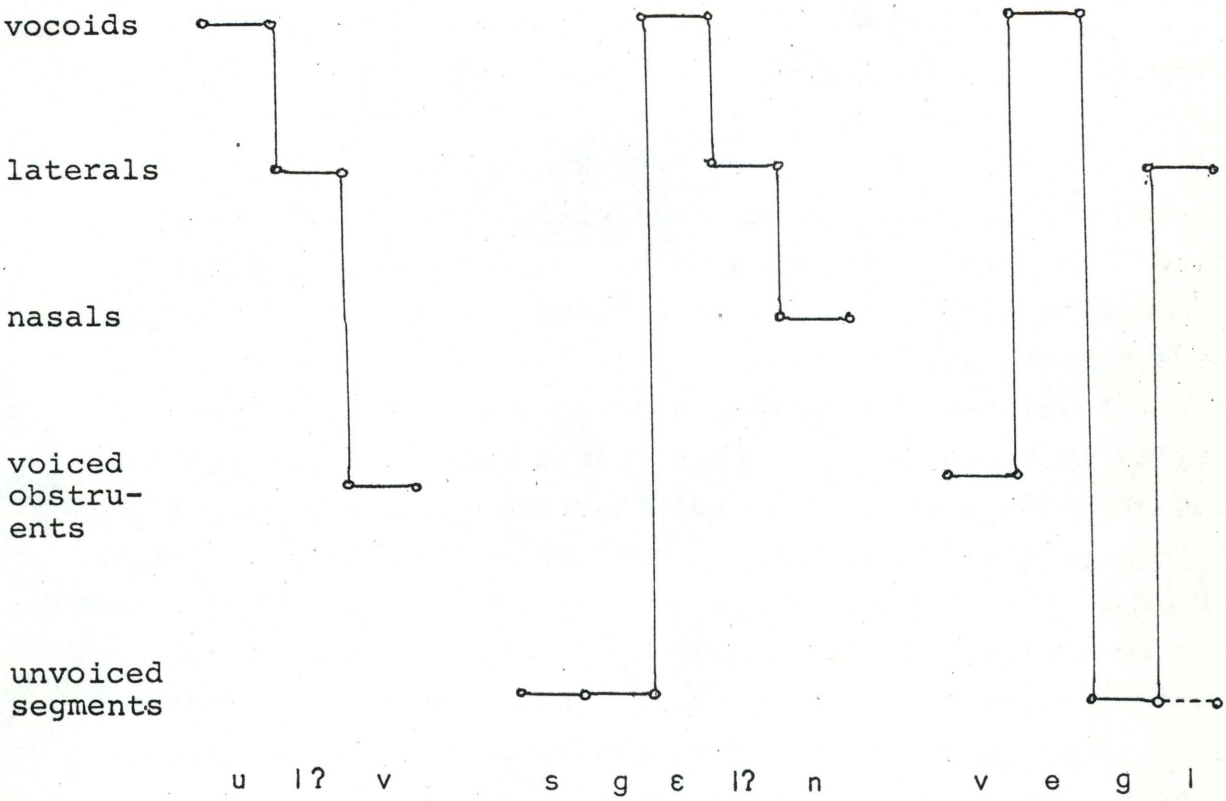


Fig. 6

The relative maximum is the syllabic peak. [ul?v] is a normal monosyllabic word with [u] as its peak (notice that stød is not considered a segment in this connection). The imperatives skeln and vikl both exhibit word-final consonant clusters which do not occur outside imperative forms. However, skeln is always pronounced as a normal monosyllable, and it is clear from the graph that it does not violate any systematic order restriction based upon the syllabic hierarchy (i.e. distance relation-restrictions). vikl, on the other hand, will, according to our model, have two syllabic peaks if it is pronounced with a fully voiced

[l] (and this violates the general principle that imperatives of verbs whose infinitives end in shwa are shorter by one syllable than the infinitive form); however, if they are pronounced with an unvoiced [l̥] instead (the dotted line in fig. 6), the number of syllables will be respected according to the model (cf. Basbøll 1970, p. 21ff). Whether the syllabic peaks predicted by the model are in fact also heard as syllabic peaks (e.g. whether vikl pronounced with a voiced [l] is a perfect rhyme word to Mikkel pronounced without the vocoid [ə]) ought to be systematically tested.

It should be emphasized, however, that the location of the syllabic peaks is not always predictable from the sequence of segments (defined by their distinctive features, except the feature [syllabic]), neither at the phonetic surface nor at the level of description which has been used throughout the present discussion. For example, a narrow unrounded palatal vocoid followed by a narrow rounded velar vocoid can be realized as [i̯u] (e.g. in just [i̯usd]) or as [iu̯] (e.g. in ivrig [iu̯ʁi]). Similarly, there is a possible contrast between [i̯ɒ] or [i̯ʌ] (e.g. in hjort, j [i̯ɒ:d, i̯ʌð]) and [iɒ̯] = [iʌ̯] (e.g. in ir [iɒ̯, iʌ̯]; the transcriptions with [ɒ̯, ʌ̯] are mere notational variants), cf. French examples like houille, oui [ui̯, u̯i] (in pays, paye [pɛi, pɛi̯] there is also a distinction in the number of syllables).[1] Examples with consonants (which differ as to number of syllables) are skeln, ellen [sgɛlʔn, ɛlʔn̩ = ɛl̩ʔn = ɛlʔl̩n] and vidn, bidden [viðʔn, biðʔn̩ = bið̩ʔn = biðʔð̩n] (the latter transcriptions, i.e. [ɛlʔl̩n, biðʔð̩n], are probably the most phonetically correct ones in such contexts, at least in younger Danish standards, if one is forced to make a segmental distinction between the forms, see the following section; it may, however, be better to say that the sequence of segments are identical in

---

1) Danish examples like naivist vs. nej vist! are also distinguished by their number of syllables (in addition to vowel quality distinctions): [naivísd, nɑi̯vésd].

the mono- and bisyllabic forms, and that the distinction in syllabicity is (phonetically) prosodic[1]).

Although the location of the syllabic peaks in the majority of cases can be predicted from the sequence of segments (defined by their distinctive features except the feature [syllabic]) occurring at the phonetic surface or at the level of description which has been used throughout the present discussion, this is thus clearly not always feasible. The conclusion is, within a generative framework, that the feature [syllabic] must be present in underlying forms. The coefficient for this feature is only changed by the rule shwa-assimilation, to be discussed in the following section, and maybe by some very late reduction processes. But apart from these exceptions, the feature [syllabic] is unchanged throughout the derivation: e.g. when underlying postvocalic voiced obstruents are turned into vocoids (viz. /ɣ, v, r/ ⟶ [i̯/u̯, u̯, ɒ̯]) those vocoids are non-syllabic, i.e. glides; and when adjacent vocoids are completely assimilated (see the following section), e.g. /di:ə/ ⟶ [di:i], the number of syllables is generally kept constant. In other words: the coefficients for the feature [syllabic] (which is redundant in the phonological representation, i.e. each underlying segment has its "natural" value for this feature, see below) are kept constant throughout the phonological rules (excluding, of course, morphological rules), apart from the rule of shwa-assimilation; this is true even when segments are changed

---

1) This agrees well with Rischel (1970c) who has found that F_o seems to be an important cue for syllabicity in Danish. This fact may also help explain why voiceless segments generally do not qualify as syllabic peaks (see the following section). Rischel (1964) also gives an interesting account of syllabicity as a level of culminative contrast below the stress levels.

in such a way that their coefficient for [syllabic] is no longer the natural one for the derived segment in the context in question. This agrees well with the idea (cf. Rischel 1970c) expressed in the preceding footnote, viz. that the feature [syllabic] is phonetically prosodic.

Hjelmslev (1951) proposed that words like *fadder*, *vammel* [faðʔʌ, vɑmʔl̩] should be derived from monosyllables: /fadr, vaml/, thus explaining the obligatory occurrence of stød (the final consonant clusters have stød-basis, since they begin with a voiced consonant). It is uncertain whether such monosyllabic underlying forms for disyllabic phonetic forms should be postulated in a generative phonology of Danish (cf. Basbøll 1972b, p. 13, 23ff). If they are, the syllabic hierarchy predicts that /r, l/ of /fadr, vaml/ will form a syllabic peak, presupposing that the final /r/ has already, at this point of the derivation, been changed to the corresponding (pharyngeal) vocoid. The underlying non-syllabic /r, l/ can thus be rewritten as the corresponding syllabic segments. Another possibility is that a rule which is sensitive to the syllabic hierarchy inserts a shwa to break up the non-permissible final cluster. (Under the presupposition that the final /r/ is a vocoid at that point of the derivation, the syllabic hierarchy explains why final clusters like /lr, ŋr/ mentioned under principle (ii) in section 3.4.1 above are *systematically* excluded, viz. excluded by virtue of the distance relation-restrictions.) Both possibilities permit the preservation of the redundancy constraint on phonological information in the lexicon that all consonantal segments are non-syllabic and vice versa, which, of course, also applies if *fadder*, *vammel* etc. end in /ər, əl/ in their most underlying forms.

As already mentioned, the feature [syllabic] is redundant in phonological representations since its coefficient is always the opposite of the coefficient for [consonantal]; i.e., all segments in the fully specified phonological representations will be [αsyllabic, - αconsonantal]. (I do not claim that the feature [syllabic] is redundant in the phonological representations of all languages, only that it is redundant in Danish.) An economical way of doing this seems to be the following: All segments which in the incompletely specified distinctive feature matrix are [-consonantal] will get the redundant specifications [+syllabic, +sonorant, +continuant, -lateral] (and probably others as well, like [-sibilant]); all segments which in the incompletely specified distinctive feature matrix have at least one of the following specifications: [-sonorant], [-continuant], or [+lateral], will get the redundant specifications [-syllabic, +consonantal] (and probably others as well, like [-round]).

4.6 Shwa-assimilation

It is well known among students of Danish phonetics and phonology that shwa can be "assimilated" in different ways to neighbouring sounds.[1] There is no general agreement on how to transcribe the results of the shwa-assimilation, however, and this is probably due to the fact that the acoustic and articulatory correlates of syllabicity in Danish are largely unknown; in fact only one very preliminary investigation of this subject has been published, viz. Rischel 1970c. According to Rischel[2],

1) E.g. Jespersen 1897-99, p. 463ff, Uldall 1933, p. 3ff, Andersen 1955 passim, Hansen 1956, p. 40ff, Fischer-Jørgensen 1962 passim.

2) Rischel emphasizes (personal communication) the highly preliminary nature of his report, particularly concerning the perceptual aspect. It is also true that the number of persons investigated is small. The investigations are planned to be continued.

some speakers of standard Danish have a consistent difference in quantity and F_0 in assimilated words of the type hårde, hårdere ([hɒ:ɒ, hɒ:ɒɒ] or the like), and a pilot experiment with a synthetic stimulus suggested that the word types can be identified correctly when the only difference in the stimuli is in F_0. All this points towards a prosodic function of syllabicity (somewhat resembling that between accent 1 and accent 2 in Swedish and Norwegian, so that Danish $C_0VC_0ə$ words correspond to accent 1, and $C_0VC_0əC_0ə$ to accent 2). A transcription indicating the location of the assimilated shwas by a hyphen seems to be rather satisfactory from this point of view, in comparison with transcriptions indicating syllabicity as occurring in one segment only: skeln, vidn, ellen, bidden [sgɛlʔn, viðʔn, ɛlʔ-n, biðʔ-n].

Recently Brink and Lund (forthcoming) have given an alternative treatment of shwa-assimilation. They suggest roughly the following: It is an option for the speaker to retain the shwa or not. If shwa is not retained, it depends on the sonority of its neighbouring segments whether it is assimilated or deleted (sonority is understood basically as in Jespersen 1897-99). If the most sonorous neighbouring segment to shwa is a voiceless obstruent, then shwa drops without leaving any trace behind (e.g. passe [pasə, pas] where the second pronunciation is also that of pas [pas]). If the most sonorous neighbouring segment to shwa is a sonorant, shwa assimilates to that sonorant (in such a way that shwa either assimilates completely to the sonorant in question with the reservation that the assimilated segment is always syllabic, e.g. koen, ballen [kʰo:ʔon, ballˌn], or in such a way that shwa is elided while making the sonorant in question syllabic, e.g. gode, pæne, tale [go:ð̩, pʰɛ:n̩, tˢæ:l̩]; Brink and Lund set up a number of principles which they claim predict which type of assimilation will be chosen in a certain context).

If there can really be given substance to the claim that shwa is assimilated to its most sonorous neighbouring segment, then data from the treatment of shwa in different contexts can give independent evidence for the sonority hierarchy, suggesting that sonority may be an independently needed auditory dimension (a multivalued linearly ordered distinctive feature), although it can be defined by distinctive features that are needed in the phonology for independent reasons. And I see no reason, after all, why a prosodic treatment of syllabicity, and a treatment in terms of a sonority hierarchy which has independent motivations, should in principle be incompatible.

References

Andersen, Poul 1955: "Dansk fonetik", in: Nordisk Lærebog for talepædagoger (Copenhagen), p. 308-354

Anderson, John M. and Charles Jones 1974: "Three theses concerning phonological representations", JL 10, p. 1-26

Austin, John 1971: Topics in Danish Phonology (unpublished thesis, Cornell University)

Basbøll, Hans 1969: "The phoneme system of advanced standard Copenhagen", ARIPUC 3/1968, p. 33-54

Basbøll, Hans 1970: "Notes on the phonology of Danish imperatives with a digression on vowel quantity", ARIPUC 4/1969, p. 15-42

Basbøll, Hans 1972a: "Some conditioning phonological factors for the pronunciation of short vowels in Danish with special reference to syllabification", ARIPUC 6, p. 185-210.

Basbøll, Hans 1972b: "Some remarks concerning the stød in a generative grammar of Danish", in: Kiefer, F. (ed.), Derivational Processes (Stockholm), p. 5-30

Basbøll, Hans 1973: "Notes on Danish consonant combinations", ARIPUC 7, p. 103-142

Basbøll, Hans 1974: "Structure consonantique du mot italien", Revue Romane 9, p. 27-40

Basbøll, Hans forthcoming: "Schwa, jonctures et syllabification dans les représentations phonologiques du français"

Brink, Lars and Jørn Lund forthcoming: Udtaleforskelle i Danmark (Copenhagen)

Cairns, Charles E. 1969: "Markedness, neutralization, and universal redundancy rules", Lg. 45, p. 863-885

Chafe, Wallace 1970: Meaning and the structure of language (Chicago)

Chomsky, Noam 1964: "Current issues in linguistic theory", in: Fodor, J.A. and J.J. Katz (eds.), The structure of language (Englewood Cliffs, New Jersey 1964), p. 50-118

Davidsen-Nielsen, Niels 1974: "Syllabification in English words with medial sp, st, sk", JPh 2, p. 15-45

Davidsen-Nielsen, Niels forthcoming: "A phonological analysis of English sp, st, sk with special reference to speech error evidence", to be published in: Journal of the International Phonetic Association

Dell, François 1973: Les règles et les sons. Introduction à la phonologie générative (Paris)

Derwing, Bruce 1973: Transformational grammar as a theory of language acquisition (Cambridge)

Ege, Niels 1965: "The Danish vowel system", Gengo Kenkyu (Journal of the Linguistic Society of Japan) 47, p. 21-35

Fischer-Jørgensen, Eli 1952: "On the definition of phoneme categories on a distributional basis", AL 7, p. 8-39, reprinted in: Hamp, E., F. Householder and R. Austerlitz (eds.), Readings in Linguistics II (1966), p. 299-321

Fischer-Jørgensen, Eli 1962: Almen fonetik (Copenhagen)

Fischer-Jørgensen, Eli 1973: "Supplementary note to Hans Basbøll's commentary on Hjelmslev's analysis of the Danish expression system", ALH 14, p. 143-152

Fudge, Erik 1969: "Syllables", JL 4, p. 47-68

Grammont, Maurice 1933: Traité de phonétique (Paris)

Greenberg, Joseph 1965: "Some generalizations concerning initial and final consonant sequences", Ling. 18, p. 5-34

Greenberg, J.H. and J.J. Jenkins 1964: "Studies in the psychological correlates of the sound system of American English", Word 20, p. 157-177

Hansen, Aage 1956: Udtalen i moderne dansk (Copenhagen)

Hjelmslev, Louis 1939: "The syllable as a structural unit", _Proc. Phon._ 3, p. 266-272, reprinted in: _Essais linguistiques II_ = _TCLC_ 14 (1973), p. 239-245

Hjelmslev, Louis 1951: "Grundtræk af det danske udtrykssystem med særligt henblik på stødet", _Selskab for nordisk filologi. Årsberetning for 1948-49-50_ (Copenhagen 1951), p. 12-24, English translation (by F.J. Whitfield) in: _Essais linguistiques II_ = _TCLC_ 14 (1973), p. 247-266

Hoard, James 1971: "Aspiration, tenseness, and syllabication in English", _Lg._ 47, p. 133-140

Hooper, Joan 1972: "The syllable in phonological theory", _Lg._ 48, p. 525-540

Jakobson, Roman 1941: _Kindersprache, Aphasie und allgemeine Lautgesetze_ (Uppsala), English translation: _Child language, aphasia and phonological universals_ (The Hague 1968)

Jakobson, R., G. Fant and M. Halle 1952: _Preliminaries to speech analysis_ (Cambridge, Mass.)

Jespersen, Otto 1897-1899: _Fonetik_ (Copenhagen)

Jespersen, Otto 1926: _Lehrbuch der Phonetik_ (4th ed.) (Leipzig/Berlin)

Kiparsky, Paul 1973: "On phonological representations", in: O. Fujimura (ed.), Three Dimensions of linguistic theory (Tokyo), p. 1-136

Kloster Jensen, M. 1963: "Die Silbe in der Phonetik und Phonemik", Phonetica 9, p. 17-38

Ladefoged, Peter 1971: Preliminaries to linguistic phonetics (Chicago)

Linell, Per 1974: Problems of psychological reality in generative phonology = ruul (Reports from Uppsala University, Department of Linguistics) 4

Malmberg, Bertil 1965: "Kring stavelseproblemet", Nyare fonetiska rön (Lund), p. 168-183

McCawley, James D. 1968: The phonological component of a grammar of Japanese (The Hague)

O'Connor, J.D. and J.L. Trim 1953: "Vowel, consonant and syllable; a phonological definition", Word 9, p. 103-122

Pertz, Doris and Thomas Bever 1973: "Sensitivity to phonological universals in children and adolescents", Working Papers on Language Universals 13 (Stanford University), p. 69-91

Pike, Kenneth 1943: Phonetics (Ann Arbor)

Pike, Kenneth 1947: Phonemics (Ann Arbor)

Pike, Kenneth and Eunice Pike 1947: "Immediate constituents of Mazateco syllables", IJAL 13, p. 78-91

Rischel, Jørgen 1964: "Stress, juncture, and syllabification in phonemic description", Proc. Ling. 9, p. 85-93

Rischel, Jørgen 1969: "Notes on the Danish vowel pattern", ARIPUC 3/1968, p. 177-205

Rischel, Jørgen 1970a: "Consonant gradation: A problem in Danish phonology and morphology", in: Benediktsson, H. (ed.), The Nordic Languages and Modern Linguistics (Reykjavik), p. 460-480

Rischel, Jørgen 1970b: "Morpheme stress in Danish", ARIPUC 4/1969, p. 111-144

Rischel, Jørgen 1970c: "Acoustic features of syllabicity in Danish", Proc. Phon. 6, p. 767-770

Rischel, Jørgen 1972: "Compound stress in Danish without a cycle", ARIPUC 6, p. 211-230

Rischel, Jørgen 1974: Topics in West Greenlandic phonology (Copenhagen)

Schane, Sanford 1973: Generative phonology (Englewood Cliffs, New Jersey)

Sigurd, Bengt 1965: Phonotactic structures in Swedish (Lund)

Spang-Hanssen, Henning 1959: Probability and structural classification in language description (Copenhagen)

Stanley, Richard 1967: "Redundancy rules in phonology", Lg. 43, p. 393-436

Uldall, H.J. 1933: *A Danish phonetic reader* (London).

Uldall, H.J. 1936: "The phonematics of Danish", *Proc. Phon.* 2, p. 54-57

Vestergaard, Torben 1968: "Initial and final consonant combinations in Danish monosyllables", *SL* 21, p. 37-66

INTERVOCALIC AFFRICATES IN PRESENT-DAY HUNGARIAN

Tamás Szende[1]

Present-day Hungarian has the following set of affricates: /ts/, /ts:/, /dz/, /dz:/, /tʃ/, /tʃ:/, /dʒ/, /dʒ:/, /c/, /c:/, /ɟ/, and /ɟ:/. All they have monophonemic status in the Hungarian sound pattern. All affricates satisfy the basic criterion in their acoustic structure in so far as they are set up as two-phase consonants where both the so-called stop (phase 1) and the explosion (phase 2) are substantial.

While the mean duration value of affricates generally exceeds that of the stops by 23 - 33 % (see e.g. the limit values of short stops: 0.065 - 0.120 s and the limits of short affricates: 0.080 - 0.160 s), the duration of [c], [c:], [ɟ], and [ɟ:] tends to that of the stops [t], [t:], [d], and [d:].

In intervocalic position certain combinations of phonemes turn up as affricates, as well, like [ts] for /t/ + /s/, [c:] for /t/ + /j/, and so on. At the same time some affricates in the speech flow correspond to different phonemes or combinations of phonemes, so e.g. [c:] to /c:/ in hattyú 'swan', to /t/ + /j/ in látja '(he is) seeing (that)', to /t/ + /c/ in hat tyúk '6 hens'.

Let us see some results of investigations carried out on intervocalic affricates.

(1) There are relevant differences between speech sounds which are realizations of (phonemic) affricates and combinations of phonemes pronounced as affricates respectively. Differences

---

1) Tamás Szende was guest research worker at the institute during 4 months in the spring of 1974.

in the acoustic features can be achieved in two ways which correspond to the two-phase structure of articulating affricates. They are distinguished either by the duration of phase 1 or by the relative intensity of phase 2, see the mean values:

[c:] for	phase 1	phase 2
/c:/	0.163 s	0.038 s
/t/ + /j/	0.160 s	0.040 s
/t/ + /c/	0.192 s	0.040 s

and [c:] for	phase 1	phase 2
/c:/		16 dB
/t/ + /j/		8 dB
/t/ + /c/		14 dB

Differences both in duration and in intensity surpass the threshold of perception. In this way the speaker is able to assert his 'phonemic intentions' by altering some of the acoustic parameters of his speech sounds.

(2) The whole articulatory formation, the several parameters, and the acoustical substance of the affricates are the function of the position within the sequence they occur in. The greater namely the relative articulatory energy of the phrase in which the affricate is found, the more stable the intrinsic articulatory features of the affricate realized in it. The features will be less marked when the affricate occurs in unstressed words and is placed in a longer phrase than those which appear in short, emphasized phrases, thus e.g.:

[tʃ] <u>kicsi</u> (long phrase, unstressed) 'little, small'

mean duration value	0.105 sec
relative intensity	11.4 dB

[tʃ] <u>csúcson</u> (short, slow phrase) 'on the peak'

mean duration value	0.135 sec
relative intensity	17.5 dB

(3) A statistical evaluation of the data shows that the affricates adapt themselves to their position in the speech flow in different ways. That means, some of them are more resistant if put to the eventualities in setting up the actual phrase than others are. So e.g. [ts] looses more in its inherent articulatory features if the relative energy of articulation decreases, while [tʃ] is more stable, and the same is valid for [dz] vs. [ts], and [c] vs. [ɟ], and so on. The rule is that the more complex the articulatory construction of an (intervocalic) affricate, the less susceptible to the influence in the speech event.

SOME AIRFLOW AND GLOTTOGRAM DATA ON DANISH WHISPER

Philip Mansell

1. Introduction[1]

In Mansell (1973) I reported on some experiments designed to look for evidence of segmental articulatory reorganization from normal speech to whisper, where such reorganization would be seen as paralleling the reorganization found at the supra-segmental level (see Trim (1973) for a review and some experimentation). What was of particular interest in the above paper was the segmental behaviour of the glottis in phonologically voiced and voiceless stops in English and the corresponding aspirated/unaspirated series in Danish.

It was expected that in the English data the three parameters measured, airflow out of the mouth, air pressure drop across the articulatory constriction, and photo-electric glottograph traces of the gross opening/closing movements of the larynx, would be sufficient to enable all relevant physical differences between voiced and voiceless cognates and between normal and whispered speech to be characterized. Any problems in the physical interpretation of the Danish data, where only airflow and glottographic traces were sampled, were to be referred to the English data for possible explanations. However, the finding

---

1) The recordings on which this paper is based were made at the Institute of Phonetics in the University of Copenhagen while the author was a visiting researcher in the Summer Term of 1971. The assistance of Prof. E. Fischer-Jørgensen and her staff, especially of B. Frøkjær-Jensen, who acted as subject, is gratefully acknowledged.

both that glottographic registration was poor during whispered speech and that in any case no great reliance could be placed upon the quantitative aspects of the glottographic traces lead inevitably to a good deal of uncertainty about the time-varying behaviour of the individual components of the sound production in both languages and for both normal and whispered speech.

The stops examined in Mansell (1973) exhausted neither the English nor the Danish material, however, and it is the purpose of this paper to present a more complete account of the Danish data. This comprises utterance medial, syllable initial [p, b, m, f, v]. It will be of great interest to discover whether those aspects of the stop traces which led to the tentative conclusion that segmental articulatory activity was to some extent reorganized during whisper will also be evidenced on the extra data considered here. Where the findings below have been reported on previously this will be noted; the great majority of the material below, however, is reported for the first time.

2. Method

A single adult male native speaker of Danish repeated the nonsense words[1] [pil], [bil], [mil], [fil], [vil] in the frame:

en lille ____ igen

The words were repeated in the order given above, each one being first spoken normally, then whispered. In the event thirteen error-free repetitions of each word were obtained.

Both the linguistic nature of the stimulus items and the experimental method were to some extent determined by the instrumentation used. Airflow out of the mouth was sampled *via* a 2-channel Frøkjær-Jensen Aerometer. Back vowels in combination with

1) The required pronunciation of these words was without stød. It has not been possible in the preparation of this paper, however, to check for this in the original tapes, and the possibility exists that the stød is at least occasionally manifested under the influence of the normal Danish pattern for such words. It is thus possible that the measure C described below is unreliable.

the front consonants of interest were excluded in order to avoid large tongue movements which might introduce volume changes in the Aerometer mask, and consequently errors in the traces. The audio was picked up by a microphone in the Aerometer mask. Glottis movements were registered via a Frøkjær-Jensen photo-electric glottograph. The mode of operation of this device excludes low vowels from the corpus since pharyngeal narrowing interferes with the photocell above the larynx.

It was found advisable to have the subject produce normal voiced/whispered pairs immediately after each other, thus making sure that the position of the light source against the neck, a potent source of artefacts with this device, was as similar as possible for both members of the pair. (For these points and a general technical description of the device, see Frøkjær-Jensen et al. 1971.)

