ANNUAL REPORT

of the
Institute of Phonetics
University of Copenhagen

ANNUAL REPORT

of the
Institute of Phonetics
University of Copenhagen

96, Njalsgade
DK 2300 Copenhagen

CONTENTS

INTRODUCTION

Eli Fischer-Jørgensen

This year we are celebrating the tenth anniversary of our
institute, and this is our report No. 10, - a suitable occasion
for looking back, tracing some lines of development and taking
stock of the situation. A further reason is that our institute
has now found a more permanent home in the new campus of the
university, and our moving about from one unsatisfactory location
to another has come to an end.

In the introduction to the first report, in 1966, I sketched
the prehistory of the institute. As that report was issued only
in a limited number, and since the introduction will not be re-
printed in re-issues of ARIPUC 1, the main points are included in
the present survey.

The teaching of phonetics has a long tradition at the Uni-
versity of Copenhagen. Vilhelm Thomsen lectured on general phon-
etics from 1881, and Otto Jespersen took up the subject repeatedly
during the years 1895-1905. It is also due to Jespersen that the
phonetics of modern languages has been taught regularly since the
beginning of this century, and was part of the requirements for
the MA in modern languages. Courses in general phonetics were,
however, not resumed until 1939, by Louis Hjelmslev. From 1943,
when a lectureship in phonetics was established, these courses
were given every year as a preliminary to the courses in foreign
languages,which were administered by the various language depart-
ments. The number of students attending these courses augmented
slowly, from three in 1943 to about 800 in 1975. However, for
some languages the general phonetics courses and the courses in

the phonetics of the foreign language are now more or less
integrated (from 1976 the general course for students of English
has been drastically cut down and will mainly be given within
the department of English).

Apart from the elementary courses special courses for ad-
vanced students were given from 1943 onwards. In 1960 a diploma
of general phonetics, based on a two years' curriculum, was
established. It has, up till now, been taken by 22 students.
From 1967 it has also been possible to take a master's degree in
phonetics, based on a four years' curriculum. This degree has
been taken by four students, and another four are presently pre-
paring for it.

Until 1963 there was only one permanent post in phonetics,
but during the latter part of the sixties the staff was rapidly
increased. In 1966 a chair in phonetics was founded, and from
1970 the staff counted seven phoneticians in all. This develop-
ment has now come to a standstill.

The possibilities for phonetic research developed more
slowly although the research was initiated earlier than the teach-
ing. The first vowel synthesizer was built in Copenhagen as early
as 1781 by C.G. Kratzenstein, and about a century later the famous
comparatist Karl Verner constructed an ingenious instrument for
enlarging the traces of the Edison phonograph. At the same time
Georg Forchhammer worked with a phonoscope, and in the early thir-
ties of this century Poul Andersen and Svend Smith received a
grant from the Carlsberg Foundation for a kymograph and a Meyer-
Schneider pitchmeter. Otto Jespersen, however, was not interested
in experimental phonetics (although he worked for a time with
Rousselot), and this was probably the main reason why the University
did not obtain a phonetics laboratory. It is true that in 1933
we got a small "room for phonetic exercises", later called the
"phonetics laboratory", but it was equipped only with an old
gramophone and a modest collection of gramophone records.

In 1956 an Institute of Linguistics and Phonetics was
established, directed by professor Louis Hjelmslev, but still
there was no room for instrumental phonetics, and the Kay Elec-
tric Sonagraph which we had received in 1953 from the Rocke-
feller Foundation had to be stationed in the Institute for Speech
Pathology. That institute had had a laboratory of experimental
phonetics since 1943, and some research could be undertaken
there as well as in various medical clinics, but mostly at night
when the laboratories were not in normal use. In 1961 we
acquired a kind of laboratory of our own: a minor set of instru-
ments were placed in a room connected with the Institute, which
was then housed in St. Kannikestræde 13. The room could not be
heated, however, and it needed repair very badly. A sheepskin
coat had to serve as "sound treated room".

It was not until 1966 that six small rooms in the basement
(which, for a change, were always overheated) could be taken in
use as a laboratory of experimental phonetics. In the same year
the Institute of Linguistics and Phonetics was divided into two
separate institutes, and in 1968 we were granted a full-time
secretary.

In 1971 the laboratory was moved to much ampler premises
in Skindergade 3, which, however, had the drawback that they were
not in immediate vicinity of the rest of the institute. At the
same time the technical staff expanded. Since 1972 it has con-
sisted of two engineers and one technician.

Finally, in the summer of 1975, the whole institute moved
to the new University buildings in Njalsgade on Amager, thus
being situated for the first time in premises which had been
built specifically for our purpose and which we could organize
and equip to fit our needs. We expect to stay here for many
years to come.

During the whole period from 1966 till now the instrumental
equipment of the laboratory has been gradually increased, partly
through University grants, partly through subventions from the

Research Council of the Humanities and the Technical Research Council, so that now we are rather content with our research possibilities. Some of the instruments, like the synthesizer and the Segmentator, have been built in our laboratory, others have been acquired commercially. The most important acquisitions have consisted in extensions of our EDB-system.

With our present staff and equipment we are able to take care of the instruction of students as well as research in a tolerably satisfactory way, and we have had quite a number of guest research workers both from other Danish institutions and from abroad. We have a particularly close cooperation with the Research Laboratory of Teletechnics and with the Institute of Phonetics in Lund (Sweden).

At present we try to centre most of our research around the following subjects: The function of the larynx in speech, mainly by means of EMG and fiberoptics; perception of speech, mainly based on experiments with synthetic speech; and (quite recently) the teaching of pronunciation and the basic research which that entails. But there is also a good deal of activity within phonology, particularly the phonology of Danish and Greenlandic and within various other areas. A list of papers published in ARIPUC 1-9 is included in the present volume.

We are glad that we were able to profit from the expansion of the sixties. That era has now come to an end. We can only hope that we shall still be able to keep up most of our activities in the years to come.

PERSONNEL OF THE INSTITUTE OF PHONETICS IN 1975

Professor: Eli Fischer-Jørgensen (director of the institute)

Associate professors:

Børge Frøkjær-Jensen, cand.mag. (seconded to the Audiologo-
      pedics Research Group from September 1)

Jørgen Rischel, dr.phil. (on leave until September 1)

Oluf Thorsen, cand.mag.

Assistant professors:

Steffen Heger, cand.mag. (until September 1, temporarily
      appointed)

Peter Holtse, cand.phil.

Birgit Hutters, cand.mag. (from October 1)

Anders Löfqvist, fil.lic. (until September 1, temporarily
      appointed)

Nina Thorsen, cand.phil.

Senior research fellow: Hans Basbøll, cand.mag. (until
      September 1)

Teaching assistants:

Peter Molbæk Hansen, stud.mag.

Steffen Heger, cand.mag.

Birgit Hutters, stud.mag.

Mimi Jakobsen, stud.mag.

Ellen Pedersen, stud.mag.

Niels Reinholt Petersen, stud.mag.

Pia Riber Petersen, cand.mag.

Engineers:

Preben Dømler, B.Sc. (on leave from February 23)

Mogens Møller, M.Sc.

Jens Holger Stellinger, M.Sc. (from March 1, temporarily
      appointed)

Technician: Svend-Erik Lystlund

Secretary: Else Parkmann

Teachers from other institutes, lecturing at the Institute
of Phonetics:

Esther Dinsen, cand.mag. (Institute for Applied and Mathe-
      matical Linguistics)

Henning Spang-Hanssen, dr.phil. (Institute for Applied and
      Mathematical Linguistics).

PUBLICATIONS BY STAFF MEMBERS IN 1975

Hans Basbøll	"Det rigsdanske stød i generativ be-lysning", Selskab for nordisk filologi. Årsberetning for 1971-73, p. 17-22
-	"Grammatical boundaries in phonology", Papers from the Second Scandinavian Conference on Linguistics, Lysebu, April 19-20, (ed.: Even Hovdhaugen), p. 35-54
Eli Fischer-Jørgensen	Trends in phonological theory, (Akademisk Forlag) København, 474pp
Anders Löfqvist	"Intrinsic and extrinsic F_o variations in Swedish tonal accents", Phonetica 31, p. 228-247
-	"Some phonetic correlates of emphatic stress in Swedish", Working Papers, Phonetics Laboratory, Lund University 10, p. 105-117
-	"A study of subglottal pressure during the production of Swedish stops", JPh 3,3, p. 175-189

Anders Löfqvist and Peter Kitzing	"Subglottal and oral air pressures during phonation - preliminary investigation using a miniature transducer system", <u>Medical and Biological Engineering</u> 13, p. 644-648
Jørgen Rischel	"Problemer og perspektiver i Hjelmslevs udtrykslære", <u>Papir</u> 1,3, p. 86-101 "Problemer ved en generativ beskrivelse af dansk tryk", <u>Selskab for nordisk filologi. Årsberetning for 1971-73</u>, p. 22-32
Nina Thorsen and Robert MacAllister	"Problems encountered in the teaching of foreign language pronunciation", <u>Papers from the Institute of Linguistics, University of Stockholm</u> (Symposium maj 1975), p. 53-62
Nina Thorsen and Oluf Thorsen	<u>Lærebog i fonetik</u>, København, 161 pp
Oluf Thorsen, Ole Kongsdal, and Karen Landschultz	<u>Fransk fonetik</u>, 5th revised edition, København, 252 pp
Oluf Thorsen, Ole Kongsdal and Isabelle Durousseau	Review of Martinet-Walther: Dictionnaire de la prononciation française dans son usage réel, France-Expansion, Paris 1973, 932 pp, <u>Revue Romane</u> 10,2, p. 437-442.

LECTURES AND COURSES IN 1975

1. Elementary courses in general phonetics

One-semester courses (two hours a week) in elementary general phonetics (intended for all students of foreign languages, except French) were given by Eli Fischer-Jørgensen, Peter Mol-

bæk Hansen, Steffen Heger, Peter Holtse, Birgit Hutters, Mimi
Jakobsen, Ellen Pedersen, Niels Reinholt Petersen, Pia Riber
Petersen, and Nina Thorsen. There was one class in the spring
semester, and 19 parallel classes in the autumn semester.

Courses in general and French phonetics, including prac-
tical exercises in the language laboratory (3 hours a week)
were given through 1975 by Oluf Thorsen.

2. Practical exercises in sound perception and transcription

Steffen Heger gave a course for beginners (two hours a
week) through 1975.

Oluf Thorsen gave a course for more advanced students
(two hours a week) in the autumn semester.

(These courses form a cycle of three semesters, and are
based on tape recordings, as well as work with informants.)

3. Phonology

Hans Basbøll gave a course in phonology for advanced
students (two hours a week) in the spring semester.

Eli Fischer-Jørgensen gave a course in phonology for
beginners (two hours a week) on the topic: Problems in pho-
nology, in the spring semester.

Jørgen Rischel gave a course in phonology (two hours a
week) on the topic: Trends in phonological theory, in the
autumn semester.

4. The physiology of speech

Anders Löfqvist gave a course in instrumental physiologic-
al phonetics (two hours a week, plus individual exercises).
Anders Löfqvist also gave a course in the physiology of speech
for advanced students and for teachers (two hours a week).
Both in the spring semester.

Birgit Hutters gave a course in the physiology of speech
(two hours a week) in the autumn semester.

5. The acoustics of speech

Mogens Møller and Nina Thorsen gave a course in the acoustics of speech and in elementary mathematics and electronics (four hours a week) in the spring semester.

Peter Holtse and Nina Thorsen gave a course in instrumental acoustical phonetics (four hours a week, plus individual exercises) in the autumn semester.

Nina Thorsen gave a course in the acoustics of speech (two hours a week) in the autumn semester.

6. Other courses

Esther Dinsen (Institute for Applied and Mathematical Linguistics) gave a course in the theory and practice of the language laboratory (two hours a week) in the autumn semester.

Eli Fischer-Jørgensen gave a course in German phonetics (two hours a week) in the autumn semester and presided at seminars on instrumental phonetics through 1975.

Peter Molbæk Hansen gave a course in Russian phonetics (two hours a week) in the spring semester.

Peter Holtse gave a course in English phonetics (two hours a week) in the spring semester.

Jørgen Rischel gave a course in Danish phonetics (with emphasis on the phonological aspect) (two hours a week) in the autumn semester.

Henning Spang-Hanssen (Institute for Applied and Mathematical Linguistics) gave a course in elementary statistics (two hours a weeks) in the autumn semester.

Oluf Thorsen gave a course in French phonetics (two hours a week) in the spring semester.

7. Seminars

Students and staff gave an account of their research projects.

Børge Frøkjær-Jensen presented two films by Minuru Hirano.

Nina Thorsen and Oluf Thorsen presented parts of the manuscript for their text book on general phonetics.

Vivi Jeel presented her study: "An investigation of the fundamental frequency of vowels after various consonants, in particular stop consonants".

Eli Fischer-Jørgensen gave a short account of F_o-differences after voiced and unvoiced stops in Hindi and presided at a discussion of possible explanations for these differences.

The participants in the Eighth International Congress of Phonetic Sciences in Leeds, August 17-23, gave an account of their impressions from the congress.

Eli Fischer-Jørgensen gave an account of the symposium on A Dynamic Aspect of Speech Perception in Eindhoven, August 4-6, and of the Phonetics Symposium in Colchester, August 25-27.

Jørgen Rischel gave an account of the symposium on The Physiology of Speech in Stockholm, August 25-26.

Nina Thorsen gave an account of the Fourth World Congress for Applied Linguistics in Stuttgart, August 25-30.

8. Participation in congresses, symposiums, and meetings

The following staff members participated in the Eighth International Congress of Phonetic Sciences, Leeds, August 17-23, 1975: Eli Fischer-Jørgensen, Børge Frøkjær-Jensen, Peter Holtse, Birgit Hutters, Anders Löfqvist, Nina Thorsen, and Oluf Thorsen.

Hans Basbøll gave a paper, "Danish diphthongs", at the Institute of Linguistics, University of Aarhus, in March, and a paper, "Grammatical boundaries in phonology", at the 2nd Meeting of Nordic Linguists in Lysebu in April and at the Institute of Linguistics, University of Uppsala in May.

Eli Fischer-Jørgensen participated in a symposium on A Dynamic Aspect of Speech Perception in Eindhoven, August 4-6 and in a Phonetics Symposium in Colchester, August 25-27. She gave a paper, "Perspectives in phonology", at the phonetics

congress in Leeds.

Børge Frøkjær-Jensen was on an extended tour to many
phonetics laboratories and speech research institutions in the
United States in June. He gave a paper, "Survey and demonstra-
tion of some instrumental possibilities in phoniatrics and pho-
netics", at the Centro Ricerche e Studi Amplifon at Milan in
October. He also gave a paper, "Registration of voice quality"
(in collaboration with Svend Prytz) at the phonetics congress
in Leeds.

Anders Löfqvist gave a paper, "On the control of aspira-
tion in Swedish", at the phonetics congress in Leeds.

Jørgen Rischel participated in a symposium on The Physi-
ology of Speech in Stockholm, August 25-26. He presented an
introduction to and presided at a seminar on Productivity in
Phonology in the Linguistic Circle in Copenhagen in May.

Nina Thorsen gave a paper, "Problems encountered in the
teaching of foreign language pronunciation" (in collaboration
with Robert MacAllister) in a symposium on Phonetics and the
Teaching of Pronunciation in Stockholm, May 15-16 and at the
Fourth World Congress for Applied Linguistics in Stuttgart,
August 25-30.

Eli Fischer-Jørgensen, Peter Holtse, Nina Thorsen, and
Oluf Thorsen and several students participated in the phonet-
ics symposium at the Institute of Linguistics, University of
Uppsala, November 27-28.

INSTRUMENTAL EQUIPMENT OF THE LABORATORY
BY THE END OF 1975

1. Instrumentation for speech analysis

2 Sona-Graphs, Kay-Electric, type 6061 A

2 amplitude display/scale magnifier units,
Kay-Electric, type 6076 A

1 contour display unit, Kay-Electric, type 6070 A

1 fundamental frequency extractor ("Trans Pitchmeter")

1 intensity meter (dual channel, with active variable
highpass and lowpass filters)

1 electro aerometer (dual channel)

2 air-pressure manometers, Simonsen & Weel, type HB 66
(modified)

1 photo-electric glottograph

1 Fabre Glottograph

1 palatoscope with complete outfit for palatography

1 segmentator, model IPO (Eindhoven)

1 Meyer-Schneider pitchmeter

1 stroboscope, Philips, type PR 9103

1 segmentator, type PT

1 electro-aerometer, four channel, type AM 508/4

1 audio frequency filter, type 445

1 vocal cords fiberscope, Olympus, type VF

2. Instrumentation for speech synthesis

1 formant-coded speech synthesizer

1 provisional vowel synthesizer

1 voice-source generator

1 larynx vibrator with power supply

3. Filters

1 LC highpass filter (with stepwise variation of cut-off frequency)

1 active RC lowpass filter

4. Instrumentation for visual recordings

1 mingograph, Elema 42 (4 channels)

1 mingograph, Elema 800 (8 channels)

1 kymograph (with electro-motor)

1 automatic frequency response and spectrum recorder, Brüel & Kjær, type 3332

1 oscilloscope, Telequipment (single beam)

1 oscilloscope, Tektronix, type 502 A

1 oscilloscope, Solartron, type CD 1400 (dual beam)

2 oscilloscopes, Tektronix, type 564 storage

1 dual-trace amplifier, Tektronix, type 3A1

1 four-trace amplifier, Tektronix, type 3A74

1 dual-trace differential amplifier, Tektronix, type 3A3

1 time-base, Tektronix, type 3B3

1 time-base, Tektronix, type 2B67

1 oscilloscope, Tektronix, type 465

1 oscilloscope, Tektronix, type 5115

2 dual-trace amplifiers, Tektronix, type 5A18N

1 time-base, Tektronix, type 5B10N

1 mingograph, Siemens-Elema, type 803

5. Tape recorders

1 instrumentation recorder, Lyrec, type TR 86

1 professional recorder, Lyrec, type TR 42

2 professional recorders, Lyrec, type TR 20

3 semi-professional recorders, Revox, type G 36

7 semi-professional recorders, Revox, type A 77

1 professional recorder, Revox, type A 700

1 portable semi-professional recorder, Uher, type 4000

4 recorders, Tandberg, type 92 SL

1 recorder, Tandberg, type 7

6. Gramophones

2 gramophones, Delphon (mono, Ortofon pick-up)

7. Microphones

1 microphone, Neuman, type KM 56

1 dynamic microphone, Sennheiser, type MD 21

2 crystal microphones, of different brands

2 microphones, Altec

1 1" microphone, Brüel & Kjær, type 4131/32

1 1/4" microphone, Brüel & Kjær, type 4135/36

1 larynx microphone

1 1" microphone, Brüel & Kjær, type 4145

1 microphone — power supply, Brüel & Kjær, type 2807

1 microphone — power supply, Telefunken

8. Amplifiers

1 microphone amplifier, Brüel & Kjær, type 2603

1 laboratory amplifier (mono/stereo, with matching for different impedances)

1 power amplifier, Brüel & Kjær, type 2706

1 microphone pre-amplifier, Brüel & Kjær, type 2627

1 measuring amplifier, Brüel & Kjær, type 2607 A

9. Loudspeakers/headphones

7 different loudspeaker systems

5 headphones, AKG, type K 58

3 loudspeakers, Beovox, type 2600

10 headphones, Sennheiser, type 414

4 headphones, Sennheiser, type 424

10. General-purpose electronic instrumentation

1 oscillator, Hewlett & Packard, type CD 200

1 function generator, Wavetec VGC III (o.oo3 c/s - 1 Mc/S)

1 frequency counter, Rochar, type A 1360 CH (5 digits)

2 vacuum-tube voltmeters, Brüel & Kjær, type 2409

1 vacuum-tube voltmeter, Heathkit, type V-7A

1 vacuum-tube voltmeter, Radiometer, type RV 23b

1 DC millivoltmeter, Danameter, type 205

1 DC nanoammeter, Danameter, type 206

1 universal meter, Philips, type P 817

1 transistor tester, Taylor, model 44

1 Piston-phone, Brüel & Kjær, type 4220

1 component bridge C/L/R, Wayne Kerr, type B 522

1 AC automatic voltage stabilizer, Claude Lyons, type BTR-5F

4 resistance decades, Danbridge, type DR 4

1 condenser decade, Danbridge, type DK 4 AV

6 stabilized rectifiers

1 multi-generator, Exact, type 126 VCF

2 stabilized rectifiers, Danica, type TPS 1d

1 stabilized rectifier, Danica, type TPS 3c

1 impulse precision sound level meter, Brüel & Kjær, type 2204

1 attenuator set, Hewlett Packard, type 350 D

1 band-pass filter set, Brüel & Kjær, type 1615

1 digital multimeter, Philips, type PM 2422

1 timer counter, Advance, type SC3

1 oscillator, Advance, type J2E

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

1 noise generator, Brüel & Kjær, type 1405

Additional oscillators, rectifiers, etc., for special purposes.

11. Outfit for photography

    1 Minolta camera SR-1 (with various accessories)

    1 complete outfit for reproduction (including 1 Liesegang UNI-RAX with frame)

    1 Telford oscilloscope camera, type "A" (polaroid)

12. Equipment for EDP

  A   computer, Digital, PDP8/E, 8k

  A_1 arithmetic unit, Digital, type KE 8-E

  A_2 bootstrap loader, Digital, type MR 8-EC

  A_3 add-on memory system 24k, Fabri-Tek, type 8/E

  B   dectape, Digital, type TD8-EM

  B_1 tape reader, GNT, type 24

  B_2 tape punch, GNT, type 34

  B_3 decwriter, Digital, type LA 30P

  B_4 teletype, Teletype, type ASR 33

  B_5 display terminal, Tektronix, type 4014-1

  B_6 real time clock, Digital, type DK8-EP

  B_7 a/d converter, Digital, type AB8-EA

  B_8 d/a converter, type PD

13. Projectors

    1 Leitz projector for slides

    1 Voigtländer Perkeo Automat - J 150

    1 16 m/m tone film projector, Bell & Howell, "Filmsound 644"

    1 projector, Leitz, type Pradovit color 250

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 (Akademisk Forlag), København 1971
Haskins SR	Status Report on Speech Research, Haskins Laboratories
IJAL	International Journal of American Linguistics
IPO APR	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. ...	Proceedings of the ... International Congress on Acoustics
Proc. Ling. ...	Proceedings of the ... International Congress of Linguists
Proc. Phon. ...	Proceedings of the ... International Congress of Phonetic Sciences
STL-QPSR	Speech Transmission Laboratories, Quarterly Progress and Status Report, Royal Institute of Technology, Stockholm
SL	Studia Linguistica

SPE   The Sound Pattern of English, Chomsky and Halle, 1968

TCLC   Travaux du Cercle Linguistique de Copenhague

TCLP   Travaux du Cercle Linguistique de Prague

UCLA WPP Working Papers in Phonetics, University of
          California

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

INTRINSIC FUNDAMENTAL FREQUENCY OF DANISH VOWELS

Niels Reinholt Petersen

Abstract: In a number of languages the intrinsic F_o of vowels
has been reported to be positively correlated with
tongue height.
Measurements of the fundamental frequencies of Danish
vowels showed a similar correlation to exist in Danish.
The F_o differences between high and low vowels were
greatest in long vowels in stressed position, but also
short vowels in unstressed position differed with re-
spect to intrinsic F_o, although to a much lesser degree.
The hypotheses advanced to account for the intrinsic
F_o differences are briefly outlined, and discussed in
relation to the results of the present and other similar
experiments.

1.   Introduction

There seems to be a tendency for the fundamental frequency
(F_o) of vowels to be correlated with their quality. Low vowels
have - other things being equal - a lower F_o than high vowels.
This tendency has been reported e.g. for English by Black (1949),
Peterson and Barney (1952), House and Fairbanks (1953), Lehiste
and Peterson (1961), Peterson (1961), Mohr (1971), for Korean by
Kim (1968) (cited in Mohr 1971), for Serbo-croatian by Ivić and
Lehiste (1963) (cited in Lehiste 1970).

Various hypotheses have been advanced to explain the con-
nection between tongue height and intrinsic F_o of vowels. The
most recent ones will be briefly outlined below:

Mohr (1971) assumes a pressure of air to be built up behind
the supraglottal constriction, which reduces the airflow through

the glottis and consequently the rate of vocal fold vibration. According to Mohr's hypothesis it is so that the greater the distance between the constriction and the glottis, the longer it takes for the air pressure to be established behind the constriction and hence for the fundamental frequency to drop. If this explanation were true, a higher F_O should be expected in [ɪ] than in [u]. What is generally found, however, is that the F_O of [u] is equal to or - more often - higher than that of [ɪ]. Furthermore, vowels of medium tongue height in the front series should have a higher F_O than [ɪ], since their constriction is less narrow than that of [ɪ]. This prediction, however, is also contradicted by data.

Another hypothesis rests on the assumption that the glottis impedance is sufficiently low for some coupling to occur between the vocal tract and the glottis (see e.g. Flanagan and Landgraf 1968, Lieberman 1970). If this is so the first resonance of the vocal tract will cause a change, i.e. an increase, of the fundamental frequency. Since such an effect becomes greater the closer the resonance in question is to the F_O range, vowels with low first formants (e.g. [ɪ] and [u]) will have a higher fundamental frequency than low vowels (e.g. [a]), which have high first formants.

The question whether the glottis impedance is sufficiently low for this effect to be active seems still open for discussion. Contradictory to this hypothesis, however, the findings reported in Beil (1962) indicate that vowels spoken in a mixture of atmospheric air and helium, which raises the formant frequencies of all vowels, have essentially the same F_O relations as vowels spoken in atmospheric air alone, although the coupling hypothesis would predict smaller F_O differences in "helium vowels" than in "normal" vowels.

A third hypothesis which has been forwarded to explain the intrinsic F_O differences assumes that the tongue, when elevated for the production of high vowels, pulls the hyoid bone and the

larynx upwards (Ladefoged 1964, Lehiste 1970). This vertical pull
is thought to be translated into an increased vocal cord tension,
which in turn leads to a higher F_o.

This explanation is contradicted by the fact that the hyoid/
larynx position always seems to be lower in [u] (and sometimes in
[i] as well) than in [a], although vowels having high tongue po-
sitions might be expected to have high larynx positions, since
larynx height and F_o are normally positively correlated. Ohala
(1973) points out, however, that the negative correlation between
tongue height and hyoid/larynx height need not invalidate the
tongue pull hypothesis; it merely suggests that the variations in
vocal cord tension, which may be assumed to underlie the intrinsic
F_o differences are not likely to be due to changes of the larynx/
hyoid position. Ohala advances another explanation according to
which the increased tongue pull in high vowels gives rise to in-
creased vertical tension in the vocal folds. The vertical tension
can be established exclusively through the mucous membrane and
other soft tissues without involving the hyoid bone and the hard
tissues of the larynx. In support of this explanation Ohala
cites data on the size of Morgagni's ventricles (van den Berg
1955, and Shimizu 1961). It appears that there is a positive
correlation between ventricle size, which is assumed to reflect
vertical tension in the vocal cords, and tongue height and in-
trinsic F_o of vowels.

The tongue pull hypothesis has been expanded by Ewan (1975).
Ewan suggests that in order for a sufficiently low first formant
to be produced in the vowel [u], the supralaryngeal cavities are
expanded by means of an active downward pull of the larynx. This
seems to account for the low larynx position of [u] in spite of
its high intrinsic F_o. Furthermore, since it may stretch the
soft tissues of the larynx and thus add to the tension created by
tongue pull, the active lowering of the larynx in [u] might also
explain why the intrinsic F_o of that vowel is often found to be
higher than the intrinsic F_o of [i].

In addition to tongue pull - or, rather, lack of tongue pull - Ewan proposes a "tongue retraction/pharyngeal constriction component" to account for the low intrinsic F_o of low vowels. He builds upon data on Arabic, from which it appears that the F_o of a vowel following a pharyngeal consonant (where the retraction/constriction component is effective) is lower than the F_o of a vowel following a stop consonant without any such component. Ewan suggests that the low F_o of low vowels, which are also assumed to involve a tongue retraction or pharyngeal constriction component, is caused by the soft tissues being pressed downwards in the direction of the larynx and thus increasing the vibrating mass of the vocal folds, which results in a decrease in F_o. The pressure of soft tissues behind the hyoid bone may also push the hyoid bone slightly upward and forward and thus explain the relatively high hyoid/larynx position of low vowels.

Although it needs further substantiation, the tongue pull theory as modified by Ohala and Ewan seems to be that one among the current theories advanced to account for the intrinsic F_o differences among vowels which shows the best agreement with available data.

The main purpose of the experiments to be reported below was to investigate whether F_o differences between high and low vowels might also be found to exist in Danish, and to estimate the order of magnitude of these differences. Such information will be relevant to investigations of Danish intonation (cp. Thorsen, this issue), as it may help to separate the F_o variation attributable to the intonation proper from variation conditioned entirely by segmental factors - in this case tongue height.

Thus, although it was not the primary purpose of the present investigation to obtain data to evaluate the hypotheses outlined above, it will, of course, be examined whether or not the results are in agreement with the predictions of the hypotheses.

2. Experimental procedure

2.1 Material

The material consisted of the Danish long vowels [i:, e̝:, ɛ̝:, æ:, ɑ:, y:, ø:, œ̝:, u:, o̞:, ɔ̟:] (henceforth referred to as [i, e, ɛ, a, ɑ, y, ø, œ, u, o, ɔ]) occurring in nonsense words of the type $p^hV'p^hV{:}p^hV$, the vowel being identical in each of the three syllables. The intention was primarily to carry out measurements of the fundamental frequency of the long, stressed vowels, but the type of words chosen made it possible to obtain measurements from short vowels in pretonic and posttonic position as well. The material also included words of the same type as above, but having [ɓ], [f], and [m] instead of [p^h] (combined with the vowels [i, ɑ, u] only), so that the relation between F_o in these vowels and the manner of articulation of the neighbouring consonants might be examined. Thus the material comprised 20 different testwords, i.e. 11 "p-words", 3 "b-words", 3 "f-words", and 3 "m-words". The words were embedded in the carrier sentence "Stavelserne i ..... forkortes." ('The syllables of ...... are shortened.').

In order to prevent a tendency towards rhythmicization during the reading of the list, 20 "dummy-sentences" were incorporated in the material. The forty sentences were arranged in five different randomizations in the reading list, the limitations to the random order being, however, that no more than three test sentences or dummy sentences should occur in immediate succession, and that each page of the list should start and end with a dummy sentence.

2.2 Recordings and speakers

The recordings took place in an anechoic chamber by means of professional recording equipment. Two female subjects (BH and JG) and three male subjects (JB, SH, and NRP) acted as speakers.
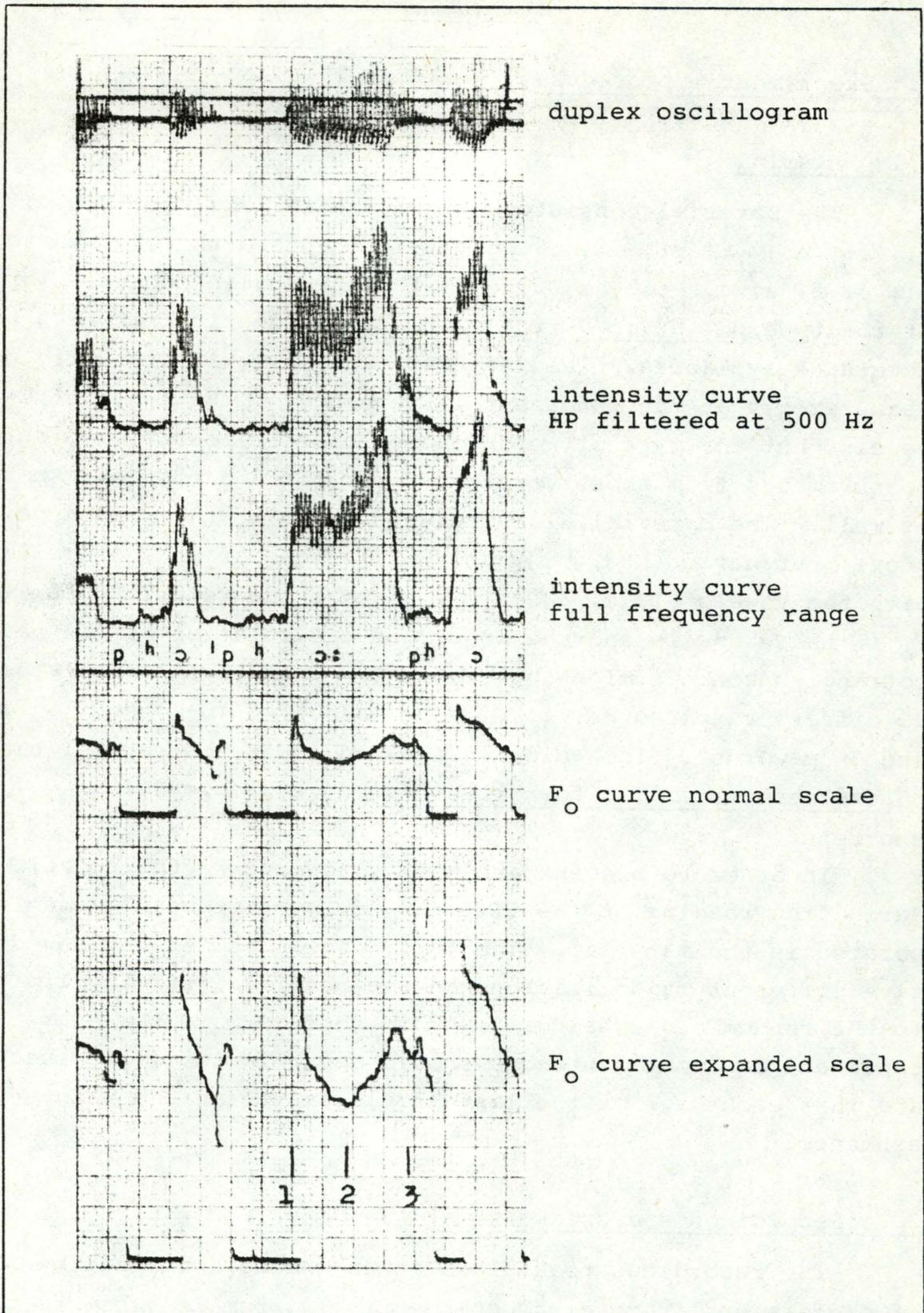
Figure 1

Example of curves of the testword [pʰVˈpʰV:pʰV]. The figures
under the expanded scale Fₒ curve refer to start of vowel,
point of measurement, and end of vowel, respectively.