All three experimental parameters were registered simultaneously on a Mingograph ink-jet recorder, running at 100 mm/sec, the audio being at the same time recorded on tape. The airflow channel for normal speech was low-pass filtered with a time constant of 40 msec to remove the voicing ripple. This filter was left in the circuit for the whispered cases so that the maximum amount of comparability could be maintained.

2.1 Measurement parameters

The two aspects of the traces which are of theoretical interest are:

(a) how the (phonologically) "same" units are variously manifested in normal speech and in whisper

(b) how the linguistic contrasts[1] are manifested in normal speech and in whisper.

---

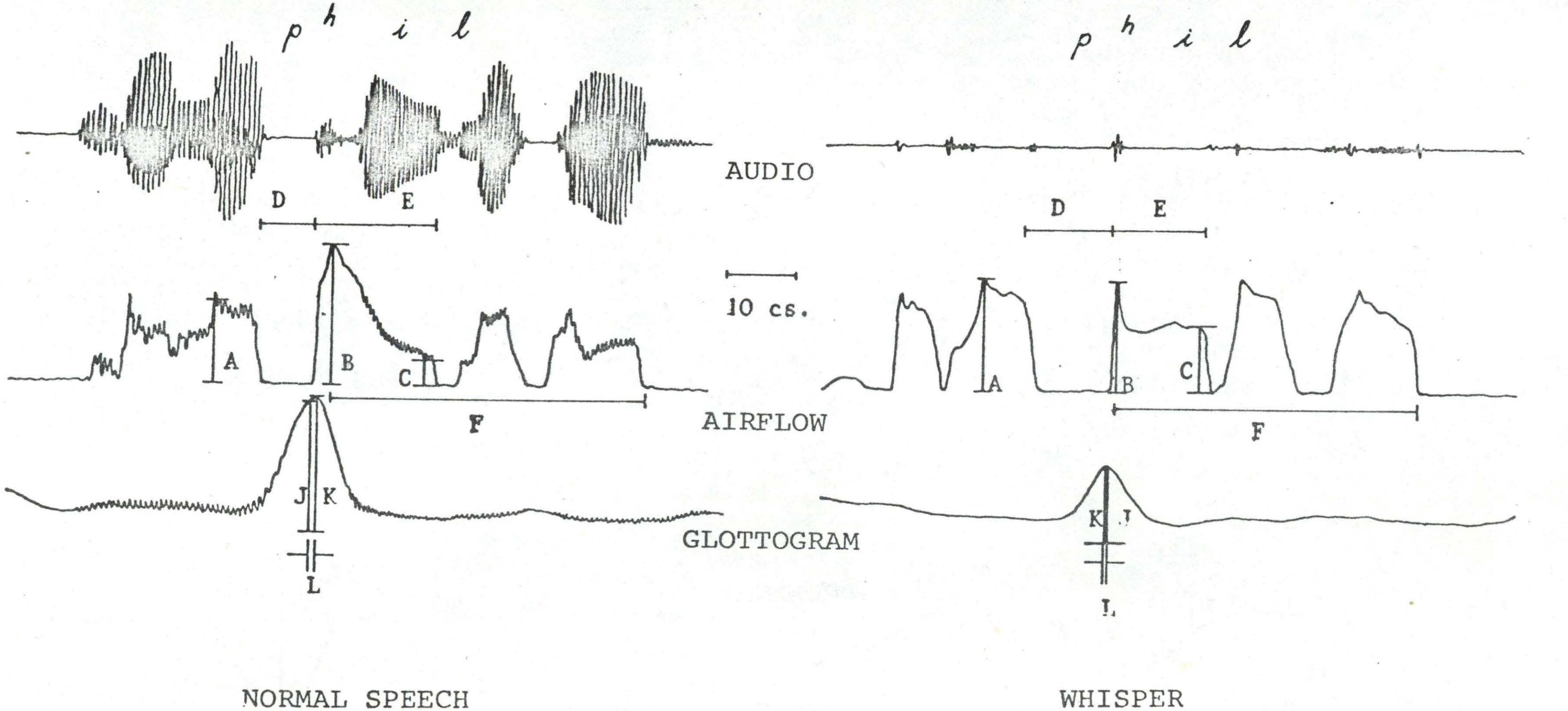
1) Aspirated vs. non-aspirated ([p]/[b]); nasal vs. non-nasal ([p,b]/[m]), voiced vs. voiceless ([f]/[v]).

Of further interest is the evidence provided by this larger corpus on the reliability of glottographic traces, a question already examined, as noted above, in Mansell (1973).

These points will be examined below in terms of the parameters depicted in figure 1 (for stops) and figure 2 (for fricatives). These parameters require some comment, since the considerations which prompted their choice are not always self-evident.

In the first place, a general measure of the effect of whisper on airflow was needed that would be independent of the consonants of direct interest. The obvious choice lay in measuring the airflow amplitude in the vowels [ə] and [i] which flank the consonants in each case (measurements A and C on figures 1 and 2). For [ə] the point chosen for measurement was the amplitude achieved immediately after the [l], rather than at the highest point, the measure which is used in Mansell (1973; see table 3.6 there). The reason for this is that this latter measure is required for other purposes in the fricative traces. Here, since in whisper the fricative cannot be segmented reliably from the preceding vowel, the highest point of pre-consonantal airflow is taken to be a measure of the airflow at the start of the fricative constriction; this measure was hence not available for vocalic purposes. For [i] the point chosen was the amplitude before the sudden drop in the trace for the following [l]. Here the difficulty of establishing any other point reliably, or once again of segmenting in the whispered traces such that measurement could be made at the onset of [i] should be evident.

The validity of this measure is, however, in some doubt, given the remarks in section 2. above about stød manifestation, since the amplitude of the airflow at this point would be greatly dependent upon the occurrence of a following stød. Since the

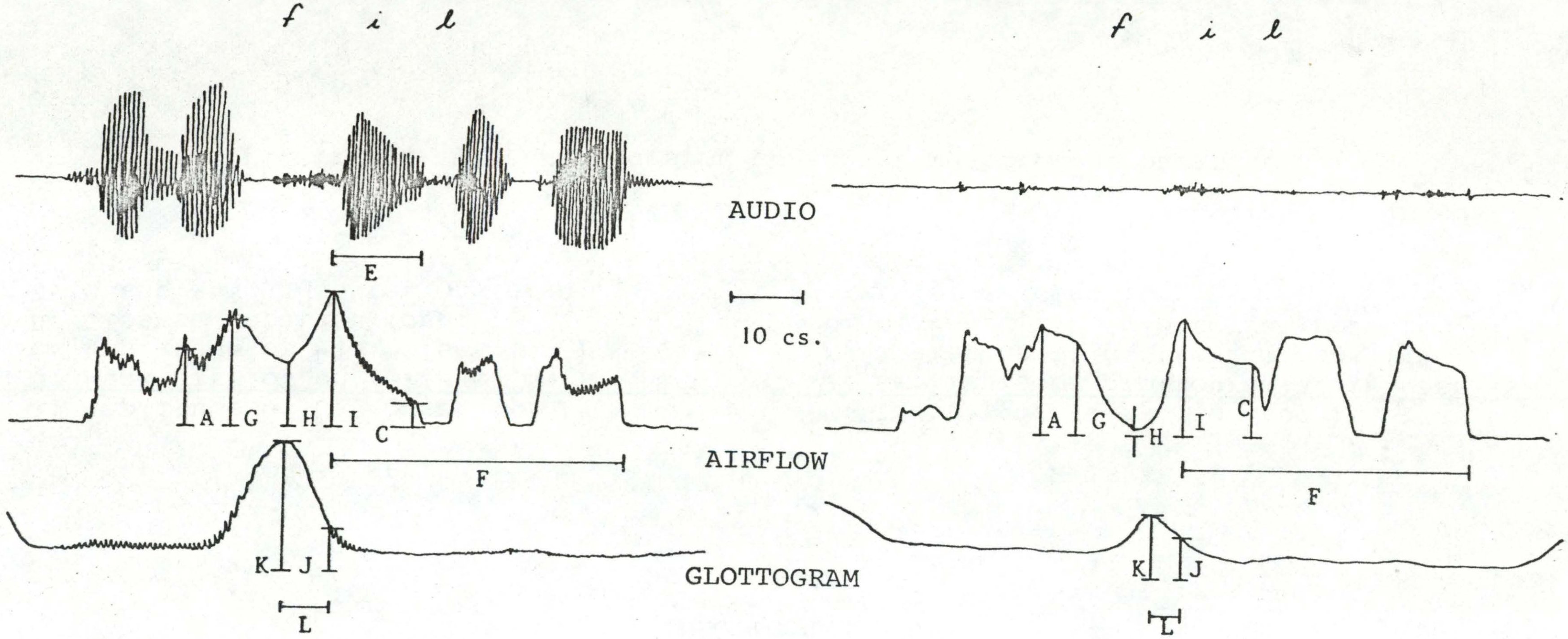


A: Amplitude of [ə] onset (mm)
B: Peak airflow on release (mm)
C: Amplitude of [i] offset (mm)
D: Closure duration (csec)
E: Burst and [i] duration (csec)

F: "Total duration" (in csec)
J: Amplitude of glottogram at release (mm)
K: Amplitude of glottogram peak (mm)
L: Distance from glottogram peak to release (csec)

Figure 1

PIL traces illustrating stop consonant measurement parameters



NORMAL SPEECH

A: Amplitude of [ə] onset (mm)
C: Amplitude of [i] offset (mm)
E: [i] duration from 2nd. peak (csec)
F: "Total duration" (in csec)
G: Amplitude of 1st. consonantal airflow peak (mm.)

WHISPER

H: Amplitude of consonantal airflow trough (mm)
I: Amplitude of 2nd consonantal airflow peak (mm)
J: Amplitude of glottogram at release (mm)
K: Amplitude of glottogram peak (mm)
L: Distance from glottogram peak to release (csec)

Figure 2

FIL traces illustrating fricative consonant measurement parameters results on this parameter do not in the event show great variation on the normal voiced traces, though, and since the theme of stød realization in normal speech vs. whispered speech cannot be gone into on the basis of the present limited material, the measure is allowed to stand.

The measures B (amplitude of airflow at release) and D (duration of the closure phase) require no particular comment beyond the acknowledgement that in B explosion and aspiration are confounded.

The measure E gives us the duration of the post-consonantal vowel in the syllable of interest, but with the duration of the explosion plus aspiration phase counted in with the vowel in the case of [p]. In part, of course, this practice is again dependent upon the difficulty of segmentation in the whisper case, but there are a number of authors (see Lehiste (1970) for a review) who would argue that such a procedure was in any case meaningful. F, labelled "total duration", is the most ad hoc parameter because the most completely tied to segmentation difficulties. What was required was some general means of estimating whether whispered speech was longer or shorter in duration than normal speech. It seemed most sensible to try to measure this on the frame rather than on any stretch including the consonants of interest, and the problem was simply to find a demarcatable stretch long enough. In the end measurement was made, as shown in figure 1, from the peak airflow on release to the point where the trace dips to the low level for the final [n] of igen.

The fricative traces shown in figure 2 share with the stop set the measures A and C for the amplitude of frame vowels. Total duration, F, is also measured as in the stop traces, the second peak of the airflow being taken, as before, as signalling the release. The measure E, duration of the post-consonantal vowel, could only be taken for the normal voiced FIL and VIL traces, and was here defined as the period when the vocal folds are vibrating for the vowel, since there was no period definable between release and onset of the vowel in the whispered cases.

This segmentation problem is also encountered when airflow behaviour in the fricative consonants is in question. It happens, however, that all traces, normal voiced as well as whispered, could be segmented into a first peak of the airflow, followed by a trough, followed by a second peak. The amplitudes of the two peaks and of the trough (G, H, I) were taken accordingly as the airflow measures for the fricative consonants. The possibility of segmentation here may have imposed more unity than actually exists. Thus in VIL, for instance, the amplitude of airflow drops sharply at the end of the pre-consonantal vowel to an extremely low level, rising again only at the onset of the post-consonantal vowel. This low level was registered as nil on the aerometer, though in all probability a complete occlusion was not reached. Such traces can be segmentally measured like the FIL trace shown in figure 2. There is no immediate indication in the results below that such measurements are illicit, but the possibility nonetheless exists.

For all the above-discussed traces tests of central tendency will be carried out, both between the normal voiced and the whispered traces, and between the linguistic types in normal voice and in whisper. In this way it is hoped to provide a fairly detailed picture of aspects (a) and (b) above of the material.

The only glottographic measure which could be automatically made for all traces was the height of the glottogram above a constant arbitrary baseline at the moment of release (J), where release is defined as for measures B and I in the above paragraphs. The reason for this lies partly in the difficulty of segmentation in whisper, partly also in the fact that in some types there is no appreciable segmental movement of the glottal curve in the consonant, and hence no means of locating a point for measurement within the occlusive or constriction stage. The above is true of MIL and VIL traces in both normal speech and whisper. The results on parameter J will be used for tests of central tendency, as with the other traces, as well as for correlation tests with the stop and fricative airflow measures B and I.

The glottographic traces for PIL and FIL both voiced and whispered, and to a lesser extent BIL also, however, do show clear segmental activity in the form of peaking during the occlusive or constriction stage of the consonant. Moreover, inspection of the traces reveal that there are certain apparently systematic distinctions between the types in normal speech with respect to this peaking. Thus, for example, PIL peaks are larger than BIL peaks, while PIL peaks occur later than FIL peaks - in fact, as Mansell (1973) showed, the PIL peaks are more or less coincident with the release; for FIL, though, the glottogram peaks much more in the middle of the constriction phase. These points form the initial justification for the measurements K, height of glottogram peak in mm of trace above the same arbitrary baseline as is used for J, and L, distance in csec between the peak and the release. K and L are measured only for the types PIL, BIL and FIL, voiced and whispered.

It is planned to carry out tests of similarity of central tendency across types and voicing conditions, in order to provide a further set of comparisons between the glottal traces for normal voice and whisper. The measures are expected to be of greater interest in the case of the timing of the peak than in the case of the amplitude, where a great diminution from normal voice to whisper has already been noted as a major finding.

It is intended also to carry out tests of correlation between the peak measures and the peak airflow on release (measures B and I) in order to test the hypothesis that it is not only the instantaneous area of the glottis (as is to be investigated in the

J/B and J/I correlation tests mentioned above) which accounts for the amplitude of the release airflow, but also the amount by which the glottis has been opened during the consonant. The examination of this aspect of the traces will be completed by correlation tests between the amplitude of the peak and the timing of the peak, this latter test looking for some internal structure to the glottograph curves which, it is hoped, might be clearly repeated or deviated from in whisper.

3. Results

It should be noted that the shorthand "voiced" is employed in the tables below to stand for "normal voiced" in contrast to whispered speech.

3.1 A note on statistical measures:

The non-parametric tests of correlation and differences in central tendency used below do not assume a linear relationship between the measured traces and reality, although, of course, the assumption that the relationships "greater than" and "less than" holding between measurements is a true representation of some aspect of reality is made. The weak assumptions of these tests seem appropriate to an exploratory investigation of this sort. In the tables of results reported on below, however, mean values for the sets of raw scores on all parameters have been included for the reader's convenience.

3.2 Glottograph results

The glottograph scores in mm of trace for the parameter J, common to all types, are shown in table 1. Mann Whitney tests for differences in central tendency across the distinctions inherent in the material are shown in table 2.

TABLE I

Height of glottogram at release constriction (J) (in mm of trace)

	PIL voice	PIL whisper	BIL voice	BIL whisper	MIL voice	MIL whisper	FIL voice	FIL whisper	VIL voice	VIL whisper
1.	18.5	8	6	4.5	2.5	2.5	4.5	4.5	3.0	5.0
2.	19	12	4	4	3	3.5	5	5.5	3.5	6.5
3.	25.5	15	4.5	3.5	3	3.5	7	7	4	5
4.	25	9.5	4.5	5.5	2.5	3	4	5.5	3.5	3.5
5.	23	11	4	6	2.5	2.5	4.5	3.5	4	3.5
6.	25	10.5	4.5	4.5	2.5	3.5	4.5	5.5	3	2.5
7.	19.5	11	5	5	3.5	3.5	5	5.5	3	4
8.	17	6.5	4.5	4.5	3	3.5	5.5	6.5	4	7.5
9.	18.5	7	4.5	4.5	2.5	3	5.5	5	3.5	5
10.	17.5	9	3.5	5.5	3	3.5	6	6.5	4	5
11.	17.5	10.5	4	5	2.5	4	5.5	5.5	3.5	4.5
12.	18.5	8.5	4.5	4.5	3	3.5	6.5	3.5	4	5.5
13.	18.5	6.5	4.5	6.5	3.5	4	6.5	6.5	4	4.5
<u>Means</u>	20.2	9.6	4.5	4.9	2.8	3.3	5.4	5.4	3.6	4.8

TABLE 2

Mann-Whitney tests over stop and fricative types on glottographic measure (J)

Comparison	U	sign.	dir.	Comparison	U	sign.	dir.
PIL_v/PIL_w	non-overlapping dists.		v > w	FIL_v/FIL_w	78.5	NS	--
BIL_v/BIL_w	9.5	.001	w > v	VIL_v/VIL_w	32	.01	w > v
MIL_v/MIL_w	36	.01	w > v				
PIL_v/BIL_v	non-overlapping dists.		p > b	FIL_v/VIL_v	3	.001	f > v
PIL_v/MIL_v	non-overlapping dists.		p > m				
BIL_v/MIL_v	1	.001	b > m				
PIL_w/BIL_w	1	.001	p > b	FIL_w/VIL_w	54	NS	--
PIL_w/MIL_w	non-overlapping dists.		p > m				
BIL_w/MIL_w	6.5	.001	b > m				

In the tests between normal voicing and whisper with the consonants held the same, the clear relationship voice > whisper shown for PIL and BIL and reported on in the previous paper is not maintained on MIL or VIL, where whisper is shown to be significantly greater than voice, or on FIL where no significant difference can be found (but see the results on parameter K below).

For the tests across consonant types, the glottograph traces show the relationships which would be expected, however, namely [p] > [b], [f] > [v], non-nasal > nasal. The latter result is given as the expected one since, at least in normal speech, the presence of an open nasal port above the glottal constriction renders unnecessary any major adjustment of the vocal folds, active or passive, such as may be seen for [b], for example, in order to assure the continuance of voicing. For the comparison MIL with PIL, on the other hand, the contrast is simply between a configuration during which voicing is to be maintained and one in which it has to be interrupted. These results hold true for both normal voicing and whisper, with the single exception that no significant result is shown for [f/v] in whisper; the value of U on this test, however, only just falls short of significance at the 5% level. The data show a trend in the expected direction ([f] > [v]).

Following the practice of the previous paper, Spearman Rank correlation tests between the glottographic measures and the peak airflow measures on release were undertaken. The results are given in table 3. It will be seen that the results fail to cluster along any physical (normal voicing versus whisper) or linguistic dimension of the data. Of the three significant results (but only at the 5% level) two are for whispered speech (PIL, BIL) while the third is for normal speech (FIL). In the whispered pair, however, while r_s for BIL is positive, for PIL it is negative.

TABLE 3

Spearman rank correlation tests between glottographic measure J and peak airflow on release (B + I)

Type	r_s	sign.
PIL_{voiced}	- 0.188	NS
$PIL_{whisper}$	- 0.62	.05
BIL_{voiced}	+ 0.235	NS
$BIL_{whisper}$	+ 0.625	.05
MIL_{voiced}	+ 0.24	NS
$MIL_{whisper}$	+ 0.14	NS
FIL_{voiced}	+ 0.56	.05
$FIL_{whisper}$	+ 0.30	NS
VIL_{voiced}	+ 0.07	NS
$VIL_{whisper}$	+ 0.19	NS

Within the insignificant results there is no trend for the results on whisper to be more significant than those for normal voice.

The results on measurements K and L, carried out for only some of the linguistic types (see section 2.1 above) are given in tables 4 and 5. Two points should be noted. First, a peak was only locatable in some of the traces for $BIL_{whisper}$. Secondly, in the measurement L, the distance from peak to release, the convention has been adopted that those traces where the peak was observed to fall before the release have been designated as "+" values, while those with a peak after the release are given "-" values. "O" means that the peak is coincident with the release. Thus, while for PIL_{voiced} the mean value on L is -0.1 csec , implying that the peak is in general coincident with or even slightly after the release, that for FIL_{voiced} is +5.8 csec , showing the peak to occur well before the release.

Mann-Whitney tests across types and conditions for measures K and L are reported on in table 6. Taking the peak measure (K) first, it will be seen that on all voiced/whisper contrasts for the linguistic types the relationship "voiced greater than whisper" is shown. The relationship between the linguistic types shown for normal speech, namely [p] > [f] > [b] is modified in whispered speech to ([p] = [f]) > [b]. It should be observed, however, that the finding [p] > [f] in normal speech is evidenced at a lower level of significance (.025) than either the finding [p] > [b] or [f] > [b].

On parameter L, which concerns the timing of the peak relative to the release, it can be seen that whereas there is no significant difference in peak position between normal voicing and whisper for PIL, for BIL and FIL a highly significant difference is found. In both cases the voiced scores are higher

TABLE 4

Raw values and illustrative means for glottogram peak (K) for selected traces (in mm of trace)

	PIL voiced	PIL whisper	BIL voiced	BIL whisper	FIL voiced	FIL whisper
1.	19.5	8.5	9.5	4.5	18	7.5
2.	19	12	7.5	-	27	9
3.	25.5	15	8	-	22	8.5
4.	25.5	9.5	6	-	16	8
5.	22.5	11	5.5	-	18.5	7.5
6.	25	10.5	6	4.5	17.5	8.5
7.	19.5	11	7	5	26	9
8.	18.5	6.5	6.5	6	16.5	11.5
9.	18.5	7.5	7.5	4.5	16	8.5
10.	18.5	9	6	6	18	10
11.	18	10.5	7	-	12.5	7.5
12.	18.5	9	8	-	16	5.5
13.	18.5	6.5	7.5	6.5	16	9.5
<u>Means</u>	20.5	9.7	7.1	5.3	18.5	8.5

TABLE 5

Raw values and illustrative means for distance from glottogram peak to release L (in csec)

	PIL voiced	PIL whisper	BIL voiced	BIL whisper	FIL voiced	FIL whisper
1.	-1	-1.5	3.5	2.5	7	5.5
2.	0	0	3.5	-	5.5	6
3.	0	-1	5	-	6.5	3.5
4.	1.5	1	4	-	5.5	3
5.	0.5	0	5	-	7	4
6.	0	0.5	4	0	7	2.5
7.	0	0	4.5	1	5	4
8.	-1	0	3.5	1.5	6	3
9.	-0.5	1	4.5	0	6	3.5
10.	-1.5	-0.5	4.5	0	5	2.5
11.	0.5	0.5	3.5	-	6	2.5
12.	-0.5	1	4	-	4.5	4.5
13.	1	0	3.5	0	4.5	3.5
<u>Means</u>	-0.1	+0.1	4.1	+0.7	5.8	3.7

TABLE 6

Mann-Whitney tests across types and conditions for measures K and L

	K		L	
	U	sign.	U	sign.
$\mathrm{PIL}_{voiced}/\mathrm{PIL}_{whisper}$	non-overlapping v > w		73	NS
$\mathrm{BIL}_{voiced}/\mathrm{BIL}_{whisper}$	9.5 v > w	.01	non-overlapping v > w	
$\mathrm{FIL}_{voiced}/\mathrm{FIL}_{whisper}$	non-overlapping v > w		13.5 v > w	.001
$\mathrm{PIL}_{voiced}/\mathrm{BIL}_{voiced}$	non-overlapping P > B		non-overlapping B > P	
$\mathrm{PIL}_{voiced}/\mathrm{FIL}_{voiced}$	39.5 P > F	.025	non-overlapping F > P	
$\mathrm{BIL}_{voiced}/\mathrm{FIL}_{voiced}$	non-overlapping F > B		9 F > B	.001
$\mathrm{PIL}_{whisper}/\mathrm{BIL}_{whisper}$	1 P > B	.001	46.5	NS
$\mathrm{PIL}_{whisper}/\mathrm{FIL}_{whisper}$	55.5	NS	non-overlapping F > P	
$\mathrm{BIL}_{whisper}/\mathrm{FIL}_{whisper}$	3 F > B	.001	non-overlapping F > B	

than the whisper scores. This is to be interpreted as showing a shift of peak positioning in whisper towards the release.

In normal voice the following significant relationships are shown between the linguistic types, [f] > (= peaks earlier than) [b] > [p]. In whisper, while again the FIL traces peak significantly earlier than those of BIL or PIL, we fail to find the expected distinction made between these last two types.

In table 7 are given the results of Spearman Rank correlation tests between glottogram peak (K) and peak airflow on release (B and I), and between glottogram peak (K) and distance to release (L). On the B/K tests, only the result for $FIL_{whisper}$ is significant, and then only at the 5% level. The value for r_s here is, moreover, positive, in common with that for $BIL_{whisper}$ while all the other results are negative. On the K/L tests there is likewise only one significant result, again at the 5% level, this time for FIL_{voiced}, where the r_s, in common with all results except that for PIL_{voiced}, is negative.

3.3 Results on stop parameters A-F

The raw scores on parameters A-F for the stop consonants are given in table 8. Amplitude measures are given in mm of the trace, duration measures in csec. The results of Mann-Whitney tests across whisper/normal voice and across stop types for these parameters are given in table 10. This table can be further analyzed in terms of three types of comparison, as presented in the following sections.

3.3.1 Whisper versus normal voice tests

There is unanimity over the stop types about the significant differences between normal voice and whisper on the vowel airflow parameters A and C, where, at high levels of significance in all cases, whisper is shown to be greater than voice, and on

TABLE 7

Spearman rank correlation tests between glottogram peak (K) and airflow on release (B + I) and between glottogram peak (K) and distance to release (L)

	B/K		K/L	
	r_s	sign.	r_s	sign.
PIL_{voiced}	-0.35	NS	+0.08	NS
$PIL_{whisper}$	-0.46	NS	-0.18	NS
BIL_{voiced}	-0.04	NS	-0.04	NS
$BIL_{whisper}$	+0.29	NS	-0.5	NS
FIL_{voiced}	-0.01	NS	-0.61	.05
$FIL_{whisper}$	+0.58	.05	-0.44	NS

TABLE 8

Raw scores and illustrative means for stop consonants (for description of parameters, see figure 1).

		A	B	C	D	E	F
PIL$_{voice}$	1.	11.5	20	4	12.5	18.5	55
	2.	9.5	18	3	8.5	17	55
	3.	9.5	20	2.5	8.5	18	57.5
	4.	10.5	20.5	2.5	9.5	17	56.5
	5.	9	20	3.5	8.5	16.5	55.5
	6.	9.5	18.5	3	8	16.5	58
	7.	10	20	3.5	8	17	56
	8.	10.5	19	3	7.5	17.5	57
	9.	11.5	21.5	3.5	8.5	16.5	56.5
	10.	10.5	20.5	4	8.5	18	58
	11.	10.5	20.5	4	9	15	55.5
	12.	9.5	21	4	8	15.5	53.5
	13.	11	20.5	3.5	9.5	16	56
Means		10.2	20	3.4	8.8	16.8	56.2
PIL$_{whisper}$	1.	16.5	17.5	12	11	16	55.5
	2.	16.5	14	10	9	14	52
	3.	18	16.5	9.5	9.5	14.5	55
	4.	16	15.5	8	10.5	15	55.5
	5.	15	15.5	8	11.5	13.5	54.5
	6.	15.5	16	7	10	14.5	54
	7.	16.5	16	9.5	10	15	54.5
	8.	13.5	16.5	10.5	10	14.5	54.5
	9.	14.5	17.5	8	9.5	15	52
	10.	16	16.5	5	9	15	52.5
	11.	15.5	17.5	6.5	10	14	54.5
	12.	14	14.5	4	11	14.5	55
	13.	14	17	10.5	10.5	12.5	53.5
Means		15.5	16.2	8.3	10.1	14.5	54.1
BIL$_{voice}$	1.	12	17	2.5	10	14	54.5
	2.	10.5	15.5	2.5	9.5	13.5	52.5
	3.	13	15.5	3.5	10.5	16	55.5
	4.	11	16	3.5	9.5	15.5	57
	5.	10	16.5	2	9.5	14.5	54.5
	6.	11	16.5	3.5	10	14.5	55
	7.	10.5	16.5	3.5	10.5	14	55.5
	8.	13	16	3	9	14.5	55.5
	9.	11.5	16.5	3.5	10	14.5	55
	10.	12	16.5	2.5	9.5	14	54
	11.	11.5	17	2	10.5	14.5	55.5
	12.	10.5	14.5	2.5	9.5	14	53
	13.	12	17	2.5	10	14.5	55
Means		11.4	16.2	2.8	9.8	14.5	54.8

TABLE 8
(continued)

		A	B	C	D	E	F
$BIL_{whisper}$	1.	18	16.5	10.5	8.5	14.5	54
	2.	18.5	15	9	8.5	14	52
	3.	14	14	4	9.5	15	52.5
	4.	19.5	16.5	11	8.5	13.5	53.5
	5.	18	15.5	9	10	12.5	53
	6.	17.5	16	6	10	14.5	55
	7.	17	15	8.5	11	13.5	55.5
	8.	14.5	15.5	9.5	10.5	13.5	55.5
	9.	15.5	15.5	5.5	10	12	52
	10.	15.5	16.5	8.5	9.5	14	52
	11.	18	16	7	9	12.5	52.5
	12.	15	14	6	10.5	13.5	53.5
	13.	16	16.5	4	12	14	55
Means		16.7	15.6	7.6	9.8	13.6	53.5
MIL_{voice}	1.	10	6.5	2	8	13.5	52
	2.	10	7.5	3	10	12.5	53
	3.	8	8.5	2	9	12.5	52.5
	4.	9	9	2.5	8	13.5	53
	5.	10.5	11	2.5	8	13	51.5
	6.	10	9	3	9	13.5	54
	7.	9	12	2.5	9.5	13.5	54.5
	8.	9.5	10.5	3	7.5	13.5	53.5
	9.	12	11	2.5	8.5	13.5	53.5
	10.	10	12	2.5	7.5	14.5	51.5
	11.	8	11	3.5	8.5	13	52.5
	12.	8.5	9	1.5	8	13.5	51
	13.	9.5	10.5	2.5	8.5	13.5	55
Means		9.5	9.8	2.5	8.5	13.3	52.9
$MIL_{whisper}$	1.	17	8	7.5	8.5	14	54
	2.	16	10	3.5	8.5	13	57
	3.	15.5	11	5	7.5	13	53
	4.	16	14.5	5.5	7	13	50.5
	5.	14.5	12	6.5	7.5	12.5	50.5
	6.	16.5	11.5	8.5	8	13	50
	7.	16	13.5	7	7.5	12	51
	8.	15.5	16.5	5.5	7.5	13	52
	9.	15	13	4	7.5	12.5	49
	10.	18.5	13.5	8.5	7	11.5	49
	11.	15.5	9.5	6.5	7.5	12	49.5
	12.	14	12.5	6.5	6.5	11.5	50
	13.	14	13.5	3.5	6.5	13.5	49
Means		15.7	12.2	6	7.5	12.7	51.1

the durational parameters E and F, where normal voice is shown to be of significantly greater duration than whispered speech.

On parameter B, the peak airflow on release, normal speech is shown to be greater than whisper (in contrast to the vowel findings above) for PIL and BIL, although for BIL this finding is only marginally significant even at the 5% level.[1] In MIL, the whispered values are shown on this parameter to be greater than those of normal speech. No pattern at all is revealed on parameter D, the closure duration. Whisper is greater than normal voice for PIL, the opposite result holds for MIL, while for BIL there is no significant difference.

3.3.2 Normal voice differences between stops

On the amplitude measures a consistent pattern is revealed only for B, the amplitude at release. Here the relationship [p] > [b] > [m] is shown extremely clearly. On parameters A and C, concerning the amplitude of the frame vowels, it would seem that the results reflect in each case the presence of unusually high amplitudes for one stop type. On A the [b] traces are of greater amplitude than either [p] or [m] while these latter are not distinguished. On C it is the [p] set that is greater than the [b] and [m] traces.

The consistent pattern [p] > [b] > [m] is revealed on both the durational parameters E and F, while on D the [b] closures are longer than those of either [p] or [m] with these not being distinguished from each other.

3.3.3 Whispered speech differences between stop types

On the amplitude parameter A, no significant differences between stop types are observed, while on C the amplitude of the [m] traces is significantly below those of either [p] or [b],

1) In Mansell (1973) this result was indeed over-conservatively given as non-significant.

these being indistinguishable from each other. This latter analysis also holds true for three other parameters, airflow on release (B), closure duration (D), and total duration (F). The relationship [p] > [b] > [m], however, is shown on parameter E.

3.4 Results on fricative parameters A,C,E,F-I

The raw scores on the fricative parameters A,C,E,F - I are given in table 9. Mann-Whitney tests across normal voicing/whisper and fricative type are given in table 11.

3.4.1 Whisper versus normal voice tests

On the amplitude of the frame vowels (A and C) whispered speech is shown to be significantly greater than normal speech for both fricatives. On F, total duration, normal speech is, again in both cases, shown to be significantly greater than whispered speech.

On the consonantal measures, only on the amplitude of the second airflow peak (I) is there agreement on the effects of whisper over the two fricative types. Here the normal traces are shown to be greater than the whispered. This relationship is also shown at the trough (H) for FIL, but on VIL no significant difference can be found. At the first peak, however, while VIL shows the relationship whisper > normal voice, no significant difference can be seen with FIL.

3.4.2 Fricative differences in normal speech and whisper

On the frame vowel amplitude measures there is agreement between normal speech and whisper. While there is a tendency (5% level) for [ə] to be higher in amplitude for VIL than for FIL (A), no significant difference is found for the post-consonantal vowel (C). On the total duration measure there is a

TABLE 9

Raw scores and illustrative means for fricative consonants (for description of parameters, see figure 2)

		A	C	E	F	G	H	I
FIL_{voice}	1.	11	3	11.5	43	14	9	19
	2.	11	4.5	13	44.5	12	1.5	19.5
	3.	10	2.5	12.5	42.5	11	2	19
	4.	9	3	11	43	9.5	2.5	17.5
	5.	10	3.5	13	44	14	3	18.5
	6.	9	3	12	43.5	11	3.5	18.5
	7.	9.5	4.5	12.5	43.5	15.5	3.5	22.5
	8.	11.5	3.5	11.5	43.5	12	4.5	19
	9.	10	3	11.5	43	12.5	5.5	21.5
	10.	11.5	4	12.5	44	14	5.5	20.5
	11.	11.5	3.5	11.5	43.5	10.5	4.5	19.5
	12.	9.5	3.5	11.5	43	8	0	20
	13.	11	4	12	43.5	11.5	2.5	21
Means		10.3	3.5	12	43.4	12	3.7	19.7
$FIL_{whisper}$	1.	16.5	8	-	43	13.5	0	15.5
	2.	14.5	10.5	-	40.5	14	2.5	15.5
	3.	16.5	8.5	-	42	13.5	1	14.5
	4.	12.5	10	-	40.5	12.5	1.5	15.5
	5.	14	8	-	41	11.5	0	15.5
	6.	17	8	-	40.5	15.5	0	17.5
	7.	15.5	10	-	42	13.5	0.5	16.5
	8.	15.5	6	-	42	10.5	0	19
	9.	12.5	8.5	-	40.5	11.5	0.5	17
	10.	14.5	6.5	-	41	13.5	0	17
	11.	15	5.5	-	41	12	1.5	16
	12.	12.5	9	-	40.5	11.5	0.5	16
	13.	16.5	10	-	43	13.5	0.5	17.5
Means		14.8	8.3	-	41.3	12.8	0.7	16.4
VIL_{voice}	1.	12	2.5	14	42.5	8.5	0	11.5
	2.	11.5	2.5	13.5	44.5	8	0	11
	3.	9.5	3	14.5	44.5	8	0	10
	4.	9.5	3	13.5	45	8	0	10.5
	5.	10	3.5	14	43	8	0	11
	6.	11	3.5	14	46	8.5	0.5	11.5
	7.	10.5	3	13	44	7.5	0	11.5
	8.	13.5	3.5	12.5	44.5	9	0	13.5
	9.	13.5	3	13	44.5	9.5	0	12.5
	10.	10.5	4.5	14.5	44	10.5	0	13.5
	11.	12.5	3	13	42	9	0	14
	12.	11.5	3.5	12.5	44.5	10.5	0	14
	13.	12	4	13.5	43.5	9.5	0.5	12
Means		11.3	3.3	13.5	44	8.8	0.1	11.9

TABLE 9
(continued)

		A	C	E	F	G	H	I
$VIL_{whisper}$	1.	14.5	9	-	43	14	O	14.5
	2.	18	9	-	41	15.5	O	13.5
	3.	15.5	9.5	-	42.5	14.5	O	15
	4.	15	9.5	-	41	13	O	12.5
	5.	16.5	9.5	-	42	14	O	13
	6.	13.5	6.5	-	42.5	10.5	O	14
	7.	18.5	7.5	-	40	15	O	14.5
	8.	15.5	10.5	-	43.5	14	O	13.5
	9.	17.5	8.5	-	41	14.5	O	15
	10.	16.5	9.5	-	40.5	14	O	14.5
	11.	16.5	3.5	-	40.5	15	O	12.5
	12.	16.5	4.5	-	41	14.5	O	14
	13.	16.5	5	-	43.5	15.5	O	15.5
Means		16.2	7.8	-	41.7	14.2	O	14

TABLE 10

Mann-Whitney tests across stop parameters

		A	B	C	D	E	F
	U	non-	non-	2	20	45	17.5
PIL_v/PIL_w	sig.	overlapping	overlapping	.001	.001	.001	.001
	dir.	w > v	v > w	w > v	w > v	v > w	v > w
	U	non-	46	non-	80.5	38	42
BIL_v/BIL_w	sign.	overlapping	.05	overlapping	NS	.01	.025
	dir.	w > v	v > w	w > v	-	v > w	v > w
	U	non-	31.5	1	25	11	34.5
MIL_v/MIL_w	sign.	overlapping	.01	.001	.001	.001	.01
	dir.	w > v	w > v	w > v	v > w	v > w	v > w
	U	29.5	non-	43.5	20.5	4	34.5
PIL_v/BIL_v	sign.	.01	overlapping	.025	.001	.001	.01
	dir.	b > p	p > b	p > b	b > p	p > b	p > b
	U	56	non-	23.5	72.5	non-	5
PIL_v/MIL_v	sign.	NS	overlapping	.001	NS	overlapping	.001
	dir.	-	p > m	p > m	p - m	p > m	p > m
	U	14	non-	62	12.5	12	21
BIL_v/MIL_v	sign.	.001	overlapping	NS	.001	.001	.001
	dir.	b > m	b > m	-	b > m	b > m	b > m
	U	51.5	54	61.5	67	38	66.5
PIL_w/BIL_w	sign.	NS	NS	NS	NS	.01	NS
	dir.	-	-	-	-	p > b	-
	U	79.5	9	35	non-	10.5	21.5
PIL_w/MIL_w	sign.	NS	.001	.01	overlapp.	.001	.001
	dir.	-	p > m	p > m	p > m	p > m	p > m
	U	57.5	13	48.5	3	33.5	28.5
BIL_w/MIL_w	sign.	NS	.001	.05	.001	.01	.01
	dir.	-	b > m	b > m	b > m	b > m	b > m

TABLE 11

Mann-Whitney tests across fricative parameters

Type		A	C	E	F	G	H	I
FIL_v/FIL_w	U	non-overlapping	non-overlapping	-	6	71	14.5	5.5
	sign.			-	.001	NS	.001	.001
	dir.	w > v	w > v	-	v > w	-	v > w	v > w
VIL_v/VIL_w	U	non-overlapping	4.5	-	13	non-overlapping	71.5	16
	sign.		.001	-	.001		NS	.001
	dir.	w > v	w > v	-	v > w	w > v	-	v > w
FIL_v/VIL_v	U	47	66	10	47.5	14	7.5	non-overlapping
	sign.	.05	NS	.001	.05	.001	.001	
	dir.	v > f	-	v > f	v > f	f > v	f > v	f > v
FIL_w/VIL_w	U	47	78	-	68.5	31.5	32.5	6.5
	sign.	.05	NS	-	NS	.01	.01	.001
	dir.	v > f	-	-	-	v > f	f > v	f > v

tendency (5% level) for VIL traces to be longer than FIL traces in normal speech; this tendency is not shown in whisper, however, where no significant difference is found. On the consonantal amplitude parameters agreement is seen between normal and whispered speech on the trough (H) and second peak (I) measures, where in both cases [f] is significantly greater than [v]. While [f] is likewise significantly greater than [v] for normal speech on the first airflow peak (G), the opposite relationship is found on this parameter in whispered speech. On the parameter (E), only measured in the normal speech cases, VIL traces are shown to be significantly greater than FIL traces.

4. Discussion

It would, I think, be fair to characterize conventional assumptions about whisper in the following terms: the glottis is throughout more open than in normal speech, and the airflow consequently, in both vowels and consonants, higher. With respect to timing in whisper, one can expect general agreement with the words of Slis and Cohen (1969) when they claim:

> "It seems justified to consider whispered speech as normal speech minus voice, leaving the time structure and probably transitional cues intact."

It will be the main task of this section to isolate those aspects of the results which cannot be accomodated under this model, or separately accounted for in ways which do not require modifying it. Having isolated them it has to be decided whether there is a consistent pattern observable in the data and a reasonable explanation available, or whether these data are better considered to be aberrant in some way. It will be seen below that glottal behaviour during consonantal constriction in whisper is possibly implicated in all traces that cannot be accomodated without alteration to the simple model above. Hence one would normally expect the glottograph records to be decisive in the evaluation of the data. Unfortunately, the arguments for and against placing reliance on the glottographic traces are quite finely balanced in the present case. This point is so central that it will be dealt with first, in 4.1 below. Those traces explainable without modification to the simple model of this section will be enumerated in 4.2, while in 4.3 the question of the exceptional traces will be considered.

4.1 The glottographic evidence

In Mansell (1973), where only the results from PIL and BIL were analyzed, the possibility of pharyngeal narrowing, as reported for some varieties of whisper by Ohala and Vanderslice (1965) was invoked to explain the fact that glottographic traces were much lower in amplitude for whisper than for normal voice. It was hypothesized that this narrowing was sufficient in the English data to extinguish the trace altogether, since the photocell had not been anchored, while in the Danish case attenuation only was the result, some light being passed to the photocell by means of the anchoring tube passing into the oesophagus. It is unfortunately the case, however, both that the airflow results from MIL, to which I return in the following section, make it unlikely that any single factor like pharyngeal narrowing is available to explain the present data, and that the relationship voiced > whisper, observed in PIL and BIL, fails to show up on any of the extra traces examined here. It is further far from clear why VIL and MIL should group together in showing whisper voice on Table 2, although the lack of any significant difference between FIL_{voice} and $FIL_{whisper}$ on glottogram height at release is probably attributable to the inappropriateness of the measurement parameter in this case, the peak of the fricative glottal movement having already been reached some time before. Note that where the peak is considered (table 6) the relationship voiced > whisper is found for all types examined.

The second aspect of the glottographic traces which causes some concern is the lack of a notable increase in amplitude in the carrier vowels in whisper as against normal speech, despite the fact that all types show marked airflow increase in these vowels in whisper.

Finally, on the negative side, there is the evidence of tables 3 and 7; the extra traces and parameters examined here serve to strengthen the conclusion from similar tests in Mansell (1973), that no reliable correlation was to be expected between glottographic traces and airflow. It would seem that neither for normal voice or whisper is either instantaneous area, as shown by parameter J, nor the maximum area reached by the glottis (K) correlated with the airflow at release. These results are so far removed from conventional assumptions about glottal action and aerodynamic results as to necessarily throw doubt on the validity of the traces. It should be noted here also that the correlation tests between amplitude and timing scores for the glottogram peak failed to show up any pattern of internal organization for the traces, and hence could not be used to look for distinctions between normal and whispered speech.

There are aspects of the glottographic results, however, which would lead us to suppose that these traces were providing a proper reflection of reality. In the first place distinctions between linguistic types which are shown at the glottis in normal speech are likewise in general shown in whispered speech. The only exceptions are for J on $FIL_{whisper}/FIL_{voice}$, where, as has been previously noted, the measurement parameter may not be appropriate, $PIL_{whisper}/FIL_{whisper}$ for parameter K, where, however, the distinction found in normal speech between PIL and FIL may be regarded with some suspicion, the finding being at a fairly low level of significance, and $PIL_{whisper}/BIL_{whisper}$ for parameter L, where a general shift of glottographic peaks towards the release can be seen in whisper, involving the collapse of BIL and PIL timing scores in whisper (see below). Further, despite the fact that significant correlations are not shown between the individual values of glottograph amplitude and airflow on release, it is nonetheless the case that the relationships shown between types on the glottograph measure J ([p] > [b] > [m], and [f] > [v]) are reflected exactly in the airflow on release results for normal speech, and in a closely approximate form in whisper (([p] = [b]) > [m], [f] > [v]). Again, despite the general amplitude diminution it can be repeated that, as the scores for K on table 4 show, there is a marked peaking of the glottographic trace for the consonants in both PIL and FIL, voiced and whisper, and also in BIL voiced and in some whispered cases too.

Furthermore, on the durational parameter E for the stops, the relationship [p] > [b] is shown for both normal and whispered speech. This result has relevance to the glottographic traces in that scores on parameter E in normal voicing at least contain a period of aspiration. If we measure the duration only of the voiced portion of the post-consonantal vowel in normal voiced PIL and BIL we find a range of 9.5 - 11.5 csec for PIL, and a range of 12.5 - 14 csec for BIL. This echoes the result VIL > FIL found on the same measurement for the normal voiced fricatives. It is thus seen that for normal speech the aspiration is the all-important factor in making [p] greater than [b] on parameter E. It is difficult to avoid the conclusion that the E measurements in whisper also contain an "aspiration" phase.

If this were so, it would be an argument on the one hand for the validity of the glottographic traces, since these - whatever the reason for the overall drop in amplitude - show greater opening for $PIL_{whisper}$ than for $BIL_{whisper}$, greater opening being by conventional assumptions associated with aspiration. On the other hand, it would also be an argument for the validity of the simple model for whisper outlined in section 4. above. For if we believe that the laryngeal gestures in whisper are essentially the same as in normal speech, and if we believe, as, for example, Rothenburg does, that the timing of stops is related principally to a glottal cyclic movement, then the results on parameter E are only what would be expected. If, on the other hand, it is believed that the laryngeal consonantal gestures in whisper are different from those in normal speech, it is a formidable problem to explain why these different gestures should have produced the same durational result as in normal speech on the post-consonantal vowel.

The only point which counts against the above is the fact that whereas in normal speech the closure phase (parameter D) for BIL is longer than for PIL, this distinction is neutralized in the present data for whisper. The results on this parameter are discussed in 4.3 below.

In summary, it will be evident that there is sufficient uncertainty about the factors involved to make it impossible without further research to decide on their validity. The addition of further data in this paper over Mansell (1973) has, it should be noted, added a number of facts to the data which make the glottographic findings more plausible. It can be suggested that a useful place to start with further research may well be in the investigation of glottographic registration in different kinds of whisper. In both the Danish case and the English case considered in the previous paper, the type of whisper used by the subjects was certainly very strong, perhaps even tending towards stage whisper. This may well have affected the results, since it is well-known that the glottis configurations for different types of whisper vary. Further, while the experimental design employed renders most of the artefacts listed in Frøkjær-Jensen et al. (1971) irrelevant here, the sensitivity of the photocell to activity at different parts of the larynx (point 2(3) in their paper) has by no means been catered for in research so far.

4.2 Results not requiring emendation of the simple model of 4.

On the whispered speech/normal speech comparisons the MIL traces, considered here for the first time, play a pivotal role in the evaluation of the other traces, for in MIL not only is airflow during the frame vowels larger in whisper, but also after the consonant, suggesting that the larynx was here held open at least at the configuration for the vowels during the consonantal closure. The results for MIL thus correspond fully with the predictions of the simple model for whisper outlined in 4. above. Further, for MIL, the duration results are maximally simple, in that the general shortening of durations observed in whisper is found on all three durational comparisons, D, E and F. Since the only suggestion I have to make about this general shortening is that it may be attributable to some relationship between air expended and duration of utterance, an explanation which does not cater for differential shortening of parts of the utterance, the MIL result is better viewed as correct than any of the other patterns observed in the results.

All stop traces share the direction of the MIL results on parameters A, C, E and F. Not falling under the MIL paradigm are the B measures of PIL and BIL, where normal voice airflow amplitudes are greater than in whisper, and the D measures for the same types, where the normal speech > whisper result is not shown. In the fricatives the only results on the normal speech/

whisper comparisons which do not fit into the MIL type model are G, H and I for FIL, and H and I for VIL. The result on the G parameter for FIL is probably attributable to the presence of a set of unusually low FIL whisper signals on this parameter. There is also the result $VIL_{whisper} > FIL_{whisper}$ on this parameter in table 11 which contradicts the finding $FIL_{voice} > VIL_{voice}$. This result can also be explained by postulating values for $FIL_{whisper}$ which are lower than to be expected.

In the contrast between stop types, it has already been noted in 4.1 that on the durational parameters the same relationships are approximately adhered to on the durational parameters E and F in normal speech and in whisper. The results on the carrier vowel amplitudes differ from normal voice to whisper, but it is probably unsafe to conclude anything from this. An explanation in terms of irregular results in one type such as was used in the presentation of results on these parameters in 3.3.2 and 3.3.3 is probably sufficient to account for the data. For the fricatives a tendency for the pre-consonantal vowel to be higher in amplitude before VIL than before FIL is shared by normal speech and whisper, while no significant difference is found for either condition in the post-consonantal vowel.

On parameter B the expected result [p] > [b] > [m] is shown in normal speech. In whisper, however, while the non-nasal stops are still greater than the oral stops, [p] and [b] are neutralized. This finding has no explanation within the simple model of 4.

On parameter D, in accordance with the findings of Fischer-Jørgensen (1966) the closure duration for [b] in normal speech is shown to be greater than for [p]. It is also found that [p] is not distinguished from [m]. In whisper, however, [b] and [p] are neutralized, with both being shown to be greater than [m]. Once again, there is no basis for this neutralization in the simple model of 4.

Finally, the only clear distinction between normal and whispered speech, besides the general diminution in amplitude, which emerges from the glottographic results is the shift in peak timing towards the release shown for BIL and FIL in whispered speech. The fact that PIL does not appear to show this trend can be perhaps simply accounted for by the fact that PIL peaks in normal voice are already coincident with or even subsequent to the release, and could not hence be shifted rightwards. This finding is not accomodated in the model of section 4.

It is thus claimed that the only results on tables 6 and 7 that require an extension of the simple model of 4, are:

PIL_v/PIL_w	:	B	D	
BIL_v/BIL_w	:	B	D	L
PIL_w/BIL_w	:	B	D	
FIL_v/FIL_w	:	H	I	L
VIL_v/VIL_w	:	H	I	

4.3 The exceptional results

First, the duration results on parameter D for the stops. MIL has normal voice > whisper on this parameter. For this pattern to be changed in the PIL_v/PIL_w comparison, the duration of the whispered PIL traces must be considerably longer than expected. This would explain also the non-significant result for the PIL_w/BIL_w comparison on this parameter. It is not unreasonable to suppose further that the non-significant result on the BIL_v/BIL_w comparison can also be explained via a lengthening of the whispered traces. I cannot at the moment, however, suggest any underlying cause for this lengthening. In the present context it is tempting, of course, to regard the durational measures as dependent upon glottal action; against this, however, speaks the fact that the same time relationships are observed in the post-closure phase (release and post-consonantal duration, parameter E) in both PIL and BIL traces in normal and whispered speech. The one set of results argues as stated in 4.1 above, for the same set of events at the glottis in normal speech and whisper, whereas the other demands that a different set be postulated. I am inclined to conclude that on the present evidence the durational results not covered by the simple model are in all probability artefactual in some manner.

Secondly, a minor point, concerning the symmetry of the amplitude results. The disagreement between the FIL_v/FIL_w comparison and the VIL_v/VIL_w comparison on parameter H arises, it seems from inspection of the raw scores in table 2, from the fact that in the normal speech traces the airflow registration sinks to the baseline during the trough, so that the possibility of there being a further diminution during whisper is ruled out.

We are thus left finally with a symmetrical set of airflow results which together suggest that the segmental activity of some part of the vocal tract may be different during these consonants in normal speech and whisper. It remains to be seen what possibilities there are for such alterations, and what arguments there are, on the other hand, to suggest that the results are artefactual. Mansell (1973), on the basis mainly of the English data, came to the conclusion that supraglottal manoeuvres were unlikely to be implicated, a view which is echoed here. This leaves subglottal or glottal parameters. Mansell (1973) was inclined to find adduction on the glottal level unlikely, and suggested that a subglottal explanation be sought. What this explanation might be was, and still must be, left vague. Note that the possibility of adduction at the glottal level is not ruled out in principle, since what is here labelled "adduction" may simply be a change in glottal configuration which happens to offer more resistance to airflow than that assumed during consonants in normal speech. Certainly, until an adequate analysis of the involvement of laryngeal muscles in forming the glottal configuration for whisper, this question cannot be decided. It is precisely at this point, however, that airflow and glottographic traces come into conflict, since, as noted above, the glottographic traces do show peaking in the constriction phase for both [p] and [f] in whisper. The only possibility of a resolution of this conflict lies in the timing of the glottal gesture; it may be that timing differences, such as those revealed here on parameter L could lead to effects such as have been observed. Consideration of this point, however, requires the context of further research.