They were all phoneticians, and all speakers of Advanced Standard Copenhagen Danish (see Basbøll 1968). Each subject read the list twice, at intervals between a few hours and a couple of days.

2.3 Registrations

The apparatus used for registration of the recorded material was a semiprofessional tape recorder, an intensity-meter, a pitch-meter, and a mingograph. The following acoustic curves were obtained:

1) duplex oscillogram
2) intensity curve, full frequency range
3) intensity curve, HP-filtered at 500 Hz.
4) F_0 curve, normal scale
5) F_0 curve, expanded scale (some 2 Hz per mm)

The oscillogram, the intensity curves and the ordinary F_0 curve were used for segmentation, whereas F_0-measurements were made on the expanded F_0 curve. The curves of one testword are shown in fig. 1.

2.4 Measurements

A preliminary examination of the F_0 curves was made in order to establish a well defined point of measurement. It was found that the long, stressed vowels in the p-, b- and f-words in almost all cases had a falling-rising F_0 movement with a minimum in the middle third of the vowel (see fig. 2a). This minimum was chosen as the point of measurement, not only because it was clearly defined but also because it occurred in a portion of the vowel where the influence from surrounding consonants may be considered to be minimal. In a few cases, however, the F_0 movement was either monotonically rising (see fig. 2b), or (very rarely) falling during the greater part of the vowel. Under these circumstances the fundamental frequency was measured at a point corresponding
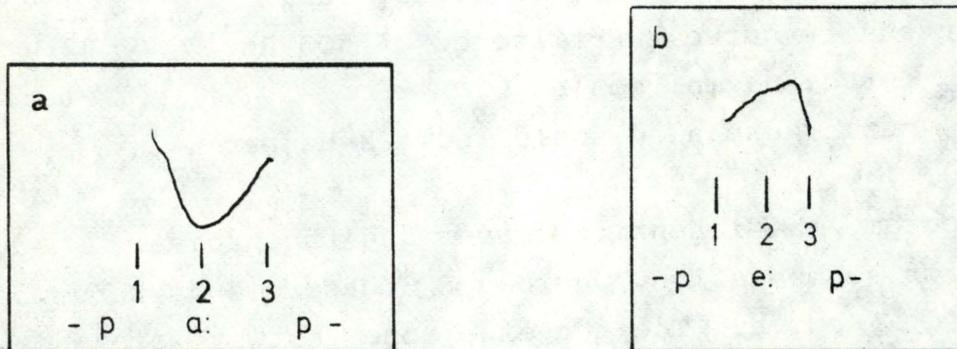
Figure 2

Examples of F_0 movements in long stressed vowels.
a) movement with a clear minimum point, b) mono-
tonically rising movement. The figures 1, 2,
and 3 refer to point of vowel start, point of
measurement, and point of vowel end, respectively.

approximately in time to the point chosen in cases where a clear minimum could be seen.

In the m-words containing the narrow vowels [i] and [u], it proved difficult, and in numerous cases impossible, to carry out a reliable segmentation, and consequently to determine a point of measurement with reasonable certainty. Therefore, it was decided to exclude the m-words from the material to be measured. Nor was the F_o of pretonic and posttonic vowels submitted to a quantitative treatment, since these vowels showed F_o movements which did not make it possible to determine a reliable point of measurement. Moreover, these F_o movements displayed considerable variation in their position on the frequency scale (see figs. 9 and 10 below).

2.5 Statistical treatment

In the statistical computations the subjects were treated individually, and for each subject the p-, b-, and f-words were kept apart.

A t-test for correlated data was applied in order to determine whether the two readings for each speaker might be pooled. If $p > .05$ the recordings were pooled, otherwise they were treated separately.

By means of a one-way analysis of variance it was examined whether the obtained F_o data for the different vowels might have been drawn from same or different populations. Since, however, the vowels differed not only in tongue height but also in place of articulation and rounding, the analysis of variance had to be supplemented by a multiple comparison procedure (Scheffé's method) in which the mean F_o of every vowel was compared with the mean F_o of every other vowel.

3. Results

3.1 Comparison of the first and second recording

With two of the speakers (BH and JG) the t-test revealed significant differences between the two recordings, the second recording being lower than the first in all cases. The differences were for BH 11.2 Hz in the p-words, 11.6 Hz in the b-words, and 10.5 Hz in the f-words, and for JG 11.6 Hz, 5.6 Hz, and 14.1 Hz, respectively.

The second recording of speaker NRP was slightly higher than the first one for the p-words. Although the difference was only 1.5 Hz it was statistically significant ($p < .01$). For the b- and f-words no significant difference between recordings could be shown with that subject.

The two recordings of the speakers SH and JB were not significantly different ($p > .05$). This applies to all three types of words.

3.2 Fundamental frequency versus vowel quality

3.2.1 p-words

The mean fundamental frequencies for each of the 11 vowels in the p-words are given in table 1 and are shown graphically in fig. 3. The computed F-values (indicating the ratio of the between group variance to the within group variance, s_b^2/s_w^2) shown in the bottom row of table 1 were in all cases statistically significant ($p < .01$).

It becomes evident from the multiple comparison results, as summarized in fig. 4, that the F_o differences between the vowels are due to tongue height rather than to place of articulation or rounding. Here the number of significant differences ($p < .05$) over all subjects are given for every vowel compared with every other vowel. It is obvious that the number of significant differences increases with the difference in tongue height between

Table 1

Average F_0 in Hz for each of the 11 vowels in the p-words. Recordings are pooled or kept apart according to the criteria stated in section 2.5. Unless otherwise stated each mean represents 5 observations. In cases of fewer observations, their number is given in brackets after the mean in question. In the bottom line the computed F values are indicated.

Subject

|	BH		JG		JB	SH	NRP	
	recording 1	recording 2	recording 1	recording 2	recording 1+2	recording 1+2	recording 1	recording 2
i	219.4	212.2	210.8	201.0	141.8	125.7	93.4	95.0
e	211.4	199.6	203.4	189.8	130.7(9)	118.8	91.2	91.6
ε	197.0	188.2	195.4	184.0	128.9	113.3	87.6	88.4
a	198.6	185.6	195.2	181.8	125.2	109.9	85.0	87.8
ɑ	188.2	174.8	185.8	176.6	118.3	110.9	83.2	87.2
y	220.0	212.2	219.2	207.8	143.1	123.5	95.8	96.2
ø	215.4	203.6	206.0	194.4	133.6	118.3	89.8	93.2
œ	199.8	188.0	195.2	185.8	132.7	114.2	88.0	87.2
u	225.2	213.0	230.2	213.8	144.5(9)	126.5	95.0	96.6
o	220.0	208.8	307.8	196.4	138.4(9)	121.1	91.8	93.2
ɔ	204.8	190.8	195.6	186.0	133.3	115.7	89.4	90.6
F	11.918	17.916	18.785	26.077	19.350	37.495	17.944	7.296
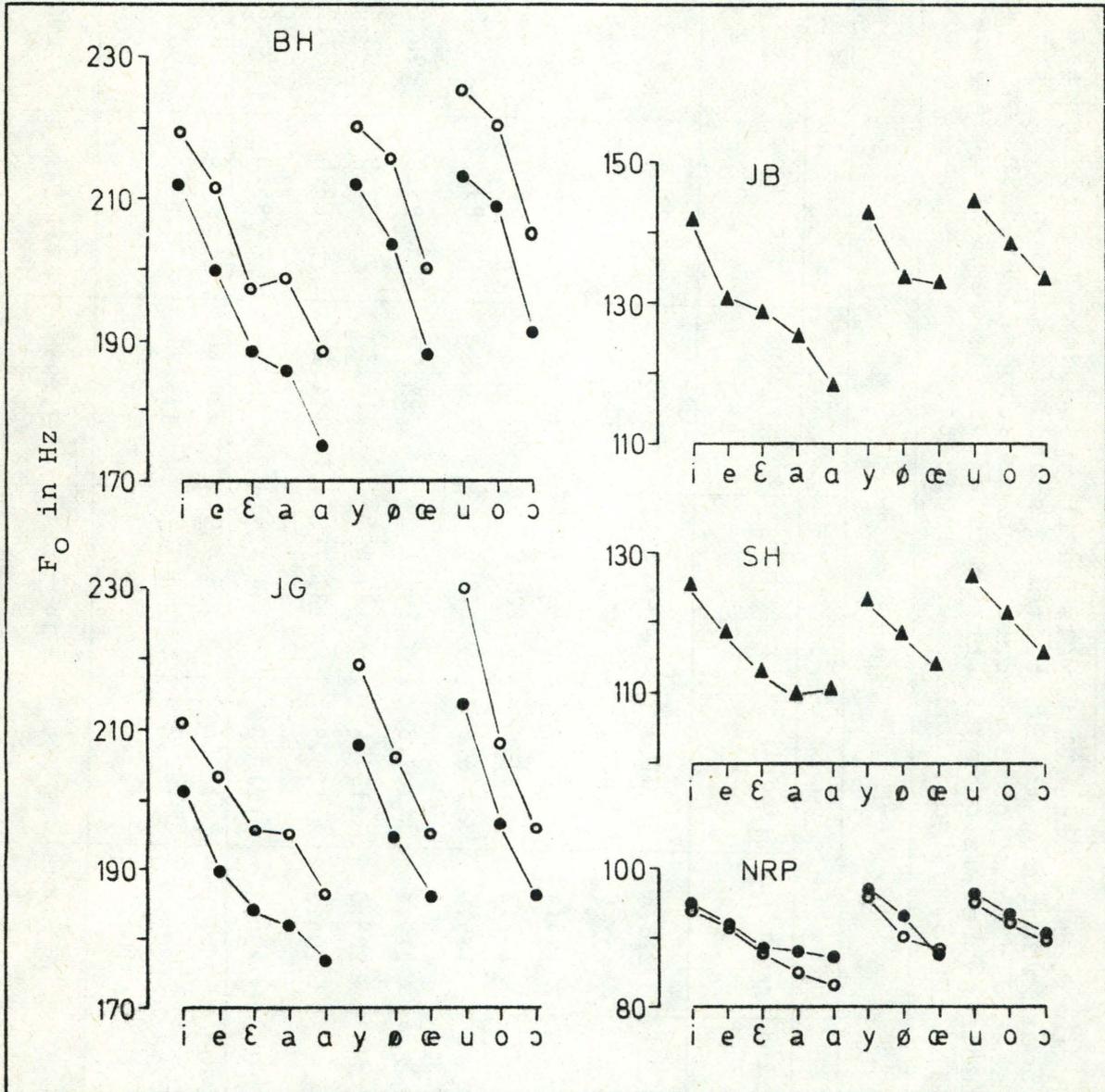
Figure 3

Average F$_0$ in Hz for each of the 11 vowels in the p-words. Open circles refer to the first recording, filled circles to the second recording. Triangles represent the averages over both recordings in the cases where they were pooled.
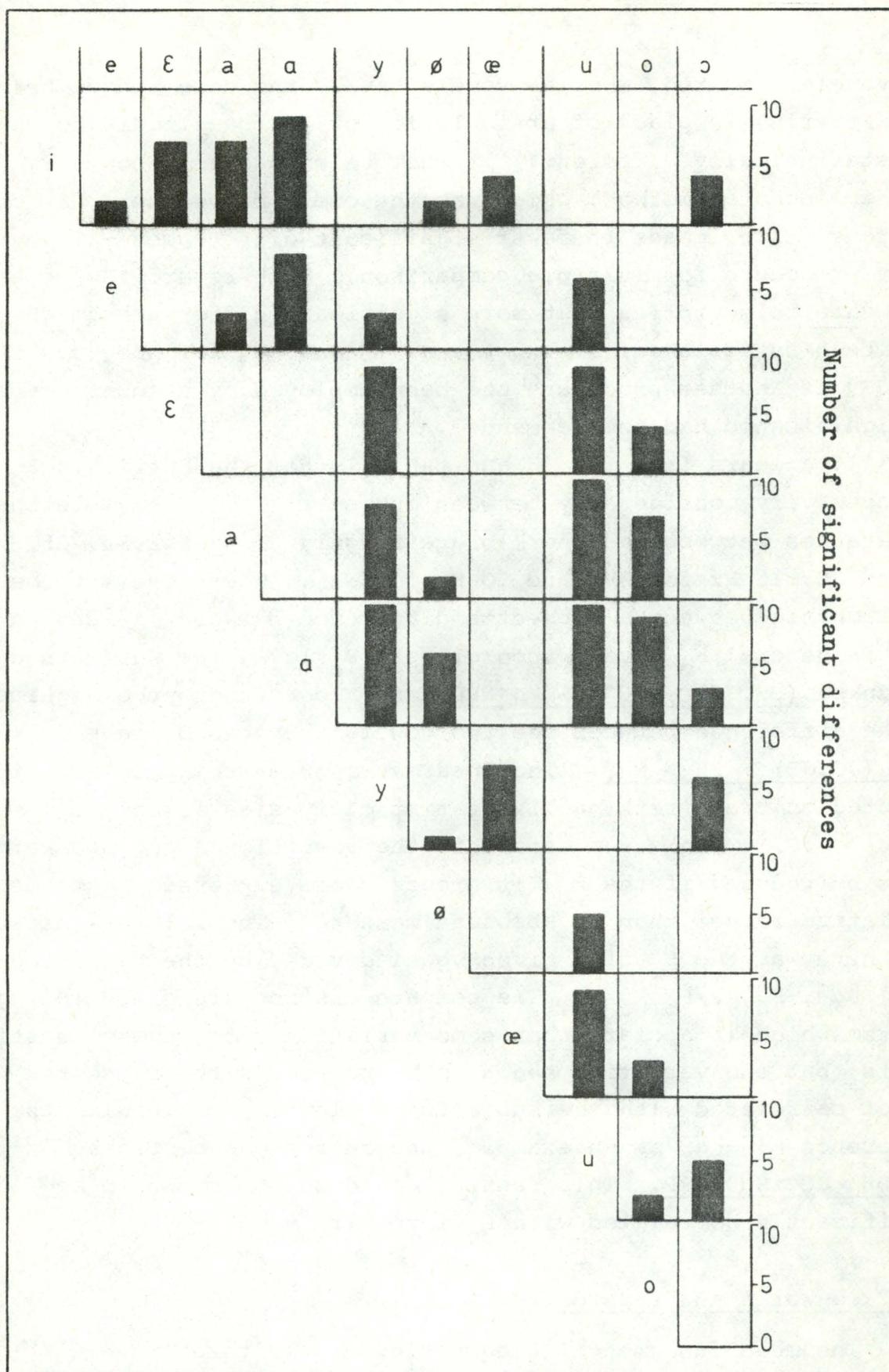
Figure 4

The number of significant differences between F-means as found by multiple comparison of all vowels in the p-words. In the cases where the first and second recordings were pooled one significant difference was given value 2, the number of significant differences for a comparison of two vowels thus being 10.

the vowels compared, whereas vowels having the same tongue height but differing in place of articulation or in rounding in no cases are statistically different. It must be emphasized, however, that the Scheffé method, which was the one employed here, is more rigorous - i.e. leads to fewer significant differences - than any other procedure for multiple comparison. With regard to the present data this implies that more significant differences might have been established, e.g. between [i] and [u] for subjects BH and JG, if another procedure had been employed or a lower level of significance had been demanded.

It appears from fig. 3 and table 1 that the intrinsic F_O differences vary considerably between subjects. As an example the differences between [ɑ] and [i] are roughly 35 Hz for BH, 25 Hz for JG and JB, 15 Hz for SH, and 10 Hz for NRP. There seems to be a relationship between a subject's difference between [i] and [ɑ] and his general F_O level: according to F_O level the subjects can be ranked (BH,JG) > JB > SH > NRP, and according to the magnitude of the difference between the two vowels they can be ranked BH > (JG,JB) > SH > NRP. The Spearman rank correlation coefficient computed on these rankings is statistically significant (r_s = .95, p < .05). This suggests that the variation among speakers might be reduced if the F_O differences were expressed in terms of a relative rather than an absolute measure. The relative measure used here was the F_O of a given vowel divided by the F_O of the vowel [ɑ], $F_{O(V)}/F_{O([ɑ])}$. The computed ratios are given in fig. 5, from which it appears that some variation among subjects still exists, but the variation seems to be random in the sense that it is not correlated with the subjects' F_O level. Again with the difference [i]-[ɑ] as an example, the ratios can be ranked JB > BH > JG > SH > NRP. This ranking could not be shown to be significantly correlated with F_O level (r_s = .67, p > .05).

3.2.2 b-words and f-words

The mean fundamental frequencies of the three vowels ([i, ɑ, u]) of the b- and f-words are given in table 2 and fig. 6. The
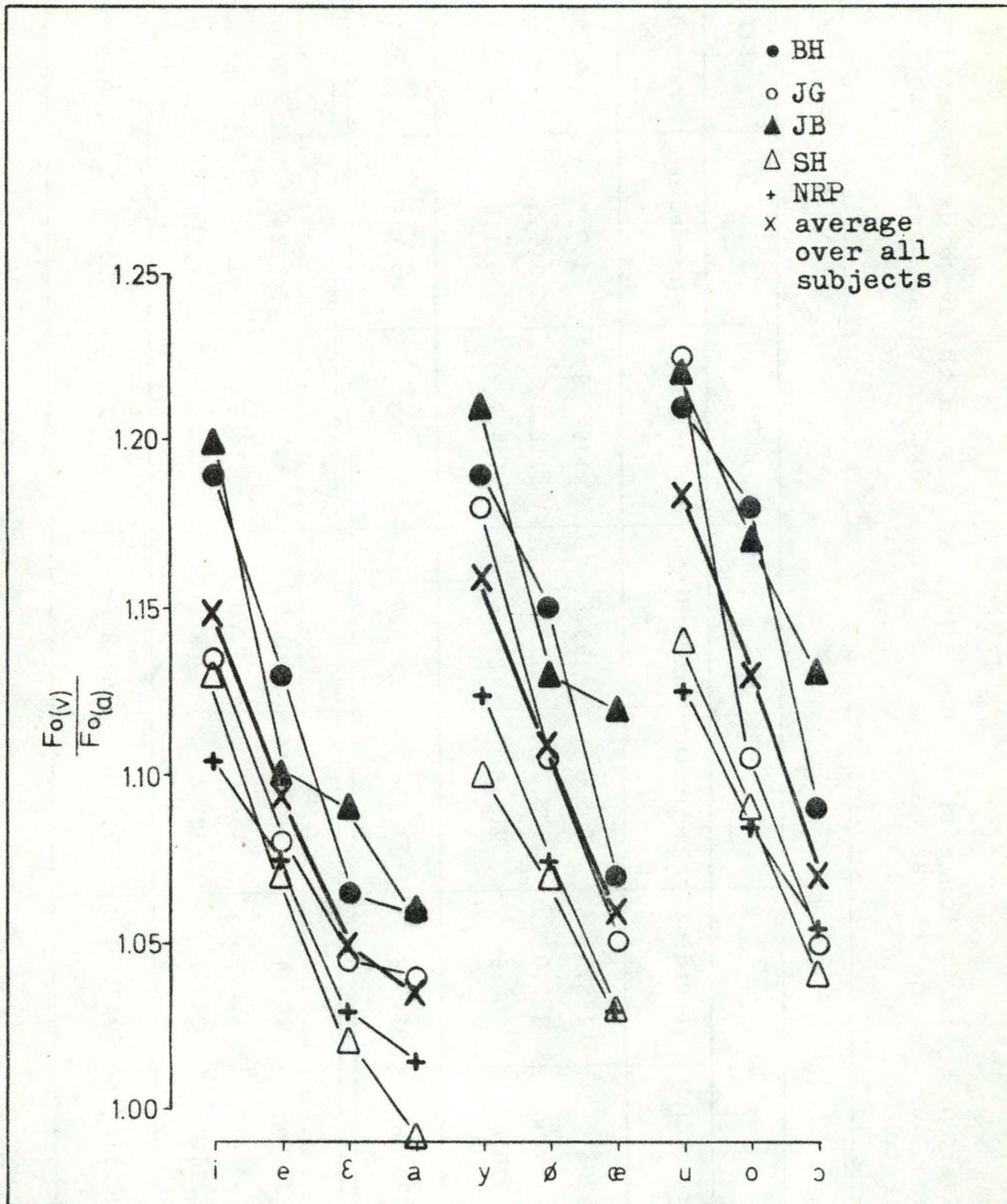
Figure 5

The relative difference between F_o of the vowel $[a]$ and the remaining 10 vowels of the p-words. The first and second recordings are pooled in all cases.

Table 2

Average F_o in Hz for each of the three vowels in the b-words (upper half) and the f-words (lower half). For further explanations, see the legend of table 1.

|	BH		JG		JB	SH	NRP
	recording 1	recording 2	recording 1	recording 2	recording 1+2	recording 1+2	recording 1+2
b-words							
i	208.8	198.8	205.0	199.8	139.0(9)	119.9	91.5
ɑ	184.6	172.0	178.4	172.3	118.1	109.5	83.0
u	220.8	208.6	216.8	210.6	139.7(9)	121.5	92.1
F	177.708	33.549	173.236	84.514	22.060	32.573	55.094
f-words							
i	219.6	209.4	215.0	201.8	142.3(9)	122.7	94.0
ɑ	187.8	176.4	181.8	170.0	119.2	108.3	82.5
u	226.2	216.2	225.4	208.2	144.4	124.0	95.1
F	236.831	107.397	65.640	167.877	124.951	69.208	120.200

Subject
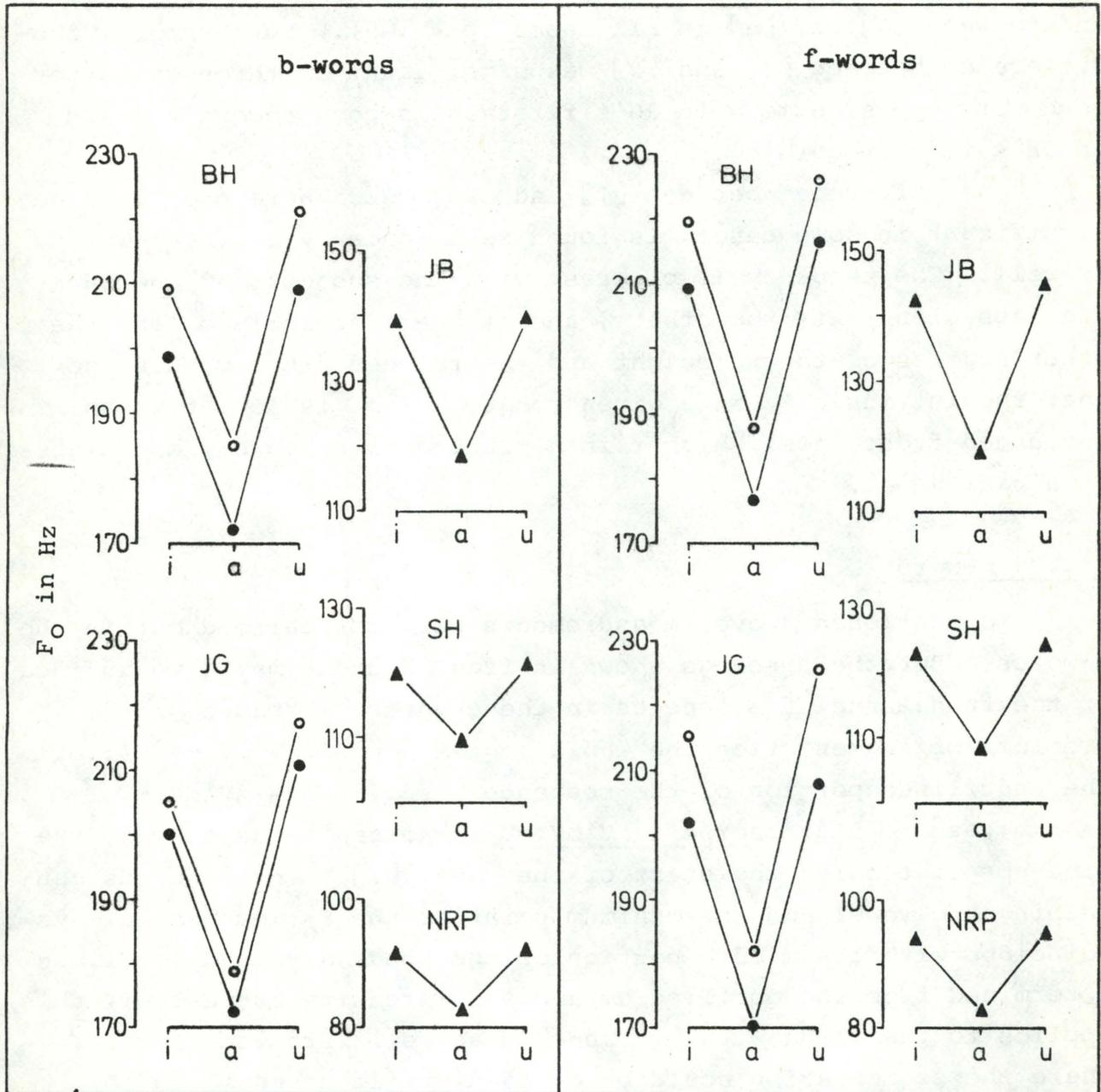
Figure 6

Average F_O in Hz for each of the three vowels in the b-words (left) and f-words (right). For further explanations, see the legend of fig. 3.

computed F-values (bottom row in the table) were all statistically significant ($p < .01$). For both word types the multiple comparison procedure showed significant differences between [i] and [ɑ] and between [u] and [ɑ] in all cases ($p < .05$). Furthermore, the difference between [i] and [u] was significant in three cases for both word types, namely in JG's first and second recordings and in BH's first recording.

The difference between [i] and [u], which here proved to be significant in some cases, is found as a tendency in the p-words as well. The tendency is clearest with the subjects JG and BH and less clear with the others, and it seems to include also the other degrees of tongue height and the rounded front vowels, so that the intrinsic F_o at a given tongue height is lowest in an unrounded front vowel, higher in a rounded front vowel, and highest in a back vowel.

3.2.3 m-words

As mentioned above, measurements were not carried out in the m-words. But the tracings shown in figs. 7 and 8 may give an idea of the fundamental frequencies in these words. Because of the problems of segmentation the whole sequence [ʌnəimV'mV:mV] (i.e. the underlined portion of the sentence ['sɡæːuls ʌnəimV'mV:mV fʌ'kʰɒːɡəs] - "Stavelserne i mV'mV:mV forkortes") was traced, the line-up point being the start of the vowel [ʌ]. In the words containing the vowel [ɑ] the minimum point of the F_o movement always coincided with the middle portion of the long, stressed vowel, as determined from the oscillogram and the intensity curves. If this applies to the words with [i] and [u] as well (from the few cases where the segmentation could be guessed at, it seems to do so), the F_o movement in the m-words appear to be in good agreement with the movement in the word types where reliable measurements could be carried out. But, of course, a more thorough examination is needed before any safe conclusions can be drawn.
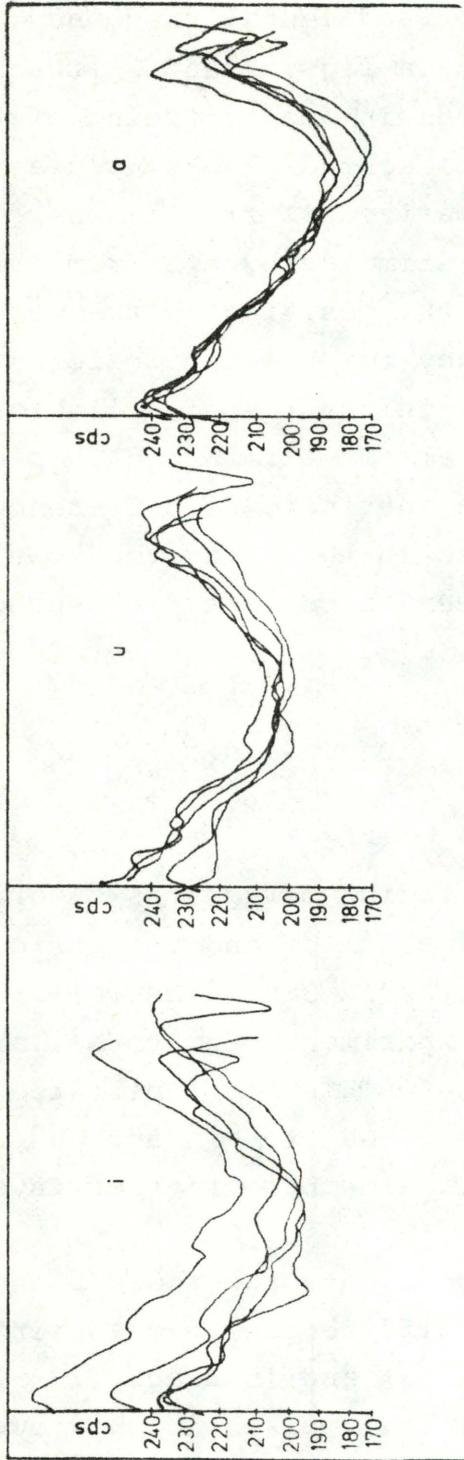
19

Figure 7

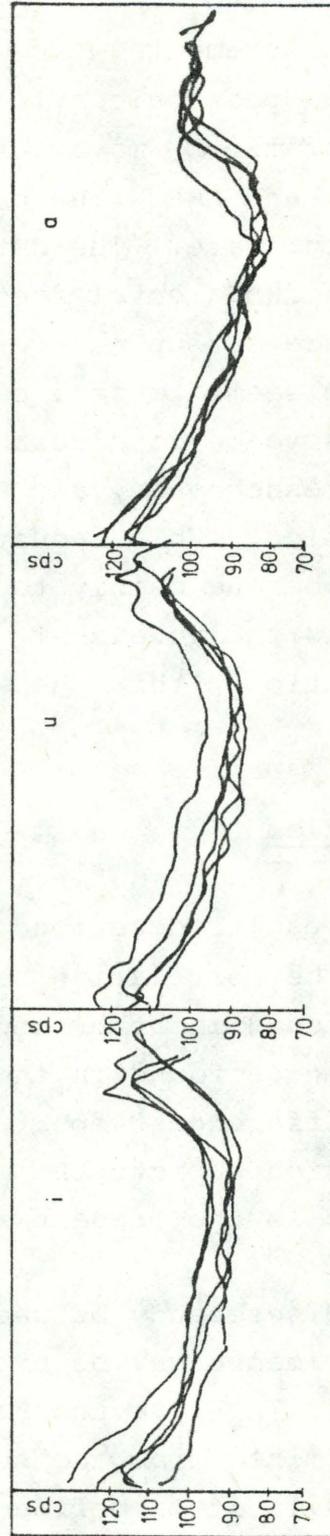F$_o$ tracings of m-words, BH's first recording.

Figure 8

F$_o$ tracings of m-words, NRP's second recording.

3.2.4 Vowels in pretonic and posttonic position

No measurements were made of the fundamental frequency in pretonic and posttonic syllables. But in figs. 9 and 10 tracings are given of the F_o movements in f-words in one recording of the subjects JG and NRP. The beginning of each vowel was used as line-up point here. The data are summarized in fig. 11 where the average fundamental frequencies (estimated by eye from the tracings) are set up as a function of the position in the word.