If we add to the above the reminder that in the airflow on release data of Lehiste (1964; see p. 168ff) the whispered consonants follow by and large the simple model of 4. above, then it will be evident that the general validity of the results treated in this section is in some doubt, especially since so many of the other results achieved can be explained by a model requiring no reorganization of articulation from normal speech to whisper. Even given this doubt, however, no conclusion seems justified here other than that if there <u>is</u> segmental reorganization in whisper, it takes place in a very restricted context (during the constriction for consonants and apparently nowhere else) and in a restricted set of segments (not, for example, in nasal consonants). It remains an urgent task for further research to provide adequate direct measures of glottal activity in whisper.

References

Fischer-Jørgensen, E. 1967: "Phonetic analysis of Danish stop consonants", ARIPUC 1/1966, p. 31-33

Frøkjær-Jensen, B., C. Ludvigsen and J. Rischel 1971: "A glottographic study of some Danish consonants", in: F&S, ed. Hammerich et al. (Copenhagen)

Lehiste, I. 1964: "Acoustical characteristics of selected English consonants", Publication 34 of Indiana University Research Centre in Anthropology, Folklore and Linguistics

Lehiste, I. 1970: "Suprasegmentals", MIT

Mansell, P. 1973: "An experimental investigation of articulatory reorganization in whispered speech", Forschungsberichte/ Working Papers II (Munich)

Ohala, J. and R. Vanderslice 1965: "Photography of states of the glottis", UCLA 2, p. 58-59

Slis, I.H. and M. Cohen 1969: "On the complex of the voiced-voiceless distinction; acoustic measurements", IPO Ms. 129, Institute for Perception Research (Eindhoven)

Trim, J. 1970: "Cues to the recognition of some linguistic features of whispered speech in English", Proc. Phon. 6, p. 919-923

THE FEATURE TENSENESS IN THE MODERN FRENCH VOWEL SYSTEM: A DIACHRONIC PERSPECTIVE

Hans Basbøll

1. Introduction

The point of departure for this article[1] is a distinctive feature analysis of the vowel system of Conservative Standard French (by this term I denote the language described e.g. in Nyrop 1951, Togeby 1951, and Grammont 1914). I follow Jakobson and Lotz (1949) in my use of the feature [tense] (section 2). This part of the paper makes no claims for originality.

The crucial part of the present paper is section 3, where we consider some language changes involving the feature [tense] in more advanced French standards, viz. the disappearance of distinctive vowel quantity and the raising of word-final [ɛ] to [e]. These changes in connection with the distinctive feature analysis given in section 2 make some specific predictions as to the possibility of distinguishing between two a-phonemes in the advanced standards in question.

The primary purpose of the present paper is not to give new data on or analyses of the French vowel system, but is of a more theoretical nature. The article aims at showing that a given distinctive feature analysis of a certain language, viz.

---

1) This is an abbreviated version of sections 2 and 4 of a manuscript (entitled 'Notes on the phonological role of the feature Tenseness in the Modern French vowel system') which I wrote during my stay in the spring 1973 at the University of Paris, Vincennes. I am indebted to Eli Fischer-Jørgensen and Oluf M. Thorsen for helpful comments on the manuscript, and to Richard Carter for stylistic suggestions.

Conservative Standard French, makes some quite specific predictions as to how this language can change. The predictions described here have the form of logical implications: Presupposing a certain analysis of the language before the change, if A changes to B, then C must also change to D. Or, in other words: the change A to B logically implies the change C to D if a certain analysis of the input language is "correct" (in the psychological sense). We can then test whether C is changed to D everywhere where A is changed to B. Such implications can, of course, only lead to falsification of the initial analyses, not to definite confirmations. Furthermore, they are only very indirect ways of "testing" the input analyses, as compared to psycholinguistic test methods. Nevertheless, I think it may be worth while to call attention to such implications of distinctive feature analyses.[1]

I think the arguments to be given below apply in basically the same way both to structuralist theories of distinctive features according to which these features build up phonemes (i.e. to the theoretical framework of Jakobson), and to generative theories according to which essentially the same distinctive features build up segments at any level of the phonological component. Throughout the paper I shall speak about the distinctive feature analysis at a level which accounts for the notion of phonetic contrast, and at which all feature coefficients are categorial (i.e. mainly binary, and always with a finite (small) number of values), cf. Rischel 1974, p. 361ff.

---

1) The intentions of the present paper are related to those of Skousen (1972, 1973). However, Skousen's treatment of the Finnish material which is the core of his papers, and the conclusions he draws from it, have been severely criticized by Kiparsky (1973, p. 92ff).

2. The feature [tense] in Conservative Standard French

Since the main function of section 2 is to serve as a basis for section 3, a survey of the vowel system in Conservative Standard French which is rather brief (except for problems related to the feature [tense]) will suffice, and the reader is referred to other treatments of the subject (e.g. Togeby 1951, p. 56ff, Delattre 1966, p. 95ff, and Malmberg 1969, p. 27ff) for more discussion and information, and further bibliography.[1]

The inventory of <u>phonemically distinct stressed vowels</u> in Conservative Standard French is different in open and closed syllables (where closed syllables in this paper means that the stressed vowel is followed by a pronounced consonant when the word is spoken in isolation):

<u>open</u> stressed syllables	i	e	ɛ	y	ø		u	o		ɑ	a	ɛ̃	œ̃	ɔ̃	ã
<u>closed</u> stressed syllables	i	ɛ:	ɛ	y	ø:	œ	u	o:	ɔ	ɑ:	a	ɛ̃:	œ̃:	ɔ̃:	ã:

All stressed vowels are long when they occur in the position before "lengthening" consonant(s), i.e. /z, ʒ, r, v(r)/, belonging to the same word.[2] Since vowel length in this position

1) The views adopted here are in several respects related to the viewpoints expressed by Oluf M. Thorsen in his treatment of the French vowel system, see Kongsdal, Landschultz and Thorsen 1973, p. 82ff and 119ff.

2) This formulation, which for the sake of brevity will be used throughout the paper, is not entirely correct: Lengthening occurs only before <u>word-final</u> /z, ʒ, r, v, vr/ (where the qualification "word-final" in fact should allow these consonants to be followed by an optional word-final shwa). It is necessary to state the environment with an optional <u>r</u>, viz. as /v(r)/, and not as X_1^2 where X is one of the consonants /z, ʒ, v, r/, in view of the existence of the word-final clusters /rv, rʒ/ which do not cause the preceding vowel to be lengthened (e.g. <u>verve</u>, <u>large</u> [vɛrv, larʒ]), as opposed to /vr/ (e.g. <u>vivre</u> [vi:v(r̥)]).

is thus completely automatic, the vowels [iː, yː, œː, uː, ɔː, aː] have not been included in the inventory, [iː] being a bound variant of [i], etc. (emphatic lengthening phenomena are, of course, disregarded in this connection).

It appears from the table that all vowels in open stressed syllables are short. Since, furthermore, only stressed vowels can be long, it follows that all long vowels occur in closed stressed syllables. In such syllables all nasalized vowels as well as all occurrences of the vowels [ø, o, ɑ] are long. All of these vowels, together with certain instances of [ɛː] (including all those which do not occur before a lengthening consonant, see below), can be called "inherently long vowels".[1] This term is in agreement with the fact that the "inherently long" vowels are long when they occur in the only position where vowel length is not neutralized (all long vowels are shortened by rule when they occur word-finally and in unstressed position).

---

1) In 1948, Hjelmslev (1970) proposed that the nasalized vowels manifest the sequence of an oral vowel phoneme plus a nasal consonant phoneme (cf. bon, bonne; bain, baigner [bɔ̃, bɔn; bɛ̃, bɛɲe] etc., also see Togeby (1951, p. 58 (2nd ed.: p. 40)). This analysis has been generally accepted within generative phonology (e.g. Rohrer 1967, Schane 1968, Selkirk 1972, Dell 1973). Schane (1968, p. 50ff) further proposes that /oː/ is derived from /al/ in cases like chevaux [ʃəvo] (cf. cheval [ʃəval]), and from /ɔs/ in cases like côte [koːt] (cf. costal [kɔstal]), cf. also Rohrer 1968. Selkirk's thesis (chapter IV, section 3) contains a number of arguments that the rules coalescing a vowel and its following consonants in cases like those mentioned above are really unitary transformations (whereby two segments become one) and not sequences of rewrite rules. Among other things, she points to the fact that the outputs of these processes show up as long vowels when they occur in closed stressed syllables, which I agree is a linguistically significant generalization, and which is in agreement with the term "inherently long vowels".

(Although Selkirk (p. 377) presupposes a rule of "L-vocalization" preceding the coalescence of /a/ and /u/ to /oː/ in e.g. chevaux, I am not convinced that the coalescence rule in question does not have /al/(instead of /au/) as its input, i.e., I doubt the existence of an L-vocalization rule. So long as a phonological rule turning /ai/ to /eː/ is not well motivated in French, Selkirk's own arguments for coalescence processes as unitary transformations should apply to the coalescence of /al/ to /oː/ too, since in all cases where L-vocalization applies, coalescence will also apply.) (continued on the next page)

Since all nasalized vowels are inherently long, their vowel quantity is most reasonably considered phonemically as an automatic consequence of their nasality.

The oral non-high vowels present particular problems of analysis, since there is an intricate combination of quantity and quality difference in closed stressed syllables before non-lengthening consonants:

open stressed syllables:	e	ɛ	ø		o		ɑ	a
closed stressed syllables:	ɛ:	ɛ	ø:	œ	o:	ɔ	ɑ:	a

The problem is now: how shall the six phonemically distinct vowels in open stressed syllables and the eight in closed ones be analyzed into phonemes? The most economical solution will be one according to which each of the triplets [e ɛ ɛ:], [ø œ ø:] [o ɔ o:] and [ɑ a ɑ:] are reduced to two phonemes, i.e., a solution which recognizes eight distinct oral non-high vowel phonemes in Conservative Standard French. All other solutions would necessarily lead to unexplained gaps in the vowel pattern. (Notice that the language investigated here obligatorily distinguishes jeûne and jeune, pôle and Paul, mâle and mal, bât and bat, dés and dais, maître and mètre; the fact that some of these distinctions are not maintained in other French standards need not concern us here, and at any rate it has no influence on the principal points made in the discussion.)

Two analytical possibilities immediately suggest themselves: that vowel quantity depends on vowel quality, or the other way round.

(continued)

It follows from the considerations above that within a generative framework all coalescing processes in French which have a vowel as their output have a vowel-consonant sequence as their input, and also that the output vowel is inherently long (i.e. it is manifested as a long vowel when it occurs in closed stressed syllables). Furthermore, it seems to be a significant generalization that the vowel and consonant of the input sequence are always homosyllabic, cf. my article on the phonological syllable in this volume, p. 63ff.

First, if vowel quality is taken to be distinctive, and vowel quantity to depend on it, then we must recognize the vowel phonemes /e ɛ ø œ o ɔ ɑ a/ (in addition to /i y u/, shwa, and the nasalized vowels). When /e ɛ/ occur in closed stressed syllables, the distinction between them will be realized as a pure quantity distinction (this problem will be taken up below). /ø o ɑ/ will be realized as long vowels when they occur in closed stressed syllables. In open stressed syllables, the opposition between /ø o/ and /œ ɔ/ is neutralized in favour of the former, a neutralization which can be argued for morphologically on the basis of alternations like sot, sotte [so, sɔt] and saut, saute [so, so:t], and - with examples where the conditioning factor for the alternation, viz. the deletion of the stem-final consonant, is due to a minor regularity - (b)oeuf, (b)oeufs [(b)œf, (b)ø]. Apart from the [e ɛ ɛ:]-vowels, where the phonemic distinction should be purely qualitative but its phonetic manifestation purely quantitative (which is certainly an unattractive feature of the solution), this does not seem an unreasonable description.

According to the other possible solution, viz. that only vowel quantity is distinctive and that the difference in vowel quality between [e ø o ɑ] and [ɛ œ ɔ a] depends on the quantity distinction, the phonemic quantity distinction will in closed stressed syllables be accompanied by a concomitant difference in vowel quality, except for [ɛ: ɛ] as in maître, mètre. The opposition between phonemically short and long round vowels is neutralized in open stressed syllables (cf. the examples sot, sotte; saut, saute; boeuf, boeufs mentioned above). This solution has the serious drawback, however, that what in open stressed syllables is supposedly a purely quantitative distinction phonemically, viz. /e:, e; a:, a/ (e.g. in dés, dais; bât, bat), is manifested phonetically as a purely qualitative distinction, viz. [e, ɛ; ɑ, a].

To avoid the situation that a purely qualitative phonemic distinction is manifested as a purely quantitative phonetic distinction, or vice versa, one thus seems forced to use a combination of the proposed solutions, viz. that there are nine oral non-high vowel phonemes, i.e.: /e ɛ ɛ: ø œ o ɔ ɑ a/. According to this solution, Conservative Standard French has one phonemically long (oral) vowel, viz. /ɛ:/. However, this solution has some important drawbacks. First of all, the stressed vowel phonemes /e/ and /ɛ:/ occur in complementary distribution, whereas all the other stressed vowels are in contrast (in unstressed syllables several other vowel oppositions tend to be neutralized, see below). Second, in the position where /ɛ:/ occurs, viz. in closed stressed syllables, the vowel phonemes /ɑ, o, ø/ are also always manifested by long vowels, but this lengthening is, within the proposal under discussion, a phonetic lengthening; however, the vowel length of fête, baisse, and the vowel length of pâte, basse, etc. clearly seem to be instances of the same phenomenon.

Are all possible solutions of this problem, then, equally unsatisfactory? No, since there is another analysis which seems, in fact, to be clearly superior to those we have discussed so far: in this analysis there are eight stressed oral non-high vowel phonemes, viz. /e:, ɛ, ø:, œ, o:, ɔ, ɑ:, a/, i.e. the distinction is a combined one of quality and quantity. If, for the moment, we accept the apparent ad hoc manifestation principle that /e:/ is realized as [ɛ:] in closed stressed syllables, then we need only make use of independently motivated principles of manifestation and neutralization in order to account for the realization everywhere in stressed syllables, viz. that all vowels in open stressed syllables are short, and that the opposition between round non-high oral vowels is neutralized in open stressed syllables in favour of the manifestation [ø o].

Which is the independent motivation, if any, for the manifestation principle that /e:/ is realized as [ɛ:] in closed stressed syllables? Notice that it is a general constraint on phonetic forms in French that the vowel [e] never occurs in closed stressed syllables, cf. alternations like céder, cède [sede, sɛd] where a phonetic form *[sed] would be structurally excluded in French. If the opposition between [e] and [ɛ] is neutralized in closed stressed syllables in favour of [ɛ], then it is not at all surprising that /e:/ is obligatorily manifested as [ɛ:] in such syllables, the pronunciation [e:] being excluded by a general surface constraint.[1]

---

1) Within generative phonology, where it is taken to be important that morphologically related forms be derived from a common underlying form by rules, it would seem to be a drawback of the solution proposed here that céder has a tense vowel in its first syllable, whereas there is a short and thus lax vowel in the stressed syllable of the present tense form cède; a similar apparent drawback is the fact that a word like fête [fɛ:t] has a tense stressed vowel /e:/, whereas there is a lax [ɛ] in the related verb form fêter [fɛte], see below in the text. There can be no question about a reclassification of [e] as lax and [ɛ] as tense within the present framework (cf. footnote 2 on p. 182f). These problems, which have been raised within a generative model, also have a solution within such a model, however. It was argued in the preceding footnote that long vowels are, within the descriptive framework of generative phonology, derived from underlying homosyllabic vowel-consonant sequences. This analysis accounts for the short vowel of cède etc. (since it is not derived from any vowel-consonant sequence) and, more importantly, it can explain why the tense vowel of fête alternates with [ɛ] and not with [e] in fêter etc.: if fête in its most abstract form has an underlying /s/ (cf. the related forms festin, festoyer etc.), then the underlying /e/ of the first syllable of fêter will be changed to [ɛ] by the rule of closed syllable adjustment because /e/ occurs in a closed syllable, viz. before a homosyllabic /s/, at the point of the derivation where closed syllable adjustment applies (on this phonological rule, see Selkirk 1972, p. 367ff, Dell 1973, p. 198ff, and Basbøll forthcoming). Within a non-generative framework, this account would probably be considered diachronic rather than synchronic.

However, in order to make the proposed solution really attractive we have to show that there can be found a non ad hoc distinctive feature which, as a consequence of its general definition, combines quality and quantity distinctions, i.e. that those two aspects of the feature have not been randomly associated for this specific purpose in French. Here the feature [tense] in the sense of Jakobson (e.g. Jakobson and Lotz 1949, Jakobson, Fant and Halle 1952, p. 36ff, Jakobson and Halle 1956, p. 30, Jakobson and Halle 1962) comes into the picture.

As already mentioned, the combination of phonetic quality and quantity distinctions is manifested in closed stressed syllables where we have the oppositions [ɑ:, a], [o:, ɔ], [ø:, œ]. These oppositions are quite similar to those which are found in closed stressed syllables of Standard German (cf. Fischer-Jørgensen 1973, p. 144-148), e.g. Mass, hass [mɑ:s, has][1] (cf. French lasse, glace [lɑ:s, glas]), Tod, Gott [to:t, gɔt] (cf. French haute, sotte [o:t, sɔt]), Söhne, könne [zø:nə, kœnə] (cf. French jeûne, jeune [ʒø:n, ʒœn]). For such German examples (with the possible exception of the two a-phonemes), there is widespread agreement that a common qualitative-quantitative distinction "Tense-lax" is relevant (although there is disagreement as to the precise phonetic nature of this distinction). Furthermore, the proposed phonemic distinction in French maître, mètre as one between /e:, ɛ/ (where /e:/ is realized as [ɛ:], see above), is the one found phonetically in German, e.g. beten, Bette [be:tn̩, bɛtə].

---

1) The classification of the two German a-phonemes used here is based upon Moulton's discussion (1962, p. 61-64). However, it seems to be the case that many German speakers do not make a consistent quality distinction between these two phonemes (for some acoustic measurements, see Jørgensen 1969).

Jakobson and Lotz (1949, p. 153) give the following general definition of the opposition TENSE/LAX: "The former are produced with walls stiffened by muscular tension and the latter by lax articulation. The stiffening of the walls of the resonance chambers causes a more definite formant ('clangs' the sound), while the damping of a lax wall is greater. The prolonged duration of the sound is an accessory effect of the tension".[1] In Jakobson and Halle 1956, the definition is (p. 30): "TENSE/LAX <u>acoustically</u>: higher (vs. lower) total amount of energy in conjunction with a greater (vs. smaller) spread of the energy in the spectrum and in time; <u>genetically</u>: greater (vs. smaller) deformation of the vocal tract - away from its rest position. The role of muscular strain affecting the tongue, the walls of the vocal tract and the glottis requires further examination."

The tense vowels [e ø o ɑ] in comparison to the corresponding lax ones [ɛ œ ɔ a] have a greater constriction, either in the palatal ([e ø] vs. [ɛ œ]), or the velar ([o] vs. [ɔ]), or the pharyngeal region ([ɑ] vs. [a]); the phonological relevance of this difference in vowel constriction was emphasized by Oluf M. Thorsen in a lecture in 1969.[2] Thus the tense vowels

---

1) Jakobson and Lotz (1949, p.153) in their footnote 17 give the following interesting quotations: "Il est très visible que les longues sont plus tendues que les brèves" - Durand [1946 p.] 151. However, "ce n'est pas tant la durée qui est en jeu que tout le déroulement de la voyelle" - <u>ibid</u>. 162. Cf. Rousselot's statement about the difference between tension and laxness: "Dans ma prononciation, il se confond avec la quantité, une voyelle <u>tendue</u> étant longue et une voyelle <u>relâchée</u> brève" [1897-1908 p.] 859."

2) The analysis adopted here of [e] being [+tense] and [ɛ] being [-tense] is at variance with the analysis of Jakobson and Lotz (1949), as shown by their matrices (p. 158) and by the following quotation (p. 154): "The tense vowels with the feature of pure or joint saturation [i.e. the non-high vowels; HB] are long when not in the word final, where French levels off the duration. The qualitative distinction is still valid though accompanied by a quantitative difference: the tense $\underline{ê}$ is opposed to the lax $\underline{e}$ as 'è ouvert' to 'é fermé' in the word final and as 'è: ouvert' to 'e moyen' elsewhere." That word-final [e] should be lax

are more distant from the neutral vowel than the corresponding lax vowels. The more extreme articulation of the tense vowels is concomitant with a longer duration. (We shall discuss some of the details of Jakobson and Lotz' use of the feature [tense] in French below.)

The main arguments for treating oppositions like [ø-œ, o-ɔ] in terms of tenseness and not of (say) vowel height can be summarized as follows: (i) Oppositions like jeûne-jeune and saute-sotte are of both a quantitative and a qualitative nature, and it seems quite arbitrary e.g. to assign the quantity distinction to one between relatively high and low mid vowels - or the other way round, for that matter - whereas the combined quality-quantity distinctions follow from the conception of the feature [tense] in the sense of Jakobson et al. (ii) The pair [ɑ-a] also exhibits a combined quality-quantity distinction (e.g. mâle, mal), in accordance with the expected behaviour of the feature [tense] (cf. some German pronunciations of Dame-Damme), and this can by no means be ascribed to the same sort of vowel height distinction. (iii) All the vowel pairs which are claimed to be distinguished by means of the feature [tense], but no others, tend to be neutralized in unstressed syllables: the vowels [e-ɛ, ø-œ, o-ɔ] can merge, whereas in normal speech vowels like [i-e, i-ɛ, i-y, y-u, ɔ-ɑ] cannot merge.

---

and word-final [ɛ] tense is in disagreement with the whole vowel pattern, as we have argued above. And in fact, in the reprint of the article in Jakobson 1962 (p. 426-434), in the matrices (p. 434) word-final [e] has been classified as [+tense] and word-final [ɛ] as [-tense] (it is, of course, out of question to analyse long [ɛ:] occurring before non-lengthening consonants as anything else than tense /e:/); and the passage quoted above has been changed to: "The tense vowels with the feature of pure or joint saturation are long when not in the word final, where French levels off the duration, while there is a constant qualitative distinction, though differently implemented" (p. 429). Jakobson made this rather important change in silence, despite his words in the Preface: "The papers contained in the present book reproduce the original text with a few abridgements and some small lexical, phraseological, and stylistic changes."

(Of course, the third argument mentioned above only suggests that there is a common feature distinction between the vowel pairs in question, but not which one it is, whereas argument (i), as well as the fact that the distinction [ɑ-a] is of the same type as the other three, shows this feature to be something like [tense].)

As already pointed out, the opposition between tense and lax vowels in French is thus manifested by a combination of vowel quality and vowel quantity distinctions. In most positions, however, only one of these aspects of the distinction is manifested: Only the quantity distinction is retained for the opposition /e:, ɛ/ in closed stressed syllables. Only the quality distinction is retained for non-round vowels in open stressed syllables, and for all vowels in closed stressed syllables before a "lengthening consonant" (i.e. the distinction /e:/ : /ɛ/ is neutralized in stressed syllables before "lengthening consonants"). Neither distinction is retained, i.e. the opposition tense/lax is neutralized, for round vowels in open stressed syllables. Furthermore, as mentioned above, there is a clear tendency towards neutralization of the opposition tense/lax in unstressed syllables, the manifestation being regulated by vowel harmony, by analogy to the occurrence of morphologically related vowels under stress, and by the distinction open-closed syllable. However, this neutralization is not carried through in Conservative Standard French.

Note further that a great number of "gaps" in the phonetic representations of French words can be found exactly where the feature [tense] is concerned: e.g. there are no words ending phonetically in [o:r, ø:r, ø:v, œ:z, ɔ:z, a:z, ø:ʒ, œ:ʒ].[1]

---

1) However, there is an isolated example with a stressed oral non-high round front vowel before [ʒ], viz. the name Maubeuge which ends in either [ø:ʒ] or [œ:ʒ]. According to Kongsdal, Landschultz and Thorsen (1973, p. 206c), [œ] never occurs before dental obstruents. As the examples show, the position before [z] is different from the others in that only the round lax

I do not pretend all these gaps to be structural (i.e. systematic), but it is significant that very often only one member of the tense/lax opposition is found exactly before the lengthening consonants (compare the fact that a great number of words vary between [a:ʒ] and [ɑ:ʒ]). The reason may be that in this position the quantity distinction (which it is argued here is an important part of the tense/lax opposition) is never manifest. (Note that within a generative framework, I have not committed myself as to whether these regularities - to the extent they are systematic at all - should in fact be expressed by means of the feature [tense], which would then have to be present in underlying forms in about the same distribution on lexical items as it has on the surface. I have only given a rather evident reason why exactly this feature should have difficulty in maintaining its distinctive power just before the "lengthening" consonants.)

---

vowels are missing (in all other cases the missing vowels are tense, or both tense and lax([ø:ʒ œ:ʒ])), just as Schane's rule predicts (1968, p. 51):

$$\begin{bmatrix} V \\ +round \end{bmatrix} \longrightarrow [-low] \; / \; \_\_\_\_ \begin{Bmatrix} \# \\ z \end{Bmatrix}$$

I agree with Selkirk's uncommented judgement (p. 392f) that Schane's rule collapsing is "purely by way of description", since the rule in the environment ___ # (accounting for sot, boeufs [so, bø], etc.) is a phonological rule which must apply after such other phonological rules as Final Consonant Deletion and Truncation, whereas the rule applying before z expresses a regularity in the lexicon (and it can therefore be formulated as a morpheme structure condition). Anyhow, Selkirk's explanation ("that only [o], and not [ɔ], appears before [z] [...] is, I think, due to the lengthening effect of /z/.[...] Yet, I do not understand at present why long /ɔ/ becomes [o] in French") is not sufficient since it would predict just the same thing for /v, ʒ, r/, in contradistinction to the facts (cf. innove, loge, Faure [inɔ:v, lɔ:ʒ, fɔ:r]). The only way to save her explanation would be to assign to the rule which lengthens vowels before /z/ a much earlier place in the ordering than the rule which lengthens vowels before the other lengthening consonants (as done by Rohrer 1968), but I know of no independent arguments pointing in that direction.

Before we set up a distinctive feature matrix for the stressed vowel phonemes in French, a few words should be said about Jakobson and Lotz' (1949) additional uses of the feature [tense] in their account of the French phonemic pattern. They also use this feature to distinguish between two sets of obstruents, viz. fortes ([+tense]) and lenes ([-tense]). Whether this distinction among obstruents can really be identified with the distinction discussed here among non-high oral vowels, is a question which has no impact on our paper, and which I shall therefore leave uncommented here.