There seems to be a clear tendency for the intrinsic F_o differences to be considerably smaller in unstressed syllables than in stressed ones, and even to be almost eliminated in pretonic position. This reduction of the intrinsic F_o differences appears to be due mainly to an increase in the F_o of the vowel [ɑ], whereas the vowels [i] and [u] seem to be less influenced by the position within the word.

4. Discussion

The results described in the preceding paragraphs clearly show that the correlation between tongue height and intrinsic F_o of vowels found in other languages exists in Danish as well. The differences found in the present experiment seem to be larger than the differences reported elsewhere. This is illustrated in fig. 12, which depicts the mean ratios of [ɑ] to [i] and [u] as established in the present study and in a number of other investigations.

The discrepancy between the present data and those found in other experiments may be explained by differences in measuring procedures. In House and Fairbanks (1953) and in Black (1949) F_o is determined from the number of glottal pulses in the vowel. In Mohr (1971) it is defined as the mean of the F_o values measured at the beginning, at the minimum point (lowest F_o), and at the end of the vowel.
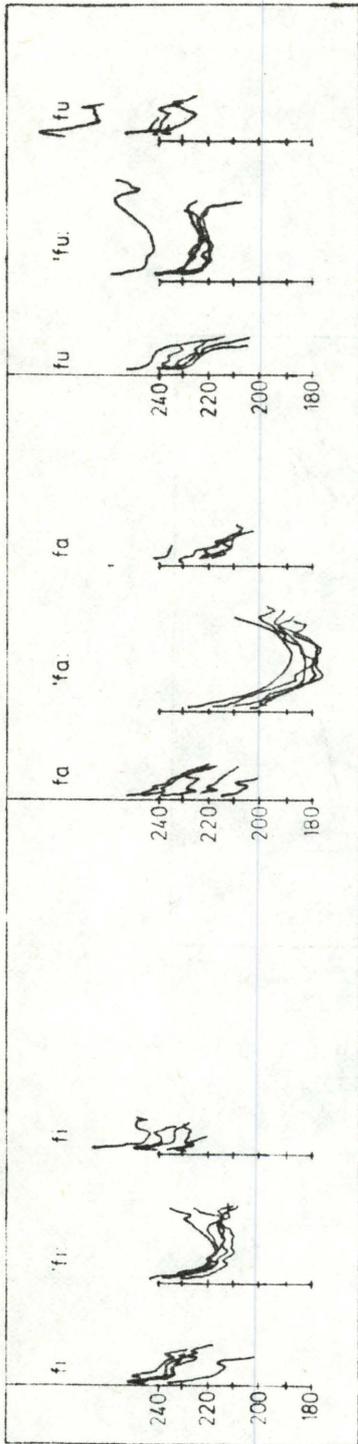
Figure 9

Tracings of F in pretonic, tonic, and post-
tonic syllables in f-words. JG's first re-
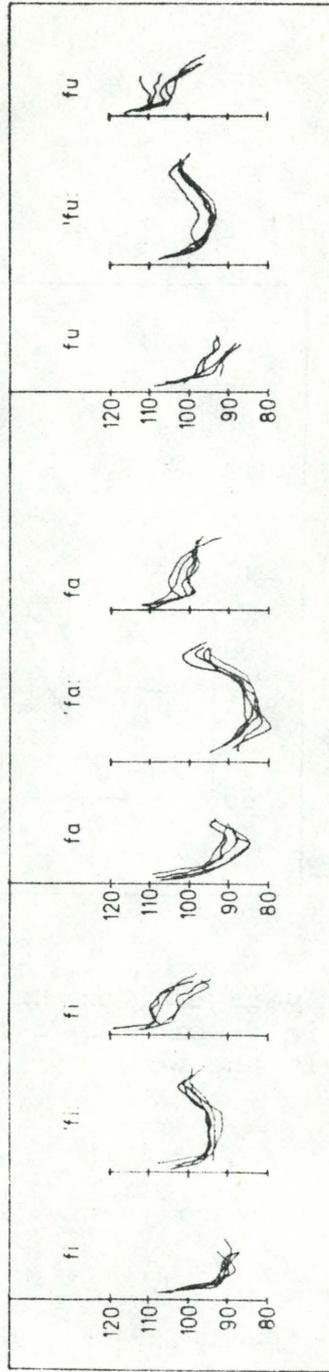cording.

Figure 10

Tracings of F in pretonic, tonic, and post-
tonic syllables in f-words. NRP's first re-
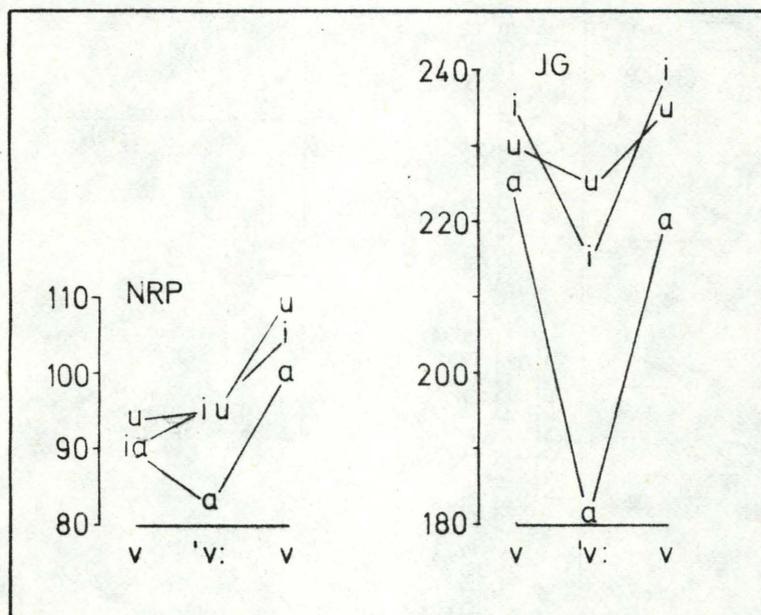cording.

Figure 11

Average F$_0$ in Hz in the vowels [i ɑ u] in f-words as a function of position in the word.
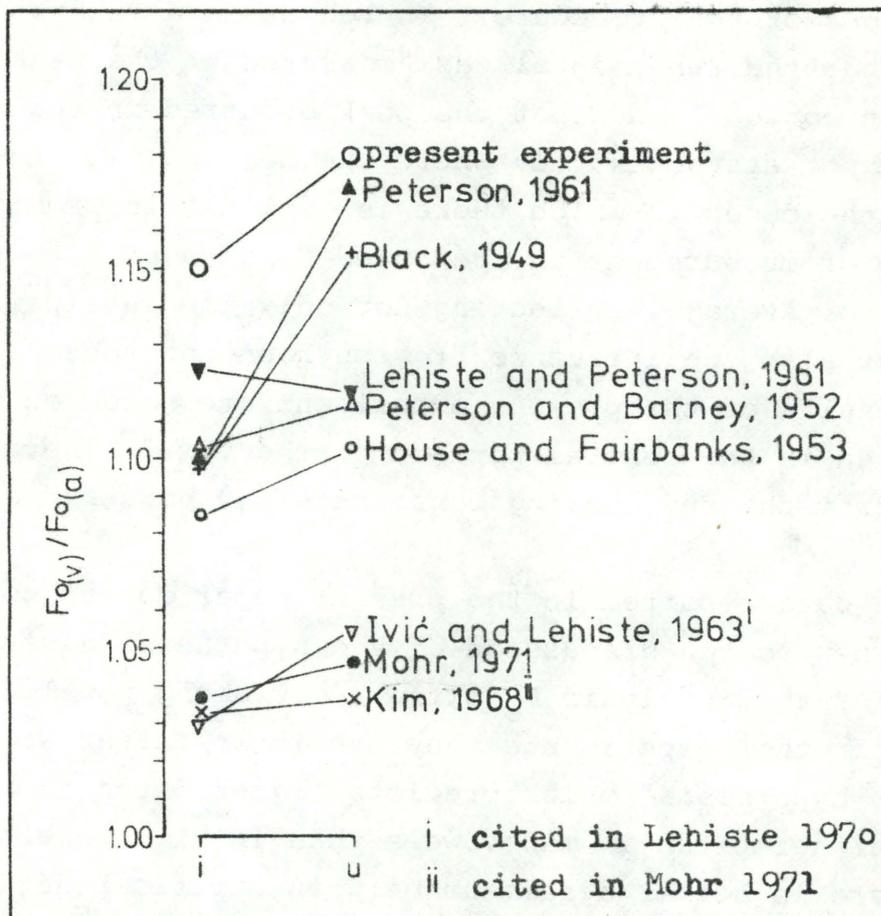
23

Figure 12

The ratios of [ɑ] to [i] and [ɑ] to [u]
computed on the basis of data from the
present and a number of other investiga-
tions.

The data from Kim (1968) were available to me only in the form of illustrations in Mohr (1971). In that paper (fig. 3) the F_o movements were given as straight lines connecting the F_o values at the beginning and end of the vowels. The points in fig. 12 were computed from the values reached at the mid point of the vowels. In Peterson and Barney (1952), in Peterson (1961), and in Ivič and Lehiste (1963) (cited in Lehiste 1970) there are no descriptions of the procedure. In Lehiste and Peterson (1961) it is said that the fundamental was "measured at the peak of the intonation contour" and that the peak occurred in the testword, but it is not stated exactly where the peak was within that word.

In the cases in which there is explicit information on the procedure of measurement the reported F_o data seem to represent some sort of average, reflecting not only the intrinsic F_o of vowels but also the influence from surrounding consonants, whereas the results of the present experiment are based on measurements taken in the central portion of the vowels where the influence from the surrounding consonants may be assumed to be minimal.

The data reported in the present paper do not contribute anything new to the discussion of the hypotheses advanced to account for the intrinsic F_o differences among vowels. Like the results of other experiments they are inconsistent with Mohr's "pressure hypothesis" which predicts higher pitch in [i] than in [u] and higher pitch in mid vowels than in high vowels.

There is better agreement between empirical data and the source/tract coupling hypothesis. However, the relationship between [i] and [u] stands out as a crux. The intrinsic F_o of [i] is generally lower than that of [u]. Now, since the hypothesis predicts an inverse correlation between F_1 and F_o, [i] should be expected to have higher F_1 than [u]. Nevertheless, we normally find the opposite relation, viz. a higher F_1 in [u] than in [i]. It may be assumed, therefore, that the intrinsic F_o differences cannot be fully accounted for by the source/tract coupling hypothesis alone.

As pointed out in section 1 of this paper, the modified
tongue pull hypothesis seems to provide the best explanation of
available data on the intrinsic F_O of vowels, particularly of
the relation between the vowels [i] and [u]. Since the results
of the present investigation are in good agreement with data
obtained in other experiments, they may be considered to speak
in favour of the tongue pull hypothesis. It must be emphasized,
however, that the data reported here do not provide conclusive
evidence. Such evidence will have to be based on physiological
rather than acoustic data.

References

Basbøll, H. 1968: "The phoneme system of Advanced
Standard Copenhagen", ARIPUC 6,
p. 33-54

Beil, R.G. 1962: "Frequency analysis of vowels pro-
duced in a helium-rich atmosphere",
JASA 34, p. 347-49

van den Berg, J. 1955: "On the role of the laryngeal ven-
tricle in voice production", Folia
Ph. 7, p. 57-69

Black, J.W. 1949: "Natural frequency, duration, and
intensity of vowels in reading",
JSHD 14, p. 216-21

Ewan, W.G. 1975: "Explaining the intrinsic pitch of
vowels", Paper presented at the
Fifth California Linguistics Asso-
ciation Conference, San Jose, May
4, 1975, p. 1-9

Flanagan, J. and L. Landgraf 1968:
"Self-oscillating source for vocal-tract synthesizers", IEEE Transactions on Audio and Electroacoustics AU-16, p. 57-64

House, A.S. and G. Fairbanks 1953:
"The influence of consonant environment upon the secondary acoustical characteristics of vowels", JASA 35, p. 84-92

Ladefoged, P. 1964:
A phonetic study of West African languages: an auditory-instrumental survey (Cambridge)

Lehiste, I. 1970:
Suprasegmentals (MIT Press)

Lehiste, I. and G.E. Peterson 1961:
"Some basic considerations in the analysis of intonation", JASA 33, p. 419-25

Lieberman, P. 1970:
"A study of prosodic features", Haskins SR 23, p. 179-208

Mohr, B. 1971:
"Intrinsic variations in the speech signal", Phonetica 23, p. 65-93

Ohala, J.J. 1973:
"Explanations for the intrinsic pitch of vowels", Monthly Internal Memorandum, Phonology Laboratory University of California, Berkeley, p. 9-26

Peterson, G.E. 1961:
"Parameters of vowel quality", JSHR 4, p. 10-29

Peterson, G.E. and H.L. Barney 1952:
"Control methods used in a study of the vowels", JASA 24, p. 118-27

Shimizu, K. 1961:
"Experimental studies on movements of the vocal cords during phonation by high voltage radiograph motion pictures", Studia Phonologica 1, p. 111-16

Thorsen, N. 1976:   "An acoustical investigation of
                            Danish intonation: preliminary
                            results", <u>this issue</u>.

SPECTRAL PROPERTIES OF GERMAN AND DANISH SIBILANTS

Preben Andersen

Abstract: The Danish sibilants s and ʃ are normally placed audi-
torily between German s and ʃ in regard to clearness
versus darkness. This distribution also seems to be
meaningful as regards their acoustic properties.
A partial overlapping is often observed between spectra
of Danish s and ʃ, whereas the German spectra are al-
ways clearly discernible. The spectral shape of the
Danish sibilants, especially that of s, seems to be
very sensitive to the parameter rounding/non-rounding
of a preceding vowel, and occasionally these spectral
differences are greater than those between sibilants
belonging to different phonemic categories. Danish and
German sibilants can be distinguished by their spectral
composition. The Danish sibilant spectra normally show
two energy maxima, one in the lower and one in the upper
part of the frequency range. German spectra normally
show but one energy maximum located in the middle of
the frequency range for s and in the lower fifth of the
frequency range for ʃ.

1. Introduction

Compared to the large number of contributions to the audi-
tory and acoustic description of vowels and stop consonants, very
little has been published on fricative consonants. The previous
descriptions of the fricative sibilants [s] and [ʃ] have been
concentrated either on acoustic cues for auditory recognition of
the sibilants or on acoustic descriptions of their spectral com-
positions, and on a specification of the most simple general
parameters needed for a satisfactory generation of synthetic
sibilant stimuli.

As for the investigation of perceptual cues of American-English and Swedish fricatives, K.S. Harris (1958) and J. Martony (1962) found that the perception of sibilants depends on the shape of the noise spectrum, not on the variation in vowel transitions, in contradistinction to the perception of other fricative consonants (for instance [f, θ]), which depends mainly on the adjacent vowel transitions. Martony further observed that a typical [f]-spectrum with a 20 dB amplification of its overall intensity was perceived as [s].

In descriptions of the spectral characteristics of sibilants the greatest importance has been attached to the lower part of the spectrum. Jassem (1962 and 1968) and Katarzyna (1968) define four fricative formants, three of which appear in sibilants within the following frequency ranges: F_2 about 1-2 kHz, F_3 about 2-3 kHz, F_4 about 3-5 kHz. The intensity relations and the density of location of these formants are decisive for the distinction between [s] and [ʃ]. Dense formant locations are seen in [ʃ], whereas dispersed formant locations are characteristic of [s]. Jassem gives the formulae: $F_4 - F_2 \geq 1.8$ kHz = [s], $F_4 - F_2 < 1.8$ kHz = [ʃ].

Peter Strevens (1960) identifies 1 kHz as the line of demarcation between formant location density for [s] and [ʃ]. As for intensity levels, Jassem finds that F4 normally has the highest level, F2 the lowest, and F3 is the varying formant in regard to intensity level; in most cases [s] had an F3 with almost the same level as F2, while F3 of [ʃ] had a level close to that of F4. A lower frequency limit is described as the lowest point on the frequency scale at which the energy reaches a level of 20 dB below the maximum level of the spectrum; this gives [ʃ] a low (not more than 2 kHz), alveolar [s] a medium (2 - 3 kHz), and dental [s] a high (more than 3 kHz) lower frequency limit. No unambiguous tendencies were found for the upper frequency limit. The energy of [ʃ] often decreases abruptly at about 7 kHz, but the spectra of some subjects exhibited intensity peaks up to 12 kHz. [s] always had continuous energy up to 12 kHz.

In several reports on synthetic sibilant production the
authors start with 4 - 5 poles and 3 - 4 zeroes but find that
auditorily acceptable sibilants can be obtained with only 2 poles
and 1 zero. Martony, Cederlund, Liljencrants, and Lindblom (1961)
mention for [s] a zero at about 3.4 kHz and poles at about 4.2
kHz and 5.3 kHz, for [ʃ] a zero at 1.4 kHz and poles at about 2.1
kHz and 3.1 kHz. Heinz and Stevens (1961) differ by using a
second pole in [s] with the center frequency 8 - 8.4 kHz and by
giving the zero in [ʃ] a center frequency of 3.4 - 4.4 kHz.
Martony also synthesized acceptable [s]-stimuli with one pole
at about 3.5 kHz and 2 zeroes at 3 kHz and 5.9 kHz.

It is evident that the details of these results depend on
the languages involved, and the more general characteristics must
be found in the relative differences between [s] and [ʃ].

2. Recording of the material for the present study

2.1 Choice of material

The present paper is concerned with an acoustic analysis of
the spectral properties of German and Danish sibilant fricatives.
On the basis of measurements of sections from sonagrams an attempt
was made to reveal some differential spectral properties, partly
among German and Danish sibilants and partly among sibilants within
each of the two languages. The latter type of comparison not
only applied to the phonemes /s/ and /ʃ/ but also comprised dif-
ferences among variants of each of these phonemes in different
environments. Needless to say, a considerable individual varia-
tion was supposed to occur.

The main part of the material consisted of one list of Danish
and one of German words. The words had the structure

(C)(C)V(:)[ʃ,s](ə)(n)

and preconsonantal vowels comprised all the important categories,
rounded-unrounded, front-back, and open-close.

It was attempted to find similar words of German and Danish in order to equalize the experimental conditions. The Danish list was spoken by 3 male persons speaking Standard Copenhagen Danish and the German list by 4 male persons, who spoke Standard German (Hochsprache) or an approximation to Standard German.

2.2 Recording procedure

The recordings of the Danish words were made in the studio of the Institute of Phonetics in Copenhagen. The instrumental equipment used was a Brüel & Kjær Condenser Microphone, type 4145, connected with a Brüel & Kjær Microphone Amplifier, type 2603, and the recordings were played on a Lyrec TR 47-2 professional recorder.

The German words were recorded in Cologne at the Institut für Phonetik. The recordings were made in a studio with a Studio Magnetofon M10 taperecorder (Studio-Mischpulteinrichtung) from Telefunken.

Subsequently, sonagrams were made of the words from the list. Equipment used was a Revox A 77 semiprofessional recorder and a Kay-Electric Sonagraph 7027. Two sets of wide band spectrograms were produced: One set with a frequency range of 160 - 16000 Hz (filter width 600 Hz), from which the sections were taken (one section approximately at the middle of each sibilant), and another set with a frequency range of 40 - 4000 Hz. Some uncertainty might be expected in the measurements of frequencies higher than 7 kHz.

3. Preliminary descriptions as a basis for the measuring procedure

## 3.1 Spectral composition as basis of description

A comparison between the continuous sonagrams and the sections showed that the sections gave a more detailed impression of the

locations of the energy concentrations and the intensity relations
between the single peaks, whereas the continuous sonagrams, and
especially those with the frequency range of 40 - 4000 Hz, gave a
more exact information about the lower frequency limits of the
sibilants and their relation to the preceding vowel. F3 of the
preceding vowel often continued as a prolonged noise formant
into the sibilant spectrum; this noise formant often represented
a lower frequency limit, varying with the presence or absence of
rounding in the vowel. On some sonagrams even prolongations of
F2 were seen, especially from preceding rounded vowels, but the
difference front - back in the preceding vowel showed no general
influence on the lower limit of the fricative spectrum. The best
information on context dependency and especially on the phonemic
and language-specific differences between the sibilants still
seemed to be obtained from the general composition of the spectra,
namely the interrelations among the various peaks both with regard
to frequency locations and intensity levels. Accordingly, the
sections were chosen as basis for the measuring procedure.

3.2 Preliminary information about spectral differences

An evidently language-specific characteristic of the sibi-
lants was observed by comparing the sections. German [s] starts
at a relatively high frequency, Danish [s] and [ʃ] have medium
lower limits, while German [ʃ] has the lowest limit. Accordingly,
German [s] and [ʃ] are clearly distinct, whereas there is more
overlapping between Danish [s] and [ʃ] and a considerable indi-
vidual variation.

Fig. 1 shows a representative example of the spectral shapes
of the German and Danish sibilants. Most of the spectra of Danish
sibilants seem to consist of two energy concentrations with a zero
in between; each of these humps of energy, henceforth called
maxima, consists of various peaks, more or less fused. The sibi-
lants of the subject NRP generally only approached this pattern.
The higher maximum was often absorbed in the long decreasing slope

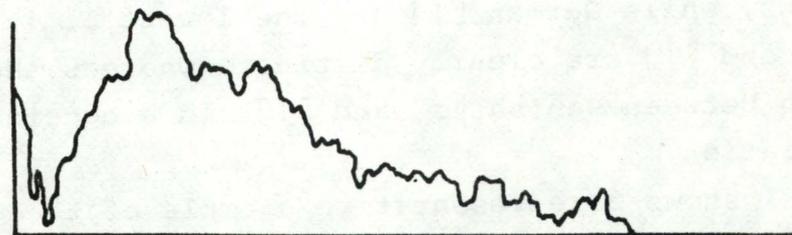0   2   4   6   8   10   12   14   16 KHZ
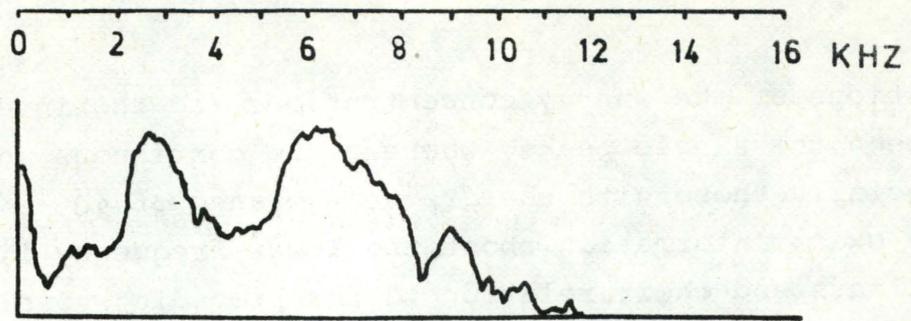
[s] in the Danish word "dus"(OTH)
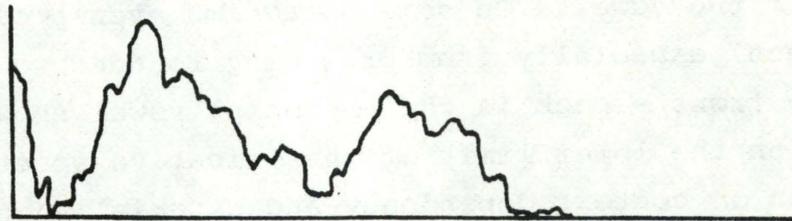
[ʃ] in the Danish word "kaleche" (PA)
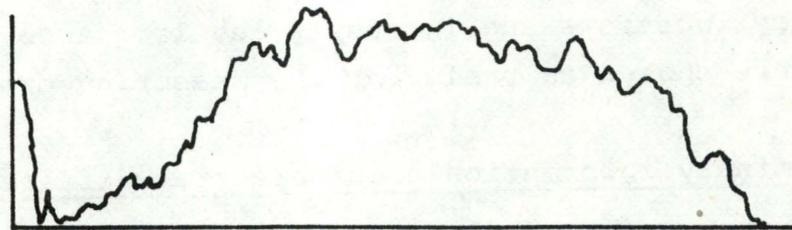
[s] in the German word "Esse" (U.S)

[ʃ] in the German word "wischen" (JS)

Figure 1

of the lower maximum. The German sibilant spectra rather show a
concentration of energy around one large maximum in the central
region of the spectrum for [s], and somewhat lower for [ʃ]. In
the lower frequency range of both Danish and German sibilant
spectra there are peaks which might be compared with the F2 -
F3 - F4 of Jassem et al.

3.3 Grouping of the spectra

The previously mentioned context dependency with regard to
the presence of rounded - unrounded vowel before the sibilant,
turned out to be important. It became clear that a distinction
between sibilants in these two environments facilitated the
general description. This further made it possible to reduce some
of the overlapping between the Danish sibilant spectra to "partial"
overlapping caused by the presence or absence of rounding in the
preceding vowel; e.g., the spectrum of an [s] following a rounded
vowel may be more similar to the spectrum of [ʃ] following an
unrounded vowel than to the spectrum of [s] following an unrounded
vowel.

The following groupings were undertaken: (a) Danish [s]
preceded by an unrounded vowel (abbreviated Da. [s] unround.),
(b) Danish [s] preceded by a rounded vowel (Da. [s] round.)[1],
(c) Da. [ʃ] unround., and (d) Da. [ʃ] round. The German sibilants
were classified in the same way although the distinctions were
less clear. Especially the two classes of [ʃ] were found to be
very similar, as German [ʃ] is often inherently rounded. As to
German sibilants in general, they seem to have a more distinct
articulation than the Danish ones, and they are therefore less
susceptible to coarticulation. Moreover, all the German vowels
were lax, and lax vowels have less pronounced rounding. Anyway,
the eight classes were maintained as a basis for the measuring
procedure.

---

1) The vowel [ɔ], e.g. in the word 'losse', was treated as a
    rounded vowel, although on some sections it was followed by
a sibilant spectrum belonging to class (a).

The difference among the groups was mainly found in the frequency locations of the lower maximum and for some Danish subjects also in the intensity relations between the two maxima. The problem will be discussed in detail in section 4.

4. Measuring procedure

4.1 Peak frequencies as primary parameters

On the basis of the previous observations the following parameters were chosen as relevant for the measurements: Center frequency of the single peaks; center frequencies of the maxima, if any; frequency range of the dominating maximum or prominent part of the spectrum; center frequencies of the zeros; upper and lower frequency limits.

There were normally 10 - 11 measurable peaks in the Danish sibilant spectra, whereas the German sibilants contained 13 - 15 measurable peaks. These peaks were chosen as primary parameters. A close comparison of the center frequencies of the primary peaks within each of the languages revealed certain frequency ranges, within each of which the occurrence of a primary peak was particularly frequent, that is: within the given ranges, there was a peak in all cases, but the relative intensities of the peaks varied for different sound classes depending on place of articulation and on rounding. For example, one peak might have a high intensity and constitute the center of the lower maximum in a [ʃ], whereas a peak of the same frequency in an [s] might appear just as a minor bulge on the rising slope of the lower maximum. Naturally there were cases of fusing or completely lacking peaks in the distribution.

4.1.1 "Basic peaks"

For each of the previous mentioned classes of spectral types (Da. [s] round., Ger. [ʃ] unround., etc.) a set of mean primary peaks were calculated by means of the mean values of the single peak measurements, first for each subject and then for the subjects combined. A comparison of the different sound types revealed that many of the peaks were practically identical and others were so close that they could be united into one, somewhat broader, peak. In this way, a final number of "basic peaks" was set up for each language, for German 14 and for Danish 15, see fig. 2. The number of "basic peaks" for Danish is larger than the number of measurable peaks (10 - 11) in the individual spectra, because the variation both among the subjects and between rounded and unrounded forms of the sibilants was relatively large in Danish.

Within the frequency ranges of the "basic peaks" one should be able to fit in the peak center frequencies of any individual kind of [s, ʃ] from the two languages. Naturally, this description does not involve the claim that there should be a fixed number and fixed location of the resonance frequencies for all sibilant articulations, and an investigation of a larger material might lead to an increased or reduced number of "basic peaks" or partly different frequency ranges for the peaks, but for the purpose of this investigation the present model turned out to be quite useful, if not as a general theory of "basic peaks", then at least as a division of the spectrum which is expedient for comparisons of intensity relations in selected ranges. A division on the basis of frequency alone, e.g. in third octaves or the like, might cut through important peaks. The "basic peak"-model was therefore used as framework for the further description.
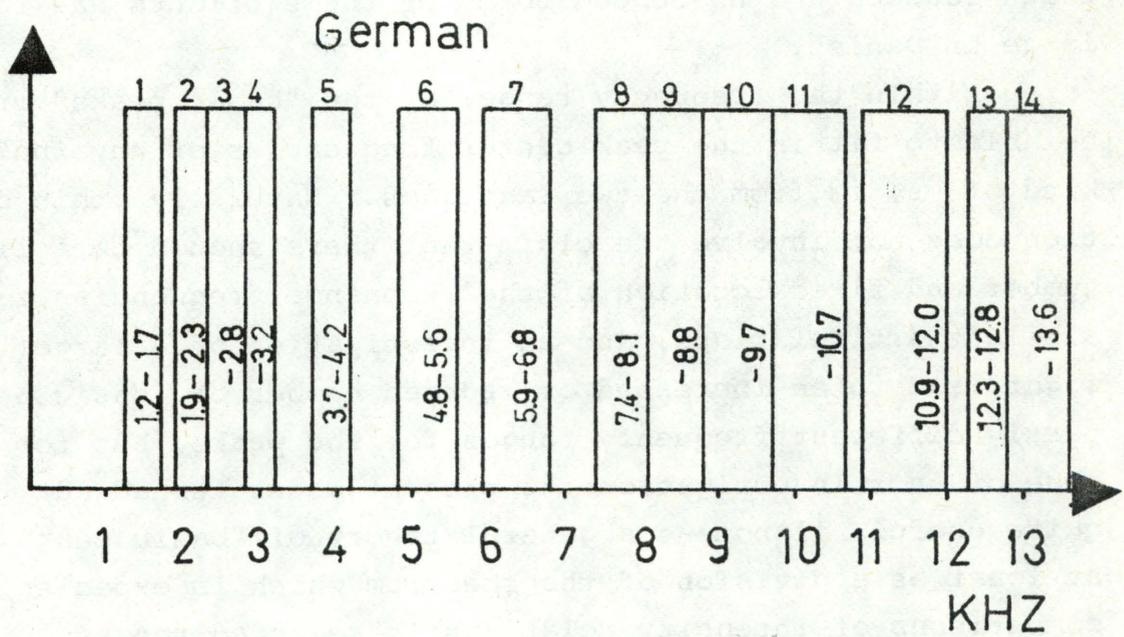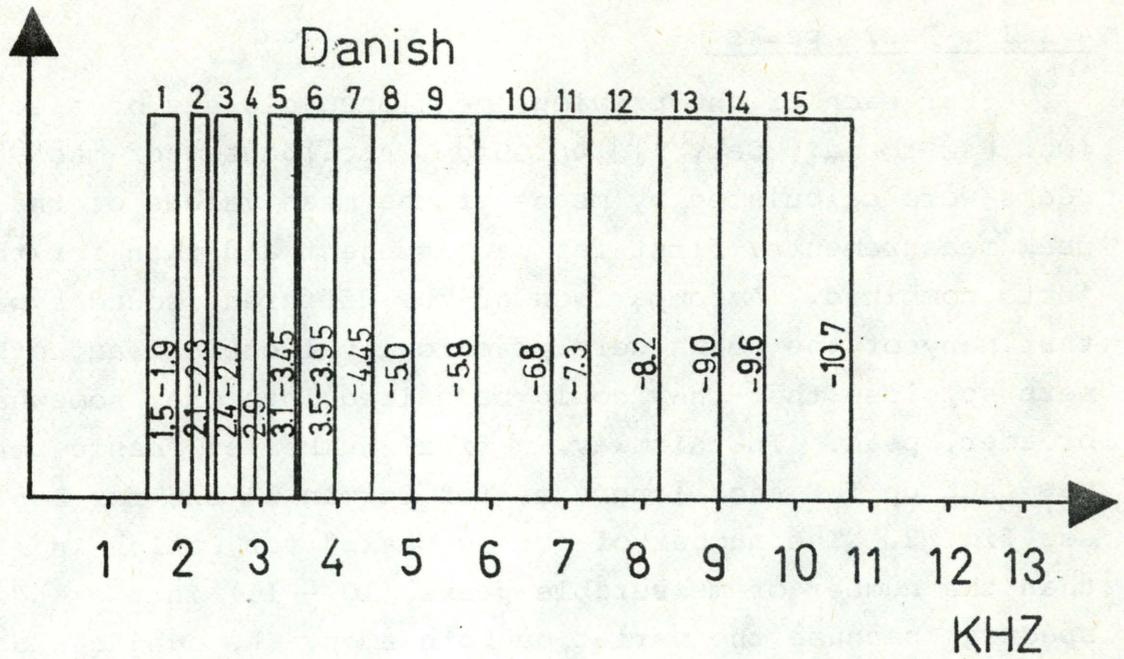
38

Figure 2
"Basic peaks".