Jakobson and Lotz also make a third use of the feature [tense] in French; they use it to distinguish between the high vowels [i y u] ([+tense]) and the semivowels [j ɥ w] ([-tense]). This proposal appears in several respects to be quite elegant:
(i) The distinction between [i y u] and [j ɥ w] seems to correspond rather well to the definition of the opposition tense/lax, the high vowels normally at the same time being longer and more precisely articulated than the corresponding semivowels.
(ii) This proposal permits [i y u] to be classified as [+tense] in agreement with the fact that these vowels have an even greater constriction than the tense vowels [e ø o] (cf. above).
(iii) The proposal permits the statement that all round vowels in word-final position are tense, thus at the same time accounting for two (apparently disjoint) sets of facts: that there is a distinction in word-final position between [e ɑ] and [ɛ a] but not between [ø o] and [œ ɔ], where only the former pair is found; and that there is a distinction in word-final position between [i] and [j] (e.g. <u>pays</u>, <u>payes</u> [pɛi, pɛj]), but not between [y u] and [ɥ w], where only the former pair is found.
(iv) The proposal seems to account for the fact that the set of vowels which have corresponding semivowels and the set of vowels which enter into an opposition between long more constricted and short less constricted vowels are non-overlapping, and that the union of these two sets equals the set of all oral vowels in French.

It is, of course, the first clause of (iv) above which has made it possible - and the second clause which has made it attractive - for Jakobson and Lotz to identify phonemically the feature [tense] of the non-high vowels and the feature which is usually called [syllabic] of the high vowels. The Jakobsonian practice of phonemically identifying features which are phonetically quite different can evidently be challenged on general grounds (cf. McCawley 1967). It is important in this connection to remember that the set of distinctive features is supposed to be universal (e.g. Jakobson, Fant and Halle 1952), i.e. the identification in question presupposes that a language can never have a three-way contrast between high tense vowels (e.g. [i: y: u:]), high lax vowels (e.g. [ɩ Y ɷ]) and semivowels (e.g. [j ɥ w]). I.e., a language with a vowel system like that of Standard German could never in addition have a set of semivowels like that of French, if the analysis of Jakobson and Lotz (1949) is to be upheld.

Even if this claim should be universally true, the justification of the identification in question can still be questioned. First of all, it cannot be a general principle for the manifestation of the Jakobsonian feature [tense] that it is one of high vowels vs. semivowels, as evidenced by the Standard German distinction [i: y: u:] vs. [ɩ Y ɷ]. Second, with regard to the rules governing vowel quantity in French, [i y u] go together with [ɛ œ ɔ a], not with the tense vowels [e ø o ɑ] (see above); thus we must distinguish between high tense vowels and non-high tense vowels everywhere where vowel quantity is concerned, and this casts doubt on the classification of [i y u] as [+tense] since exactly the vowel quantity phenomena were crucial in the arguments given above for the use of the distinctive feature [tense]. Furthermore, the principle which accounts for the distribution of high vowels and the corresponding semivowels (basically that we have semivowels in the position before vowels) has no relation to the principles of vowel quantity (see above). Third, I think the generalization (cf. Jakobson and Lotz 1949, p. 154) mentioned under (iii) above is only a spurious one, since the general restriction on the occurrence of the semivowels [ɥ w] is that they only occur before vowels, and the fact that they do not occur in word-final position is only one consequence of this general restriction; the restriction on [œ ɔ] is quite different, on the other hand, and it really concerns word-final position, cf. jeune, sotte [ʒœn, sɔt], etc.

As a consequence of these considerations I shall in the matrix below not specify the coefficient for the feature [tense] in the high vowels. The matrix includes all stressed oral vowels in French (for shwa and the nasalized vowels, see below).

	i	e:	ɛ	y	ø:	œ	u	o:	ɔ	ɑ:	a
HIGH	+	−	−	+	−	−	+	−	−	−	−
BACK	−	−	−	−	−	−	+	+	+	+	+
ROUND	−	−	−	+	+	+	+	+	+	−	−
TENSE		+	−		+	−		+	−	+	−

The four nasalized vowels [ɛ̃ œ̃ ɔ̃ ɑ̃] have the same features as the oral vowels [ɛ œ ɔ a], except that they are [+nasal], of course. They are all very low vowels (even lower than suggested by the transcription), and the fact that they are always long when they occur in closed stressed syllables may be dependent on that. Within a more abstract generative framework, however, their "inherent length" can be connected with the fact that they are derived from underlying homosyllabic vowel-nasal-sequences (see the footnote on p. 176f).

I said above that the matrix includes the stressed vowels only, but the set of vowel phonemes occurring in unstressed syllables is, in fact, identical to, or a subset of, the inventory of stressed vowels (the main difference being that the feature [tense] may be neutralized in unstressed position, see above). However, the vowel symbol [ə] ("shwa, e caduc") is often used in phonetic transcriptions in addition to the stressed vowel symbols. But [ə] in fact means a vowel identical either to unstressed [œ], or to a neutralization product of [ø/œ], and it is thus already defined in terms of distinctive features.[1] The reason why a special symbol, viz. [ə], is often

---

1) Thus I do not follow Schane (1968, p. 30ff) who uses the feature [tense] at the phonetic level to distinguish between shwa ([-tense]) and all other vowels ([+tense]). At the phonological level,Schane uses [tense] as a "diacritic" feature distinguishing between two classes of underlying vowels classified according to whether they undergo certain alternations. This description presupposes a phonological rule which tenses all vowels except shwa. Selkirk does not explicitly discuss the role of the feature [tense] in her thesis, but scattered remarks indicate that the specification [-tense] characterizes the shwas which drop, whereas all vowels which are manifested phonetically are [+tense]. The following quotation is particularly informative: "The rule which prevents /ə/ from dropping before an h aspiré (call it H-EX and understand it to cause tensing of a pre-"h" schwa) operates, needless to say, before the operation of the rule of final schwa deletion" (p. 374).

Selkirk's use of the feature [tense] is more reasonable than Schane's, but it is still an "abstract" use, i.e. the feature is not defined phonetically (but only as characterizing 'the vowels which drop'). It is possible that the underlying shwas should be defined as [-tense] in distinction to all full vowels (and furthermore as [-high, -back, -round]) within a generative framework (see Basbøll forthcoming), but the shwas which are manifested phonetically must then be changed to nonhigh round front vowels by a phonological rule, cf. Dell loc. cit. Forms where shwa occurs under stress in an open syllable, like prends-le, parce que! [prãlœ́/prãlǿ, parskœ́/parskǿ], support the existence of such a rule. Furthermore, the possibility of both [œ́] and [ǿ] in such forms, as opposed to "normal words" where only final [ǿ] occurs, can be described by means of rule ordering: If shwa-rounding applies before raising of word-final [œ,ɔ], we get forms like[prãlǿ], but if the order is the reverse, we get [prãlœ́], etc.

used for this vowel is probably that it alternates with zero - partly in free, partly in bound variation - in distinction to the vowel /œ/, cf. Dell p. 196f.[1]

3. Some evolutionary tendencies in more advanced French standards, involving the feature [tense]

It will be remembered that the language used as basis for the preceding discussion was a rather "conservative" variety of Modern French, and it is well known that younger standards deviate from the more conservative norms on (among other things) several of the points which were discussed in section 2, and which provided the arguments for the phonetic role of the feature [tense]. The import of these evolutionary tendencies for the present analysis will be taken up below. It should be added, however, that the sound changes mentioned in the present section are viewed in isolation, and if they are in fact only specific cases of more general changes in French grammar, my remarks below may in certain cases miss the point.

In many younger French standards the quantity of oral vowels is completely predictable from other traits of the phonetic surface without regard to quality. Thus the phonetically long vowels in non-emphatic speech are exactly those which occur in closed stressed syllables and which are nasalized

---

1) However, I do not wish to exclude a priori the possibility of an extra-weak stress on certain [ə]-syllables as opposed to (unstressed) [œ]-syllables.

and/or followed by a lengthening consonant.[1] Since the main reason for postulating the (phonetic) distinctive feature [tense] in the French vowel system was the interrelationship between quality and quantity among the oral vowels (see section 2), there is thus no longer any strong reason for maintaining this low-level feature in these younger standards.

The speakers in question pronounce maître and mètre, fête and faite, etc. identically, whereas they still maintain a quality distinction between haute and sotte, Beaune and bonne, etc. This means that the vowels [u o ɔ], and generally also [i e ε] (see below), are distinguished phonemically (whereas the phonemic distinction between [ø œ] is often not

---

1) However, the effect of this latter vowel lengthening seems in certain young standards to be so little pronounced that it may, in fact, be explainable by the physiologically conditioned (and thus universal) effect lengthening vowels before voiced obstruents, particularly fricatives (note that [r] is phonetically a voiced fricative in Modern French). In this case the vowel lengthening rule will, of course, have disappeared from the French grammar (a particularly clear indication that this is the case would be the "outfading" or even more the disappearance of the apparent particularity that sequences of voiced fricatives have exactly the opposite effect in the case /vr/ which lengthens the vowel, e.g. vivre, versus /rv, rʒ/, which do not, e.g. verve, large; note, however, that /vr/ is a possible syllable-initial cluster as opposed to /rv, rʒ/). If the duration of the nasalized vowels by the speakers of the young standard in question should also be explainable by physiological factors alone, then there will be no principles specific to the grammar of French which govern vowel quantity in that language (and in that case there would be no reason for indicating (non-emphatic) vowel length in phonetic transcriptions of French).

maintained in these younger standards[1]. Since the main argument for the existence of the distinctive feature [tense] in these innovating dialects has disappeared, another feature must distinguish between [o ɔ] (and [e ɛ]), and it seems evident that "degree of opening" or (with a better terminology, since [ɑ] is, strictly speaking, a narrow (pharyngeal) vowel) "vowel height" is the relevant choice here. If we say that the feature [tense] has been replaced by the feature [low] in the younger standards in question, or - equivalently within the present framework - that the feature [tense] has disappeared while the feature [high] has become ternary, we can thus account for the evolution from Conservative Standard French. Within a binary framework, the following distinctive features thus seem necessary and sufficient to account for all oppositions among stressed oral vowels in the Modern French Standards in question: [high, low, back, round]. (The vowel pairs which may be neutralized in these standards, i.e. [e-ɛ, ø-œ, o-ɔ], are still distinguished by one feature, namely [low].)

Now observe that it is impossible to distinguish between two a-phonemes within this framework of distinctive features (without introducing, of course, additional features or coefficients of features which will then necessarily be entirely ad hoc; cf. Westring Christensen 1969, p. 120f). Thus the distinctive feature analysis of the French vowel system made in

---

1) The merger of /ø/ and /œ/ which can be observed in many speakers, is probably to be explained in terms of the universal tendency to reduce oppositions among round front vowels (and non-round back vowels) as opposed to non-round front and round back vowels. Such universal tendencies have been studied notably by Jakobson (e.g. 1941) and Martinet (e.g. 1955), and the tendency in question is found independently in many languages, e.g. Danish.

section 2, together with the observed change in the principles of vowel quantity in more advanced French standards, makes the prediction that those standards will not distinguish between two a-phonemes in any environment.[1] As far as I know, this prediction is borne out.

Thus we argue for the following distinctive feature analysis of the French vowel system after the sound-change mentioned. This analysis is identical to that of Dell 1973, p. 284, whose dialect, actually, seems to represent exactly the stage of evolution discussed here. Only the stressed oral vowels are included in the matrix:

	i	e	ɛ	y	ø	œ	u	o	ɔ	a
HIGH	+	-	-	+	-	-	+	-	-	-
LOW	-	-	+	-	-	+	-	-	+	+
BACK	-	-	-	-	-	-	+	+	+	+
ROUND	-	-	-	+	+	+	+	+	+	-

There appears to be in Advanced Standard Parisian a coalescence between dés and dais which are both pronounced [de], fée and fait both [fe], etc., i.e. (in diachronic terms) word-final [ɛ] has been raised to [e]. This process could be a generalization of at least two different regularities in French phonology.

---

1) I here refer to speakers who have given up a real phonemic distinction between two a-phonemes in their own language. However, it is well known that a few minimal pairs like mâle-mal and pâte-patte are systematically taught in school; but it appears to me that the very fact that a new pronunciation of a few words can be learnt should not force the analyst to postulate a restructuring of the phoneme system. An investigation of the speech of pre-school children would probably be relevant in this context. In the final analysis, it is of course a psychological question whether (or: under what conditions) people restructure their sound pattern when they learn new words which contain a new "taxonomic phoneme" or a hitherto unknown phoneme combination.

The first of these is the principle that [e] and [ɛ] tend to occur in complementary distribution in unstressed syllables, [e] occurring in open syllables and [ɛ] in closed ones. If this principle is generalized so that it applies to stressed and unstressed syllables alike, it will have the effect that both dés and dais are pronounced [de], etc. If this is in fact the explanation for the sound change in question, then this sound change makes no predictions at all concerning the phonological role of the feature [tense] after the change has been carried through. Since there is thus one perfectly reasonable explanation for the sound change which does not involve the feature [tense] in any crucial way, the case to be made below, which is built upon another explanation of the sound change, really seems a weak one. However, since the present paper is mainly an illustrative one, I shall demonstrate below which consequences can be drawn if the second explanation for the raising of word-final [ɛ] is correct.

According to this second explanation, the raising of word-final [ɛ] to [e] is a generalization of the rule which says that the distinction between [ø o] and [œ ɔ] is neutralized in word-final position in favour of the manifestation [ø o] (e.g. sot, boeufs [so, bø]). If this rule is generalized to cover [e ɛ] too, it predicts that all word-final [ɛ]s will be raised to [e].

Now, this generalization could theoretically occur either (i) while the feature [tense] was still in operation in the French vowel pattern, or (ii) after the phonological role of the feature [tense] has been taken over by the feature [low] (see above).

In case (i), the generalization of the rule which predicts that [ø o] will occur in word-final position, but not [œ ɔ], would have to involve the feature [tense], and simply consist of removing [+round] from the definition of the input segment;

but there would then be no non ad hoc means of also preventing [ɑ a] from being neutralized in favour of the manifestation [ɑ] (see the distinctive feature matrix on page 188). I.e., if the raising (in the diachronic sense) of word-final [ɛ] is a generalization of the raising (in the synchronic sense) of word-final [œ ɔ], and if the language still has a phonemic distinction between [ɑ:] and [a] in closed stressed syllables, and thus two a-phonemes and the distinctive feature [tense], then there should be a neutralization of the opposition [ɑ/a] in word-final position in favour of the manifestation [ɑ].[1] Such a neutralization is completely unknown in French, as far as I know.

We thus conclude that if the raising of word-final [ɛ] is a generalization of the raising of word-final [œ] and [ɔ] (in the synchronic sense), then the raising of word-final [ɛ] does not take place until the feature [tense] has ceased to play the phonological role ascribed to it in section 2 above, i.e. until the two a-phonemes have merged in any environment. As far as I know, this prediction is also borne out by the facts (notice that this prediction does not follow logically from the facts of the first sound change mentioned in section 3, i.e. the disappearance of distinctive vowel quantity).

---

1) If, furthermore, Jakobson and Lotz' (1949) analysis of the distinction between [i y u] and [j ɥ w] as [+tense] vs. [-tense] were correct, and if the feature [tense] still played the phonological role ascribed to it in section 2 at the time of the raising of word-final [ɛ], and if this raising were due to a generalization of the rule accounting for the fact that lax round vowels (including semivowels) do not occur word-finally, then we should expect a concomitant change of word-final [j] to [i] (i.e. a tensing in word-final position, parallel to that of [ɛ] to [e]). Such a change does not occur in French at all, as far as I know. All this is, of course, very hypothetical.

One final remark. If the tendency (which is found in certain dialects) to merge the vowel pairs [e-ɛ, ø-œ, o-ɔ] in any environment - except for purely phonetic variation (including bound variation) - should be carried through in Standard French too, one aspect of the evolution from the language considered in section 2 above could be stated briefly as follows: The feature [tense] in the Jakobsonian sense has ceased to be relevant for the vowel pattern of French, without being substituted by any other feature(s). The vowel system of such a variety of French could thus be given the following distinctive feature analysis:

	i	e	y	ø	u	o	a
HIGH	+	-	+	-	+	-	-
BACK	-	-	-	-	+	+	+
ROUND	-	-	+	+	+	+	-

References

Basbøll, Hans forthcoming: "Schwa, jonctures et syllabification dans les représentations phonologiques du français"

Delattre, Pierre 1966: Studies in French and comparative phonetics (The Hague)

Dell, François 1973: Les règles et les sons. Introduction à la phonologie générative (Paris)

Durand, M. 1946: Voyelles longues et voyelles brèves (Paris)

Fischer-Jørgensen, Eli 1973: "Perception of German and Danish vowels with special reference to the German lax vowels /I, Y, U/", ARIPUC 7, p. 143-194

Grammont, Maurice 1914: Traité pratique de prononciation française (Paris)

Hjelmslev, Louis 1970: "Le système d'expression du français moderne" (summary by Eli Fischer-Jørgensen of a lecture held in 1948), Bulletin du Cercle linguistique de Copenhague 1941-1965 (Copenhagen), p. 217-222 (discussion p. 223-224)

Jakobson, Roman 1941: Kindersprache, Aphasie und allgemeine Lautgesetze (Uppsala); English translation: Child language, Aphasia and phonological universals (The Hague 1968)

Jakobson, Roman 1962: <u>Selected writings</u> I (The Hague)

Jakobson, Roman, G. Fant and M. Halle 1952: <u>Preliminaries to speech analysis</u> (Cambridge, Mass.)

Jakobson, Roman and M. Halle 1956: <u>Fundamentals of language</u> (= <u>Janua linguarum</u>, Series minor 1) (The Hague)

Jakobson, Roman and M. Halle 1962: "Tenseness and laxness", in: Jakobson 1962, p. 550-555, and as an appendix to the 1962ff-editions of Jakobson, Fant and Halle 1952

Jakobson, Roman and J. Lotz 1949: "Notes on the French phonemic pattern", <u>Word</u> 5, p. 151-158; reprinted in Jakobson 1962, p. 426-434

Jørgensen, H.P. 1969: "Die gespannten und ungespannten Vokale der norddeutschen Hochsprache mit einer spezifischen Untersuchung der Struktur ihrer Formantfrequenzen", <u>Phonetica</u> 19, p. 217-245

Kiparsky, Paul 1973: "On phonological representations", in: Fujimura, O. (ed.), <u>Three dimensions of linguistic theory</u> (Tokyo), p. 1-136

Kongsdal, O., K. Landschultz and O.M. Thorsen 1973: <u>Fransk fonetik</u> (Københavns universitets offsettrykkeri, Copenhagen)

Malmberg, Bertil 1969: <u>Phonétique française</u> (Malmö)

Martinet, André 1955: *Economie des changements phonétiques* (Bern)

McCawley, James 1967: "Le rôle d'un système de traits phonologiques dans une théorie du langage", *Langages* 8 ("La phonologie générative", S.A. Schane, ed.), p. 112-123

Moulton, William 1962: *The sounds of English and German* (Chicago)

Nyrop, Kr. 1951: *Manuel phonétique du français parlé* (6th ed., Copenhagen)

Rischel, Jørgen 1974: *Topics in West Greenlandic phonology* (Copenhagen)

Rohrer, C. 1967: "Die Behandlung der französischen Nasalvokale in der generativen Phonologie", in: Hamm, J. (ed.), *Phonologie der Gegenwart*, p. 287-296; reprinted in English in: Fudge, E. (ed.), *Phonology*, Penguin Education (Harmondsworth 1973), p. 232-241

Rohrer, C. 1968: "Das französische Vokalsystem", in: Brekle, H.E. and L. Lipka (eds.), *Wortbildung, Syntax und Morphologie. Festschrift zum 60. Geburtstag von Hans Marchand* (= *Janua linguarum*, Series maior 36), p. 190-202 (The Hague)

Rousselot, L'abbé 1897-1908: *Principes de phonétique expérimentale* (Paris)

Schane, Sanford 1968: French phonology and morphology (Cambridge, Mass.)

Selkirk, Lisa 1972: The phrase phonology of English and French (unpublished Ph.D.-thesis, MIT)

Skousen, Royal 1972: "On capturing regularities", Papers from the Eighth Regional Meeting, Chicago Linguistic Society, p. 567-577

Skousen, Royal 1973: "Evidence in phonology", in: Kisseberth, C.W. (ed.), Studies in Generative Phonology (Edmonton), p. 72-103

Togeby, Knud 1951: Structure immanente de la langue française = TCLC 6 (2d edition Paris 1965)

Westring Christensen, B. 1969: Review of Schane 1968, Revue Romane 4, p. 116-125

NOTES ON WORK IN PROGRESS

Electromyographic investigation of Danish consonants, stress, and stød

As part of a physiological investigation of Danish consonants, stress, and stød a series of electromyographic (EMG) recordings from some selected laryngeal muscles have been made in cooperation with Hajime Hirose, who visited our laboratory in May-June, 1974. These studies constitute a continuation of the experiments started by Hajime Hirose and Eli Fischer-Jørgensen at the Haskins Laboratories in 1972 (see Haskins SR, forthcoming) and also a continuation of the glottographic investigation undertaken by Frøkjær-Jensen, Ludvigsen and Rischel (1971, see ARIPUC 7). The material has not yet been processed, and this processing may take some time, but some preliminary conclusions can be drawn on the basis of visual inspection of the individual raw EMG traces and the integrated curves recorded on the mingograph.

1. Consonants

Most of the consonants in question were placed in word-initial position in a stressed syllable in vocalic surroundings. We obtained very good recordings from the interarytenoid (INT) for two Danish subjects and a tolerably good recording for one subject. All three showed a dip in INT curves for the consonants, and the general tendency is that the dip is most pronounced in the aspirated stops p, t, k and in f, s, h, slightly smaller for (unvoiced) b, d, g, somewhat smaller for v, and almost non-existing for l and m. For the same subjects we obtained good recordings from the posterior cricoarytenoid (PCA), showing a peak of short duration corresponding to the dip in INT, and with the same tendency in the amplitudes, i.e., highest for p, t, k, f, s, h, slightly lower for b, d, g, v, and rarely present for l and m.

These findings are in agreement with the results obtained for one subject in the Haskins experiment, and this means that the hypothesis tentatively advanced by Frøkjær-Jensen, Ludvigsen and Rischel in 1971, on the basis of glottographic recordings (viz. that the relatively small glottal opening found in the production of b, d, g might be accounted for solely by the aerodynamics of the production of these plosives) is not corroborated: there is a pronounced activity in PCA for b, d, g.

Moreover, the activity of the vocalis muscle (VOC) was recorded for seven Danish subjects. For one subject, the VOC pattern was not very clear with respect to the consonant production. The VOC tracings of the remaining subjects showed a pattern of suppression of EMG activity in the consonants similar to that observed for INT, the relationship between different categories of consonants being of the same kind, but with a larger degree of individual variation.

The cricothyroid (CT) activity was recorded from some of the subjects, but its behaviour for consonants does not show any consistent picture.

2. Stress

As far as word stress is concerned, the preliminary impression is that INT tends to show higher peaks for vowels with main stress than for those with secondary or weak stress, whereas VOC and, particularly, CT show EMG peaks in connection with high pitch irrespective of the degree of stress. The text material contained a number of disyllabic words with stress on the first syllable (e.g. ['p^hanə]), which were often pronounced with relatively low pitch on the stressed syllable and a rise on the unstressed syllable. Both VOC and CT showed higher activity at the beginning of the weak syllable in these cases.

3. Stød

Recordings from VOC for test words with and without stød were made for seven Danish subjects in order to see whether there was any difference in the activity of the laryngeal muscles. (The stød has been assumed by S. Smith (1944) to be due to an increased activity of the respiratory muscles, cp. also the report by P. Riber-Petersen in ARIPUC 7 on the acoustic characteristics of the stød.)

Four of the seven subjects showed a very clear peak in vowels with stød. Most of the subjects pronounced the stød syllables on a higher pitch than corresponding stød-less syllables, but the peak in the VOC curves cannot be due to the rise in pitch alone. The VOC peak for stød is generally very sharp and of a short duration (of a kind similar to that found for subject PM in the Haskins experiments), whereas some subjects had only a modest activity of VOC in high pitched syllables without stød. In cases where the stød is produced in a syllable with secondary stress, spoken at a relatively low pitch, the peak is usually lower, but at least one of the subjects has a clear peak also in this case, e.g. for [ˈma:ləbø:ʔɣʁ] VOC showed a peak in the third syllable, but CT had its peak in the first syllable.

The remaining three subjects did not show any obvious difference in VOC activity between stød and stød-less words. For one of the subjects the VOC curve did not seem very reliable, but for the other two it seemed all right. For one of the two subjects a short list of stød and stød-less words was recorded in two separate sessions with almost the same result. Thus, it seems as if VOC activity is subjected to individual differences. Both of these latter subjects had a very pronounced VOC activity for high pitch, approximately similar to the CT activity.

For one of the subjects a simultaneous recording of subglottal pressure was made. There seems to be a peak in the subglottal pressure at the beginning of a vowel segment with stød. Unfortunately, this subject was one of the three subjects who did not show any clear difference in VOC activity in stød and stød-less words, so that the timing cannot be compared with certainty. For the same subject an attempt was made to record activity from the intercostal muscles by means of subcutaneous needle electrodes, but the recording was not successful.

More details will be given in later reports.

E.F.-J.

ACOUSTICAL AND PERCEPTUAL PROPERTIES OF THE DANISH STØD

Nina G. Thorsen

In the preceding volume of ARIPUC (vol. 7, 1973) Pia Riber Petersen published a report on the acoustical characteristics of the Danish stød. However, a couple of problems were left unsolved. In particular, it was intriguing that one of her subjects showed little or no visible difference between stød and stød-less vowels in neither spectrograms nor in the oscillogram and intensity curves, whereas there was a clear difference in the pitch contour, and, rather surprisingly, his stød-words were identified better in a listening test than those of another subject, whose stød was clearly visible in the spectrograms, oscillogram, and intensity tracings.

If one could make (trained) listeners pin down a point in a stød-vowel where they felt the stød-phase to set in, we would know where to look, if not what to look for, since there has got to be something in the acoustical signal that distinguishes a stød vowel from a stød-less one. It might be, of course, that pitch is the only cue with some speakers, but more research is needed before we can establish this as a fact.

A pilot experiment was performed using the author as a subject.[1] The following two word pairs were recorded on tape: [blɛ:sʌ - blɛ:?sʌ] and [vi:sʌ - vi:?sʌ] ('(a) blower - (he) blows' and '(a) hand (on a clock, meter, etc.) - (he) shows') which differ only in respect to the stød, the vowel qualities in each pair being identical. The words were consciously produced on the same pitch contour. - By means of a segmentator

1) I speak a form of advanced standard Copenhagen Danish and have a clearly identifiable stød.

(Thorvaldsen, 1970) the words were cut into pieces of varying duration. The first consonant was always included; the shortest stimulus had a vowel duration corresponding to a rather short vowel (8-10 cs), and the longest stimulus included the beginning of the following [s]; there were 20 stimuli for each word, the difference in duration between neighbouring stimuli being 1 cs. - It is possible to adjust the (rise-time and) fall-time on the segmentator in steps from 5 to 60 ms. A fall-time of 20 ms was chosen since shorter fall-times gave a slight impression of a stop consonant or glottal stop! following the vowel.