4.2 Ranking of intensity relations

A ranking of the intensity values based on the relation between the peaks of the various classes was preferred to physical scales, partly because a valid calibration of the sections was not available, and partly because of the relatively large differences in over-all intensity among the recordings of the various subjects.

The rank relations among the basic peaks were set up for each individual example, each "basic peak" frequency of the section being given an intensity rank between 1 and 14 or 15, depending on the language. As previously mentioned, only 10 - 11 peaks were measurable in each of the sections of Danish sibilants, therefore one could not expect to find a peak in everyone of the 15 "basic peak" frequency ranges, and in the instances where a basic frequency range contained a slope, the mean level of intensity of this sloping portion was ranked in relation to the other "basic peak" areas of the section.

4.3 Stylized average spectra of sibilants

Now it became possible to calculate the mean intensity relations for a whole group of spectra. The median was chosen as a statistic measure for calculating the mean rank of the "basic peaks" of any desired number of sections. Accordingly, each of the classes (Da. - Ger. [s]-[ʃ] round. - unround.) could be acoustically described by the average rank of the intensity in the "basic peaks". Figs. 3 - 11 show stylized average spectra constructed on the basis of the average intensity ranks of the "basic peaks". The average for each "basic peak" range had to be reproduced as a horizontal line; this was necessary because the (exact) frequency location of the peak within a "basic peak" range might vary for each individual sound. The empty spaces between some of the "basic peaks", especially those of the German sibilants, are supposed to be minor zeros or at least frequency domains with no obvious peaks.

Figure 3

Stylized average spectra
(Danish).   Subject PA.

41

Figure 4

Stylized average spectra
(Danish). Subject: OTH.
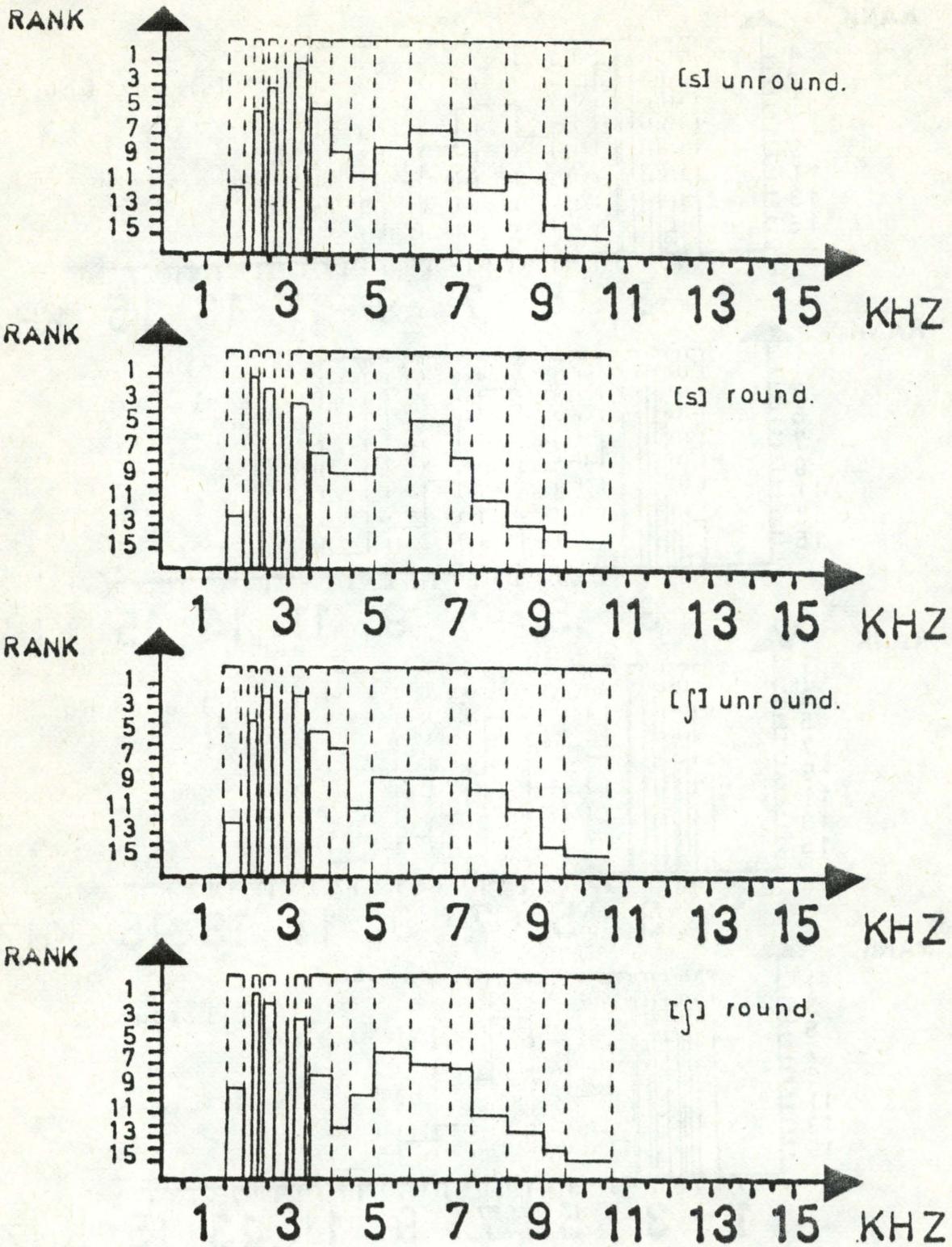
Figure 5

Stylized average spectra
(Danish).   Subject: NRP.

Figure 6

Stylized average spectra
(Danish). Average across
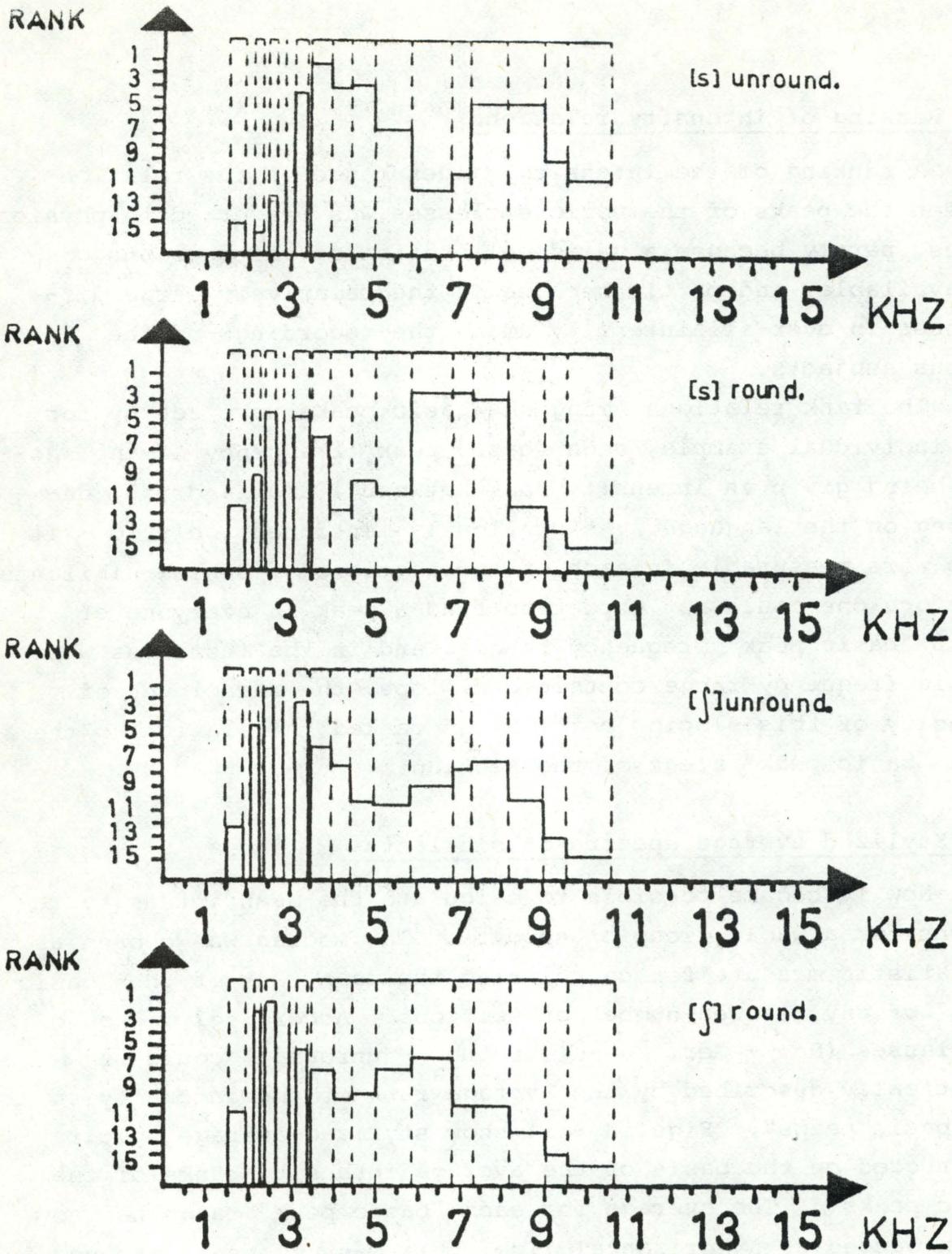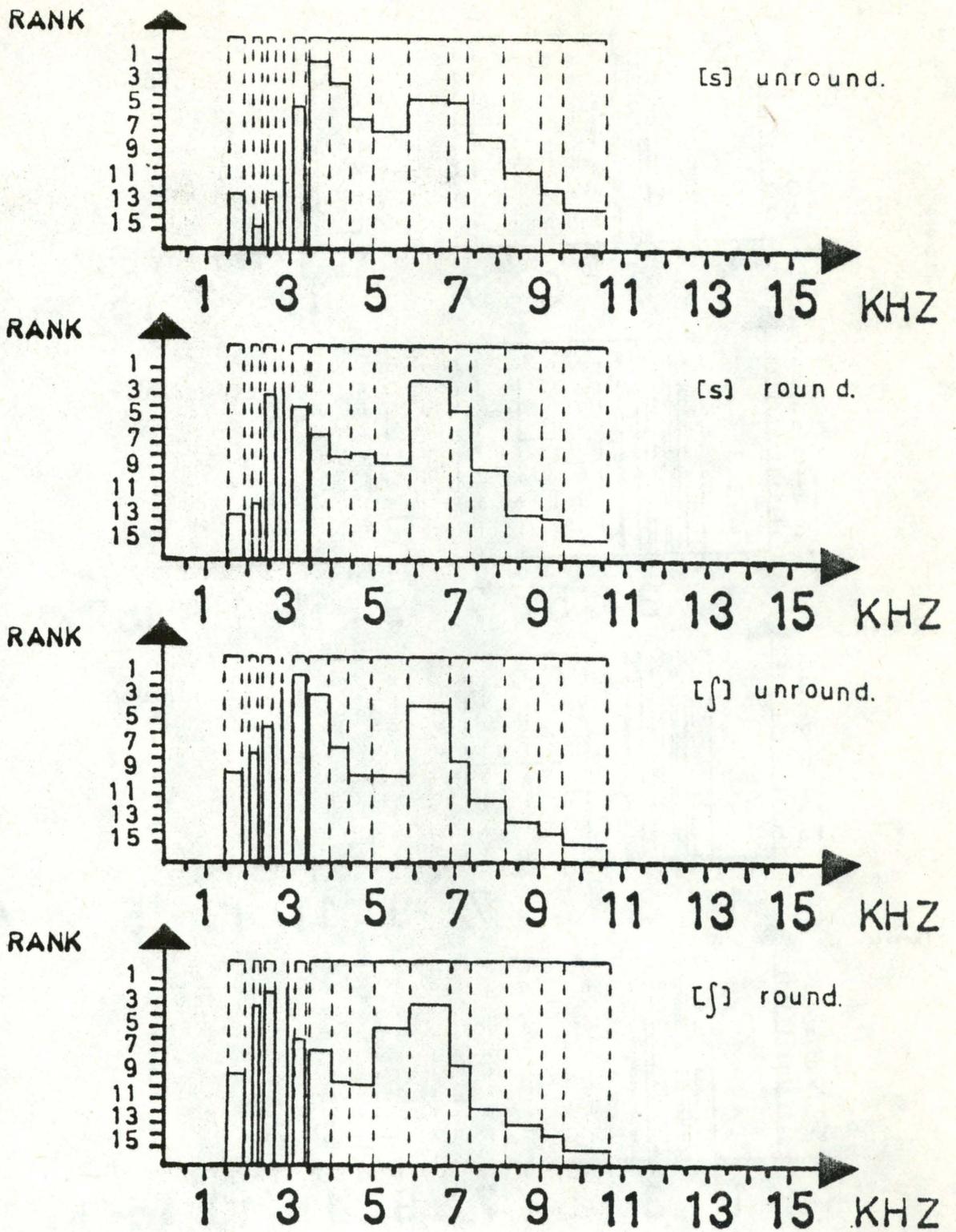subjects PA, OTH, and NRP.

44

Figure 7

Stylized average spectra
(German). Subject: PS.
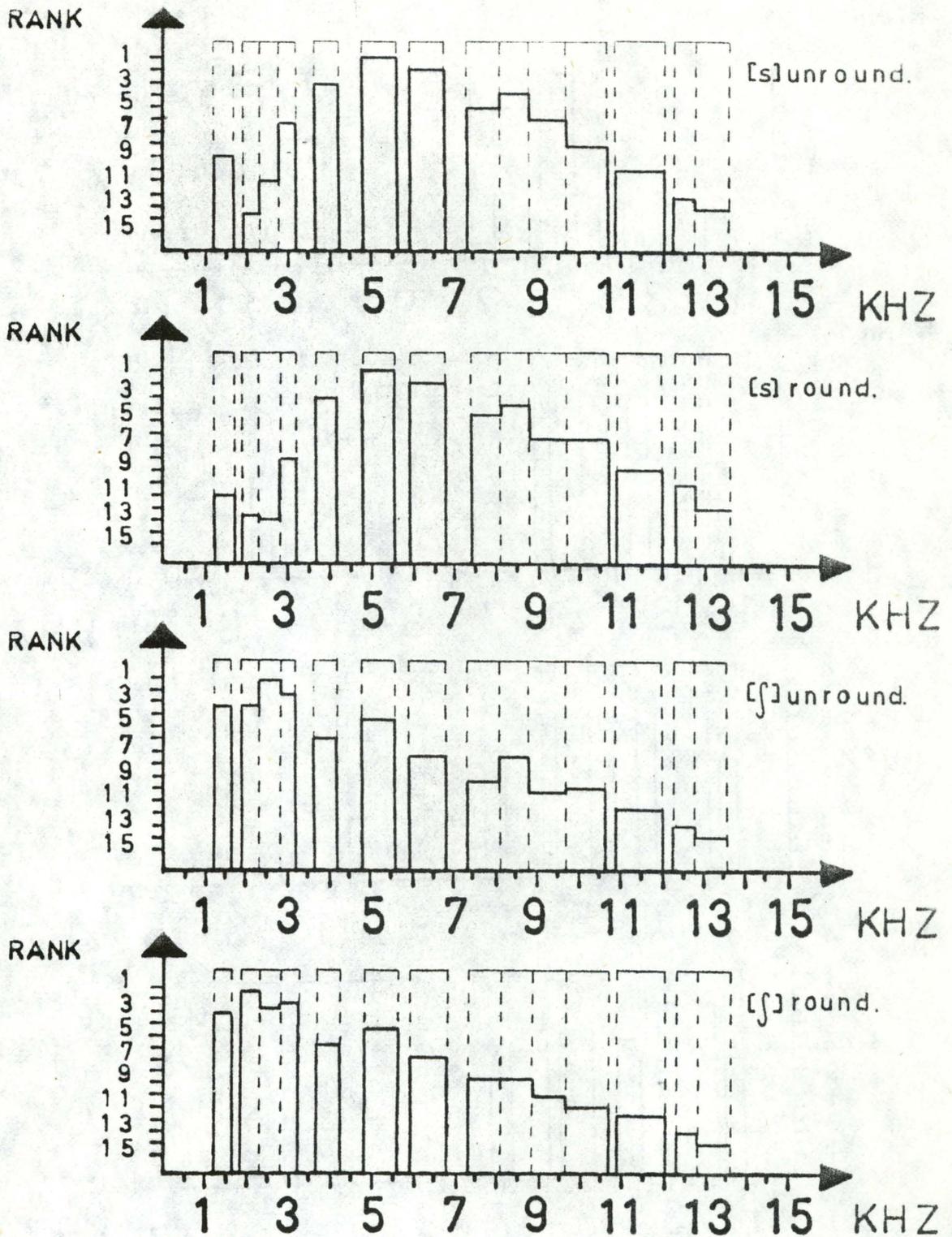
Figure 8

Stylized average spectra
(German). Subject: HH.

46
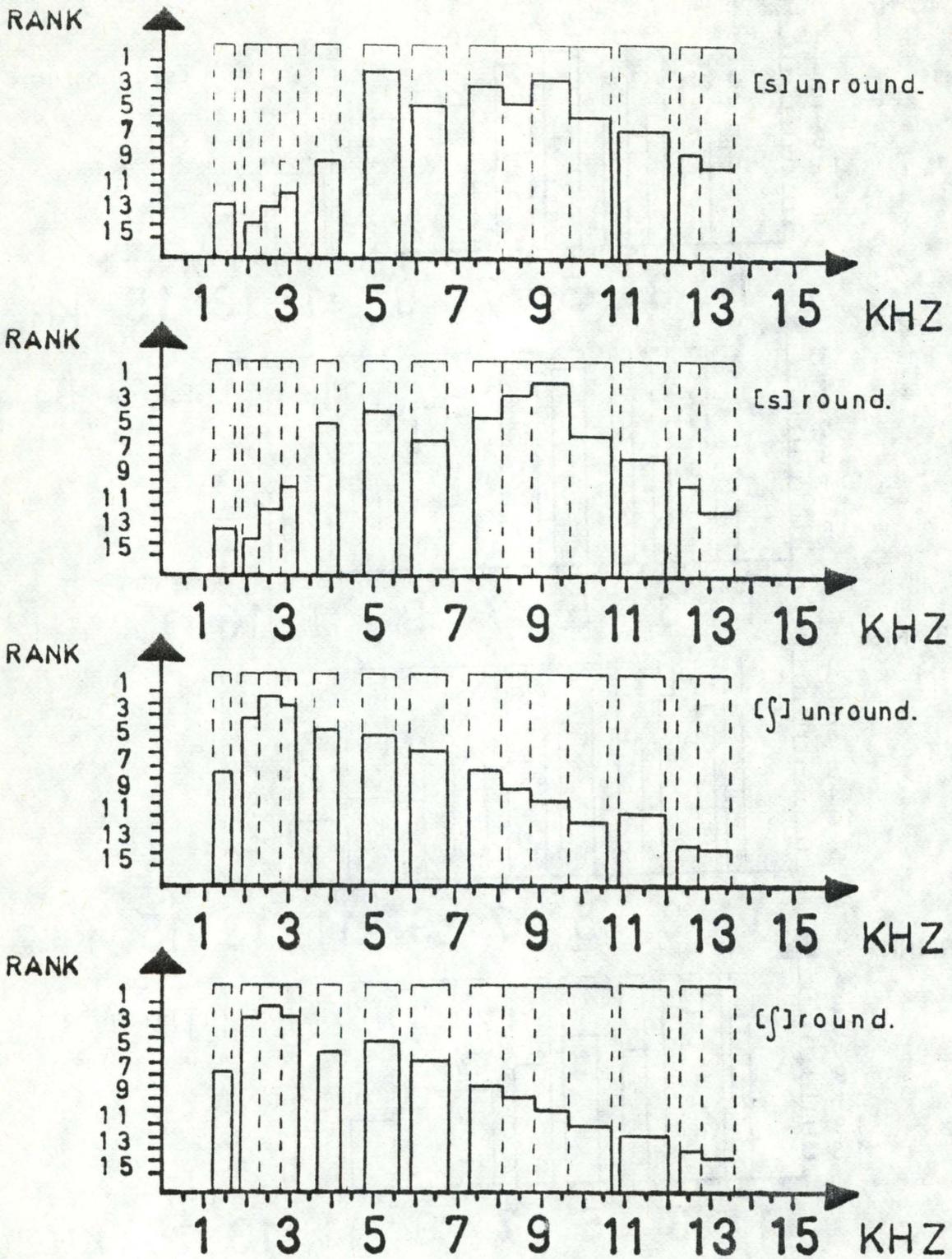
Figure 9

Stylized average spectra
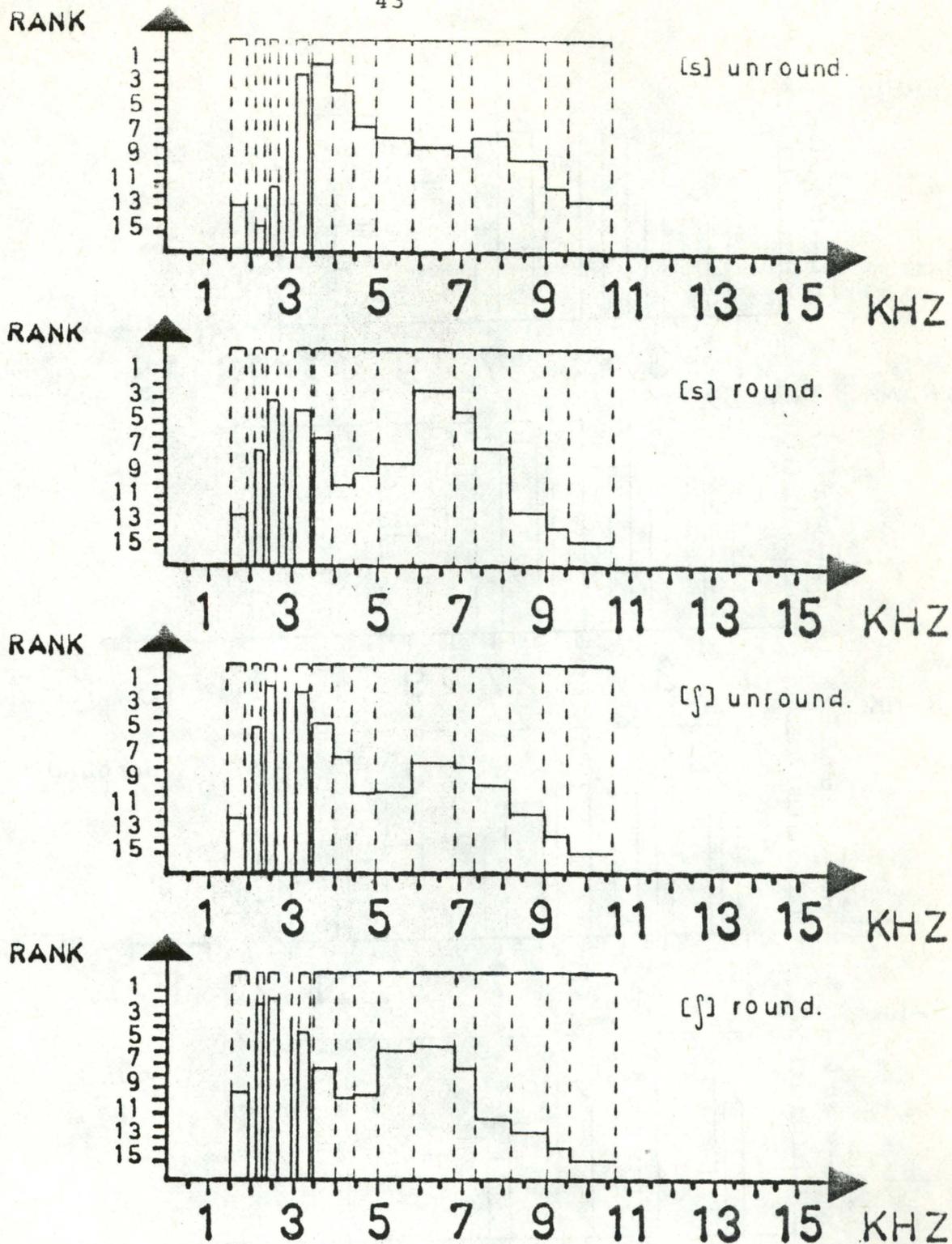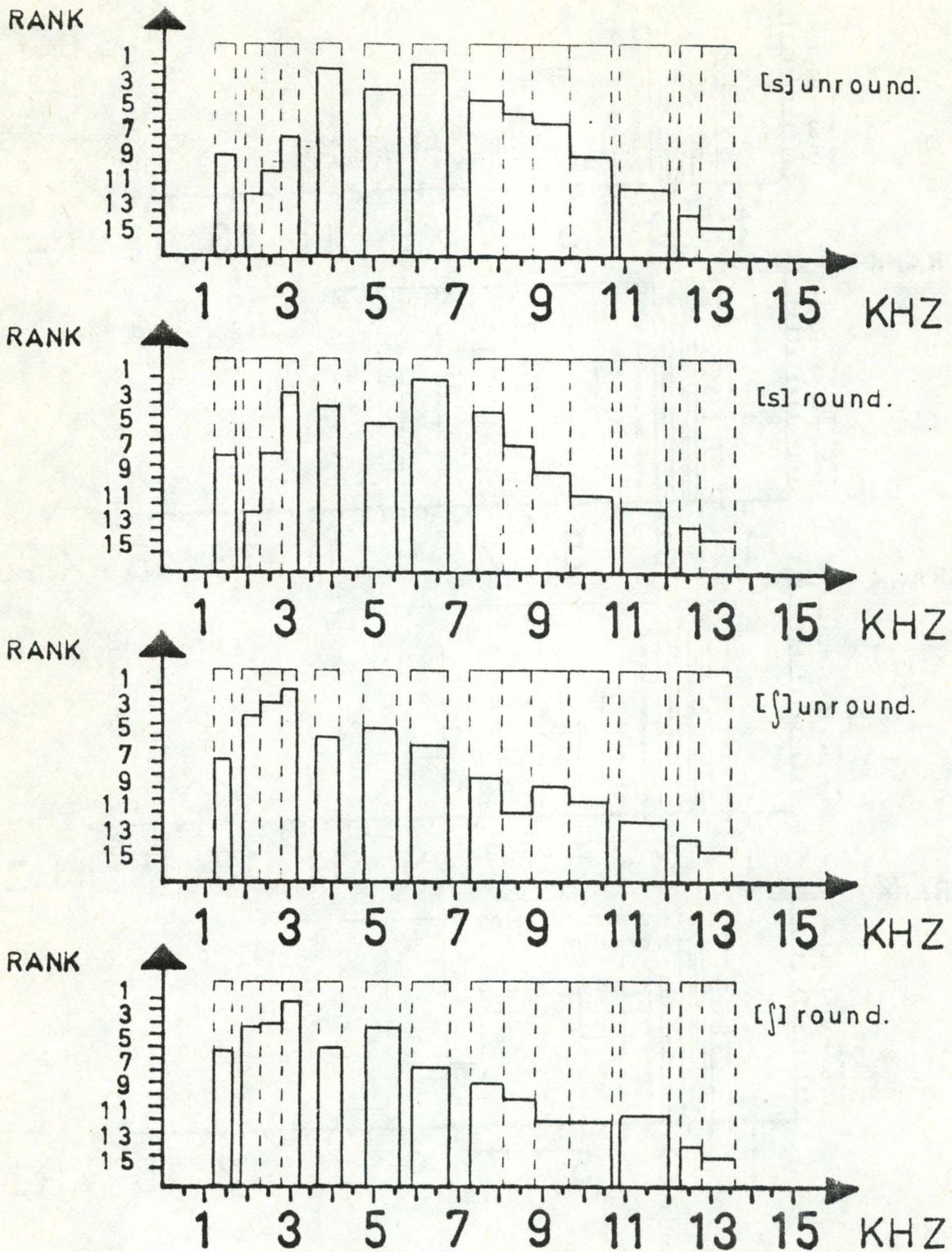(German). Subject: US.
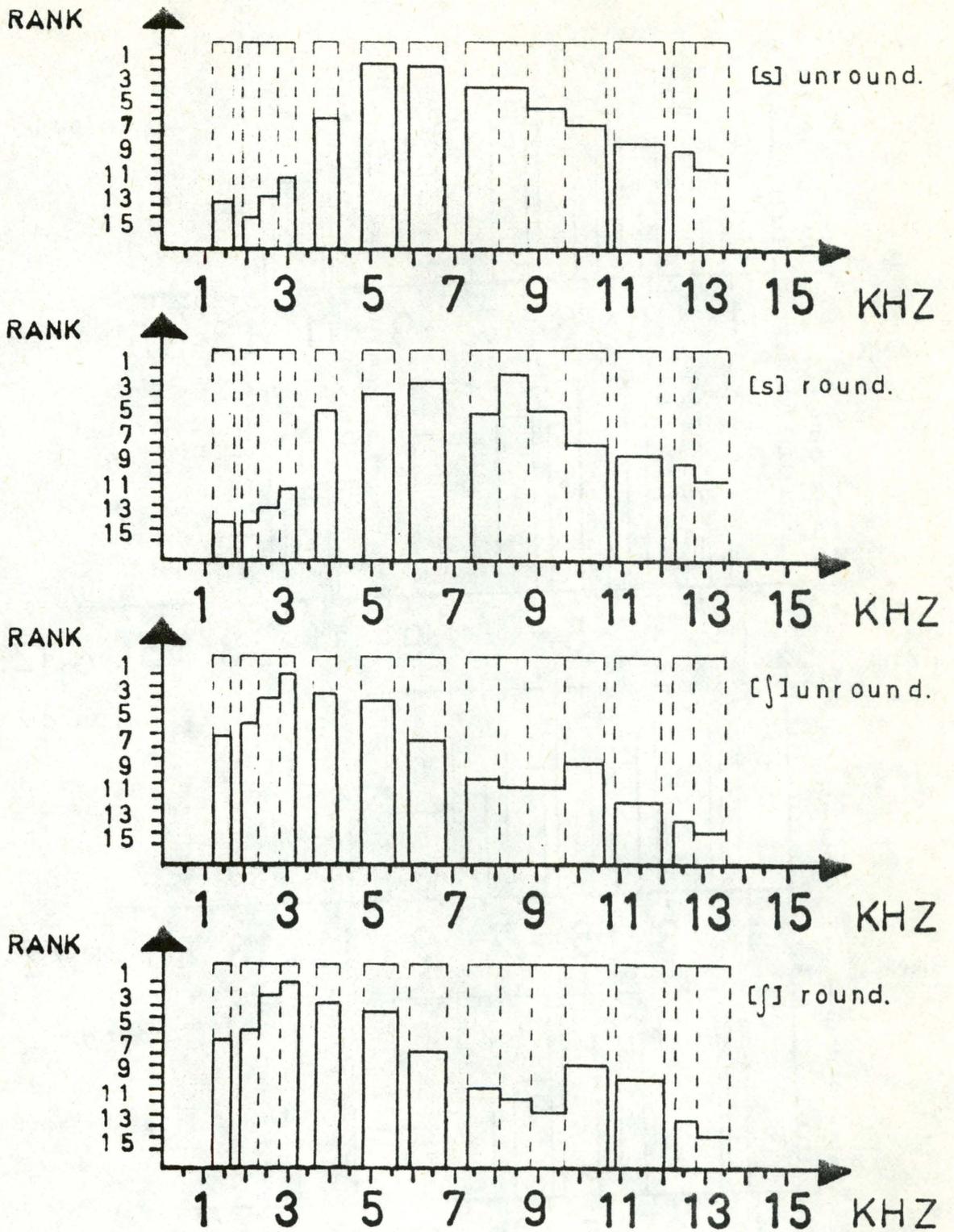
Figure 10

Stylized average spectra
(German). Subject: JS.

48
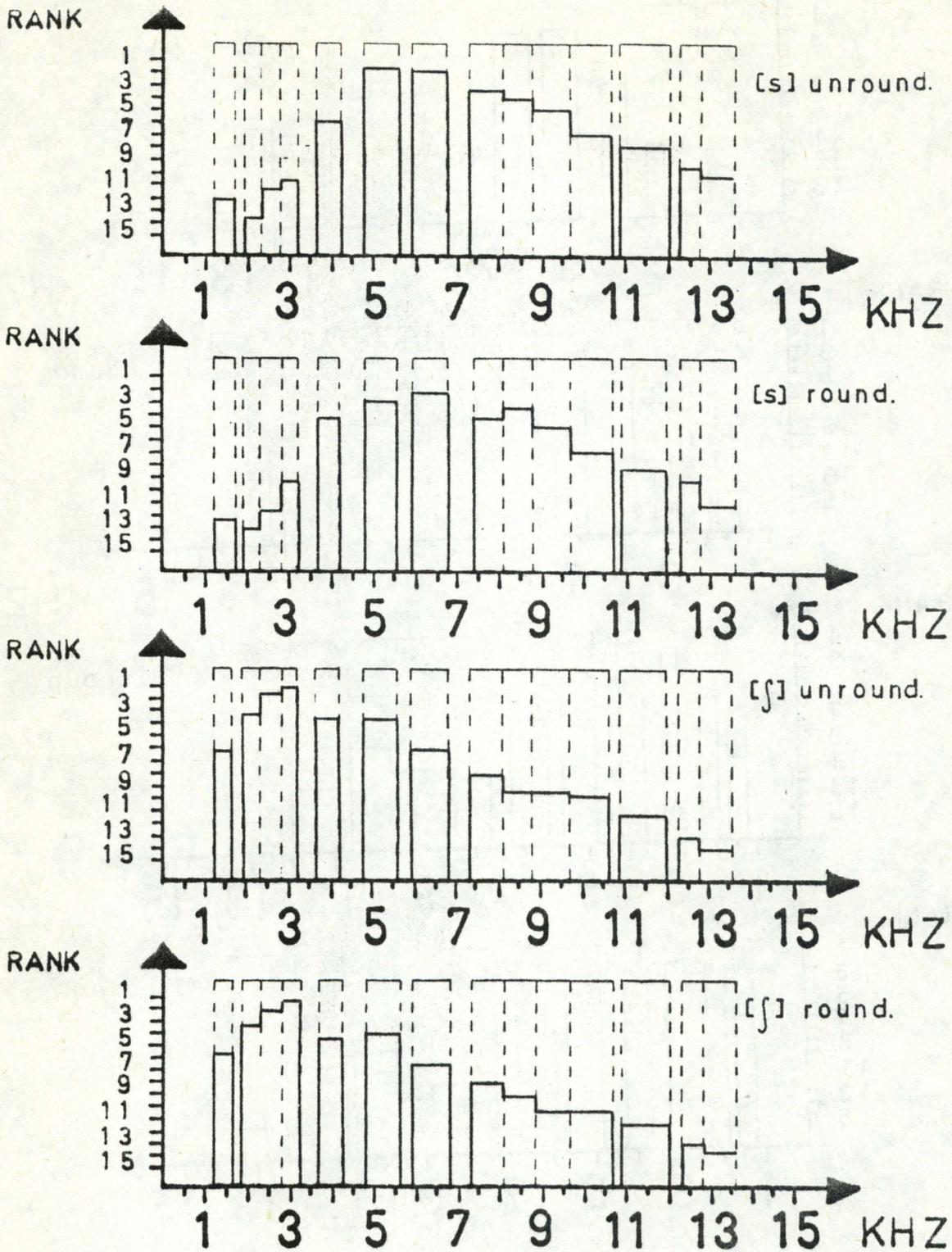
Figure 11

Stylized average spectra
(German). Average across
subjects PS, HH, US, and JS.

5. Results and discussion

On the basis of the stylized average spectra and the pre-
liminary observations supported by the physical measurements
the following tentative general characteristics of the classes
can be put forward:

5.1 Spectral properties of Danish sibilants

5.1.1 Sibilant preceded by rounded versus unrounded vowel

This difference is manifested most clearly in the [s]-
spectra and appears in the following ways:

When the sibilant is preceded by a rounded vowel the in-
tensity of the lower maximum is relatively weak, while the in-
tensity of the higher maximum is relatively strong (see figs.
3 and 4).
In several cases the upper maximum is dominant, whereas the
lower maximum is usually dominant before unrounded vowels.

The center frequency of the lower maximum is lower after
a rounded vowel (see figs. 3-6).
Average differences are greatest for the [s]-spectra: [s] unround.
3800 Hz - [s] round. 3000 Hz; [ʃ] unround. 3000 Hz - [ʃ] round.
2800 Hz.

The frequency range of the increasing slope of the lower
maximum is smaller after rounded vowels.
Here the measurements show the averages:
[s] unround. 2250 - 3300 Hz - [s] round. 2100 - 2400 Hz;
[ʃ] unround. 1650 - 2450 Hz - [ʃ] round. 1600 - 1900 Hz.

Decrease of the energy in the upper part of the spectrum
starts earlier after rounded vowels (see figs. 3-5).
This tendency can hardly be observed for the subject OTH
Average for all subjects:
[s] unround. 8450 Hz - [s] round. 7750 Hz; [ʃ] unround. 7250 Hz
- [ʃ] round. 6050 Hz.

The valley (zero) between the two maxima has a lower center frequency after rounded vowels (see especially fig. 3).

This is only obviously true for subject PA.

On the whole, a general shift downwards of the spectral energy takes place after a rounded vowel, but at the same time the central point of energy shifts upwards on the frequency scale, that is, the peaks within the two maxima are shifted upwards. The rather dubious measurements of absolute upper and lower frequency limits show the shift downwards:

[s] unround. 1550 - 11950 Hz - [s] round. 1300 - 11600 Hz
[ʃ] unround. 1300 - 11000 Hz - [ʃ] round. 1200 - 10400 Hz

5.1.2 General differences between Danish [s]- and [ʃ]-spectra

Here the interesting problem arose that it seemed almost impossible to find general acoustic differences in [s]- versus [ʃ]-spectra, especially for the two subjects OTH and NRP. One pronounced difference seems to be:

[ʃ] has a decisive decline of upper spectral energy at lower frequencies than [s].

This is obvious in most of the spectra, and the average measurements are:

[s] 8050 Hz - [ʃ] 6850 Hz.

On the whole, the [ʃ]-spectra seem to be shifted somewhat toward lower frequencies, but this information must be taken with reservation, as the differences in environment (viz. preceding unrounded versus rounded vowel) causes quite a dispersion in the total averages of the measurements of each sibilant.

A comparison of [s] and [ʃ] preceded by the same vowel category shows clearer differences, especially after unrounded vowels, but in several cases the differing features seem to be individual. The [s]- and [ʃ]-spectra of subject NRP are so similar that a general distinction between them can hardly be made, even when they are paired according to vowel categories.

There seems, however, to be a slight tendency for the [ʃ]-spectra to have a slightly higher amount of energy concentrated in the lower maximum, a little less in the upper maximum which seems to be shifted a little down the frequency scale (see fig. 5).
Measurements for NRP are:
Frequency range of the dominant part of the spectrum:

    [s] unround. 2000 - 4350 Hz - [ʃ] unround. 1600 - 4650 Hz

    [s] round. 1850 - 4350 Hz   - [ʃ] round. 1550 - 4150 Hz.

Frequency range of the decisive increase of intensity leading up to the lower maximum:

    [s] unround. 2000 - 2850 Hz - [ʃ] unround. 1500 - 2450 Hz

    [s] round. 1850 - 2100 Hz   - [ʃ] round. 1350 - 1950 Hz.

Center frequency of the lower maximum as a whole:

    [s] unround. 3350 Hz - [ʃ] unround. 3050 Hz;

    [s] round. 2900 Hz   - [ʃ] round. 2800 Hz.

For the other two subjects the same differences can be observed, but more pronouncedly. The [ʃ]-spectra contain more energy in the lower part of the spectrum and the increase of intensity leading up to the lower maximum begins at lower frequencies than for [s] (the difference may amount to 1300 Hz), and a general weakening of the upper spectral energy (see figs. 3 and 4). The measurements are as follows:

Frequency range of the decisive increase of intensity leading up to the lower maximum:

    [s] unround. PA: 2400 - 3500 Hz OTH: 2250 - 3450 Hz

    [ʃ] unround. PA: 1600 - 2300 Hz OTH: 1700 - 2650 Hz

    [s] round.   PA: 2100 - 2500 Hz OTH: 2200 - 2600 Hz

    [ʃ] round.   PA: 1550 - 2100 Hz OTH: no measurements.

Center frequency of the lower maximum as a whole:

    [s] unround. PA: 4000 Hz   OTH: 3800 Hz

    [ʃ] unround. PA: 2900 Hz   OTH: 3200 Hz

    [s] round.   PA: 3100 Hz   OTH: 2950 Hz

    [ʃ] round.   PA: 2750 Hz   OTH: no measurements.

It is interesting to notice the previously mentioned partial overlapping between the phonemic categories: Danish [∫] preceded by an unrounded vowel may have a higher center frequency of the lower maximum than Danish [s] preceded by a rounded vowel.

Beside the previously mentioned general difference between [s] and [∫], namely the decrease of upper spectral energy located at lower frequencies in [∫], also the frequency range 1500 - 2500 Hz seems to contain crucial information as to the general differences between [s] and [∫] in Danish.

5.2 Spectral properties of German sibilants

5.2.1 Sibilant preceded by rounded versus unrounded vowel

This problem seems easier to handle for German sibilants. For the spectral shape of [∫] it apparently makes no difference whether [∫] is preceded by a rounded or an unrounded vowel. As mentioned in a previous section, [∫] is frequently articulated with rounded lips, and besides German sibilants have a more distinct articulation than Danish ones and are therefore less affected by coarticulation.

The fact that the German rounded lax vowels are frequently articulated with no pronounced lip rounding also seems to neutralize the influence of coarticulation to a certain degree. One of the subjects (PS) has long tense rounded front vowels in two of the words, and this fact has made his spectra of [s] unround. and [s] round. differ more than those of the other subjects, who articulated a lax [ʏ] in this position (see fig. 7 vs. figs. 8-10). Accordingly, one might expect that long tense vowels preceding the sibilant would bring out a clearer distinction between [s] unround. and [s] round. Moreover, the small differences that can be observed between [s] unround. and [s] round. in this material exhibit individual variation. A varying amount of extra energy is seen in the frequency ranges of "basic peak" no. 4 or 5 (3000 - 4000 Hz), furthermore, coarticulation from a preceding

rounded vowel seems to cause a slight increase of intensity at about 8000 - 9500 Hz, but generally this material showed no consistent distinctive spectral properties. The individual variation is very small; the [ʃ]-spectra, especially, are similar for all the German subjects.

5.2.2   The difference [s] - [ʃ] in German

In contradistinction to Danish the German [s]- and [ʃ]-spectra differ fundamentally without regard to the preceding vowel. All of the [ʃ]-spectra have their main energy in the lower part of the spectrum with a maximum centered at about 3000 Hz, from which one may observe a slight decrease of energy up to "basic peak" no. 7 (about 5800 Hz) and a steeper decrease above that frequency region. All the [s]-spectra have almost no discernible energy in the [ʃ]-maximum, whereas their maximum of energy is concentrated in the central region of the spectrum, in the "basic peaks" nos. 6-7 (4800 - 7600 Hz) (see fig. 11).

5.3   The difference between Danish and German sibilants

Auditorily the Danish sibilants [s] and [ʃ] are normally placed in between German [s] and [ʃ] with regard to the clearness versus darkness of the noise (clear - dark: German [s] Danish [s] Danish [ʃ] German [ʃ]) and, not surprisingly, an overlapping of acoustic distinctions was found only in Danish, not in German, sibilant spectra. The acoustic distinction between on one hand Danish and German [s] and on the other hand Danish and German [ʃ] is rather obvious.

The clearest language-specific acoustic type is represented by the double-peaked Danish sibilant spectra (as for subject NRP, see section 3.2). The tendency to a configuration: lower maximum - zero - upper maximum is most apparent for Danish [s]-spectra. The German [s]-spectra generally show one single maximum near the center of the frequency range. The German [ʃ]-spectra likewise contain a single maximum, but it is shifted downwards and has a more slowly declining upper intensity slope.

Further, the German sibilants show energy at higher fre-
quencies than the Danish ones. The measures of absolute upper
frequency limits are:
Da. [s] 11800 Hz - Ger. [s] 14100 Hz; Da. [ʃ] 10800 Hz - Ger.
[ʃ] 12900 Hz.

In the lower part of the spectrum the decisive increase of
energy of German [s] starts at about 2500 Hz higher than for
Danish [s], that is somewhere near the zero between the two
maxima of the Danish average spectra. The absolute lower fre-
quency limit differs only slightly for Danish and German [s]-
spectra. German [ʃ]-spectra show energy at lower frequency
locations than do Danish [ʃ]-spectra. Values for absolute lower
frequency limits are:
Ger. [ʃ] 950 Hz - Da. [ʃ] 1300 Hz.
Frequency locations of points above which there is a decisive
increase of lower spectral energy:
Ger. [ʃ] 1050 - 1800 Hz - Da. [ʃ] 1600 - 2250 Hz.

5.4 General differences among sibilant spectra

5.4.1 Spectra of sibilants preceded by rounded versus unrounded vowels

Generalizations of common acoustic properties of [s]- and
[ʃ]-spectra naturally have to be made with strong reservations.
Still, there are some differences between sibilants preceded by
rounded versus unrounded vowel which are common to Danish and
German and might reflect a general tendency. For both languages
a rounded vowel before the sibilant (if the sibilant itself is
not characterized by obvious liprounding) causes an amplification
of the energy immediately below the dominant energy maximum of
the sibilant spectrum. This may be due to the fact that lip-
rounding causes a lowering of the formant frequencies (the poles
of the transfer function).

The difference rounded - unrounded vowel before the sibilant seems to affect [s] more than [ʃ]. (As previously mentioned, German [ʃ] does not seem to be affected at all.) An explanation of this may be the fact that the influence of rounding on formant frequencies depends on the place of articulation (constriction of the resonator). F4 is thus more affected by rounding if the primary constriction is situated in the front of the mouth (as in [s]) and less affected if the primary constriction is made somewhat further back (as in [ʃ]). The place of articulation of [ʃ] may vary, but the center of the constriction will always be situated further back than that of [s].

5.4.2 General differences among [s]- and [ʃ]-spectra

Even here the tendencies must be stated with strong reservations. In the present material two features were observed that separate [s] from [ʃ] acoustically: [ʃ] had stronger energy in the lower part of the spectrum relative to [s] and a steeper decline of energy in the upper part. The strong reservation to the generalizations is due to the large individual variation among the Danish subjects; for instance, the spectra of [s] and [ʃ] unround. by the subject NRP showed acoustically greater similarity than did the spectra within the same class produced by the other Danish subjects.

References

Harris, K.S. 1958: "Cues for the discrimination of American English fricatives in spoken syllables", LS 1, p. 1-7

Heinz, J.M. and K.N. Stevens 1961: "On the properties of voiceless fricative consonants", JASA 33, p. 589-596

Hughes, G.W. and M. Halle 1956: "Spectral properties of fricative consonants", JASA 28, p. 303-310

Jassem, Wiktor 1962: "The formant patterns of fricative consonants", STL-QPSR 3, p. 6-16

Jassem, Wiktor 1962: "Noise spectra of Swedish, English, and Polish fricatives", Speech Communication Seminar, Stockholm 1962, and IVth international Congress of Acoustics, Copenhagen 1962

Jassem, Wiktor 1968: "Acoustical description of voiceless fricative consonants in term of special parameters", Speech Analysis and Synthesis (W. Jassem, ed.), (Warsaw), p. 189-206

Katarzyna, Kudela 1968: "Spectral analysis of Polish fricative consonants", Speech Analysis and Synthesis (W. Jassem, ed.) (Warsaw), p. 93-188

Martony, J. 1962: "On the perception of Swedish voiceless fricatives", STL-QPSR 2, p. 25-28

Martony, J. 1962: "On the synthesis and perception of voiceless fricatives", STL-QPSR 1, p. 17-23

Martony, J. 1962: "Some experiments on perceptual cues for Swedish fricatives", Speech Communication Seminar, Stockholm 1962

Martony, Cederlund, Liljencrants, and Lindblom 1961: "On the analysis and synthesis of vowels and fricatives", Proc. Phon. 4, p. 208-213

Strevens, Peter 1960: "Spectra of fricative noise in human speech", LS 3, 1, p. 32-49

IDENTIFICATION AND DISCRIMINATION OF VOWEL DURATION

Niels Reinholt Petersen

Abstract: The identification and discrimination of vowel duration
        was investigated. The experimental results could not
        be unambiguously interpreted in favour of either cate-
        gorial or continuous perception of the acoustic variable
        under study.
        Furthermore, the subjects' response bias in the dis-
        crimination test was examined. It turned out that the
        number of "false alarms" (i.e. "different" responses to
        pairs of physically identical stimuli) varied in a
        systematic manner, being considerably higher near the
        phoneme boundary (as established by identification tests)
        than within the phoneme areas. It is attempted to ex-
        plain the systematic variation of response bias in terms
        of Fujisaki and Kawashima's model of the decision pro-
        cess in discrimination tasks.

1.   Introduction

It is commonly reported that consonants and vowels are per-
ceived differently by human listeners. In a great number of ex-
periments on the identification and discrimination of speech
sounds it has been demonstrated that stop consonants in particular
are perceived in a categorial manner, i.e. listeners can discrimi-
nate between different sounds only if they can identify them as
belonging to different phonetic (or phonemic) categories. Vowels,
on the other hand, seem to be perceived in a more continuous manner,
similar to that of non-speech stimuli, i.e. listeners can discri-
minate also between sounds which are identified as belonging to
the same category.

Whereas the categorial perception of consonants is fairly
well established (see e.g. Abramson and Lisker 1968, Liberman,
Harris, Hoffman, and Griffith 1957, Liberman, Harris, Kinney, and

Lane 1961, Liberman, Harris, Eimas, Lisker, and Bastian 1961, Pisoni 1971), the perception of vowels seems to be more susceptible to influence from experimental conditions. If vowels are presented in CVC syllables (Stevens 1968) or are followed by another vowel (Fujisaki and Kawashima 1968, Pisoni 1973) they tend to be perceived more categorially than isolated vowels. Similarly, short vowels seem to be less continuously perceived than long vowels (Pisoni 1971 and 1973, Fujisaki and Kawashima 1968).[1]

If the differences between the vowels to be compared are very small, this also seems to evoke a more categorial perception of the vowels (Holtse 1973).

With respect to the influence of test format Pisoni (1971 and 1973), who compared the commonly used ABX procedure to a four interval test of paired similarity (4IAX), found that the perception of vowels was more continuous with the former procedure than with the latter one, whereas the two procedures could not be shown to influence the perception of stop consonants differently.

The different modes of perception of vowels and consonants and the dependence of the perception of vowels on experimental conditions may be accounted for in terms of a model for the discrimination process developed by Fujisaki and Kawashima (here quoted from Pisoni 1971). According to their model the acoustic signal is converted to an auditory representation and a categorized representation in short term memory. When the listener is to decide whether the two stimuli in a pair are same or different he first compares the categorized representations. If they are different, i.e. belonging to different phonetic categories, he answers "different" right away. If they are same, i.e. belonging to the same category, he turns to the auditory representations and compares those. His answer will then depend on the extent to which the auditory representations are preserved in short term memory.

In the case of stop consonants it seems that their auditory representations are lost from short term memory as soon as the

---

[1] However, from the discrimination curves given by these authors the tendency towards categorial perception does not appear to be very marked.

categorization has taken place, leaving only categorial informa-
tion for the discrimination process. Auditory information on
vowels, on the other hand, may or may not be available in the
discrimination process, depending on experimental conditions:
in some cases vowels seem to be perceived in a nearly continuous
manner and in other cases in a more categorial manner.

The bulk of experiments on the perception of vowels has
dealt with vowel quality, whereas the duration of vowels has been
investigated to a very limited extent. Bastian and Abramson
(1962) report briefly that Thai vowels of different duration
presented to Thai and American listeners were perceived in a
continuous fashion. The opposite is reported for Japanese by
Fujisaki, Nakamura, and Imoto (1973), who found a clear discri-
mination peak at the boundary between short and long o̱.

The results of previous experiments carried out by the
present author (Reinholt Petersen 1974) point in the same direc-
tion. In those experiments subjects adjusted the duration of the
vowels i̱ and ɛ̱ in the synthetic words ['lV(:)sə] in accordance
with the following instructions: a) adjust a "natural duration"
of a phonemically short vowel, b) adjust a "natural duration" of
a phonemically long vowel, c) reduce the duration of a long vowel
until a phonemically short vowel is heard, and d) increase the
duration of a short vowel until a phonemically long vowel is heard.

The adjustments made around the phoneme boundaries (i.e.
instructions c) and d)) seemed to show a statistical variation
which was smaller than the variation of the adjustments of "natural
vowel duration". Although this finding may be taken to speak in
favour of categorial perception of vowel duration, the evidence is,
at best, indirect. It was considered desirable, therefore, to
seek further information about the perception of vowel duration by
repeating the earlier experiments in the field.

2. Stimuli and procedure

2.1 Stimuli

The vowels chosen were Danish /ɛ/ and /ɛ:/ (IPA [ɛ̞]). In
order to avoid the stimuli to be perceived as non-speech sounds
the vowels were embedded in the frame ['ɪ_sə]. Since the dif-
ferences between the formant frequencies of short and long ɛ
are very small in normal spoken Danish (see e.g. Fischer-
Jørgensen 1972), the Danish words /'ɪɛsə/ læsse 'to load' and
/'ɪɛ:sə/ læse 'to read' could be generated by merely varying the
duration of the vowel.

The words were synthesized on the parallel speech synthe-
sizer of the Institute (see Rischel 1969 and Rischel and Lystlund
1972). The acoustic structure of the synthetic words were based
on spectrograms of four Danish speakers. Information about the
formant frequencies and levels of the vowel ɛ was kindly provided
by Peter Holtse, and correspond to formant data which rendered
100% /ɛ/-identification in an identification test carried out by
him (Holtse 1973). The fundamental frequency was rising through-
out the stimulus word at a rate of 4 Hz per 100 ms, starting at
92 Hz at the beginning of the [ɪ].

The duration of the vowel was defined as the distance be-
tween the programmed change of parameter values from [ɪ] to [ɛ],
i.e. at the start of the consonant-vowel transitions, and the
programmed cessation of periodic energy in the vowel.

In the course of the experiments it turned out that the
actual value of vowel duration may deviate from that specified by
the function generator of the speech synthesizer, even if the
latter performs correctly. This can be explained as an inherent
source of error due to the fashion in which the voice source
amplitude is controlled: the point in time at which the sound is
programmed to start or a change in amplitude is to occur, i.e.
the point at which the voice source amplitude gate is activated
to produce an increasing amplitude, is independent of the repeti-
tion rate of the voice source pulses. Since the voice source

amplitude gate is placed before the formant filters, the first pulse to produce an appreciable excitation of the formant filters or an amplitude change may occur with varying time lags relative to the programmed onset or change, i.e. signals differing up to one period in duration (e.g. 10 ms at F_o = 100 Hz) can be generated by exactly the same programming of the function generator. There is a similar possibility of error at the end of a synthesized vowel.

In the stimuli used in the tests the rise in amplitude from the [ǀ] to the vowel and the fall at the end of the vowel before [s] were generated partly by the voice source amplitude gate and partly by the formant amplitude gates. Since the latter are placed after the formant filters they may, to some extent, reduce the durational errors mentioned above.

In order to investigate the possible distortion of the identification and (particularly) discrimination results the durations of all stimuli were measured on mingographic tracings and compared with the discrimination results. The deviations from the programmed durations were generally very small, less than 2 ms, and they could not be shown to have influenced the shape of the discrimination functions.

2.2   Test procedure

2.2.1   The identification test

In the identification test the vowel duration was varied in 10 ms steps in the range from 100 to 200 ms. An informal pilot test taken by the same subjects who were to participate in the subsequent more formal tests showed that the phonemic boundary between /ɛ/ and /ɛ:/ in all cases was well within the limits of the range chosen.

The stimuli were arranged in ten different random orders on the test tape, and were presented one at a time with a response pause of 5 secs. between consecutive stimuli. The test was played

back to the listeners via earphones (Sennheiser HD414) on semi-professional REVOX tape recorders. The listeners were asked to note on answer sheets whether they heard the word læsse (short vowel) or the word læse (long vowel). Each listener took the test twice, so that 20 answers were obtained for each stimulus value. There were six subjects, who were all phoneticians. One (NT) was a staff member of the Institute, the others (BH, EBC, MJ, PA, and JJ) were university students of phonetics.

The phoneme boundaries were computed for the subjects individually. The phoneme boundaries were defined as the arithmetic means of the cumulative distributions described by the identification functions (see Guilford 1954 p. 120ff). In contrast to the more commonly used linear interpolation between the data points adjacent to the point of 50 per cent identification this method takes advantage of all information available in the identification functions and in addition, it permits the standard deviation to be computed as a measure of the statistic reliability of the identification results.

2.2.2 The discrimination test

In a discrimination test like the present, where the discrimination is to be measured at various points along a physical continuum, the physical differences between stimuli to be compared should correspond to perceptually equal differences throughout that continuum. Furthermore, they should be of an order of magnitude approximately equal to the differential threshold for the variable in question. If the differences are too small or too large relative to the threshold they may fail to reveal possible discriminatory discontinuities along the continuum.

A survey of the literature on differential thresholds for sound duration did not contribute to the determination of differences between the stimuli in the present experiment, partly because most of the research in the field has been carried out on pure tones or isolated vowels, and partly because of a very great dispersion of the results.

Therefore, instead of attempting to conjecture some function describing the relation between physical and perceptual differences, it was decided to have physically equal steps between the stimuli throughout the 100-200 ms range, and then later to determine the step values in a pilot experiment. (As a matter of fact, this decision caused troubles later on, when the discrimination results were to be interpreted. However, an attempt to keep equal perceptual differences along the continuum might also have caused problems, problems which might not reveal themselves as such in the experimental results.)

In the pilot discrimination test only one reference duration, viz. 125 ms, was used. That duration was compared with test durations of 125 ms (control), 130 ms, 135 ms, 140 ms, and 145 ms.[1] The scores across the five listeners participating (i.e. those mentioned in section 2.2.1 above, minus JJ) appeared to be slightly above chance when the difference between the vowels was 5 ms, and almost 100 per cent correct when the difference was 20 ms. Intermediate score values were obtained for 10 ms and 15 ms discriminations, the scores of the latter being slightly higher than those of the former.

Judging from the results of the pilot experiment, an appropriate difference between stimuli seems to be 10 ms. Since it was considered undesirable to have only one value of interstimulus difference, and since, furthermore, differences of 20 ms might be too large (cp. above), it was decided to include 15 ms interstimulus differences in the discrimination test in addition to the 10 ms ones. Thus vowels of the durations 100, 110, ..., 190 ms were compared to vowels of the same durations (control condition) and to vowels being 10 ms and 15 ms longer.

---

1) The reference duration of 125 ms was chosen because this duration and the ones it was compared with were thought to fall in the vicinity of the phoneme boundary, i.e. in a duration range where discrimination may be expected to be most acute.

The stimuli were arranged for discrimination in a so-called AX test, where they are presented pairwise to the listeners, who are to judge whether the two stimuli in a pair are same or different. The reason for choosing this test format rather than the more frequent ABX format was that the latter procedure seems to make heavy demands on short term memory (see e.g. Harris 1952 and Pisoni 1971), so that what is actually measured is memory rather than discrimination.

On the test tape the 30 AX pairs, viz. 10 AA pairs, 10 AB_{10} pairs (i.e. 10 ms difference between A and B), and 10 AB_{15} pairs (i.e. 15 ms difference between A and B), were arranged in four different random orders. Each AX pair was presented three times in succession followed by a 5 secs. pause for response. Each listener took the test three times, thus giving 12 responses to each of the discrimination pairs.

The subjects, the same as those who took the identification test, were asked to rate the certainty of their responses by choosing between four response categories: (f+) 'I am certain they are different', (f) 'I am uncertain, but I believe they are different', (e) 'I am uncertain, but I believe they are same', and (e+) 'I am certain they are same'.[1] As the test was rather long (about 35 minutes) the listeners were instructed to take a break of at least five minutes in the middle. The initial five stimuli on the test tape were "dummies".

For each discrimination pair the area under the ROC-curve, P(A), was computed as a measure of the discriminability at that point in the continuum. The curve relates the probability of a given response (f+, f, e, e+) to AB pairs to the probability of that same response to the corresponding AA (control) pairs. Since correct discriminations of AB pairs and "false alarms" (i.e. the responses "different" to control pairs) alike are taken into account in the computation of the P(A) value the discrimination can be estimated independently of the listeners' response bias

---

1) The letter f stands for Danish forskellig 'different', and e stands for ens 'same'.

(see e.g. McNicol 1973 and Robinson and Wattson 1972).[1] Due to the limited number of responses the non-parametric measure P(A) was used rather than the parametric measure d' or $\triangle$ m.

3. Identification results

Identification functions for short and long ε are shown in fig. 1, and the phoneme boundaries and corresponding standard deviations are given in table 1. The listeners, except for NT, and perhaps JJ, seem to agree quite well on the position of the boundary. With respect to NT, the high value of her phoneme boundary is in agreement with her results in other experiments on the perception of vowel duration (see Reinholt Petersen 1974).

4. Discrimination results

The discrimination functions, i.e. P(A) as a function of vowel duration, for the six listeners are shown in fig. 2. With all subjects there is a pronounced tendency for P(A) to decline as a function of vowel duration. This is, undoubtedly, caused by the physical equidistance between stimuli throughout the continuum. This equidistance, however, cannot fully explain the shapes of the curves. If it were so, monotonically declining functions should be expected, whereas the discrimination functions obtained here deviate from monotonicity by having discrimination peaks in the range 110 to 140 ms, and by showing tendencies to rising P(A) values at the longest vowel durations.

---

1) As a matter of fact, P(A) is not entirely independent of response bias. The P(A) turns out to be slightly lower if the subject is strongly biased towards either extreme of the rating scale than if he distributes his responses more evenly along the scale (see Robinson and Wattson 1972, p. 125).
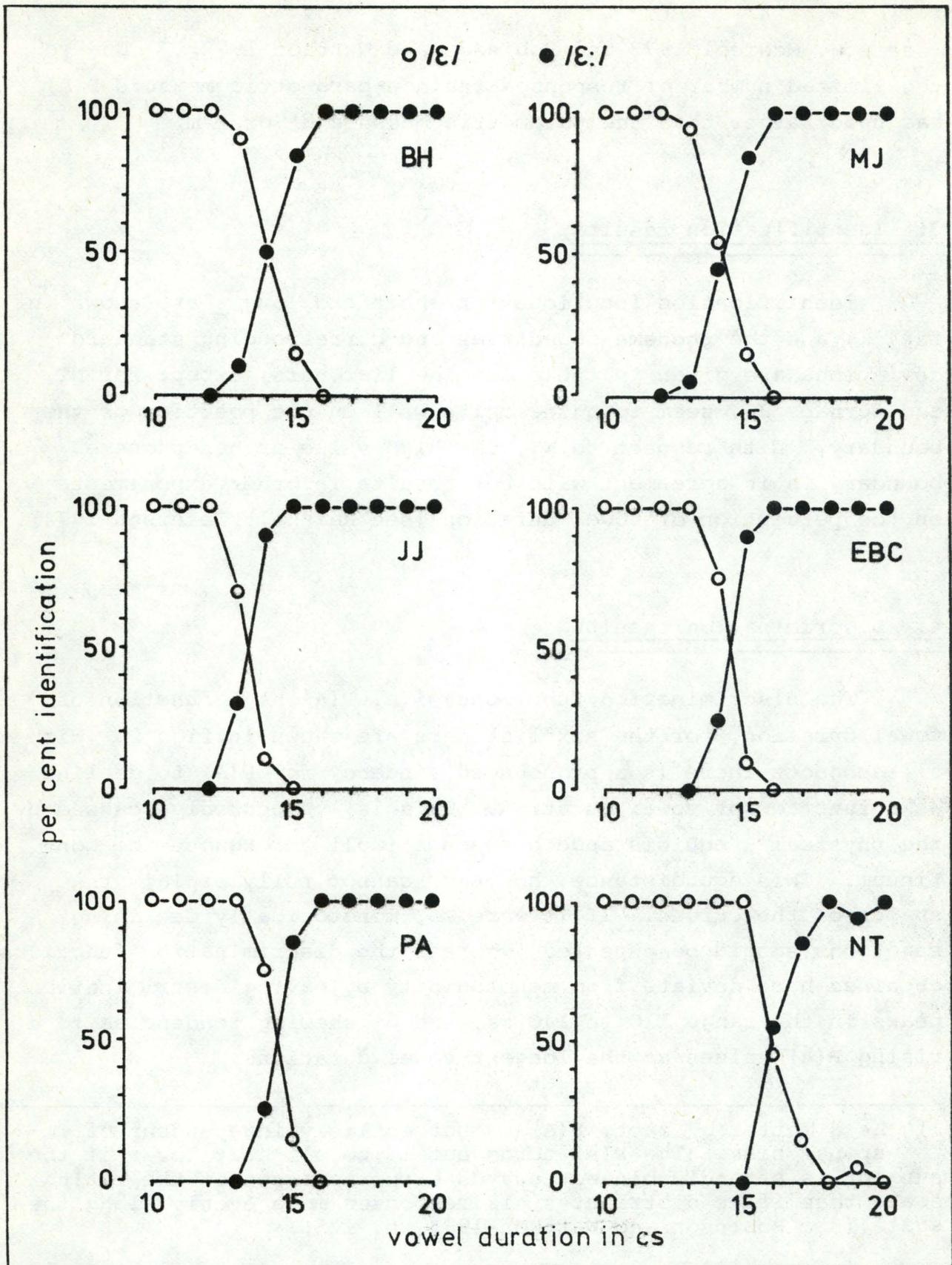
Figure 1

Identification functions for the six listeners.