The stimuli were randomized and arranged in two separate listening tests, one containing 40 segments of [blɛ:sʌ - blɛ:ʔsʌ], the other 40 segments of [vi:sʌ - vi:ʔsʌ].

5 trained phoneticians (including the author) were supplied with answering sheets and asked to decide whether or not they heard a stød in the mutilated words. They all took the test 3 times. (An untrained listener also took the test, but his answers were completely inconsistent, and it may be that naive listeners cannot perform a task like this, but it is, of course, premature to say anything definite about this - more subjects will have to be tried. However, it <u>is</u> a common observation among teachers in the general phonetics classes attended by students in their first year of university that these students cannot easily identify a stød word when they hear one; i.e. they are not aware of the difference between stød and stød-less words.)

The results were, from the author's point of view, surprisingly good. First of all, nobody gave "stød-answers" to stimuli that did not originally have the stød - which is important, since the cutting technique cannot then be made responsible for the perception of stød in any of the stimuli. Secondly, there was next to universal agreement on the point in the sequence short-to-long (stød)stimuli where the stød was perceived; that is, the 9 ([vi:ʔsʌ]) and 5 ([blɛ:ʔsʌ]) shortest stimuli (having vowel durations of 18 and 14 cs, respectively) were heard as having no stød, and all the remaining (longer) stimuli were judged as having stød. Now, this point coincides very nicely with the point in the tracings (marked with an arrow at the top) where an obvious change takes place (see fig. 1,a+b - only tracings of [blɛ:sʌ - blɛ:ʔsʌ] have been shown): the higher formants drop out, oscillations decrease in amplitude and are highly irregular, intensity decreases sharply and shows sharp peaks.

From PRP's recordings two words spoken (in a carrier sentence) by the aforementioned subject were chosen: [lɛ:sʌ - lɛ:ʔsʌ] ('(a) reader - (he) reads'). A listening test was constructed in the same way as before, but even before it was run, the results were suspected to be different from those reported above. - (It was very clear that the stød-word was produced on a higher pitch contour than the stød-less word (see fig. 2) and, furthermore, there seems to be a difference in voice quality in the two words (not visible in the tracings). This is true not only of this word pair, but of all the words in PRP's recordings of this subject.)

True enough, even the shortest mutilated stød-word was identified as having stød by the author and the only other phonetician who has taken the test so far, although we tried, purposely to disregard the pitch difference. This is in itself a dubious affair, particularly since the two words were recorded (and played back to the listeners) immediately before the test was run. Thus it may well be that the two words really have been identified as high-tone/low-tone words, and not as stød/stød-less words. It is interesting to note, though, that the pitch contour in the "stød-words", as spoken by this subject, is rising, whereas the general tendency with the other subjects in PRP's material is to produce stød-words on a falling contour. Also, as mentioned before, the "stød-words" seemed

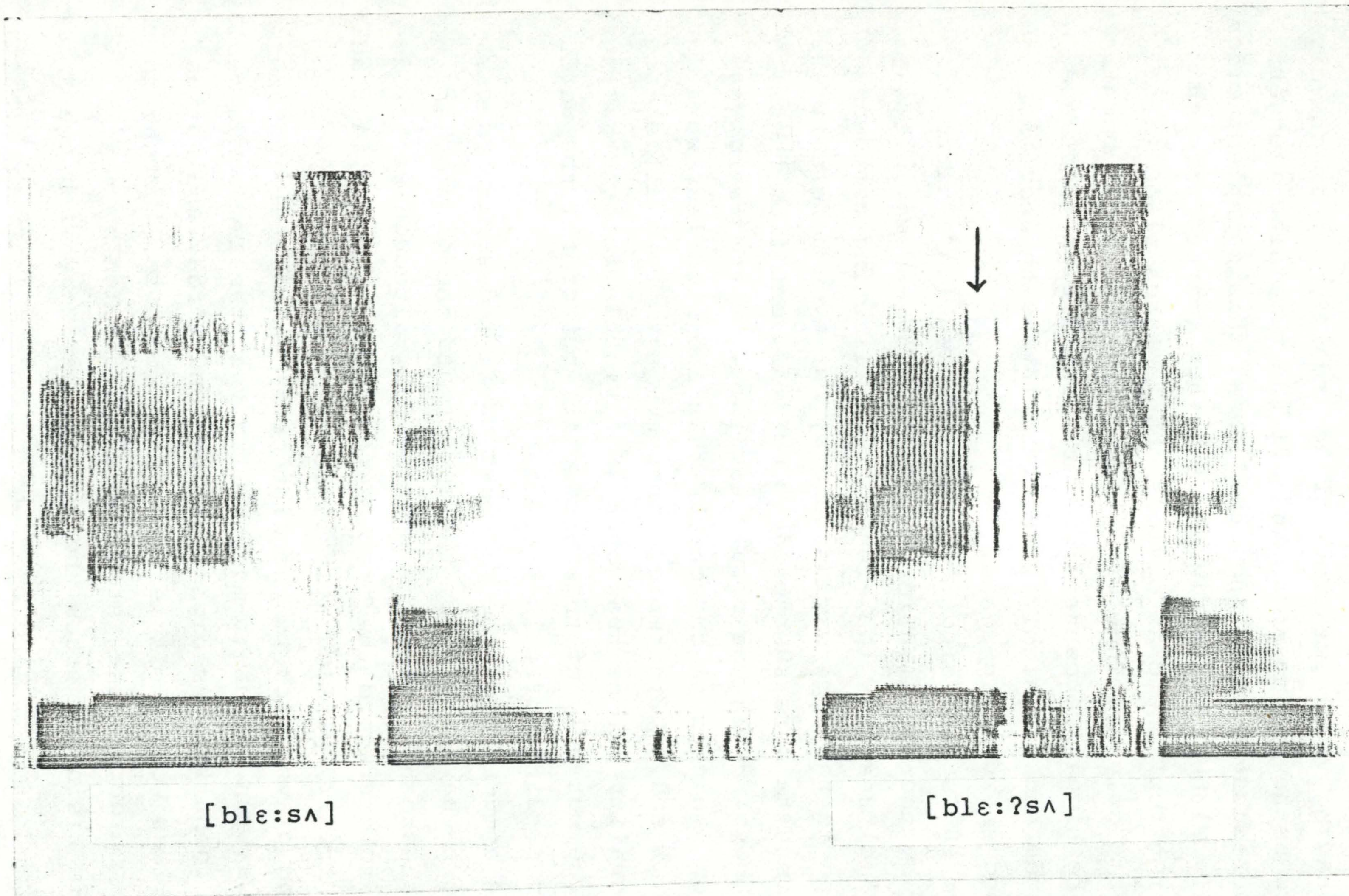
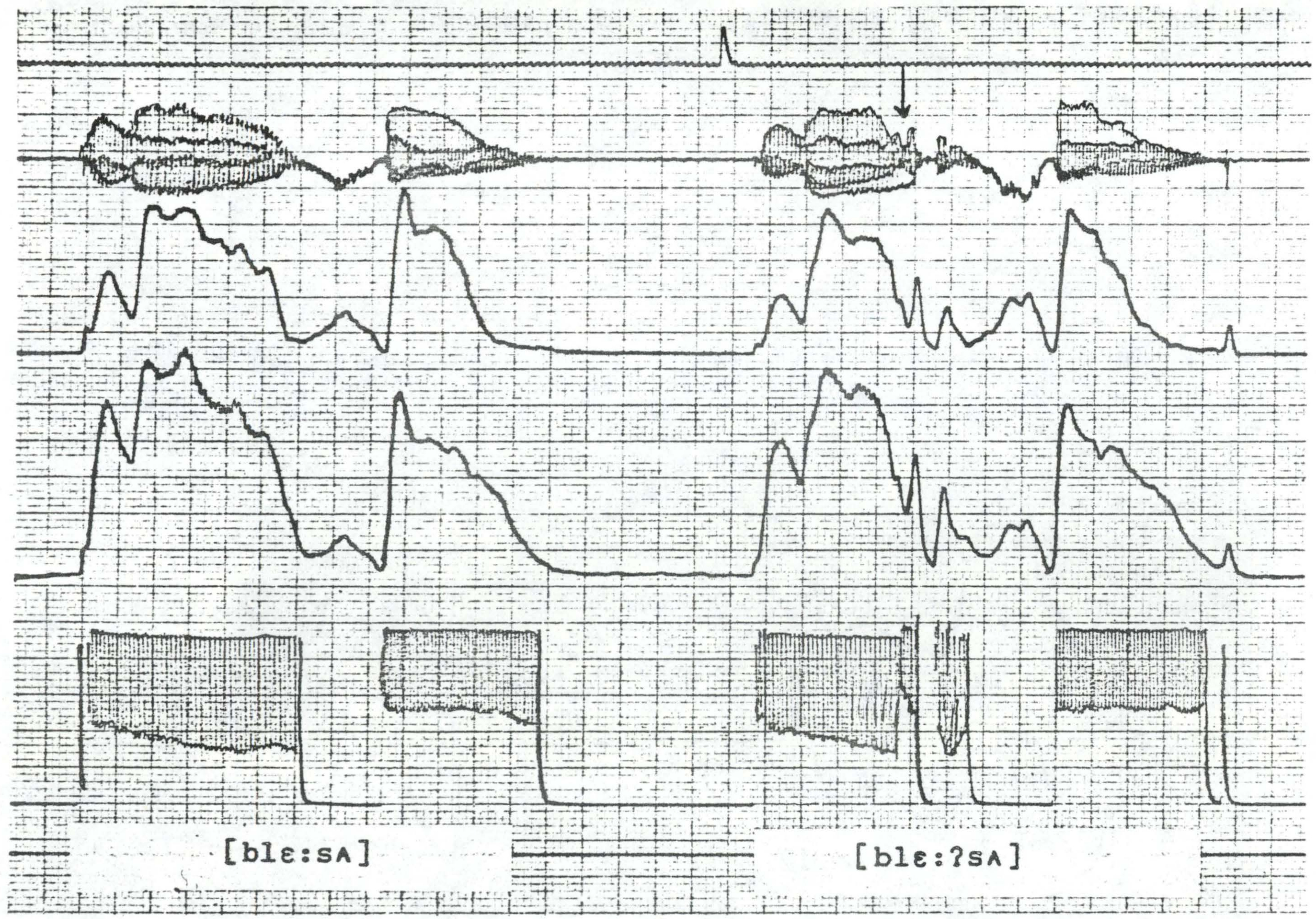


Figure 1a

Spectrogram of the author's pronunciation of [blɛ:sʌ - blɛ:ʔsʌ]



<u>Figure 1b</u>

Mingogram of the author's pronunciation of [blɛ:sʌ - blɛ:ʔsʌ]. The traces are, from top to bottom: duplex oscillogram, high-pass filtered (500 Hz) intensity (linear display), full frequency range intensity (linear display), and pitch.

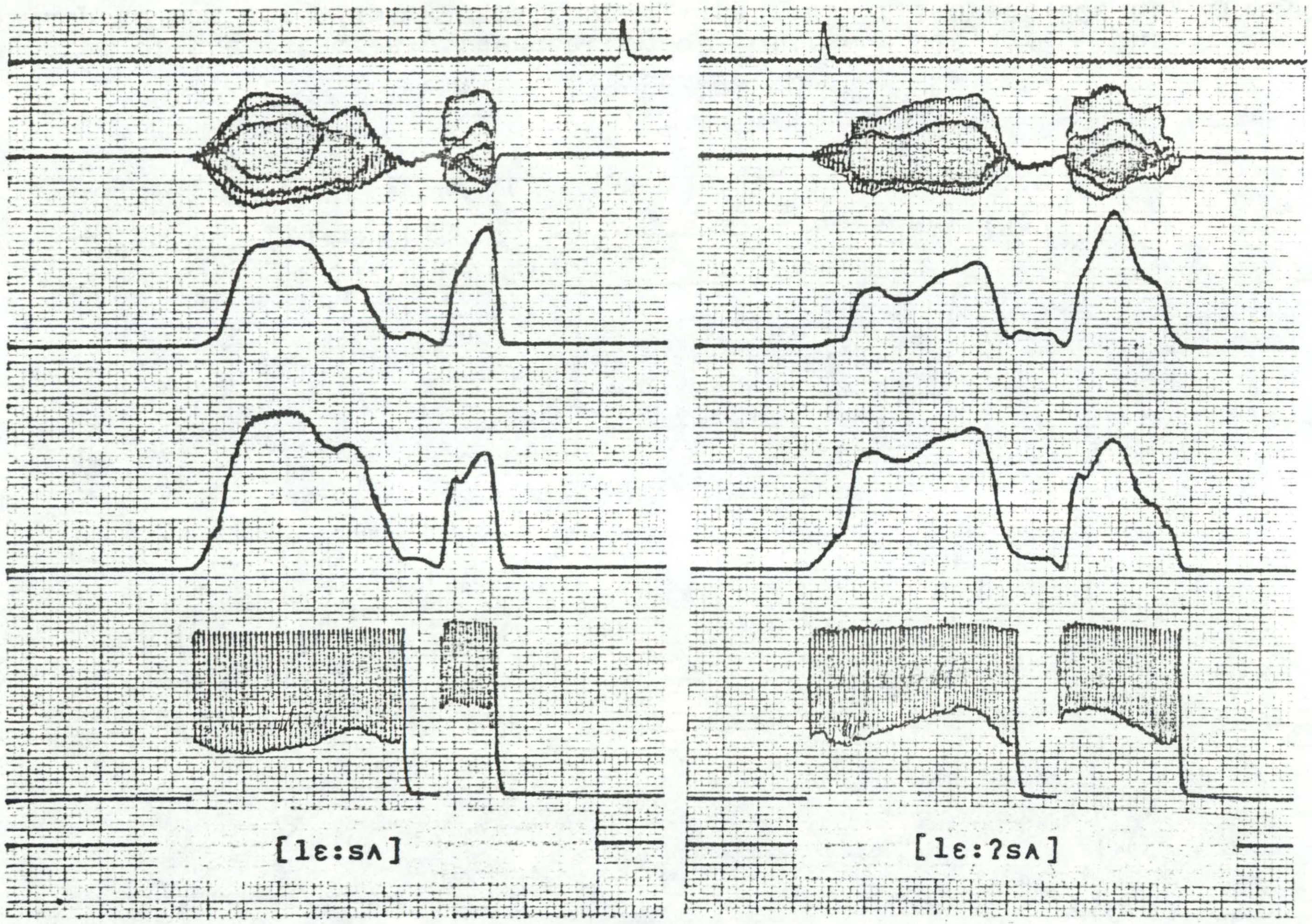


Figure 2

Mingograms of PRP's subject's pronunciation of [lɛ:sʌ - lɛ:ʔsʌ]. For an account of the traces, see fig. 1 b.

to have a different voice quality (slightly strained-sounding or compressed) - but whether this last factor is decisive as a cue for stød apart from pitch contour can only be determined when more experiments have been made.

It is evident, especially of course from PRP's investigations, that there are more ways than one to produce the stød and the task for the future will be to try to determine just how many different "stød"s are there, and is there a common denominator underlying them all? If there is not (and it is hardly to be expected), then how do we characterize these different acoustic signals, that are all perceived under the heading "stød"?

References

Petersen, P.R. 1973: "An instrumental investigation of the Danish "stød"", ARIPUC 7, p. 195-234

Thorvaldsen, P. 1970: "Construction of a segmentator", ARIPUC 4/1969, p. 8-15

PRELIMINARY WORK ON COMPUTER TESTING OF A GENERATIVE PHONOLOGY OF DANISH

Hans Basbøll and Kjeld Kristensen[1]

1. Introduction

The purpose of the project to be reported here is to test a part of a generative phonology of Danish (which had been worked out by one of the authors (HB) before this project started). An abstract phonological representation (i.e. a systematic phonemic representation) of a given word is used as input, and the program together with different sets of "background data" (see below) changes this input form to one or several phonetic representations. These phonetic representations can then be compared to actual pronunciations of the input word. If there is no agreement between the computed phonetic representation and the corresponding actual pronunciation, then there must be some error, either in the input form or in the background data (or in the program, of course). This error should then be found and corrected. Thus the ultimate purpose of this project is to improve the part of a Danish phonology being tested. It is evident that this method <u>only</u> allows finding <u>lack of observational adequacy</u>, i.e., it can be used <u>neither</u> as a <u>proof</u> of observational adequacy, nor for finding lack of <u>descriptive adequacy</u> (in the Chomskyan sense).

The project was initiated at the third Scandinavian summer school for research workers in the field of computional linguistics, held in Copenhagen from July 29th to August 10th,

---

1) Kjeld Kristensen is an engineer and cand.phil. (in Danish). He is a teaching assistant at the university of Copenhagen.

1974.[1] The program has only been tested with a very small sample of (quite arbitrary) data, but its general structure may nevertheless be briefly reported here. A more detailed account is planned for the forthcoming volume of ARIPUC.

2. Strategy

The program together with the three sets of background data (see section 2.1 below) is called the "grammar". The grammar may be considered to consist of a linearly ordered set of phonological rules together with the principles governing their application (see section 2.3 below). The input form of a given word is first submitted to rule no. 1 and is possibly changed by this rule. The output form from rule no. 1 is the input form to rule no. 2, etc. The output form from the last rule is the output of the grammar (see section 2.5).

2.1 General organization of program and data

As shown in fig. 1, the grammar consists of the MAIN PROGRAM together with three sets of "background data", viz. UNITMATRIX, RULEMATRIX and RULEINDEX. The main program operates with integers corresponding to the different vowels, consonants and boundaries. Thus we need a subroutine which translates a string of IPA-symbols (the input to the grammar) into the corresponding string of integers, and another subroutine which translates a string of integers into the corresponding string of IPA-symbols (the output from the grammar), see section 2.5 below.

---

1) We are indebted to the teachers at the summer school, Martin Kay and Richard Rubinstein, for much good advice during the initial stage of the project. The programming language used is UNIVAC ALGOL, and the program has been run at the regional computer center of the university of Copenhagen (RECKU).

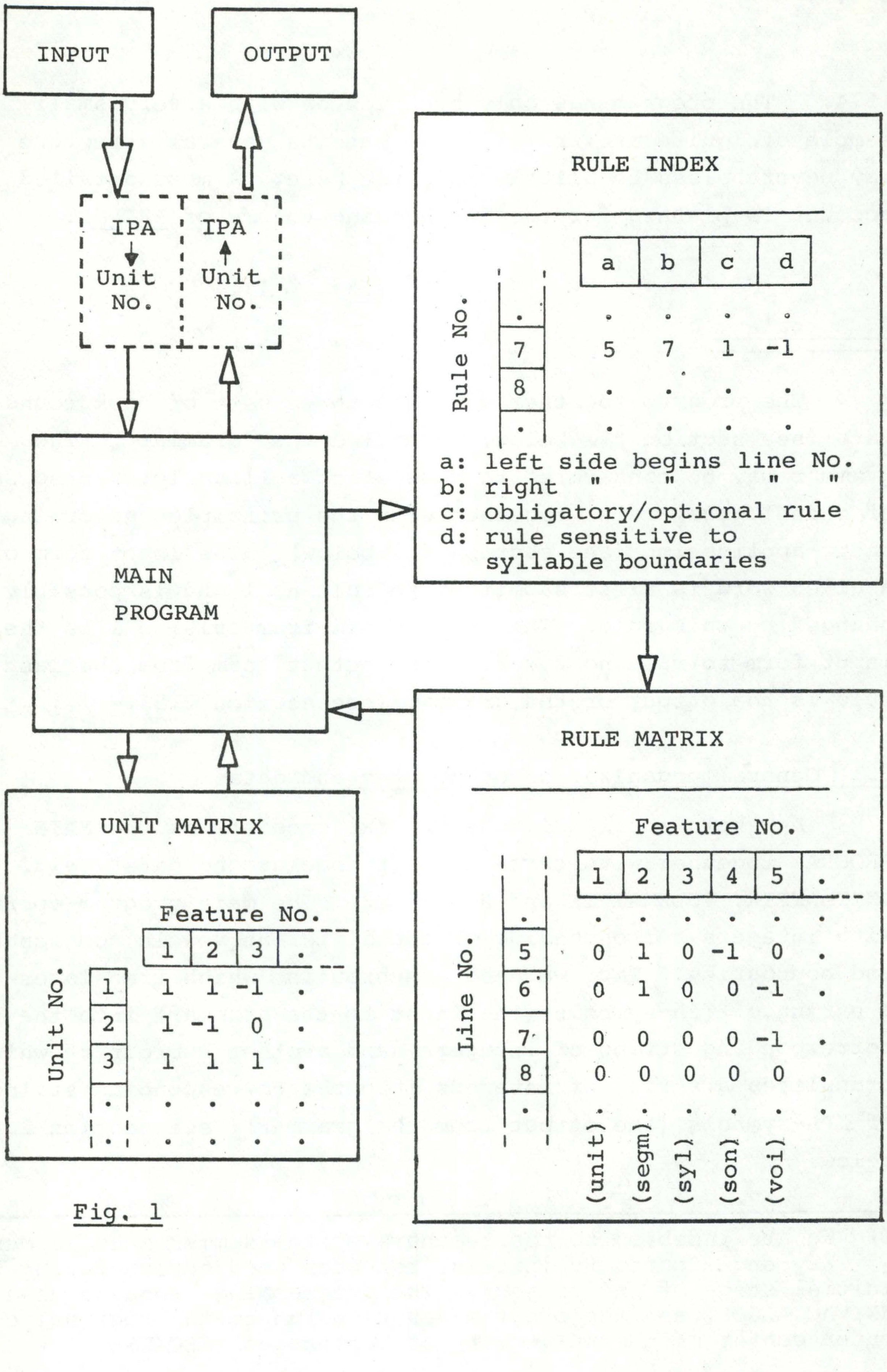

INPUT
OUTPUT
IPA
Unit No.
IPA
Unit No.
MAIN PROGRAM
RULE INDEX
a b c d
Rule No.
7 5 7 1 -1
8
a: left side begins line No.
b: right " " " "
c: obligatory/optional rule
d: rule sensitive to syllable boundaries
RULE MATRIX
Feature No.
1 2 3 4 5
Line No.
5 0 1 0 -1 0
6 0 1 0 0 -1
7 0 0 0 0 -1
8 0 0 0 0 0
(unit) (segm) (syl) (son) (voi)
UNIT MATRIX
Feature No.
1 2 3
Unit No.
1 1 1 -1
2 1 -1 0
3 1 1 1

Fig. 1

UNITMATRIX is a two-dimensional integer array, the dimensions being unit-no. (i.e. the integers corresponding to every vowel, consonant, and boundary) and feature-no. (i.e. the integers corresponding to every distinctive feature, like "segment", "syllabic", "sonorant", etc.). Binary features may have the coefficients 1 or -1, a multivalued feature like "height" may have the features 1, 2, 3, etc. The coefficient 0 (zero) is used in cases where the feature is phonetically undefined (e.g. the feature "lateral" is undefined for a boundary, i.e. for a unit which has the coefficient -1 for the feature "segment"). The matrix is thus not redundancy-free, but contains only phonetically meaningful specifications.

RULEMATRIX is a two-dimensional integer array, the dimensions being line-no. and feature-no. To each rule corresponds a number of consecutive lines in RULEMATRIX (e.g. to rule no. 7 correspond lines no. 5-8, see fig. 1). Each line number refers to a unit (i.e. a segment or a boundary in the phonological sense) in the left side or right side of a rule (see section 2.2 below). Only the feature coefficients which are crucial for the correct application of a rule, i.e. (by and large) the features which are mentioned in the standard notation of a phonological rule (see below), are specified in RULEMATRIX, all other coefficients are 0 (zero).

RULEINDEX is a two-dimensional integer array with rule-no. as one of its dimensions. The first number of the other dimension indicates the line number (in RULEMATRIX) where the left side of the rule begins, the second number indicates the line number where its right side begins, the third number indicates whether the rule is obligatory or optional, and the fourth and (for the moment) last number indicates whether the rule is sensitive to syllable boundaries or not. Thus RULEINDEX contains two types of information: it governs the data of RULEMATRIX (this has the consequence that the feature specifications of a rule can be stored everywhere in RULEMATRIX), and it contains all information which considers the rule as a whole, i.e. its domain and other general conditions for its application.

2.2 The format of phonological rules

(i) is the traditional notation of a phonological rewrite-rule which devoices obstruents before voiceless segments:

(i) [-son] ⟶ [-voi] / ___ [-voi]

(i) can be rendered in the format of a phonological transformation instead, viz. as (ii):

$$\text{(ii)}\quad \underset{1}{[\text{-son}]}\ \underset{2}{[\text{-voi}]} \Longrightarrow \underset{1}{[\text{-voi}]}\ \ 2$$

(i.e. if an input string consists of a [-son]-segment (1) followed by a [-voi]-segment (2), then the non-sonorant, i.e. obstruent (viz. 1), gets the specification [-voi], while the voiceless segment (viz. 2) remains unchanged). This is the format we use for storing phonological rules in RULEMATRIX.

Each unit of the rule (i.e. segment or boundary in the phonological sense) is represented by a line in RULEMATRIX, and we impose the restriction that there must be the <u>same</u> number of lines (in RULEMATRIX) corresponding to the two sides of a rule. This restriction does not mean, however, that we cannot handle deletion, since we use the feature "unit" as a common denominator for segments ([+unit, +segment]) and boundaries ([+unit, -segment]), in agreement with a suggestion of <u>SPE</u> (p. 359, footnote 14). Thus [-unit] means a blank (Ø in the rule algebra), which in UNITMATRIX is defined as -1 "unit", all other features being unspecified, i.e. zero. A deletion rule is a rule whereby a unit of the left side changes into a blank.

Let us now look closer at rule (ii), and let it be rule no. 7 in the ordering. It is stored in RULEMATRIX in four consecutive lines (i.e. lines no. 5-8 in fig. 1). The two latter lines, which correspond to the right side of (ii), only contain the specification that segment no. 1 of the rule gets the coefficient -1 for the feature "voiced", all other coefficients are unspecified (i.e. zero). The two first lines contain the specifications -1 "sonorant" and -1 "voiced", of course, but also +1 "segment" for both units (otherwise a boundary (which is unspecified as to the features "sonorant" and "voiced") in the input form to the rule would, incorrectly, be compatible with any of the segments of the left side of the rule (ii); any blank in the input form to a rule must be ignored, see the following section).

2.3 The application of an obligatory rule to a form

Let us see what happens when we apply rule (ii), which is rule no. 7 in the ordering, to a form. The output form from rule no. 6 is a string of integers where each integer is a unit-no. First of all, RULEINDEX is consulted to see whether the rule is optional or obligatory. If it were optional, we would go to the next rule in the ordering (see further section 2.4 below), but let us say that it is obligatory. RULEINDEX also indicates whether the rule in question is sensitive to the occurrence of syllable boundaries. If this is not the case, the syllable boundaries (which have been inserted by earlier rules) of the input form to the rule are ignored when the program determines whether the rule can be applied (see below). The program must also ignore any blank which the input form to a rule may contain (resulting from the application of earlier deletion rules); otherwise deletion rules could never create new environments for later rules to apply in.