Table 1

Boundary between /ɛ/ and /ɛ:/ and
corresponding standard deviation
for each of the six listeners.

subject	boundary	standard deviation
BH	140.5	8.65
JJ	133.0	6.00
PA	144.0	6.41
MJ	141.5	7.92
EBC	143.5	5.72
NT	161.5	10.70

Figure 2

Discrimination curves (P(A) as a function of vowel duration) for the six listeners. The vertical lines in the graphs indicate the phoneme boundaries.

The peaks in the 110-140 ms range may be associated with the boundary between short and long vowel. (This matter will be treated more thoroughly in section 5 below.)

The upward slopes of the discrimination curves at the longest vowel durations may be explained by the possible existence of a third durational category, "unnaturally long". A similar phenomenon has been reported by Liberman, Ha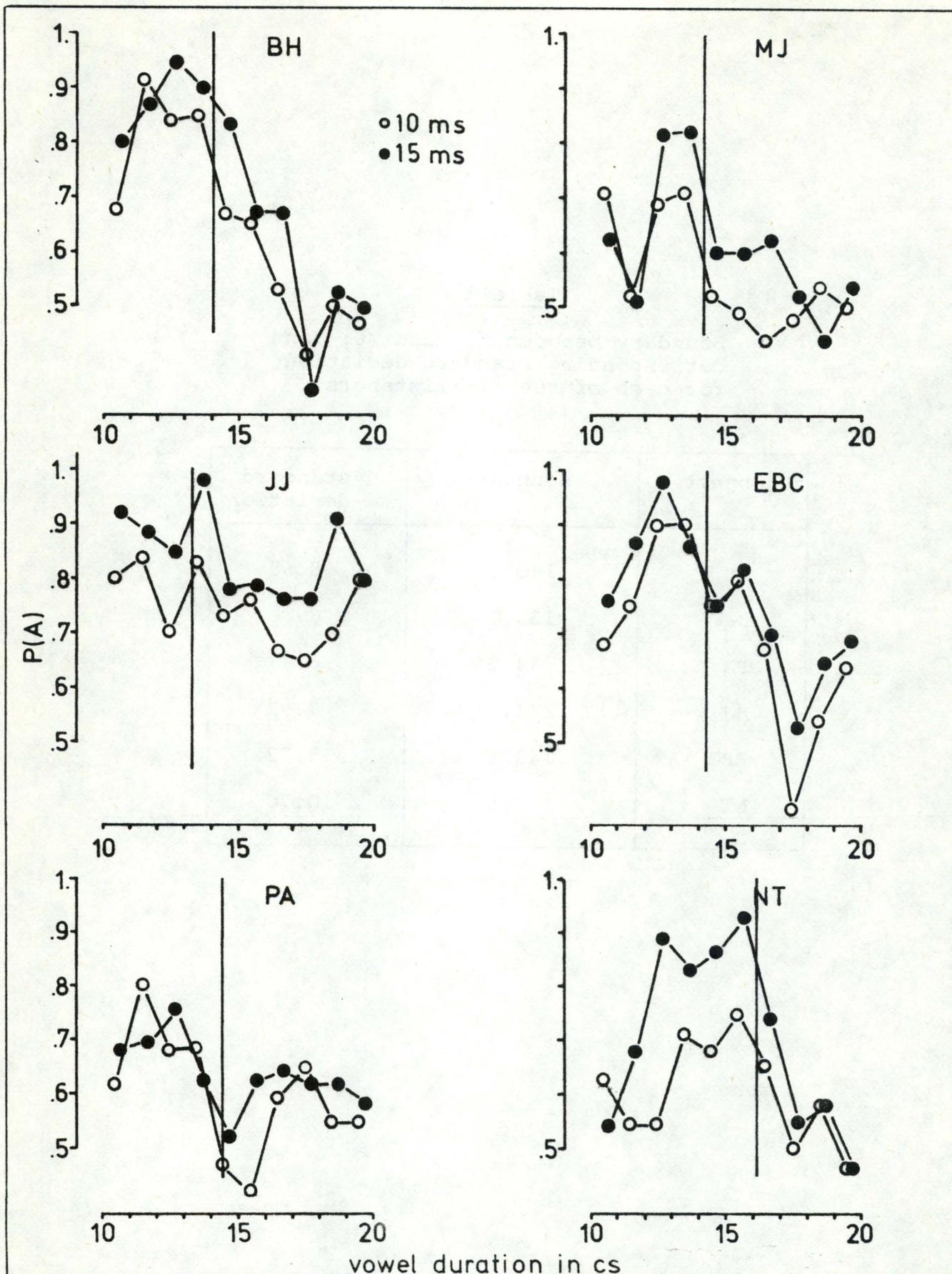rris, Eimas, Lisker, and Bastian (1961) for very long p-closures. In order to be tenable, this explanation presupposes, apart from vowel duration being perceived in a categorial fashion, that the range of durations from 150 to 170 ms corresponds to the range of what may be termed natural long vowels. However, in a previous experiment (Reinholt Petersen 1974) one of the tasks of the subjects (the same as those taking part in the present experiments) was to adjust the duration of the vowel /ɛ:/ in the word læse in accordance with a criterion of "naturalness". It appeared that the subjects preferred durations around and above 200 ms for a natural long vowel. On the other hand, the perception of vowel duration is presumably quite sensitive to external factors, such as experimental design, so the results referred to should not be considered to preclude with certainty the possibility of a third durational category.

5. Comparison between identification and discrimination

In the discrimination functions of fig. 2 the phoneme boundaries computed from the identification scores are indicated by vertical lines. In all cases (except for JJ's 15 ms discriminations) the discrimination peaks (defined as the highest P(A)) are found at lower stimulus values than the phoneme boundaries. It is not unusual, however, for discrimination peaks not to correspond exactly to the phoneme boundaries. But if the peaks are to represent a more acute discrimination across the boundary, a

random distribution of peaks around the boundary should be expected. With the present results a one-tailed t-test ($H_0: \mu_{boundary} - \mu_{peak} \leq 0$) revealed the displacement of peaks towards lower stimulus values to be systematic ($p < .01$ for 10 ms discriminations and $p < .05$ for 15 ms discriminations).

Two explanations may be hypothesized to account for the displacement of the discrimination peaks relative to the phoneme boundaries:

1) The displacement is due to the use of physically equal differences between stimuli. The discrimination functions obtained may be the result of two factors, viz. an "equidistance effect" which causes the discrimination to decline as a function of stimulus duration, and a "phoneme boundary effect" which causes the discrimination to be sharpest at the phoneme boundary and to decline as a function of the distance to the boundary. Thus, in the range of durations beneath the phoneme boundary the two effects will counteract (the "equidistance effect" causing a falling and the "phoneme boundary effect" a rising discrimination as a function of vowel duration in that range) with the result that they may more or less neutralize each other, whereas at stimulus values above the boundary the two effects will act together to produce a steep decline in the discrimination functions.

If this is tenable in principle, the probability of the highest P(A) occurring at stimulus values below the phoneme boundary will exceed the probability of its occurrence at values above the boundary, i.e. at the steeply declining slope.

On the basis of the obtained discrimination curves it is, of course, not possible to evaluate the contributions of the two effects separately so as to obtain an estimate of the linguistically relevant information in the curves. As it seems pretty certain, however, that the "equidistance effect" will cause a monotonic decline of P(A) values, a normalization of the discrimination curves based on such a monotonic function might give a rough indication of the contribution from the "phoneme boundary

effect", the greatest deviations from the function chosen assumedly reflecting the strongest phoneme boundary effect.

The function used for normalization was a straight line fitted to the data points of the discrimination functions by the least squares method. In fig. 3 the normalized curves are shown for 10 and 15 ms discriminations for the individual subjects. It must be emphasized that this normalization procedure is a very crude one, indeed, and theoretically unsatisfactory, too. But, supposedly, it gives at least a hint of the effect to be expected from more sophisticated procedures, preferably to be based on independent data on the influence from equal physical differences.

From fig. 3 it appears that the peaks are still located at stimulus values beneath the phoneme boundaries (indicated by vertical lines in the graphs). A one-tailed t-test $(H_0: \mu_{boundary} - \mu_{peak} \leq 0)$ proved the mean difference between boundaries and peaks to be statistically significant for the 10 ms discriminations $(p < .05)$ and non-significant for the 15 ms discriminations. Since in the non-normalized functions the differences were significant in both cases $(p < .01$ for 10 ms and $p < .05$ for the 15 ms discriminations) it can be tentatively concluded that a more sophisticated normalization procedure might further reduce or perhaps eliminate the displacement of the discrimination peaks relative to the phoneme boundaries.

2) Another reason for the displacement might be the difference in experimental design between the discrimination test and the identification test. In the discrimination test the stimuli in a pair were separated by .7 secs. of silence and the pause between repetitions of the pair was 1.7 secs., while in the identification test stimuli were presented one at a time, separated by the 5 secs. response pause. The relatively rapid repetition of stimuli in the discrimination test may have caused a shift of the perceptual phoneme boundaries towards lower stimulus values. In order to test this hypothesis an additional
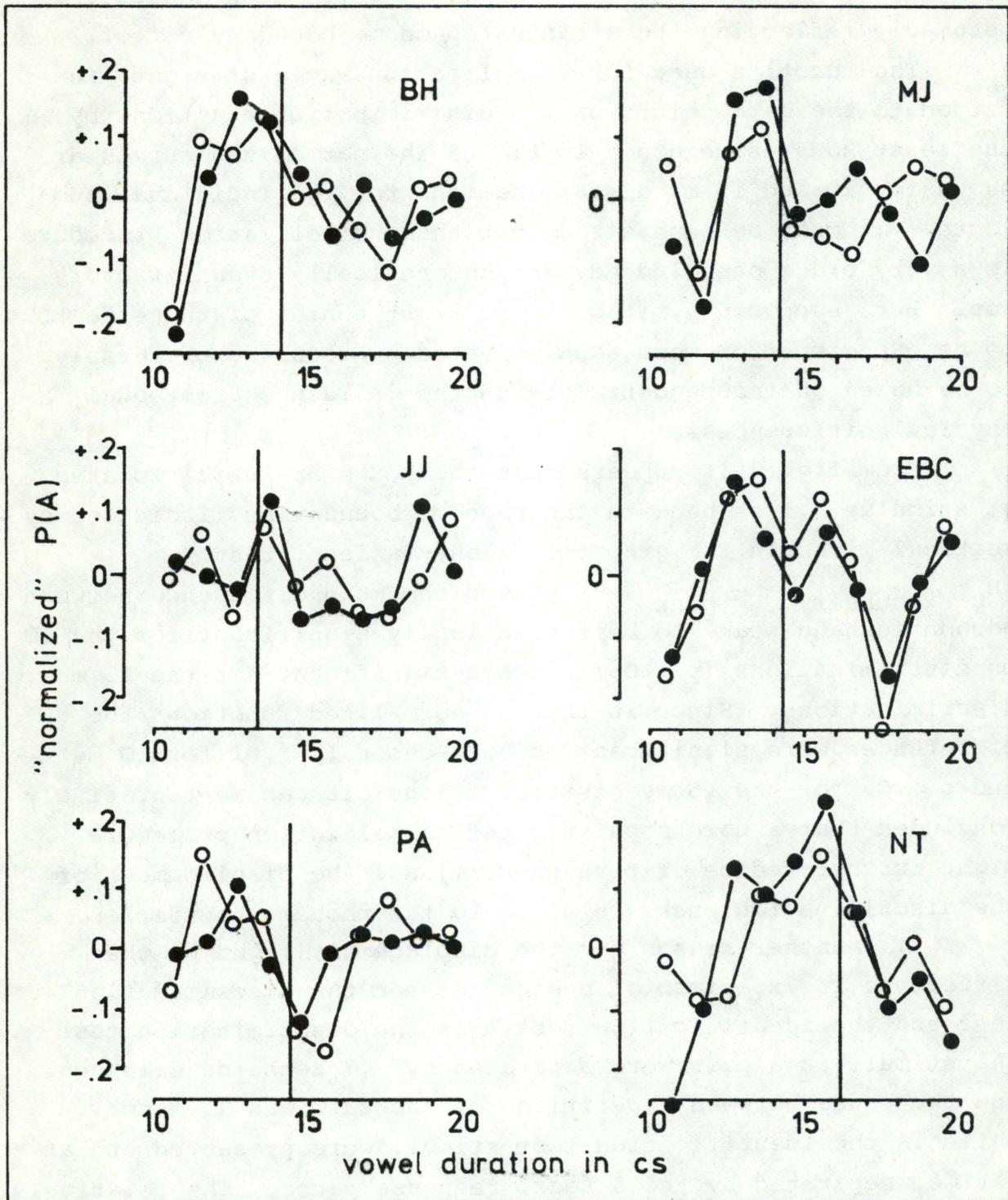
Figure 3

"Normalized" discrimination curves for the six listeners. The vertical lines indicate the phoneme boundaries. For further explanations, see the text.

listening test was arranged, in which three subjects (BH, JJ, and NT) were to identify the stimuli of the discrimination tape.

The results of the supplementary identification test did not, however, conform very well with predictions of the hypothesis, namely that the phoneme boundaries should be shifted to stimulus values lower than those of the original identification test. The listeners BH and JJ showed statistically significant shifts ($p < .05$) of 5 ms, but in opposite directions; for NT the phoneme boundary shift (1.5 ms in the direction contrary to the one predicted) could not be proved significant ($p > .05$).

If these results are valid for the other subjects as well the displacement of peaks in the discrimination test probably cannot be ascribed to the different manners of presentation in the discrimination test and the identification test.

6. Investigation of response bias

It was mentioned in section 2.2.2 above that the P(A) as a measure of discrimination is largely unaffected by the listener's response bias. However, information about response bias may contribute to the description of the listeners' perceptual behaviour. It was therefore considered desirable to collect such information, particularly with a view to revealing possible variations in response bias along the stimulus continuum.

The measure of response bias employed was the "false alarm" probability, i.e. the probability of obtaining an "f+" or "f" response to an AA stimulus pair (henceforth referred to as P("D"/s)).

In fig. 4 the computed P("D"/s) values are shown as a function of vowel duration. It is evident that the response bias varies along the continuum, the variation being more pronounced with the subjects JJ, PA, MJ, and EBC, and less pronounced with BH and NT. Moreover, the variation seems to be systematic in
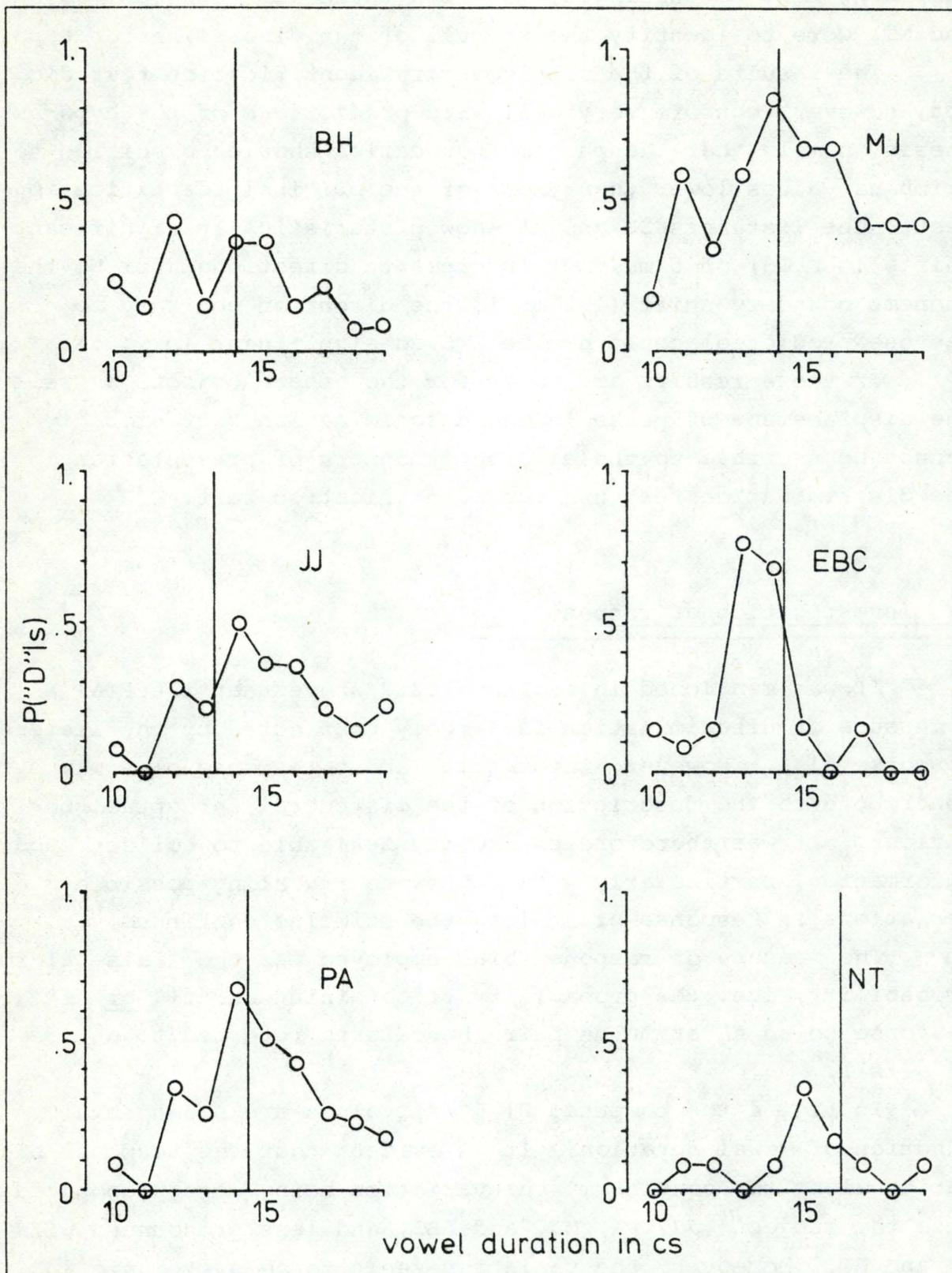
Figure 4

P("D"/s) as a function of vowel duration. The
vertical lines indicate phoneme boundaries.

the sense that P("D"/s) is generally low at the marginal stimulus values and high at stimulus values in the vicinity of the phoneme boundary (indicated by vertical lines in the graphs). Only one subject, BH, does not conform to this pattern.

The question is whether the systematic variation of response bias is part of the subjects' perceptual behaviour or whether it is simply an artifact introduced by the stimulus inaccuracies described in section 2.1 above: Measurements revealed that the vowels in the AA pairs were not exactly equally long in all cases. However, the differences between stimuli in the AA pairs were rather small (between 1.0 and 1.5 ms at all stimulus values except the 190-190 ms pair, where the difference was 3 ms), and in order to explain the P("D"/s) variation the differences between stimuli in AA pairs should have been large in the range 130 to 150 ms and small at the margins of the continuum. As this was not the case the systematic variation in response bias can hardly be an experimental artifact due to stimulus inaccuracies.

If the variation is part of the subjects' perceptual behaviour a similar systematic variation may be expected to occur in the results of other identification/discrimination experiments. The only data available to me in a form permitting P("D"/s) to be computed were the raw data from a series of experiments on identification and discrimination of vowel quality in Danish carried out by Peter Holtse (Holtse 1973). The format of Holtse's experiments was essentially identical to the design of the present tests, and four of his subjects (BH, MJ, EBC, and NT) also served as listeners in the present experiment. Thus, Holtse's material seems very well suited for a comparison.

The P("D"/s) values computed from Holtse's data are shown in figs. 5 and 6 as a function of stimulus number. In this case P("D"/s) indicates the probability of obtaining a "++" or "+" response (corresponding to "f+" and "f", respectively) to an AA stimulus pair. It appears that variations in response bias along

Figure 5

P("D"/s) as a function of stimulus number for two
listeners in a discrimination test on vowel quality.
The vertical lines indicate the phoneme boundaries.
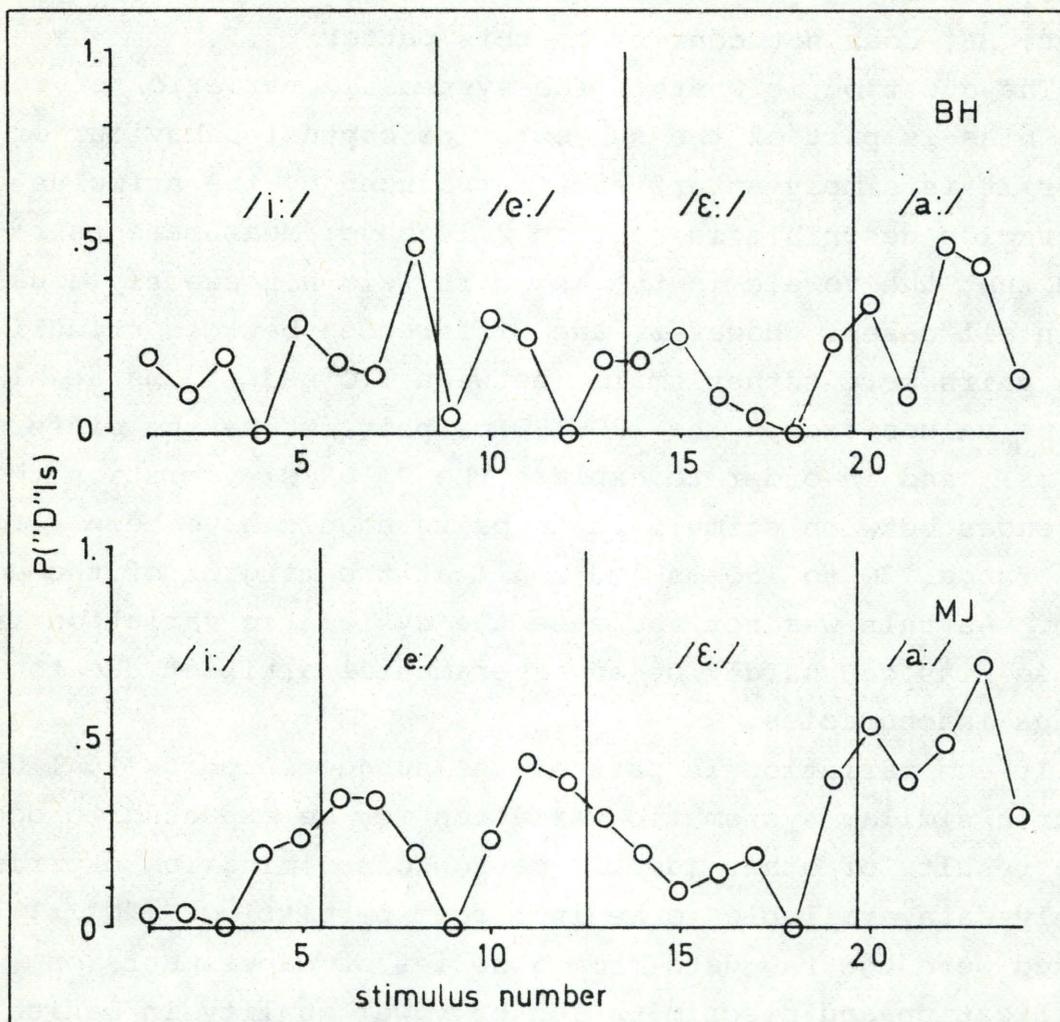
Figure 6

P("D"/s) as a function of stimulus number for two
listeners in a discrimination test on vowel quality.
The vertical lines indicate the phoneme boundaries.
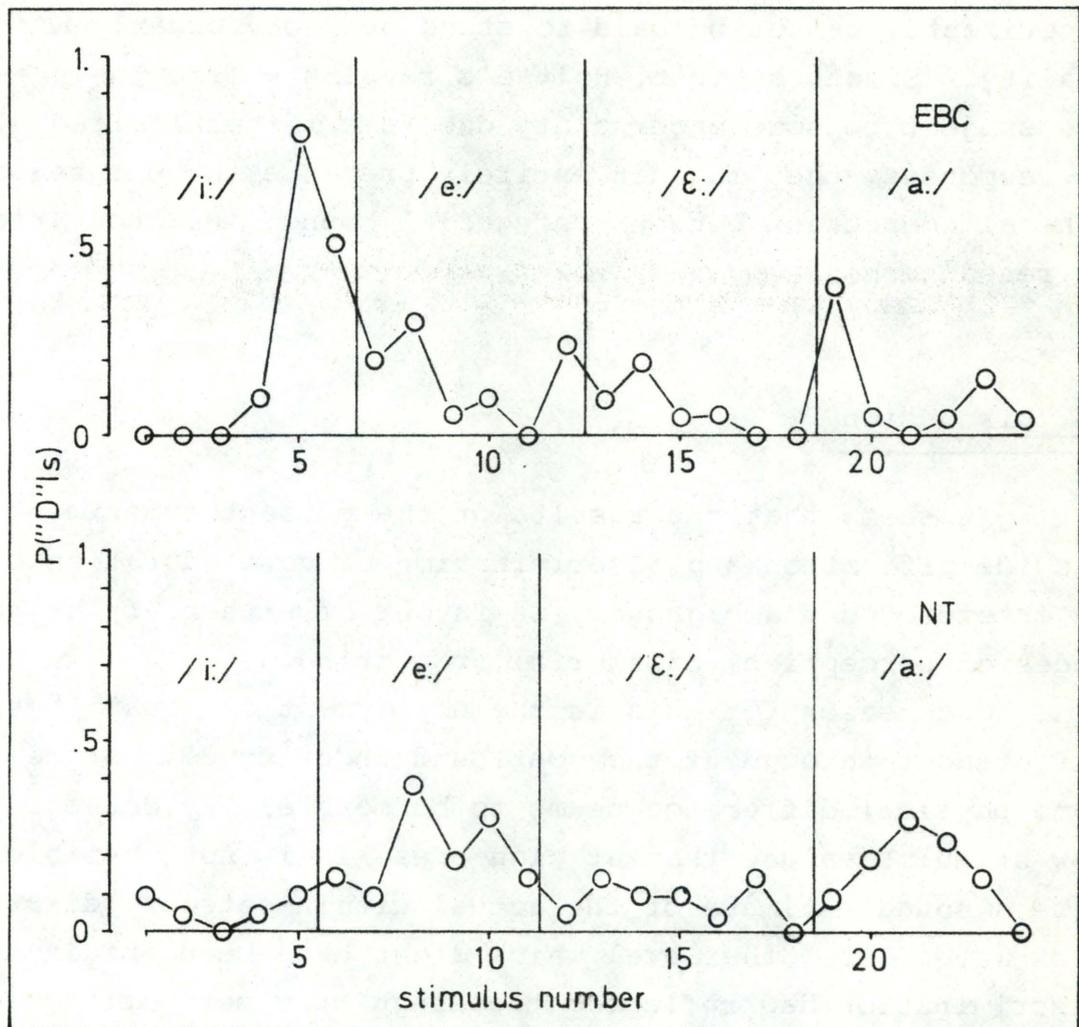
the continuum are present also when the variable discriminated
is vowel quality. But the strong tendency for the response bias
to vary systematically, which was found in the vowel duration
experiments, cannot be said to stand out very clearly with vowel
quality. Since, however, Holtse's results - like the present -
are subject to some uncertainty due to a rather limited number
of responses, they may not entirely preclude the existence of a
general connection between categorial boundaries and variations
of response bias.

7. Discussion

It seems that the results of the present experiments on
the identification and discrimination of vowel duration cannot
be interpreted unambiguously in favour of either of the two
modes of perception, categorial or continuous.

One reason for this is the employment of equal physical
differences throughout the continuum under study. Since the
same physical difference seems to be more easily detected at
low stimulus values than at high ones, it is not possible to
give a sound estimate of the actual within-category discrimina-
tion level, i.e. the level which might have been obtained if the
discrimination had reflected the linguistic perception alone.
This is most unfortunate, because, as pointed out by Pisoni
(1971), it is the discrimination level within categories rather
than between categories which serves as an indicator of categorial
versus continuous perception.

The evaluation of the perception of a given acoustic variable
in terms of the results of the identification/discrimination test
paradigm rests on the assumption that peaks in the discrimination
functions do reflect the between-category discrimination level.
In the present experiments it was found that the peaks did not
coincide with the phoneme boundaries established in the identifi-
cation test, but were displaced towards lower stimulus values.

The displacement could not be attributed to stimulus inaccuracies or to the different manners of stimulus presentation in the two tests. However, an attempt to compensate for the influence of equal physical differences between stimuli seemed to reduce the degree of peak displacement relative to the phoneme boundaries.

If, on this basis, the discrimination peaks can be assumed to reflect the between-category discrimination level, it may be concluded that vowel duration is not perceived in a continuous manner. On the other hand, the lack of a reliable measure of the actual within-category discrimination level prevents an estimation of the degree to which the perception of vowel duration approaches the categorial perception. It may be said with some certainty, however, that vowel duration is not perceived in the same nearly-categorial manner as stop consonants. Were that the case one might have expected a less pronounced influence from physical equidistance between stimuli, since the perception of stop consonants apparently is highly resistant to the influence of experimental conditions.

Apart from the analysis of vowel discrimination as such, the subjects' response bias was examined. It turned out that the listeners had, in a number of cases, responded "different" to AA stimulus pairs. That is, of course, not very remarkable. A number of false alarms should always be expected. What is more interesting was the tendency for the false alarm probability, P("D"/s), to be considerably higher near the phoneme boundaries than within the phoneme areas. This is surprising; one might expect essentially the same rate of false alarms at all stimulus values.

The systematic variation of P("D"/s) may perhaps be explained in terms of Fujisaki and Kawashima's model of the decision process in discrimination tasks (see section 1 above). According to their model the acoustic signal is translated into an auditory representation and a categorized representation in short term memory.

Now, if we expect P("D"/s) to be invariant along the continuum the underlying assumption is that the two stimuli in AA pairs are always given identical categorized representations in short term memory. This assumption, however, is doubtful: Errors of categorization (i.e. the two AA stimuli being categorized as different) may occur, and in such cases the P("D"/s) will increase. This can be illustrated by an example:

Let us assume first that the AA stimuli are never categorized as different. Under these circumstances comparison of categorized representations will never provoke "different" responses, and the decision will be based entirely on the auditory representations. Here some "different" responses are likely to occur in, say, 25 per cent of the trials. The percentage of false alarms will be 25 per cent in all, and because of the non-categorial nature of the auditory representations that percentage will be invariant along the continuum.

If, on the other hand, the two stimuli in an AA pair are categorized as different in, say, 20 per cent of the trials, comparisons of categorized representations will contribute that percentage to the total false alarm rate. In the remaining 80 per cent of the trials the false alarm rate will depend on the auditory representations. If, as above, a quarter of the remaining 80 per cent of the trials, i.e. 20 per cent of the total number of trials, give rise to "different" responses, the total percentage of false alarms amounts to 40 per cent (20 + 20).

If categorizations of AA stimuli as different occur at all they assumedly occur more frequently near the phoneme boundary than well within the phoneme areas.

In an informal identification experiment on vowel duration in which subjects were to label both stimuli in AA and AB pairs it sometimes happened that the two stimuli in AA pairs were given different labels. And this different labelling was found only near the phoneme boundary, never within the phoneme areas. However, the number of AA pairs in the test was very small and it

was not possible to test statistically the reliability of the evidence.

The explanation outlined here for the systematic variation of response bias along the duration continuum is weakened by the fact that a similar systematic variation could not be unambiguously established in the case of another acoustic variable, namely vowel quality. Thus, it will require further investigations to decide whether the systematic variation of response bias reported here can be said more generally to be part of the perceptual behaviour of listeners in the discrimination of linguistic stimuli.

References

Abramson, A.S. and L. Lisker 1968: "Voice timing: Cross-language experiments in identification and discrimination", Haskins SR 13/14, p. 49-63

Bastian, J. and A.S. Abramson 1962: "Identification and discrimination of phonemic vowel duration", JASA 34/5 (abstract), p. 743-44

Fujisaki, H. and T. Kawashima 1968: "The influence of various factors on the identification and discrimination of synthetic speech sounds", Reports of the 6th international congress on acoustics

Fujisaki, H., K. Nakamura and T. Imoto 1973: "Auditory perception of duration of speech and non-speech stimuli", Paper presented at the Symposium on Auditory Analysis and Perception of Speech, Leningrad 1973, p. 1-27

Guilford, J.P. 1954: Psychometric Methods (New York)

Harris, J.D. 1952:   "Remarks on the determination of
                                     a differential threshold by the
                                     so-called ABX technique", JASA 24,
                                     p. 417

Holtse, P. 1973:   "Identification and discrimination
                                     of closely spaced synthetic vowels",
                                     ARIPUC 7, p. 235-64

Liberman, A.M., K.S. Harris,   "The discrimination of speech sounds
H.S. Hoffman, and B.C.   within and across phoneme boundaries",
Griffith 1957:   J. Exp. Psychol. 54,5, p. 358-68

Liberman, A.M., K.S. Harris,   "The discrimination of relative
J.A. Kinney and H. Lane   onset-time of the components of
1961:   speech and nonspeech patterns",
                                     J. Exp. Psychol. 61,5, p. 379-88

Liberman, A.M., K.S. Harris,   "An effect of learning on speech
P. Eimas, L. Lisker and   perception: the discrimination of
J. Bastian 1961:   durations of silence with and
                                     without phonemic significance",
                                     LS 4,4, p. 175-95

McNicol, D. 1972:   A Primer of Signal Detection
                                     Theory (Sydney)

Pisoni, D.B. 1971:   "On the nature of categorical
                                     perception of speech sounds",
                                     Supplement to Haskins SR

Pisoni, D.B. 1973:   "The role of auditory short term
                                     memory in vowel perception",
                                     Haskins SR 34, p. 89-118

Reinholt Petersen, N. 1974:   "The influence of tongue height
                                     on the perception of vowel dura-
                                     tion in Danish", ARIPUC 8, p. 1-10

Rischel, J. and
S.-E. Lystlund 1972:

"A formant coded speech synthe-
sizer", ARIPUC 6, p. IX-XXX

Robinson, D.E. and
C.S. Wattson 1972:

"Psychophysical methods in modern
psychoacoustics", Foundations of
Modern Auditory Theory, vol. II
(J.V. Tobias, ed.), p. 99-132

Stevens, K.N. 1968:

"On the relations between speech
movements and speech perception",
Zs.f.Ph. 21, p. 102-106

AN ACOUSTICAL INVESTIGATION OF DANISH INTONATION:

PRELIMINARY RESULTS

Nina Thorsen

Abstract: The relationship between linguistic stress and funda-
mental frequency in short declarative sentences and
the intonation contours of short declarative sentences,
interrogative sentences, and non-terminal clauses in
Advanced Standard Copenhagen Danish are investigated.
A tentative model for F_o patterns and -contours in
such short simple (i.e. non-compound) sentences is
proposed. The model predicts the following (which is
to a large extent confirmed by the data):
A stressed syllable and all succeeding unstressed syl-
lables within the same simple sentence constitute a
stress group, irrespective of intervening morphological
and syntactical boundaries.
Sentence initial unstressed syllables constitute a
separate entity and are lower than the first stressed
syllable which is always relatively high.
A stressed syllable is lower than the immediately suc-
ceeding unstressed syllable (in the same stress group).
The unstressed syllables in the stress group describe
a falling F_o course. The more unstressed syllables
there are, the more likely the last ones are to be
lower than the stressed syllable in the succeeding
stress group.
The stressed syllables describe a falling contour in
declarative sentences and an almost level one in inter-
rogative sentences where the only cue distinguishing
the interrogative from the declarative sentence is the
intonation. The rest of the interrogative sentences
and the non-terminal clauses are situated between these
two extremes.

1. Introduction

Danish intonation is an under-developed province in the world
of phonetics. Research in this area will, therefore, increase our
knowledge of Danish and it may also contribute to our knowledge of
intonation as a universal phenomenon. Furthermore, the results

can be applied in the teaching of Danish to foreigners, and - by
constituting a basis for comparative studies - in the teaching of
foreign languages to Danes. Finally, the results will be useful
in rule-synthesis of Danish.

1.1 Delimitation of the subject

An exhaustive description of Danish intonation is, of
course, a formidable task, which could keep several people busy
for years.

This paper deals with Advanced Standard Copenhagen Danish
(see Basbøll 1968) and focusses on two phenomena:

1) the relationship between linguistic stress (in the
following referred to simply as: stress), and fundamental fre-
quency (abbreviated: F_o) in short[1] declarative sentences (state-
ments), and

2) the intonation in short
a) declarative sentences,
b) interrogative sentences (questions), and
c) non-terminal clauses.

1.2 Terminology

The terms will be employed in accordance with the following
definitions:
To a single syllable (stressed as well as unstressed) is attached
a _pitch_ or a _pitch movement_ which is realized as an F_o or an F_o-
movement, modified by intrinsic properties of the segment or chain
of segments. - A stressed syllable and all succeeding unstressed
syllables within the same short simple (i.e. non-compound) sentence
constitute a _stress group_, regardless of morphological and syn-
tactical boundaries. (This is not a definition in the strict sense
of the word but, rather, a rationalization of the results, cf.
section 3.1.3.) - To a stress group is attached a _pitch pattern_,

---

1) 'short' meaning: containing three or four stress groups.

realized as a (modified) F_O pattern. One or more stress groups
may form a simple sentence which may be led by one or more un-
stressed syllables. To a simple sentence is attached an intona-
tion with a specific pitch contour, realized as an F_O contour,
which is that figure, or shape, or line one can 'draw' through
the stressed syllables and upon which the unstressed syllables
are superimposed. (This is a rather narrow definition of sentence
intonation but it appears to be justified by the results, cf.
section 3.2.3.1.) - Presumably, there is a rather close connection
between the actual, physical F_O course and the perceived pitch
course ('course' is employed indiscriminately about 'movement',
'pattern', and 'contour'), especially as far as the slowly varying
pitch/F_O contours are concerned, wherefore the term intonation
contour is employed in the following for both pitch- and F_O con-
tour. When no distinction is made between physical (F_O) and
auditory (pitch) phenomena, the term 'tonal' is employed. Un-
stressed syllables before/after the stressed syllables are occa-
sionally, for the sake of brevity, called pretonic and posttonic,
respectively.

1.3   The relationship between stress and F_O

Pitch is one (the most important, perhaps) of several audi-
tory cues for the perception of stress in Danish, or at least in
Standard Copenhagen Danish. To the foreigner (or synthesizer)
who desires to speak intelligible and authentic Danish it is
thus important, among other things, to attach the right pitch
patterns to the stress groups. Those pitch patterns will there-
fore have to be mapped - at first by describing the F_O patterns
of the stress groups and then, later, by establishing their per-
ceptual correlates.

This will permit us, in a given signal (an F_O-tracing), to
separate such variations which are due to the distribution of
stressed and unstressed syllables from the intonation contour
proper, which will then appear more clearly and simpler, to obvious
descriptive and pedagogical advantages.

In the teaching of Danish and in connection with synthesis by rule it works the other way round: Given a certain (simple) sentence with a specific intonation contour the student or the synthesizer (the computer) will, by way of rather simple rules, add the pitch/F_O patterns proper to the stress groups in the sentence.

It seems reasonable, in the light of the above remarks, to include an investigation of intrinsic F_O (F_O differences) pertaining to single segments and of those variations that arise from the interaction between neighbouring segments. The procedure would then be as follows: A given signal is cleansed or corrected for intrinsic and coarticulatory F_O variations, and next for variations due to the distribution of stressed and unstressed syllables, thus producing a clean intonation contour.

And the other way round: to the intonation contour are added first the pitch/F_O patterns pertaining to the stress groups, next the intrinsic and coarticulatory variations at the segment level. The last step is, of course, unnecessary in the class-room situation, F_O variations at the segment level being universal phenomena. However, an investigation of F_O at the segment level is a rather extensive project, and it was decided to omit this aspect from the investigation. (Niels Reinholt Petersen has just completed an investigation of intrinsic F_O in Danish vowels (this issue); his results are turned to account in the treatment of the present material, cf. section 3.1.2.) Besides, it is possible to some extent to dodge the problems introduced by intrinsic and coarticulatory F_O variations, cf. below, section 2.1.

1.4 Advanced Standard Copenhagen Danish

This dialect/sociolect was selected partly because it is the most thoroughly described variety of Danish in recent literature partly because it is that type of Danish which most foreigners will probably need to learn. (Besides, it happens to be the author's own language and one for which subjects are easy to get hold of.)

2. Method

2.1 Material

One corpus is aimed at the stress/F_O relationship, and one at intonation contours in short simple sentences/clauses.

2.1.1 Stress and F_O

Pitch/F_O movements, -patterns, and -contours are probably exceedingly sensitive to the speaker's attitude to what he is saying, which is one reason why nonsense words are of dubious value in an investigation of this kind but they are unavoidable if one wants to introduce changes in one and only one parameter, namely the position of the main stress, in words (otherwise identical) of two and three syllables. There are a few word pairs in Danish, distinguished solely by stress placement, but we do not have five words as below (they are, however, possible words in Danish, - the acute accents mark the stressed syllables):

<u>pípi</u>

<u>pipí</u>

<u>pípipi</u>

<u>pipípi</u>

<u>pipipí</u>

The vowel being <u>i</u> ([i]) in all cases and the consonant being <u>p</u> ([b̥ʰ]), one can disregard intrinsic and coarticulatory F_O variations, so that observed differences in F_O (F_O movements) in the words must be ascribed to different stress distributions. (Still: the aspiration of <u>p</u> is stronger (of longer duration) before a stressed vowel than before an unstressed vowel which may well influence F_O, at least in the beginning of the following vowel.)

An aspirated consonant was chosen (rather than an unaspirated one or a sonorant) in order to obtain as many clearly discernible line-up points as possible for the subsequent averaging

procedure. It was also thought that the shortest possible voiced stretch within each syllable would be the closest approximation to the perceptual model, i.e. to the perceived pitch pattern, thus, at least to some extent, eliminating the difficult question of what part of the voiced stretch in a syllable is perceptually relevant.

One cannot, beforehand, assume that a given word will produce identical F_o patterns in different positions in a sentence. Therefore, each of the five nonsense words were placed initially, medially, and finally in short declarative sentences, which are as much alike semantically and syntactically as possible:

'_____ giver kórtere stávelser.' (_____ has shorter syllables.)

'Stávelserne i _____ forkórtes.' (The syllables of _____ are shortened.)

'Det giver kórtere stávelser med _____.' (There are shorter syllables in _____.)

2.1.2 Intonation contours

This corpus consists of two principal statements:
'Der går mánge bússer fra Tíflis.' (There are many buses out of Tiflis.)

and

'Der er for mánge tímer i statistík.' (There are too many classes in statistics.)

From these two main themes were made up two times five different types of questions and two times three different types of non-terminal clauses, and in such a way that the underlined sequences in the statements above were unchanged in all 8 sentences/clauses of each set. In this fashion it will be possible, later on, to perform identification tests on the intonation contours of those sequences:

(A1)   'Der går månge bússer fra Tíflis?'   (There are many buses
                                                      out of Tiflis?)
       Interrogative sentence: no word order
       inversion, no interrogative particle,
       the intonation alone signals the question.
       (This is a rather unusual type of question.
       It is more in the nature of a polite
       expression of doubt about somebody else's
       statement, or a way to seek confirmation
       of an opinion of one's own than it is
       a proper information seeking question.)
       It may be termed an intonation question.

(A2)   'Mon der går månge bússer fra Tíflis?' (Would there be many
                                                      buses out of Tiflis?)
       Interrogative sentence: no word order
       inversion, interrogative particle
       'mon'.

(A3)   'Går der månge bússer fra Tíflis?'   (Are there many buses
                                                      out of Tiflis?)
       Interrogative sentence: word order
       inversion, no interrogative par-
       ticle.

(A4)   'Hvórfor går der månge bússer   (Why are there many buses
       fra Tíflis?'   out of Tiflis?)

       Interrogative sentence: word order
       inversion, interrogative particle
       'hvórfor'. Note that this sentence
       contains four stress groups.

(A5)   'Går der månge bússer fra Tíflis,   (Are there many buses out
       eller går der kún få?'   of Tiflis or are there
                                            only a few?)
       Interrogative clause: word order
       inversion, no interrogative par-
       ticle, - followed by a coordinate
       interrogative clause.

(A6)   'Der går månge bússer fra Tíflis,   (There are many buses out
       så vi kan gódt lade bílen stå.'   of Tiflis, so we may well
                                            leave the car.)
       Non-terminal main clause, -
       followed by coordinate main
       clause with 'så'.

(A7) 'Hvis der går mánge bússer fra   (If there are many buses
      Tíflis, kan vi gódt lade bílen   out of Tiflis, we may well
      stå.'   leave the car.)

      Non-terminal conditional clause
      with conjunction 'hvis', -
      followed by main clause.

(A8) 'Går der mánge bússer fra Tíflis, (The translation would be
      kan vi gódt lade bílen stå.'   identical to A7.)

      Non-terminal conditional main
      clause with word order inver-
      sion, - followed by main clause.

(B1) 'Der er for mánge tímer i stati-   (There are too many classes
      stík?'   in statistics?)

      (See A1)

(B2) 'Mon der er for mánge tímer i   (Would there be too many
      statistík?'   classes in statistics?)

      (See A2)

(B3) 'Er der for mánge tímer i sta-   (Are there too many classes
      tistík?'   in statistics?)

      (See A3)

(B4) 'Hvórfor er der for mánge tímer   (Why are there too many
      i statistík?'   classes in statistics?)

      (See A4)

(B5) 'Er der for mánge tímer i șta-   (Are there too many classes
      tistík, eller er der for få?'   in statistics, or are there
                                         too few?)
      (See A5)

(B6) 'Der er for mánge tímer i sta-   (There are too many classes
      tistík, så vi bliver nødt til   in statistics, so we will
      at strýge et par stýkker.'   have to leave out a couple.)

      (See A6)

(B7) 'Hvis der er for mánge tímer i   (If there are too many
      statistík, bliver vi nødt til at   classes in statistics, we
      strýge et par stýkker.'   will have to leave out a
                                         couple.)

      (See A7)

(B8) 'Er der for mánge tímer i sta-   (The translation would be
      tistík, bliver vi nødt til at   identical to B7.)
      strýge et par stýkker.'

      (See A8)

Furthermore, there were two times two statements where the
key-words Tíflis and statistík occur sentence initially and
-medially, in order to permit a comparison with the nonsense
words pípi and pipipí in three positions in the sentence:

'Tíflis må lígge i Káukasus.'   (Tiflis should be situated in
                                   the Caucasus.)

'Bússerne fra Tíflis var for-   (The buses from Tiflis were be-
sínkede.'   hind schedule.)

'Statistík er i første seméster.' (Statistics is in the first
                                   term.)

'Eksámen i statistík er frívillig.' (The examination in statistics
                                     is non-compulsory.)

Tíflis and statistík also occur in isolation as declarative
and interrogative utterances, and as members of a reeling off,
i.e. non-terminal words. This was thought to make differences in
intonation contours easy to ascertain and interpret.

In order to ensure a natural performance, the sentences
were incorporated in small dialogues. In the reading list these
dialogues were interrupted by the statements containing the non-
sense words.

The reading list was recorded six times in all, but only
five recordings were included in the subsequent treatment of the
material.

2.2 Subjects, recordings, and registration

The subjects were four phoneticians or students of phonetics between 20 and 35 years of age, two males, NRP and SH, and two females, BH and JG, who all speak Advanced Standard Copenhagen Danish.

The recordings took place at the Institute of Phonetics in a quasi-damped room via a condensator microphone (B&K 4131) and an amplifier (B&K 2607) on a professional tape recorder (Lyrec TR47-2) at a speed of 7½ ips on Agfa HiFi Low Noise tape (PE 36).

The interval between subsequent readings of the list ranged between a couple of hours (in rare cases) and several days. No attempt was made to control the rate of speech. Each subject was allowed to read at the speed most comfortable to her/him.

The tapes were processed by an intensity meter (F-J Electronics) and a pitch meter (F-J Electronics), the outputs of which were registered on a mingograph (Elema 803) at a paper speed of 100mm/sec. Fig. 1 is an example of the tracings. Note that there are two editions of the F_0 registration: one ordinary and one amplified, which will permit an accuracy of measurement of about 1 Hz for the male voices and 2 Hz for the female voices (corresponding to about ½mm on the scale to the left) in the central portion of their respective F_0 ranges.

2.3 Processing of the tracings

The beginning of every stressed vowel was marked, and also the beginnings of all unstressed vowels after unvoiced consonants. The F_0 course was then touched up with a solid line, but only in the vowels and the (voiced) sonorants where these could be delimited safely from the surroundings (which was not always the case), assuming that the F_0 course in voiced obstruents is irrelevant for the perception of pitch patterns and -contours.
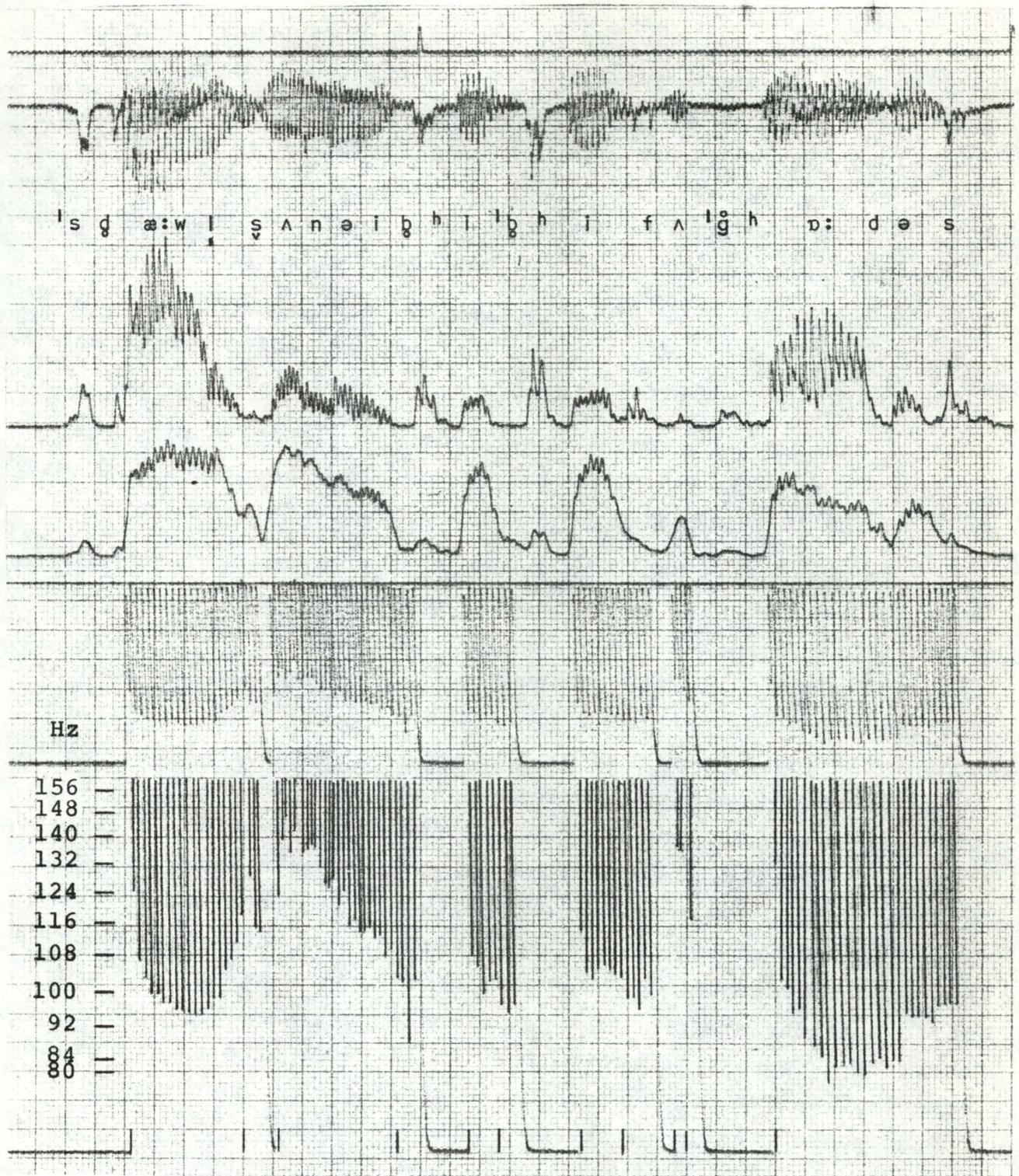
Figure 1

Mingogram of the sentence 'Stávelserne i pipí forkórtes.' (sub-
ject NRP). The traces are, from top to bottom: 1) duplex oscil-
logram, 2) intensity, HP 500Hz, linear display, integration time
5ms, 3) intensity, HiFi, linear display, integration time 10ms,
4) 'ordinary' F_o, 5) amplified F_o (the amplification causes
the ink traces to be almost invisible - they have been touched
up in crayon).

This assumption will have to be tested in perceptual experiments but, so far, it seems justified in the light of a case like the one below (which is but one out of many):

The word stávelserne may appear in any of the three shapes depicted in fig. 2:

(a): a rise in the first two syllables, continuing through the [ş], followed by a fall in the last two syllables,[1]

(b): a rise in the first two syllables, a fall and a rise in the [ş], and a fall in the last two syllables, which corresponds to the fall in case (a), and

(c): a rise in the first two syllables, 'silence' during the [s], a jump upwards to the fall in the last two syllables, which corresponds to the falls in cases (a) and (b).

It seems reasonable to assume that the 'basic production model' looks somewhat like case (a) and it may be realised as such (in rapid speech) when the glottis remains 'closed' and the vocal cords vibrate during the [s], and, probably, when at the same time the constriction at the alveolar ridge is fairly loose. It may be modified as in case (b) when the vocal chords vibrate around a somewhat more open position, a modification which is not voluntary but due to mechanico-acoustico-aerodynamical factors. Finally, the model may be realised (in fairly slow speech) as in case (c) if the glottis is wide open during the [s], impeding vocal chord vibrations. Unless one's attention is specifically drawn to the detection of voicing in an /s/, the three editions are probably perceived in the same fashion, i.e. there is as much information for the listener in case (c) as in case (a), and the F_o course in the obstruent is thus irrelevant in the sense that it passes unnoticed.

---

1) Danish does not have an opposition between unvoiced and voiced
    alveolar fricative, but /s/ may be voiced between voiced
sounds, especially in rapid speech.

Figure 2

F₀ tracings of three recordings of the
word 'Stávelserne ..' (subject JG).

In vowels one very often finds, after an aspirated consonant or an (unvoiced) /s/ or /f/ (i.e. consonants produced with an open glottis and with forceful airflow), one or two vibrations which are between 10 and 40 Hz above the succeeding ones. Such vibrations are left out, cf. the line of argument above for /s/.

The pruning of the F_o course is, regrettably, not uncontroversial. The question soon arises: how much of a movement (a rise or a fall) in a vowel can be ascribed to influence from neighbouring consonants? Or, in other words, what does the intended F_o course - the production model - look like before it is superimposed by modifications introduced by mechanico-acoustico-aerodynamical factors, resulting in the rather complicated F_o course observed in the tracings?

One can also pose the question this way: what is perceptually relevant? Ideally, one might trim the traces in accordance with their perceptual correlates (which are not necessarily identical to the production model). But before one can do that, a good deal more must be known about pitch perception in connected speech, i.e. signals which are far more complicated and contain far more information than do e.g. sinusoidal waves.

One example will suffice to illustrate the schism: The first syllable in the word pipí in the sentence 'pipí giver kórtere stǎvelser.' produces, with all subjects, a continuous but rather steep fall (cf. figs. 4-7). The beginning of this falling movement may be ascribed to an influence from the preceding aspirated consonant and the end of it may be ascribed to the articulatory transition to the following stop consonant. The intended F_o (movement) may be a level F_o at a frequency corresponding to the central part of the fall, but we cannot be sure (see also section 3.1.1.2). How is the actual movement perceived? As a fall? And in that case: of which extent? As a level pitch corresponding to the beginning? or to the middle? or to the end of the fall?
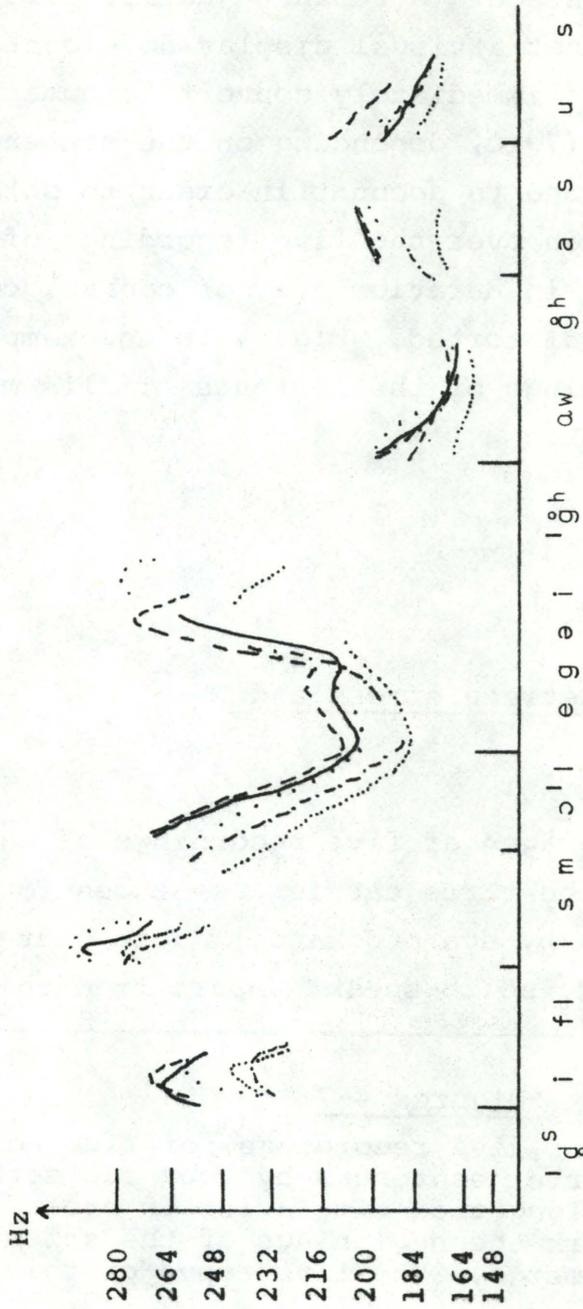
Hz

280
264
248
232
216
200
184
164
148

^s_id i f l i s m ^l_jc l e g e i ^{l}gh ɑw ɡh a s u s

Figure 3

F_o tracings of five recordings of the sentence 'Tíflis mǎ líg-
gě i Káukasus.' (Subject BH).

The various types of lines indicate the recordings thus:

1st recording: _____
2nd   — — —
3rd   · · · · ·
4th   · · · · · ·
5th   — · — · —

Because of this uncertainty the touched-up tracings have
been transferred to tracing paper and have been evaluated quali-
tatively, rather than quantitatively. I.e. no measurements have
been performed at this stage. A further justification for this
procedure is the fact that a visual display does contain more
information than one can immediately convert to numericals.
All the line-up points (7-10, depending on the sentence, cf.
section 2.1.1) were turned to account in order to obtain the
least possible dispersion over the five recordings of each item.
In this way differences in duration are, of course, conceiled,
and the 'time-axis' is distorted. Fig. 3 is an example of
tracings of five recordings of the sentence 'Tíflis må lígge i
Káukasus.' by subject BH.

3. Results

3.1 The relationship between stress and F_o

3.1.1 Nonsense words

Figs. 4-7 are averages of five recordings of each of the
five words in each of the three carrier sentences for each subject.
(The averaging was done by eye and hand.) Note that the time-
axis is twice distorted, so to speak: Apart from the distortion

---

Figures 4-7

F_o tracings (averages of five recordings) of five nonsense words
in three different carrier sentences by four subjects. The trac-
ings have been moved along the 'time-axis' in order that the
line-up points that mark the beginnings of the stressed syllables
in the nonsense words merge. The five types of thin lines indi-
cate the words thus:

```
pípi      ———————
pipí      · · · · · ·
pípipi    – – – – –
pipípi    –··–··–··–
pipipí    ·············
```

Figure 4   (NRP)
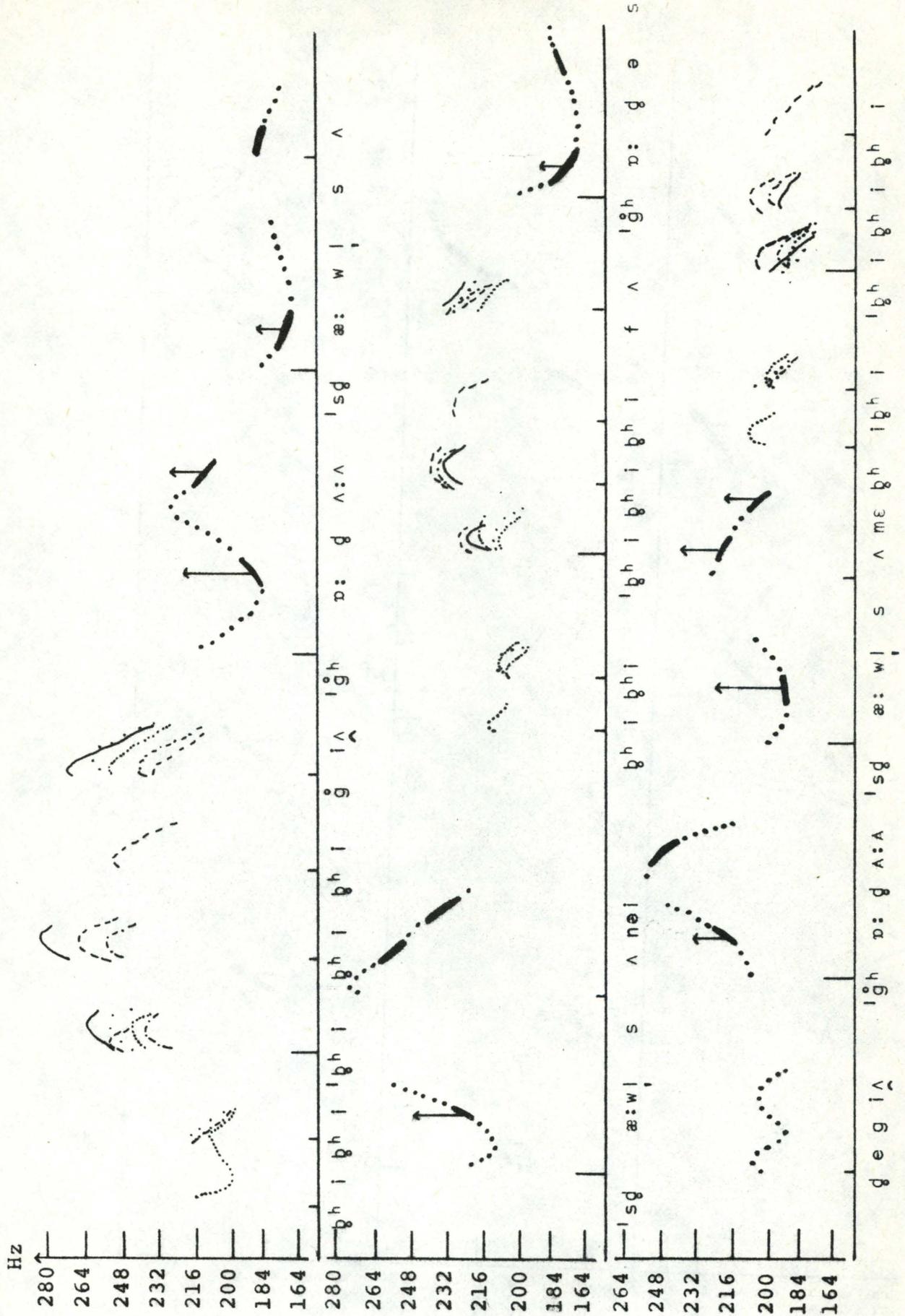
Figure 5 (SH)

Figure 6 (BH)

Figure 7   (JG)

introduced by the adjustment to each line-up point through the
sentences (cf. above), the five nonsense words were also moved
along the axis in such a way that the line-up points marking the
beginnings of the stressed vowels merged, which is why there
seem to be five pi-syllables in the sentences. - The heavy dotted/
full lines represent the averages of 25 recordings (five recordings
times five nonsense words) of the carrier sentences. (The expla-
nation for the use of dotted/full lines and for the arrows will
be given later, in section 3.1.2.) (There are occasional heavy
full lines in the nonsense words (NRP and JG), representing a
complete merging of 20 and 25 recordings, respectively.) The dis-
persion in the carrier sentences over the 25 recordings was no
greater than the dispersion within each set of five recordings,
and the carrier sentences are thus a good frame of reference for
the nonsense words.

3.1.1.1 Stressed followed by unstressed syllables

A general tendency shows up in the figures:
<u>the stressed syllable in the nonsense words has a lower F_o than
the immediately succeeding unstressed syllable</u> (which may belong
to the next word, as in the case of pipí and pipipí in sentence
initial and -medial positions).