Furthermore, RULEINDEX indicates at which line-no. the left and the right side of the rule begins, and because of the restriction that the two sides of the rule must be stored in the same number of lines in RULEMATRIX, and because all lines of a rule are adjacent, we know the position of each unit of the rule in RULEMATRIX. The program now examines whether any substring of the input form is compatible with the structural description (i.e., the left side) of the rule. This comparison is carried out between two ordered sets of feature numbers at a time, viz. line-no. (in RULEMATRIX) and unit-no. (in UNITMATRIX). There is compatibility between a line of the left side of the rule and a given unit in the input form of the rule if it is true for all features either that the coefficients of the feature are the same, or that one of them is zero. (It is this function of zero which guarantees that zero can never be used improperly as a third value of a binary feature, or a n+1th value of a n-ary feature.) If there is compatibility between a substring of the input form to the rule and all lines of the left side of the rule, then there is "full compatibility".

If no substring of the input form has full compatibility with respect to the structural description of the rule, then rule no. 7 is quitted and we turn to rule no. 8. But if there is full compatibility, then the units of the substring in question of the input form get the feature coefficients which are specified in the right side of the rule. All feature coefficients which are unspecified in the right side of the rule are kept unchanged in the input form. The program then consults UNITMATRIX in order to find the unit-no. corresponding to each of the units which have been changed by the rule (we impose the restriction that all derived units must be defined in UNITMATRIX). An ordered set of feature coefficients is defined as a certain unit-no. in UNITMATRIX if it is true for any feature either that the coefficients of the feature are identical or that one of them is zero. (Thus a unit which has been deleted, i.e. which has got the specification [-unit] by the rule, will be defined as zero (a blank) in UNITMATRIX since blank in the matrix is unspecified for all other features than [-unit].) The output form of the rule is thus a string of unit numbers, and this output form is then the input form to the following rule in the ordering, i.e. rule no. 8.

2.4 Optional rules

That a rule is optional means that it need not be applied even though its structural description is satisfied. Two alternative ways of handling optional rules within a program like ours present themselves: either we could indicate some stylistic value together with the input form to the grammar, and then specify the stylistic conditions for the application of each optional rule; or we could follow all possible paths through the derivation. We have chosen this latter procedure since it cannot destroy useful information, in contra-distinction to the former alternative (see below).

We adhere to the following strong working hypothesis on optional rules: <u>all optional phonological rules are stylistic</u>, i.e., the non-vacuous application (see below) and the non-application of an optional rule give output forms which differ as to level of style (formality, or the like), and, furthermore, <u>the non-vacuous application of an optional phonological rule gives an output form which belongs to a "lower" (less formal, etc.) level of style than the non-application</u>.

Each time a form is input to an optional rule, the rule is first skipped, and the form is input to the following rule (if this is also an optional rule, then this rule is skipped too, etc.). When such a path of derivation has been followed through, then the program returns to the last optional rule and this time tries to apply it, and so forth. If the grammar contains n optional rules, each input form will thus follow 2^n different paths of derivation. However, most of these will only differ in uninteresting ways (e.g. when the distinction is one between skipping an optional rule or trying to apply it in cases where its structural description is not met, i.e. is not satisfied for any substring).

The interesting case occurs when the structural description of an optional rule is met and its application will be non-vacuous (i.e. when the input form and the output form of the rule are distinct, i.e. do not consist of the same string of unit numbers). In such cases the program adds the designation L (for "lavsprog", i.e. "low style") to the output form from the rule if it is applied, and H (for "højsprog", i.e. "high style") if it is not applied. The printout contains (among other things, see below) the information L or H concerning the application or non-application of each optional rule whose application to the input form in question would be non-vacuous.

This stylistic information concerning the output opens up exciting perspectives, in addition to the fact that it could falsify our hypothesis on optional rules: Are some or all output forms with both L and H in their derivational history impossible since they are stylistically "incompatible" (e.g. is a form like [le:v̩] of leve impossible since shwa-assimilation is low style, and the keeping of the postvocalic [v] (instead of [u̯]) is high style, compare the normal high style and low style pronunciations [le:və, le(:)o], where [o] in [le(:)o] is derived via [u̯ə])? Can the optional rules be arranged into a "stylistic hierarchy" so that the L-application of a certain rule excludes the H-(non-)application of another rule but not inversely? Does a greater number of L-applications correspond to a more markedly low style pronunciation (as judged by native speakers of the language)? And so on.

2.5 Input and output

The dotted box in fig. 1 indicates the subroutines which translate from IPA-notation (in the input to the grammar) into unit-no., and from unit-no. into IPA-notation. Thus every printout will be in IPA-notation.

The printout corresponding to a given input form to the grammar will consist of this input form together with all the different output forms, and selected information on the derivational history of any path through the grammar. This derivational history consists of all intermediate forms which differ in interesting ways (see the preceding section), together with a letter for each rule in relation to each form: L, H (on these two letters, see the preceding section), O (meaning "the rule has not been skipped, but there was not full compatibility"), V (meaning "vacuous application", i.e. "there was full compatibility, but the input form equals the output form"), A (meaning "non-vacuous application of an obligatory rule", i.e. "there was full compatibility, and the input form was distinct from the output form"; A for obligatory rules thus corresponds to L for optional rules).

3. Concluding remarks

The preliminary nature of this report has already been emphasized. Our first task will be to fill a lot of relevant data into UNITMATRIX, RULEMATRIX and RULEINDEX, and to test the grammar with as many and as varied input data as possible.

This program ought to be coordinated with Peter Holtse's work on speech synthesis by rule of Standard Danish (see his report in the present volume of ARIPUC). The output forms of our program should be used as input forms to the rule synthesis program, and at present we (i.e. PH, HB and KK) are trying to make this possible. The perspective of this cooperation seems very interesting to us: we may then get an "external test" of the relevance of our phonetic representations, and the borderline between the two programs may turn out to give substance to a distinction between "phonological rules" (contained within the program reported here) and purely "phonetic rules" (contained within the rule synthesis program).

EVALUATION OF SPEECH DISORDERS BY MEANS OF LONG-TIME-AVERAGE-SPECTRA

Børge Frøkjær-Jensen and Svend Prytz[1]

1. Introduction

As part of a project dealing with acoustic parameters for objective estimation of voice disorders (Buch and Frøkjær-Jensen 1972) some pilot registrations have been carried out by means of long-time-average-spectra (LTAS) (Winckel 1967, Winckel and Krause 1965, Blomberg and Elenius 1970, Jansson and Sundberg 1973, Fritzell, Halén and Sundberg 1974) of 20 patients with unilateral paralyses of the recurrens nerve.

Our LTAS-analyses had two purposes: (1) to test the applicability of a new commercially available "Real-time Narrow Band Analyzer" (type 3348, Brüel & Kjær)[2] in speech research, and (2) to study the changes in the spectral distribution of speech intensity which occur during a period of voice therapy.

2. Instrumentation

To our knowledge, long-time-average-spectra of human speech have only been employed sporadically in the last 25 years, partly due to the complicated instrumentation needed for these analyses, especially when performed in real time. Only a few laboratories have been able to set up their own LTAS-analyzing system.

---

1) Svend Prytz is a doctor at the ENT-clinic, University Hospital, Copenhagen.

2) We thank the firm Brüel & Kjær, Nærum, Denmark, for kindly having placed a type 3348 analyzer, a type 2307 Level Recorder, and a type 7003 DC tape recorder at our disposal.

This paper does not attempt to describe the new instrumentation in detail. We only intend to demonstrate the utility of the equipment when applied to speech research.

The analyzer is a 400 channel hybrid measuring system which consists of (1) a spectrum analyzer, (2) an averager and interface control, and (3) a 12" CRT display unit. Furthermore, (4) a level recorder is used for the paper curve recordings.

The LTAS-spectra shown in this paper have been analyzed by means of the spectrum analyzer which was adjusted to a squared cosine filter slope (Hanning weighting which gives the best selectivity) with a 3 dB bandwidth of 18.75 Hz and a resolution (distance between the centre frequencies of adjacent filters) of 12.5 Hz. A full spectrum analysis in the frequency range 0-5000 Hz was carried out every 80 ms.

In the averager and interface control the spectra have been averaged over a period of 45 seconds with a 50 dB dynamic range of the resulting average spectrum. Since there are pauses in the read text, whose effect are to lower the average level, it has been necessary to subject the averaged spectrum to +12 dB "post averaging gain" before displaying it. This suffices to give a 40 dB dynamic range in the long-time-averaged-spectra for normal subjects if there are only few pauses in the reading.

The averager and interface control unit contain digital memories for storing two spectra. The contents of either memory may then be continuously displayed, or the contents of both may be displayed alternately. In this way comparisons of LTAS before and after speech therapy are easily made.

In the display unit (the CRT) each of the 400 channels are updated 22 times per second. A special feature is a 4-digit display showing the signal level in dB (resolution 0.2 dB) of any preselected channel. Outputs are available at the rear of the cabinet for read-outs in BCD code for further dataprocessing or in analog form for a YT-recorder.

3. The Real-Time-Narrow-Band-Analyzer used for long-time-averaging of speech spectra

3.1 Test of the instrumentation with speech input

The instrumentation has been tested with speech input from 30 different speakers, some of these speaking slowly with long pauses, some speaking fast without pauses, some having a good voice quality and some having a weak voice with few harmonics.

To sum up our results: The analyzer seems to be an expedient instrumentation for LTAS analyses of speakers not making too long pauses in the speech flow. For a pathological speaker who must often talk with long pauses, the dynamic range of the averaged spectrum is too reduced.

If 50 % of the averaging time is used for pauses the dynamic range is reduced with a factor 2 which equals 6 dB. Furthermore, the mean level of sustained vowels is about half of the peak level, i.e. 6 dB below the peak level, and the consonants are of considerably lower mean level. The mean level of the speech flow is 9-18 dB below the peak level when we do not take the pauses into account.

If we average over a read text including pauses the mean level will be about 12-24 dB below the peak levels, depending upon the pause/speech-ratio and the spectral composition of the speaker's voice.

Fig. 1 shows an LTAS of a female patient with unilateral recurrens paralysis. The voice sounds dull and weak with a reduced ability to modulate the pitch and intensity.

The dynamic range of the instrumentation was not sufficient for registration of amplitude levels above 1200 Hz. If the dynamic range of the LTAS could have been expanded down to -20 dB most of the relevant information in the spectrum could have been displayed.

3.2 Changes in LTAS during speech therapy

Pilot analyses have been made in order to test the validity of LTAS analyses as an objective means for the evaluation of the spectral changes that occur during the recovery period in patients suffering from speech disorders.

Typically, the LTAS analyses of recurrens paralyses can be divided into two groups, based upon the spectral changes:

(1) Patients who get a reduced energy spectrum above the first formant region during the recovery, and

(2) patients who get an increased energy spectrum above the first formant region during the recovery.

Both groups consist of patients having a dyscoordination between the subglottal airpressure and the medial compression of the vocal folds. In the first group the result is a phonatory hyperfunction which emphasizes the higher part of the spectrum and diminishes the stability of the fundamental frequency (rough voice quality), often to such a degree that diplophonia occurs.

The second group, on the contrary, is characterized by a phonatory hypofunction where the vocal folds cannot close very well, which results in an acoustic spectrum with weak higher harmonics. Often some breathy noise is heard, caused by turbulent air passing through the glottis.

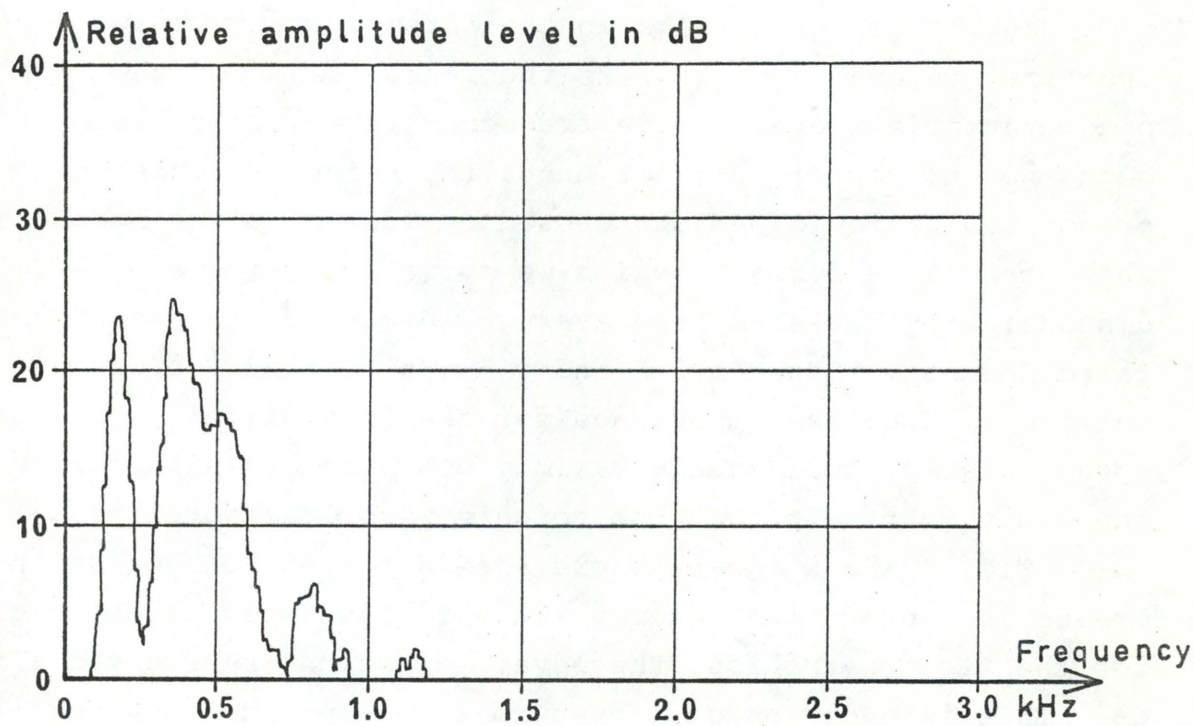


Figure 1

Spectral distribution of speech amplitude levels averaged over 45 seconds for a female patient with unilateral paralysis of the recurrens nerve.

Fig. 2 and Fig. 3 are illustrations of the two types of changes in the LTAS. In fig. 2 (female voice suffering from recurrens paralysis recorded before and during speech therapy) we observe that the acoustic spectrum below 1000 Hz has not been changed very much during therapy. Above 1000 Hz, however, the higher harmonics of the spectrum are weaker than before treatment. The three peaks in the first formant region depict the first three harmonics. If the voice is a good one with pitch variations over a wide frequency range, this discontinuity ought not to appear, but for unhealthy voices we observe a pronounced harmonic pattern caused by lack of pitch modulations. Above the first formant region we normally notice some spectral discontinuity caused by the average levels of the second and third formant. Between 3.5 and 4.5 kHz we find a broad energy maximum without any importance for the intelligibility. During speech therapy this energy maximum has been diminished with the auditive result that the voice sounds less shrilly.

Fig. 3 shows a case where speech therapy causes an increase in the spectral energy (in this case about 4 dB) except for the first harmonics, the levels of which are reduced a few dB. This is a common case of normal therapy progress for recurrens paralysis.

4. Further investigations

4.1 Averaging the LTAS analyses

Some preliminary attempts have been made with computer averaging of the LTASes.[1] The result of such averagings will be information concerning the mean distribution of the long-time averaged spectra for a certain group of speakers.

---

1) The averagings of long-time-average-spectra (recorded on a DC tape recorder) have been carried out on a PDP 8/lab computer at the ENT-clinic, University Hospital of Copenhagen.

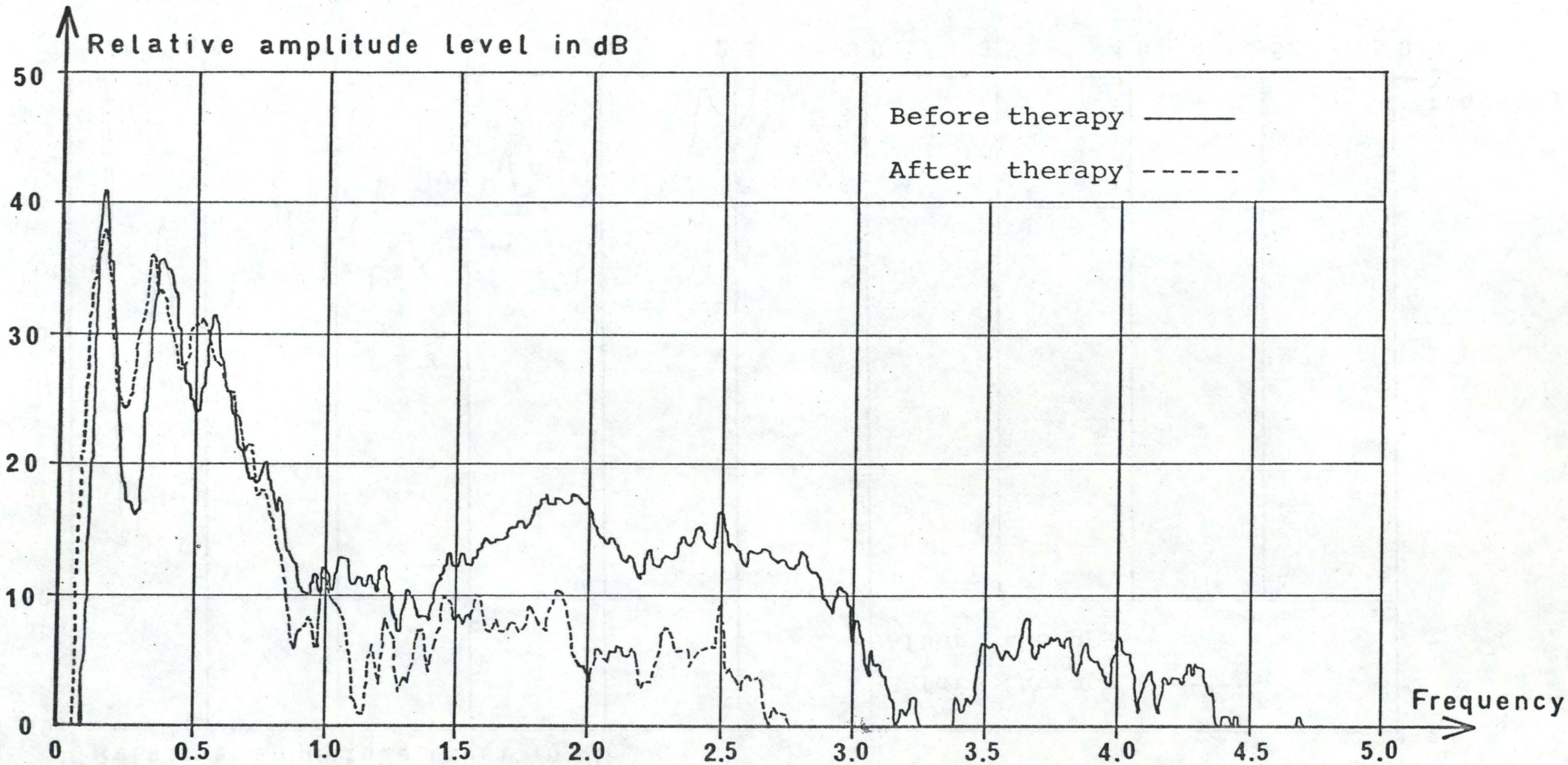


Figure 2

Spectral distribution of speech amplitude level averaged over 45 seconds.
Female patient with unilateral recurrens paralysis.

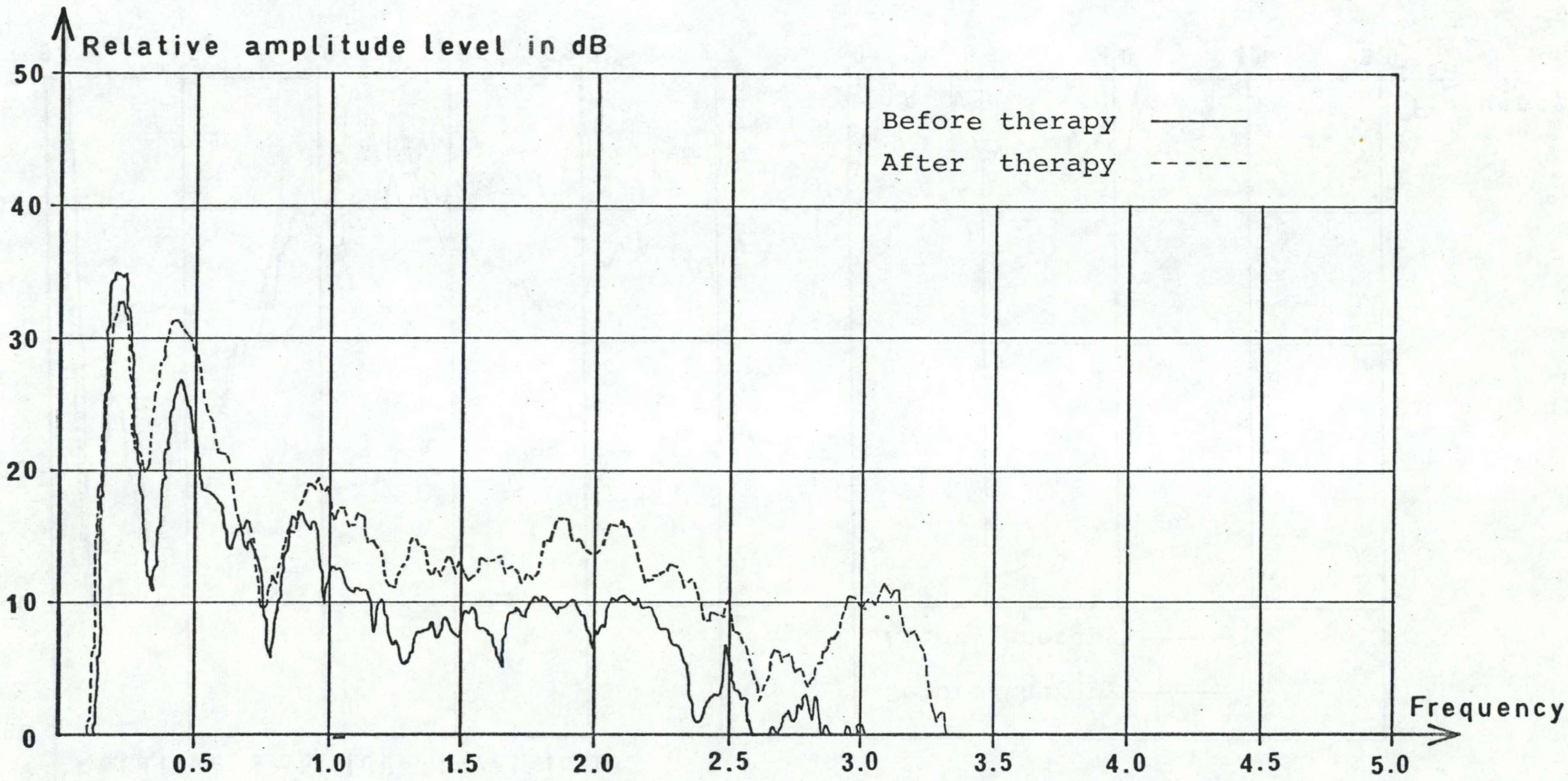


Figure 3

Spectral distribution of speech amplitude level averaged over 45 seconds.
Female patient with unilateral recurrens paralysis.

In general, these LTAS-averagings show for the normal voice:

(1) that the "normal" amplitude spectrum of speech has a slope of about 12-14 dB per octave, and

(2) that the first formant emphasizes the spectrum round 400-500 Hz an extra couple of dB,

both phenomena have been demonstrated previously by other authors (Dunn & White 1940, Winckel & Krause 1965, Fritzell, Hallén and Sundberg 1974).

In particular, the LTAS-averagings of a homogeneous and well documented group of patients suffering from a certain disorder yield information concerning the general spectral tendencies for this disorder, and may thus be useful in establishing criteria for the evaluation of the deviations from the normal spectrum.

For example: during a period of voice therapy of voices suffering from recurrens paralysis the spectral changes in the LTAS-averagings show that the therapy mainly affects the level of harmonics above 1000 Hz (increase or decrease as explained above, see Buch & Frøkjær-Jensen 1972).

The contribution of the unvoiced sounds to the total spectral energy is nearly negligible as far as we have experienced. However, more detailed studies of the spectral contribution from the unvoiced and from the voiced sounds will be carried out separately. A phonation detector (sensing when the first formant is present) with a gate has been constructed. This circuit will be used for gating out the unvoiced and voiced sounds, respectively, before making the LTAS analyses.

4.2 Variations in the averaging time

All our analyses have been made with an averaging time of 45 seconds. A few LTAS have been made with shorter averaging times, and it seems that even with an averaging time as short as 20 seconds the average spectrum will differ only slightly from average spectra with a longer averaging time.

Future investigations will, among other things, concern the relation between the averaging time and the variations of the LTASes.

5. Summary

The LTAS analysis can now be performed easily in real time with a new type "Real-Time-Narrow-Band-Analyzer". The instrument has been tested on speech material. Preliminary results show that the LTAS analysis seems to be a good way of obtaining information about the acoustic changes in voice quality, e.g. as a diagnostic equipment in speech clinics or as an objective means for the evaluation of voice quality.

References

Blomberg, M. and K. Elenius 1970: "Statistical analysis of speech signals", QPSR 4, p. 1-8

Buch, Nils H. and B. Frøkjær-Jensen 1972: "Some remarks on acoustic parameters in speech disorders", ARIPUC 6, p. 245-259

Dunn, H.K. and S.D. White 1940: "Statistical measurements on conversational speech", JASA 11, p. 278-288

French, N.R. and J.C. Steinberg 1947: "Factors governing the intelligibility of speech sounds", JASA 19, p. 90-119

Fritzell, Bjørn, O. Hallén, and Johan Sundberg 1974: "Evaluation of teflon injection therapy for paralytic dysphonia", QPSR 1, p. 18-24

Jansson, E.V. and Johan Sundberg 1973: "Organ stops analyzed by means of Long-Time-Average-Spectra", QPSR 4, p. 48-53

Winckel, F. 1967: "Darstellung des Sprechverhaltens als Statistische Tonhöhen und Formantverteilung mittels "Langzeitsanalyse"", Proc.Phon. 4, p. 1031-1035

Winckel, F. and M. Krause 1965: "Ermittlung von Spracheigenschaften aus Statistischen Verteilungen von Amplitude und Tonhöhe", Proc. Acoust. 5 (Liege) paper A41.

PRELIMINARY EXPERIMENTS WITH SYNTHESIS BY RULE OF STANDARD DANISH

Peter Holtse

1. Introduction

Since the autumn of 1973 a system for synthesis by rule of Standard Danish has been under development as a joint project between the Telecommunications Research Laboratory (TFL) and the Institute of Phonetics, Copenhagen.[1]

The aim of the project is to develop a system which will generate acceptable spoken Danish from a table of stored parameter values. The system should provide a basis for testing hypotheses about the perceptual relevance of various acoustic features, and it is hoped that it will help in formulating hypotheses on subjects like temporal organization and intonation of Danish.

2. Hardware equipment

The rule system is implemented on the RC 4000 computer system of TFL. The hardware equipment consists of an RC 4000 with 32 k bytes of core and conventional peripheral equipment. The actual synthesis is performed by a digitally controlled OVE IIIc synthesizer. Control information from the central

1) The practical problems of writing the computer programs are handled by B. Bagger Sørensen of TFL.

computer is transferred to a PDP-8/I minicomputer working as a buffer station. The buffer program controls the synthesizer by updating the fifteen control parameters of OVE once every 10 milliseconds. (Fig. 1.)

3. Synthesis strategy

The synthesis program consists of a complex of four independent computer programs communicating via the operating system of the central computer. This means that the synthesis programs are run in time sharing with any other jobs under execution. Therefore the time needed to synthesize a given string of speech may vary depending on the number of other jobs running. During the writing of the programs we have been more concerned with the development of a useful tool for phonetic research than with reducing execution time or program size.

The program synthesizes a sample of speech by chaining together a string of "segments" stored on the disc. A "segment" is a characteristic acoustic event, e.g. the explosion of a stop consonant, the steady state of a vowel, or the fricative phase of an [s]. Each segment has associated with it a table of information consisting of (1) a head and (2) a matrix of values for the central parameters of the synthesizer.

The control parameters are:

AO: amplitude of voice source
Ac: amplitude of friction noise
Ah: amplitude of hiss through formants
An: amplitude through nasal branch
F1: frequency of first formant
F2: " " second "
F3: " " third "
Ak: pole/zero ratio of fricative formants
K1: frequency of first fricative formant
K2: " " second " "

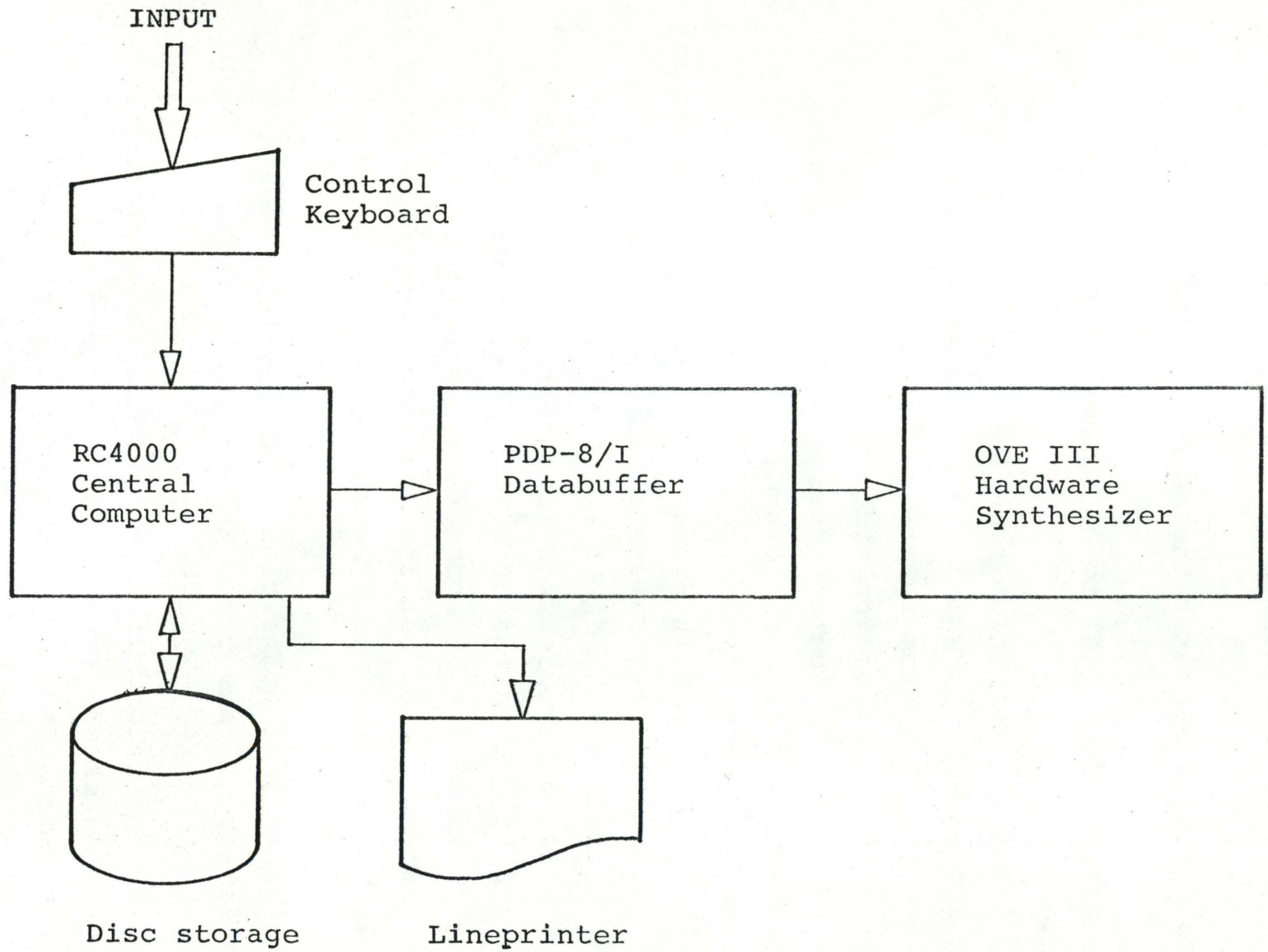


Figure 1

Hardware organization of the synthesis system.

The synthesizer needs a twelfth control parameter, FO: fundamental frequency, but the segment tables contain no information on this parameter. The band widths of the three (vowel) formants are kept at a constant value.

For each control parameter the matrix contains three entries:

(1) The target value (tg) to be reached during the segment

(2) The internal transition time (ti), i.e. the part of the segment during which the parameter is moving towards or away from the target value.

(3) The external transition time (tx), i.e. the part of the neighbouring segment during which the parameter is moving away from or towards the target value of the neighbouring segment.

All target values are listed in Hz for the frequency parameters and in dB for the amplitude parameters. The transition times are listed in centiseconds.

The transition times are used to interpolate straight lines between the target values of adjoining segments, as shown in figure 2. Straight line interpolation has been chosen as being easy to conceptualize, but if experience should demonstrate the need for it the program could be changed to use other strategies.

Figure 3 shows that transition times may have negative values. This feature allows the boundary between adjacent segments to be treated as a mere reference line.

The head of the information table for each segment tells the chaining procedures of the synthesis programs how the data of the parameter matrix should be treated. The first entry of the head is the name of the segment. The name must be two alphanumeric characters. Furthermore, the head contains the standard duration and the rank of the segment. When two segments

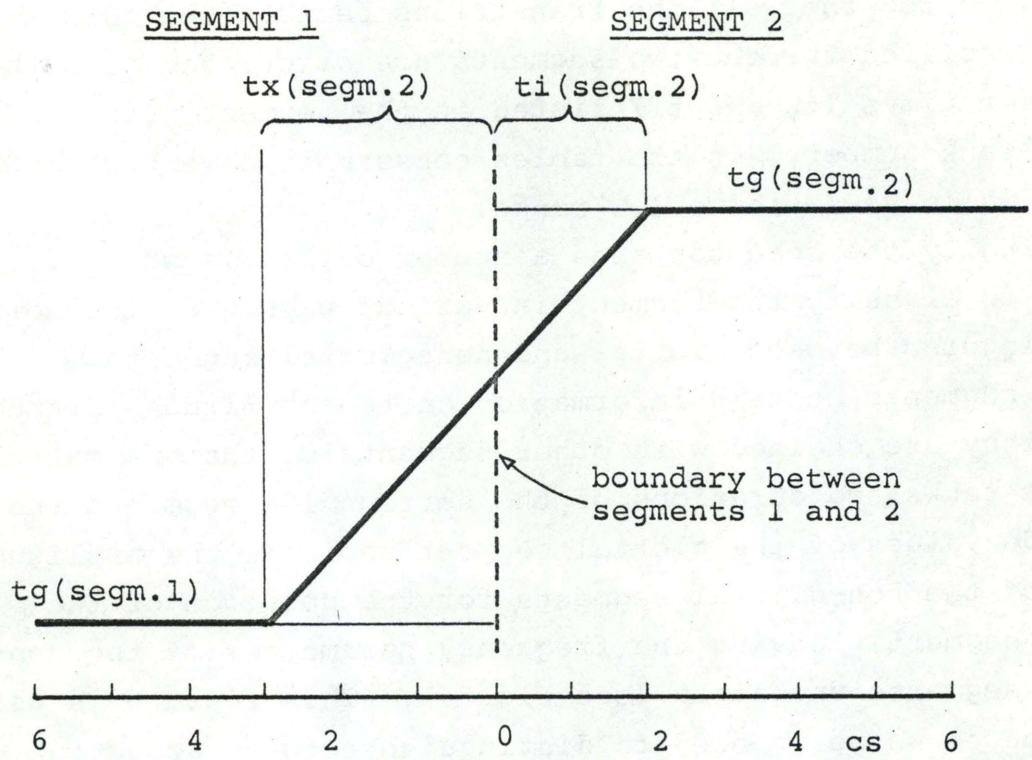


Figure 2

Straight line interpolation between two segments. Transition times in this example are taken from segment 2.

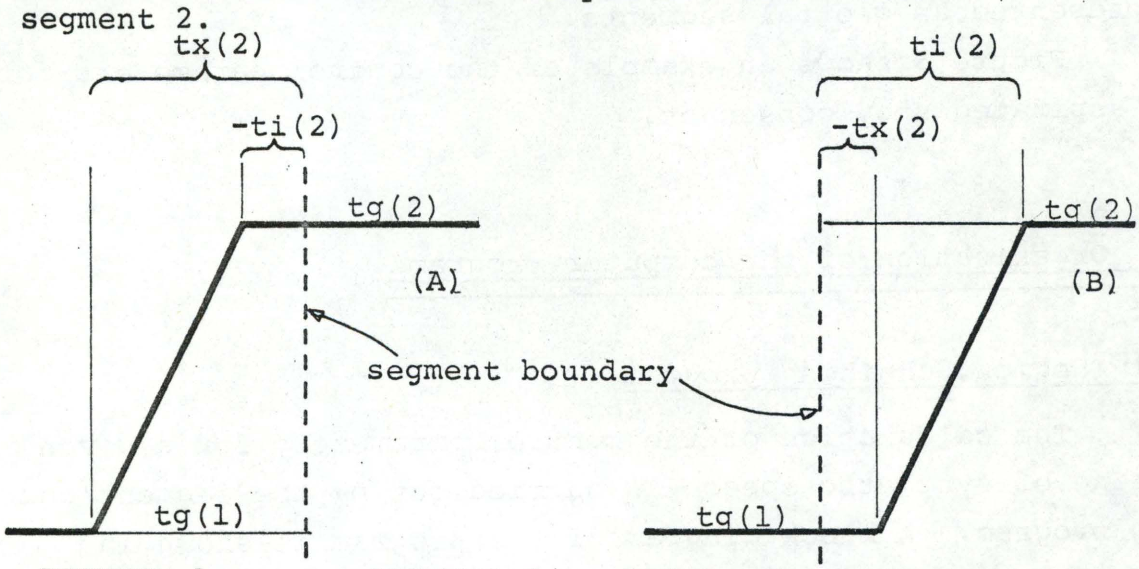


Figure 3

Examples of negative values of internal (A) and external (B) transition times. The entire transition may be moved to either side of the segment boundary line.

are chained together all the transitions between the target values specified for the two segments are calculated from the transition times (tx and ti) listed in the segment with the highest rank number. In the tables consonants have high rank numbers while vowels have low ranks.

Finally the head contains a number of labels which may be used to classify the segment in various ways. At the moment we distinguish between glottal and non-glottal segments. Glottal segments contain information on the amplitude parameters only. They are chained with other segments so that no matter what the ranks and durations of the surrounding segments the amplitude values of the glottal segment replaces the amplitude values of the non-glottal segments for the duration of the glottal segment, leaving the frequency parameters of the non-glottal segments unchanged (see fig. 4). This feature is useful in cases where we need to distinguish between acoustic parameters governed mainly by the upper articulators as opposed to acoustic parameters governed by the larynx. Thus aspiration of stop consonants and the Danish "stød" are, for the moment, represented as glottal segments.

Figure 5 shows an example of the control parameters for an aspirated stop consonant.

4. Organization of the computer programs

4.1 Segment chaining program

The calculation of the control parameters for a given string of synthetic speech is carried out by the segment chaining program. A block diagram of the program is shown in figure 6.

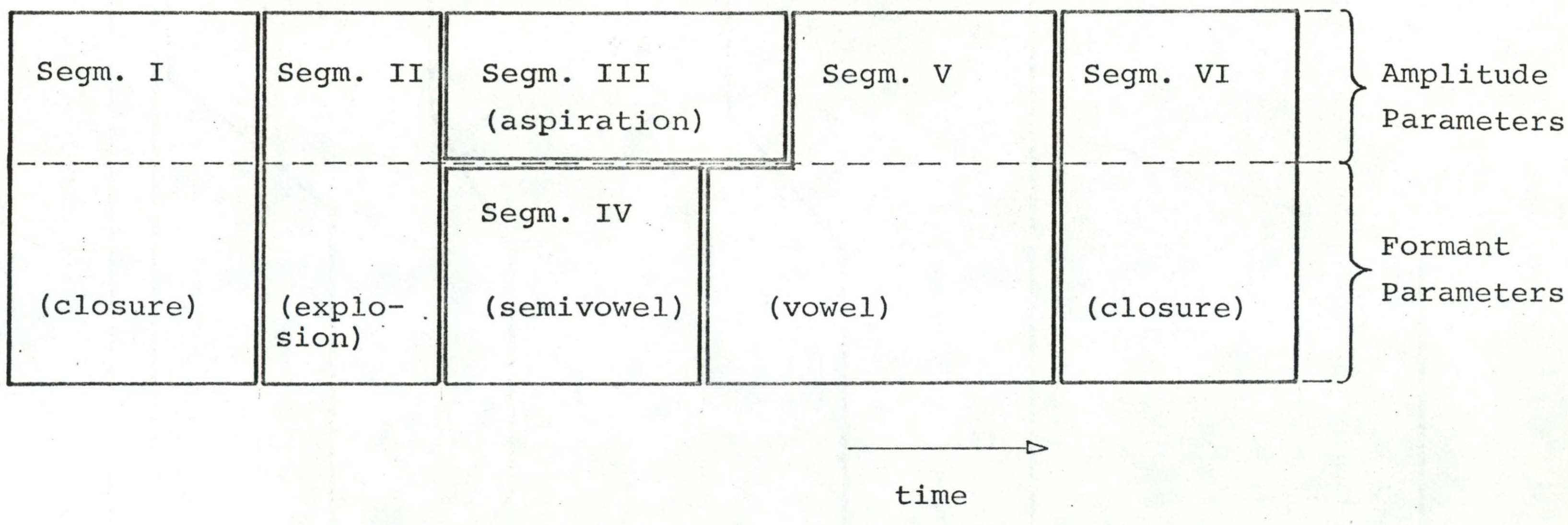


Figure 4

Schematic drawing showing how a glottal segment assumes control over the amplitude parameters and leaves the formants unchanged.
The example shows the chaining of the segments for the sequence /pjat/ = [pçad] . The segments will be listed in the input string as /closure/, /explosion/ /aspiration/,/semivowel/, /vowel/ etc.

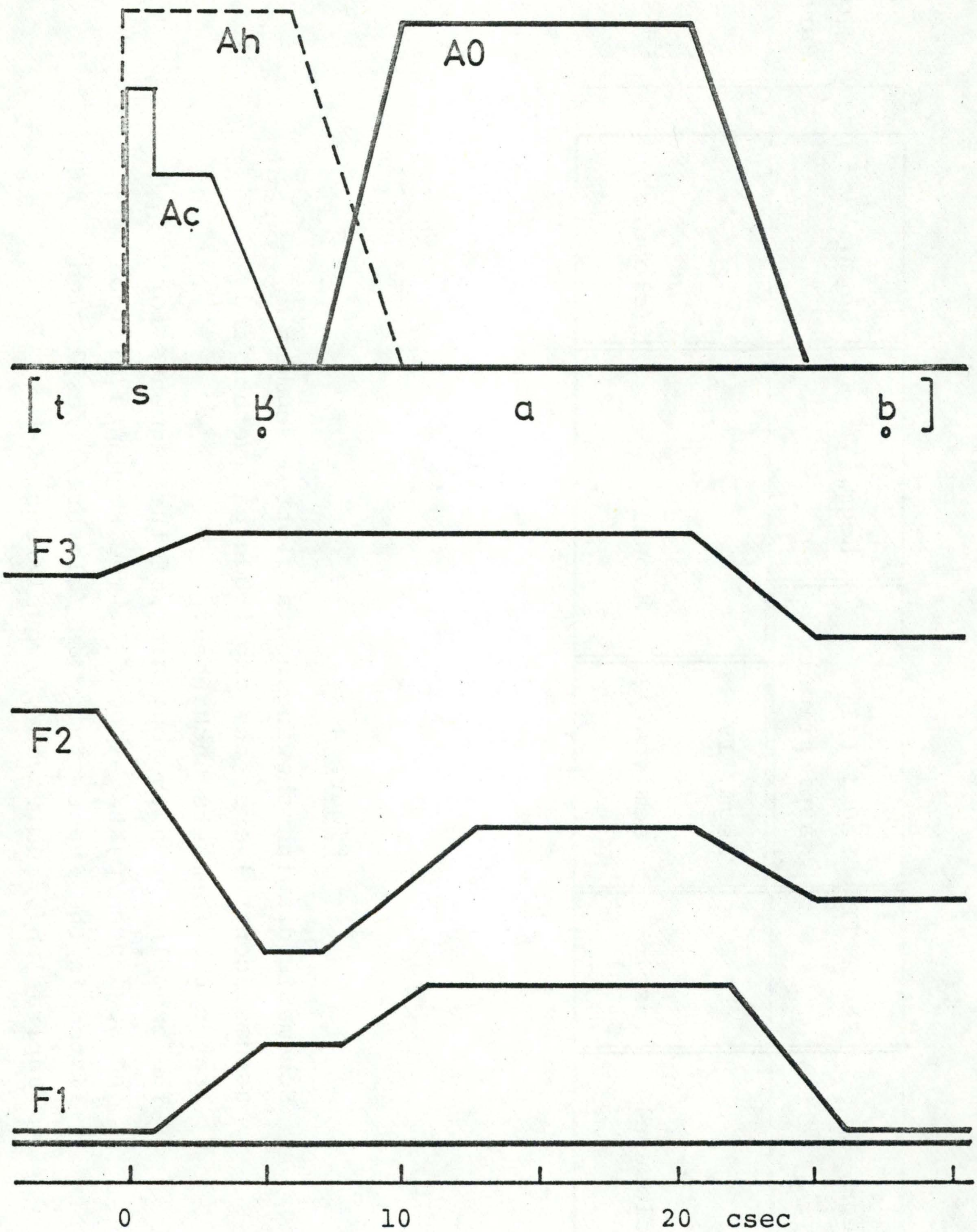


Figure 5

Example of the control parameters for the sequence /trab/.

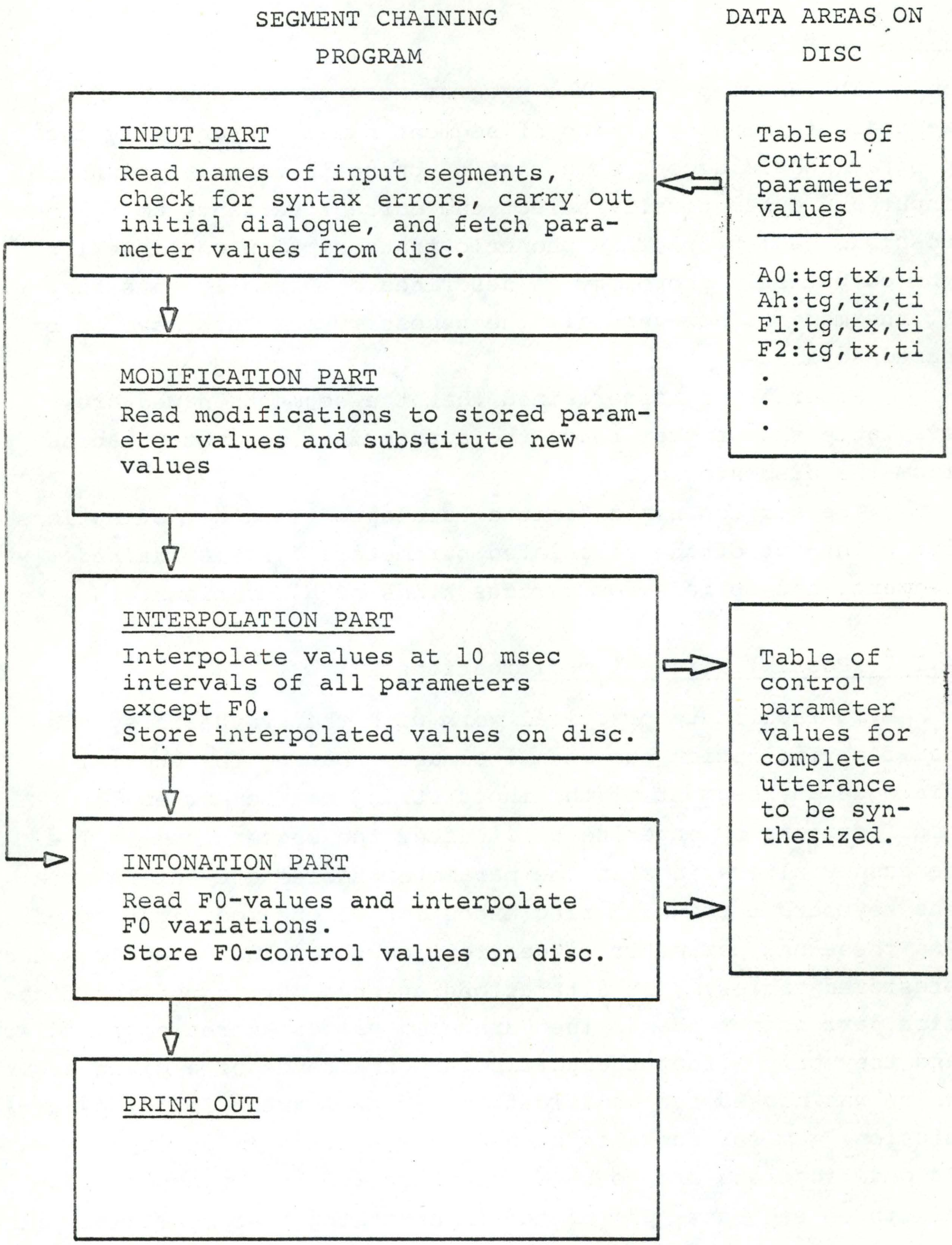


Figure 6

Block diagram of synthesis program.

4.1.1 Input part

The input part of the program accepts as input from a keyboard terminal a string of segment names. Eventually the program should accept some kind of phonetic transcription as input and automatically select the correct variants among the possible segments. This phonetic transcription could well be the output of a phonology as described by Basbøll (this issue). At the moment, however, all the necessary segments have to be named separately.

After having ascertained that the segments named are available the program reads the appropriate parameter tables from the disc storage.

The section named "initial dialogue" is a set of options for print-out of the calculated parameters for the chained segments and options for various kinds of alterations.

4.1.2 Possibilities of modifications

To facilitate practical work with the synthesis system modification option has been included. During the initial dialogue any segment of the input string may be marked for modifications by entering a '!' after the segment name. The necessary alterations in the parameter tables are entered via the keyboard terminal during execution of the modification part. And these new parameter values replace the values read from the prestored tables. The alterations entered during the modification part do not change the parameter values stored on the disc, and they only affect the particular occurrence of a given segment which was marked for modification. Thus segment 'p2' ([p]-explosion) may for instance be used three times in an input string. If modifications are to have effect on all three occurrences all three segments 'p2' of the input string must be marked with a '!'. And the same alterations must be typed three times.

(In this way it would, of course, also be possible to make three different editions of the same segment in one synthesis run.)

When all modifications are entered the program interpolates straight lines between adjacent segments as described in section 3. The interpolated values, quantized in steps of 10 msec, for all control parameters except F0 are stored on the disc.

4.1.3 Fundamental frequency control

The last part of the program calculates the values needed for the control of the fundamental frequency. At the moment there are two F0-options: Either a detailed description of the intonation pattern may be entered via the keyboard or an automatic intonation contour generator may be called. The automatic F0-variation is a slightly falling curve whose only justification is that it avoids a complete monotone which is uncomfortable.

Any detailed, or indeed natural sounding, F0-variation must be entered by hand. This is done by typing a series of points in time and their corresponding fundamental frequency values into the computer. The program then interpolates straight lines between the specified points. After interpolation the F0-control values are stored on the disc together with the other control parameters.

It is possible to make repeated changes to the F0-pattern of a given synthetic utterance without recalculating all the control parameters. This is done by typing an '=' as input string during the input part. Program control then goes directly to the intonation part of the program, bypassing the modification and interpolation parts. This allows the user to enter a new intonation pattern to the old edition of the synthesized string.

4.1.4 Print-out of control values

The last part of the program has options for printing out the complete list of control parameter values on the line printer. For visual reference a special option draws a diagram of the variations of the control parameters on the line printer.

4.2 Utility routines

Besides the segment chaining program there are three utility routines. One, the execution program, reads the output from the synthesis program on the disc, converts the frequency and amplitude values to the binary format demanded by the synthesizer, and transmits these control values to the PDP-8 buffer. Using the PDP-8 as a buffer allows quite long strings of continuous speech to be synthesized, the upper limit being given only by the storage capacity of the disc.

The other two utility routines are: "GETPAR" which has options for printing the contents of the permanent segment tables on the disc storage, and "READPAR" which enters modifications into the permanent tables.

5. Conclusion

So far the emphasis of our work has been on developing an adequate library of segment tables. This work is still in progress, and we are beginning to consider the problems of segment duration as a function of number and quality of surrounding segments. Later on we hope to develop algorithms for the control of intonation, stress, and rhythm. This work will be done in cooperation with the group working on a computer based phonology of Danish (cf. Basbøll, this issue).

References

Basbøll, H. and K. Kristensen 1974: "Preliminary work on computer testing of a generative phonology of Danish" (this issue p. 216-226)

Holmes, J.N., I.G. Mattingly, and J.N. Shearme 1964: "Speech synthesis by rule", LS 7, p. 127-143

Nooteboom, S.G., I.H. Slis, and L.F. Willems 1973: "Speech synthesis by rule: Why, what and how?", IPO 8, p. 3-13