SH (fig. 5) has not a single exception to this rule. But
the remaining subjects do (although not in all recordings of a
given word):

JG (fig. 7): two words in two positions, namely medial
and -final pípi and pípipi, are exceptions to the rule in 17
(out of 20) recordings, but then in both words the stressed syl-
lable is clearly lower than the immediately preceding unstressed
syllable, and thus the stressed syllable is probably marked al-
ready as being stressed. Further, in the case of the sentence
medial words, and sentence final pípi, one may wonder whether
these words are exceptions perceptually. We are dealing with
movements of the falling (in the stressed vowel) - rising (in

the unstressed vowel) kind. It is possible that such a movement
is perceived as a sequence 'low-high'.

(Bertil Malmberg (1961) has investigated the relation be-
tween F_o movement and pitch perception in various words in
Yoruba and Ibo and found that he perceived movements like the
ones below:
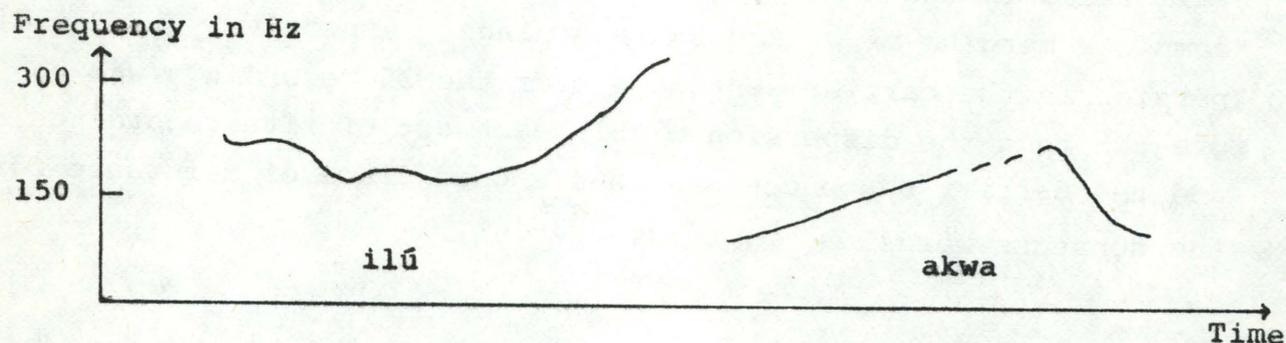
as sequences of 'low-high' (ilú) and 'high-low' (akwa), respec-
tively.)

BH (fig. 6): BH is deviant in sentence initial pipí (2
recordings), sentence initial pipipí (3 recordings), sentence
final pípi (4 recordings), and sentence final pipípi (3 recordings).
In the first three cases the tracings are open to the same inter-
pretation perceptually as in JG's case. The sentence initial words,
however, also hold another possibility, see below about NRP.

NRP (fig. 4): NRP deviates in one (maybe two) case(s),
namely sentence initial pipipí giver[1] (5 recordings). Most of
the F_o movement in the stressed syllable lies above the succeeding
unstressed syllables, and it is presumably also perceived as being
higher. In sentence initial pipí giver the same may be true.

---

1) giver is pronounced in one syllable ([g̊i∧]).

These are instances where also BH deviates, and an explanation
may be sought in the fact that
sentence initial unstressed pi-syllables are (considerably)
lower than the first stressed syllable, which is always rela-
tively high.
The stressed syllable in the sentence initial words is thus
always marked, namely in relation to the preceding unstressed
syllable(s). Therefore, it may not be an absolute necessity
to continue further up the scale on the succeeding unstressed
syllable, especially when, as in these instances, there is only
one unstressed syllable before the next stressed syllable in
the sentence, a stressed syllable which has to be lower than
the first stressed syllable, since
F_o of the stressed syllables is falling from the first to the
last stressed syllable in short statements (cf. sections 3.1.2,
3.1.4, and 3.2.2).
In cases like these, where there is only one unstressed syllable
between two stressed syllables, the choice of strategy is
seemingly free:
NRP has giver lower than pipipí (and maybe also pipí).
BH has giver higher in two recordings, lower in 3 recordings
than pipí and pipipí. JG and SH have giver higher than pipipí .
and pipí in all instances.
(Incidentally, there are other factors besides pitch which con-
tribute to the correct identification of stress: the unstressed
word giver is pronounced (in one syllable) with a short stød-
less vowel, and p of the stressed syllable in (pi)pipí has a
stronger aspiration than p in the preceding unstressed syllable(s).)

3.1.1.2 Stressed preceded by unstressed syllables

The F_o relations between a stressed syllable and the pre-
ceding unstressed syllable(s) are less clear than the relations
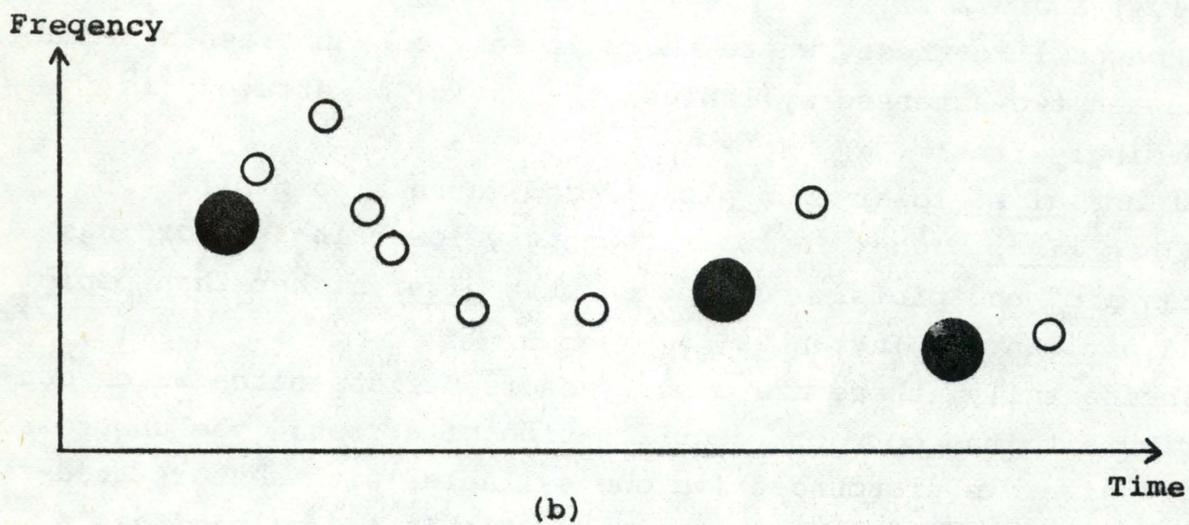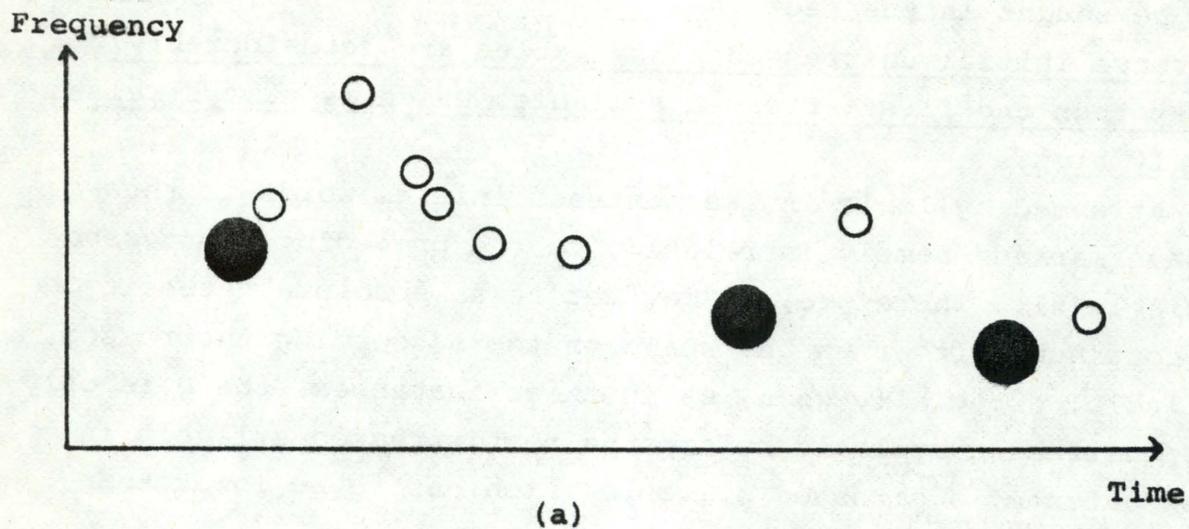to succeeding unstressed syllable(s).

(a)

(b)

Figure 8

F relations between stressed (large filled circles) and un-
stressed (small empty circles) syllables in the sentence
'Stá-vel-ser-ne i pi-pi-pí for-kór-tes.' (a): subject SH,
(b): subjects NRP, BH, and JG.

Sentence initial unstressed <u>pi</u>-syllables are, as mentioned above, always lower than the succeeding stressed syllable.

With SH the remaining stressed syllables are always lower than the preceding (as well as succeeding) unstressed syllables, and his F_o patterns are thus the simplest to describe (cf. below). With NRP, BH, and JG the situation is slightly more complicated: When the number of unstressed syllables between two stressed syllables is three, or less than three,[1] the unstressed syllables are higher than the succeeding stressed syllable (i.e. the pattern that is general for SH). But when there are more than three[1] unstressed syllables between the stressed ones, as in the sentence 'Stá-<u>vel-ser-ne i (pi)-pi</u>-pí forkórtes.', number four and following unstressed syllables are as low as or lower than the succeeding stressed syllable. The difference between SH on one hand and NRP, BH, and JG on the other may be depicted graphically as in fig. 8. (The somewhat peculiar position of '-<u>vel</u>-' will be dealt with in section 3.1.2.)

SH's statements did sound a little more insisting than did the statements of NRP, BH, and JG, and this impression could be attributed to the fact that his stressed syllables are always lower than the unstressed syllables (excluding the sentence initial unstressed syllables). This is corroborated by informal listening to Copenhagen Danish in various situations: the lower the stressed syllables in relation to the unstressed syllables, the more obstinate a speaker sounds.

---

1) The limit may not be three exactly. It may also, it seems, be two or four.

3.1.1.3 F_o movement in individual syllables in the nonsense words

There are variations in F_o movement in individual syllables, but the picture is complicated by the fact that there are inter- and intra subject variations as well as variations seemingly due to position in the sentence and to the difference between stressed and unstressed syllables. However, the following statements hold true, not only of the averages but also of individual recordings.

(i) Sentence initial position

The pretonic syllables are generally falling.

The stressed syllable is rising with NRP, rising-falling with BH, slightly rising or rising-falling with SH, and rising slightly or falling with JG.

The (first) posttonic syllable (of two) is very slightly rising-falling with NRP, and rising-falling with SH, BH, and JG. However, with SH the fall is greater than the rise, with BH the rise and fall are of almost equal extent, and with JG the rise is greater than the fall.

The second of two posttonic syllables as well as the following word giver are falling with all subjects.

The very rough general picture is one of falling unstressed syllables and rising stressed syllables.

(ii) Sentence medial position

The pretonic syllables are falling with all subjects.

The stressed syllables are very slightly rising-falling with NRP, falling or falling-rising with SH, rising-falling with BH, and falling with JG.

The description of the posttonic syllables in medial position is exactly the same as in sentence initial position.

The rough general picture is one of falling unstressed and stressed syllables alike.

(iii) Sentence final position

Stressed and unstressed syllables alike are generally falling, but only slightly so with NRP, and with exceptions in the posttonic syllable for NRP (pípi) and JG (pípi and pipípi).

The very crude (too crude, perhaps) picture is thus one of falling unstressed syllables in all positions in the sentence and of rising stressed syllables in sentence initial position, falling stressed syllables in sentence medial and -final positions.

It should be remembered, however, that much of the falling movement in stressed and unstressed syllables may be due to in-fluence from preceding aspiration and following closure, - and it is hard to say how much of the falling movement is inherent in the production of [i]. The rise in the sentence initial stressed syllables can hardly be ascribed to surrounding sounds, and it is not likely to be an inherent feature of [i]-production either (since this is the only position where stressed [i] is rising, and the rise is not found in all words with all subjects, cf. above). - This rise, when it occurs, may simply be a glide towards the high start of the intonation contour proper.

In order to get a clearer picture of F_0 movements in indi-vidual syllables one must look at other vowels besides [i] and other surrounding sounds besides [b̥ʰ]. Niels Reinholt Petersen (this issue) investigated eleven long stressed vowels in words of the type [b̥ʰV$_1$'b̥ʰV$_1$:b̥ʰV$_1$] in sentence medial position. He found that the F_0 movement was almost identical in all vowels, being generally falling-rising (the fall was of greater extent than the rise). Three vowels, [i:, ɑ:, u:] were also investigated between [b̥], [f], and [m] where the same falling-rising movement showed up but with a (not very pronounced) tendency for the extent of the fall to be greater after [b̥ʰ] and [f]. Thus, it does not, at present, seem as if different vowel qualities carry inherently different F_0 movements in Danish, and different surrounding con-sonants do not seem to influence the direction of the movement,

only the extent. The falling-rising movement in Reinholt Petersen's material as opposed to the falling movement in the present material may reflect a difference between long and short vowels, and both movements may reflect a property specific of Danish vowels.

3.1.2 The natural words in the sentences

The F_o relations in the nonsense words reappear in the real, or natural, words in the sentences (stávelser(ne), kórtere, forkórtes), but it is not quite so easy to demonstrate from the tracings. In most cases the F_o course is greatly varying: which portion in a given course is then to be compared to what portion in another course?

If we assume that the tonal patterns in the (pi)pipi words shall recur in the other words (an assumption which can be verified or falsified in perception experiments), the state of affairs is greatly simplified (thereby acting as an indirect support for the assumption). One can compare the courses in the underlined sequences in the six pairs of sentences below, where the nonsense words have been chosen to give the best fit to the natural words of interest in each pair:

(a)   'pípipi giver kórtere stávelser.'
       'Stávelserne i pípi forkórtes.'

(b)   'pipípi giver kórtere stávelser.'
       'Det giver kórtere stávelser med pípi.'

(c)   'Stávelserne i pípi forkórtes.'
       'pípipi giver kórtere stávelser.'

(d)   'Stávelserne i pípipi forkórtes.'
       'Det giver kórtere stávelser med pípi.'

(e)   'Det giver kórtere stávelser med pípi.'
       'pipípi giver kórtere stávelser.'

(f)   'Det giver kórtere stávelser med pípi.'
       'Stávelserne i pípi forkórtes.'

Note that the best fit to kórtere in (a) and (e) is (pi)pípi rather than pípipi, although kórtere is a three syllable word, and, likewise, the best fit to ståvelser in (b) and (f) is pípi rather than pípipi.

Phonemically, kórtere and ståvelser are /ˈkɔʁdəʁə/ and /ˈsdaːvəlsəʁ/, respectively, phonetically they are [ˈg̊ʰɒːɟʌːʌ] and [ˈsɟæːwˌlsʌ], respectively, i.e. the third syllable in kórtere is assimilated to the second syllable (but it still counts as a syllable), and the second syllable in ståvelser has /ə/ elision, resulting in a syllabic lateral, [ˌl]. It may be that sequences of two syllables, where the second is weakened through /ə/ assimilation or -elision and separated from the preceding syllable by zero or a sonorant consonant, behave tonally as one syllable, i.e. they show what may be termed tonal assimilation. - In fig. 8 was seen that the -vel- ([wˌl]) of ståvelserne occupied a position on the frequency axis between the preceding stressed syllable and the succeeding unstressed syllable.[1] - This is somewhat unusual since the first posttonic syllable throughout the rest of the material is the highest one in the stress group, which is a further indication of a peculiarity with such types of unstressed syllables. - More research is needed in order to establish this point with any certainty and to formulate the exact conditions (segmental and otherwise) under which the tonal assimilation takes place.

If the assumption is valid, namely that the same tonal patterns are intended in the nonsense and natural words (in these particular sentences which are of approximately the same semantic and rhythmical structure), one can prune the F_o course in the natural words if, at the same time, one takes intrinsic F_o differences between different vowel qualities into account. This procedure was facilitated by the investigation of Niels Reinholt Petersen who used the same subjects. The relevant figures from his table 1 are repeated below in table 1, expressed as differences in Hz between [iː] and the other 10 vowels of his material.

---

1) The exact location is difficult to establish on the mingograms but it is clearly below the level of the vowel in -ser- ([ʌ]).

Table 1

Differences in Hz (rounded off to the nearest integral) between the (long) vowel [i:] and various other (stressed) vowels (in sentence medial position). The difference is positive if [i:] has the highest frequency of the pair, negative otherwise. For subjects NRP, BH, and JG the figures represent the mean between two values, as observed in two different recordings in Reinholt Petersen's material. [æ:] is identical to [a:] in his notation.

Subject	NRP	SH	BH	JG
Vowel pair				
[i:] – [e:]	3	7	10	9
[i:] – [ɛ:]	6	12	23	16
[i:] – [æ:]	8	16	24	17
[i:] – [ɑ:]	9	15	34	25
[i:] – [y:]	-2	2	0	-8
[i:] – [ø:]	3	7	6	6
[i:] – [œ:]	7	12	17	15
[i:] – [u:]	-2	-1	-3	-16
[i:] – [o:]	2	5	1	5
[i:] – [ɔ:]	4	10	18	15

(NRP did not measure [ɒ:], but presumably it behaves like [ɑ:] (cf. that there is only a negligible difference between [i:] and [y:]).)

The heavy full lines in figs. 4-7 represent those parts of the F_o courses in the natural words which merge with the courses in the nonsense words when the tracings are moved up the frequency axis to an extent indicated by the arrows (which tally reasonably well with Reinholt Petersen's results, cf. table 1). The heavy dotted lines represent the remainder of the F_o courses in the natural words. - In consequence of the remarks above about the [wḷ] of stávelserne, a fit to this syllable has not been attempted and, likewise, not to the third syllable of kórtere either. - Fittings are missing in some instances where it proved impossible to perform them. - In several instances it seemed unnecessary to move the unstressed syllables at all to obtain a fit.

Reinholt Petersen measured only long stressed sentence med-
ial vowels. Presumably, his results (or at least the tendencies,
if not the exact figures) can be generalized to short vowels as
well (the differences may be smaller), and to sentence initial and
-final positions. However, in order to obtain the best possible fit
the intrinsic differences will have to be slightly greater in
sentence initial position, and slightly smaller in sentence final
position than in medial position, and although this point needs
to be verified, it was deemed justifiable to adjust the procedure
accordingly.

Reinholt Petersen did not measure the unstressed vowels,
but it seems as though the intrinsic F_o differences are smaller
and the dispersion of measurements greater in unstressed than in
stressed position, cf. Reinholt Petersen's section 3.2.4. The
moving of the unstressed syllables ([ʌ, wi, ə, i, ɛ]) along the
frequency axis was performed simply till the best fit was obtained,
which proved always to be a smaller distance than for the stressed
vowels nearest to the unstressed ones in quality, which would be:
[ʌ∼ɔ-ɑ, wi∼u-i, ə∼ɛ-œ, i∼i, ɛ∼ɛ].

It appears that there are several factors involved which
may vitiate the procedure with possible errors, and the outcome
of the trimming must be regarded with some reservation.

Keeping those reservations in mind, however, it seems none-
theless as if the former statements about F_o relations in the
nonsense words can be extended to be valid in short statements
in general:

Sentence initial unstressed syllables are always lower than the
first stressed syllable, which is always relatively high.
(The difference between JG's sentence initial unstressed syllables
and the first stressed syllable is smaller than for NRP, SH, and
BH, however.)
A stressed syllable is lower than the immediately succeeding
unstressed syllable.

One unstressed syllable between two stressed ones, however, may be lower than the preceding stressed syllable.

The unstressed syllables in the stress group describe a falling F_o course. The more unstressed syllables there are, the more likely the last ones are to be lower than the stressed syllable in the succeeding stress group (the limit seems to be two to four unstressed syllables).

Furthermore, it seems as if the difference in Hz between a stressed syllable and the immediately succeeding unstressed syllable is greatest in sentence initial and smallest in sentence final position.

F_o of the stressed syllables (the intonation contour) is falling from the first to the last stressed syllable.

3.1.3 The segmental domain of F_o patterns

An interesting problem is that of the structure of the F_o patterns in relation to the chain of segments. Are the F_o patterns related to word boundaries? Or to the boundaries between larger entities within the simple sentence? Or do the stressed syllables alone determine the F_o patterns? (The answer to the last question seems to be affirmative, a fact that has already been turned to account in the preceding pages.) In this case, as in numerous cases mentioned above, a special material should be prepared, aimed specifically at this problem, and the following is more in the nature of speculations than of results proper:

One can search for those portions of the F_o course that vary and look at the extent of the variation in each of the three sentence types where the nonsense words occur initially, medially, and finally:

In sentence initial position in '_____ giver kórtere stávelser.' it is evident that the variation in F_o is not limited to the (pi)pipi words but extends over the following unstressed giver ([g̊iʌ]), in such a fashion that giver is higher the closer it is to the preceding stressed syllable in time, i.e. the fewer

unstressed pi-syllables there are: giver is lower after pípipi than after any of the other nonsense words, and it is highest (not quite so clearly so) after pipí and pipipí. The same tendency appears medially, but only clearly so for NRP, and not at all for JG. - This points in the direction of word boundaries being non-decisive for the F_O patterning. However, the boundary between (pi)pipi and giver and between (pi)pipi and forkórtes is not just a word boundary but also a noun phrase + verb phrase boundary, i.e. a substantial syntactic boundary within the simple sentence. Thus it seems as if F_O patterns and their re-currence are governed by the stressed syllables alone. A further, but rather weak, support for this claim is found in the pretonic pi-syllables in sentence medial and -final positions which show less F_O variation than do the stressed and posttonic pi-syllables. These pretonic pi-syllables may be part of the preceding (in-variant) course, thus: 'Stávelserne i (pi)pi ...'/'(Det giver kórtere) stávelser med (pi)pi ...'. - This means that a stressed syllable and all succeeding unstressed syllables within the same simple sentence constitute the unit for the F_O patterning, irre-spective of intervening morphological and syntactical boundaries. This unit may be termed a stress group. Sentence initial un-stressed syllables constitute a separate entity.

The restriction 'in the same simple sentence' above is inserted because then one can be reasonably sure, on the basis of the present material, that the statement holds true. (But the material is, admittedly, limited and future investigations may, of course, show that the definition of 'stress group' should be changed or modified.)

There are indications, however, that a stress group may also sometimes extend across the boundary between two simple sentences (clauses) in a compound sentence: There are two times four compound sentences in the material, namely A5-8 and B5-8 (cf. section 2.1.2). Only the initial clauses of these sentences have been subjected to processing because they constituted the focal points of interest. But when one looks at the mingograms the following facts appear:

In the majority of a total of 160 recordings (8 clauses times 4 subjects times 5 recordings) the initial unstressed syllables in the terminal clauses are higher than the succeeding

stressed syllable, contrary to the low position of sentence initial unstressed syllables throughout the rest of the material. This is universally true of the terminal clause in sentences A7,8 and B7,8, i.e. main clauses preceded by conditional clauses. Whether the high position of the unstressed syllables is a property of this particular combination of clauses or whether it should be ascribed to an 'across clause boundary stress group effect' is debatable. - In favour of the first solution speak the facts that in 8 instances (A7: 2, A8: 4, B7: 1, and B8: 1) a pause intervenes between the two clauses, thus breaking up what seems to be the normal temporal relations within the stress group. And in 29 out of 40 instances in B7 and B8, where the last syllable of the initial clause is stressed (statistík) the unstressed syllables (two in both cases) in the succeeding terminal clauses are lower than the preceding stressed syllable and thus do not follow the normal F_o pattern of a stress group. In A7 and A8 there are 6 instances where the first unstressed syllable of the terminal clause is higher than the preceding unstressed syllable (Tíflis), thus breaking the normal stress group pattern. But the second solution would probably be correct for 11 out of 40 instances in B7 and B8 where there is no intervening pause and where the clause final stressed syllable is lower than the clause initial unstressed syllable and for 22 out of 40 instances in A7 and A8 where there is no intervening pause and where the 2 clause initial unstressed syllables together with the clause final unstressed syllable form the smooth falling course characteristic of the unstressed syllables of a stress group.

In A6 and B6 (two coordinate main clauses) both high and low (in relation to the succeeding stressed syllable) clause initial unstressed syllables (three in both cases) are found, and in almost equal numbers. A pause may intervene before high and low unstressed syllables alike. There are only 3 instances where the final syllable(s) of the initial clause and the initial syllables of the terminal clause may be said to constitute a stress group, according to the criteria mentioned above.

In A5 (two coordinate interrogative clauses) subjects unanimously produce the four clause initial syllables on the same low level in relation to the succeeding stressed syllable, and in 13 out of 20 instances without an intervening pause. - The second clause in A5 contains two stressed syllables. The second clause in B5 is extremely short, containing only one stressed syllable, which may be the reason why it is treated differently from A5. SH and BH in 9 out of 10 instances produce what looks like an across clause boundary stress group, but NRP and JG do not.

The conclusion seems to be that the initial unstressed syllables in the terminal clause in a compound sentence may be high or low in relation to the succeeding stressed syllable, whereas sentence initial unstressed syllables always are low. In compounds of conditional and main clauses the initial unstressed syllables of the terminal clause are always high, which

may be a signal for a close link between such clauses. In compounds of coordinate (interrogative) clauses both high and low unstressed syllables are found initially in the terminal clause, but the low ones take the upper hand, which could be an expression of the fact that coordinate clauses are not as closely linked as are compounds of conditional and main clauses. If a stress group is a temporal and tonal entity it may (in about 40% of the instances) extend over the conditional + main clause boundary, whereas this happens only rarely (in 15% of the instances) across the boundary between two coordinate clauses.

Finally, the hypothetical nature of all of the preceding section should be stressed. The properties of the stress group will have to be investigated in much further depth, taking also temporal relations into account.

3.1.4 Conclusion

To recapitulate the results, a model for the F_o course in short statements, which contain only stressed and unstressed syllables (i.e. no secondary stresses), is proposed, cf. fig. 9. It cannot be rejected on the basis of the present material, but its corroboration must await further experiments.

Any sentence initial unstressed syllables are relatively low. The first stressed syllable is relatively high. The first posttonic syllable of several constitutes the F_o maximum in the stress group and the sentence (with possible exceptions if this first posttonic syllable is 'weak' and separated from the preceding stressed syllable by zero or a sonorant consonant, cf. section 3.1.2). A single posttonic syllable may be lower than the preceding stressed syllable, but it is always higher than the succeeding stressed syllable. Three[1] or less posttonic syllables follow the course of the full line. Four or more posttonic syllables may follow the course of either the full line or the broken line, in which last case the first three syllables lie above the heavily dotted line connecting the stressed syllables. This dotted line constitutes the intonation contour proper. The posttonic syllables in the second and, probably, third stress group(s) describe the same figure as those of the

---

1) See footnote to section 3.1.1.2.
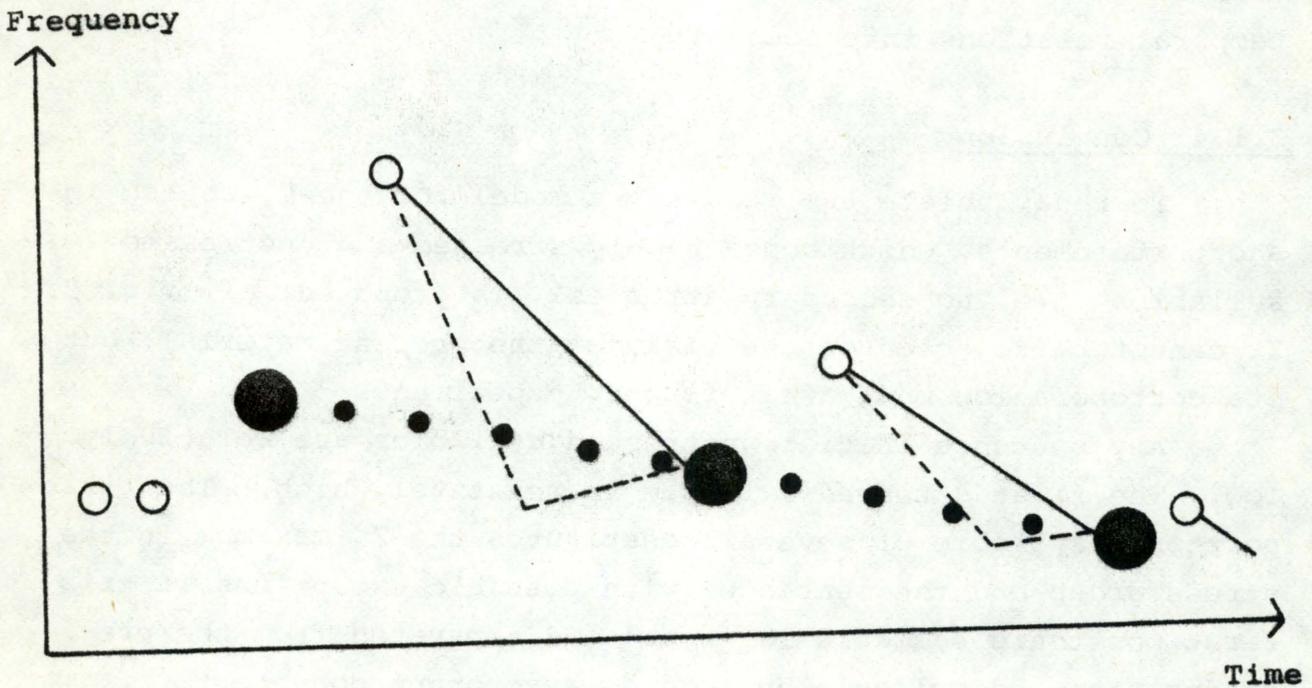
<u>Figure 9</u>

A model for F$_o$ in short declarative sentences. Large filled and small empty circles indicate stressed and unstressed syllables, respectively. - The small filled circles indicate the intonation contour. For an account of the thin full and broken lines, see the text.

first one, but the distance in Hz between the stressed and the
first posttonic syllable decreases progressively. There seems
to be the same freedom in choice of strategy, i.e. full or broken
line, in the second stress group. (NRP is the only subject who
actually follows the broken line in this case.)

The intonation contour describes a smooth fall from be-
ginning (the first stressed syllable) to end (the last stressed
syllable).

3.1.5 A postscriptum about normalizing/non-normalizing the tracings

The question whether a normalization (a zero offset adjust-
ment, strictly speaking) of the traces from the outset would not
have facilitated the subsequent processing is a very reasonable
one. One might, for each subject, move the traces up or down
the frequency scale, as the case might be, so that, for instance,
F_0 minima merged in all five recordings of a given sentence, and
maybe also in all sentences of the same type. - However, Reinholt
Petersen found in the course of his work that the dispersion does
not reduce in this way. He attempted a normalization based on
F_0 minima in the stressed vowels in the pV'pV:pV words. -
It is evident, also in this material, when one looks at individual
recordings that such a normalization may well reduce the disper-
sion in one instance but increase it in another: In the state-
ments the dispersion is generally greatest on the initial stress
group, smaller on the medial and smallest on the final stress
group, regardless of sentence type and nonsense vs. natural words
(fig. 3 is an example). Thus, if one normalizes after the minima
in the first stressed syllable, a dispersion is introduced on the
last stressed syllable, which is not present in the recordings.
If one normalizes after the last stressed syllable, the dispersion
on the first stressed syllable is not substantially reduced.
A normalization which will reduce the dispersion all over the
sentence must therefore be performed for each individual stress

group, as it turns out. In future experiments one might try
a normalization along these lines.

3.2 Intonation contours in short statements, questions, and non-terminal clauses

The material was accounted for in section 2.1.2.

3.2.1 Isolated words and reeling-off members

JG is left out in this section. Her recordings were not
quite successful in that she did not treat the statements as
separate utterances, and her questions sounded overly dis-
believing or surprised.

Figs. 10 and 11 are averages of five recordings of the
two words in each of the three functions by NRP, SH, and BH.

3.2.1.1 Stress and F_O in the words Tíflis and statistík

(i) Statements

(i, a) Tíflis

With NRP and SH the stressed (first) syllable of Tíflis
is lower than the unstressed syllable, i.e. a relationship which
is found with pípi in all positions in the sentence with these
two subjects (cf. figs. 4 and 5).

With BH the two syllables show no appreciable F_O differ-
ence and this fact, together with their position on the frequency
axis, makes them resemble her sentence final pípi closely.

Both syllables in Tíflis are falling with all three sub-
jects.

(i, b) statistík

With NRP the two unstressed syllables are clearly lower
than the stressed (third) syllable. The same is true only of
the second unstressed syllable with SH and BH, but we may still
conclude that the relationship between unstressed and stressed
syllables resembles that of sentence initial pipipí, and since

123
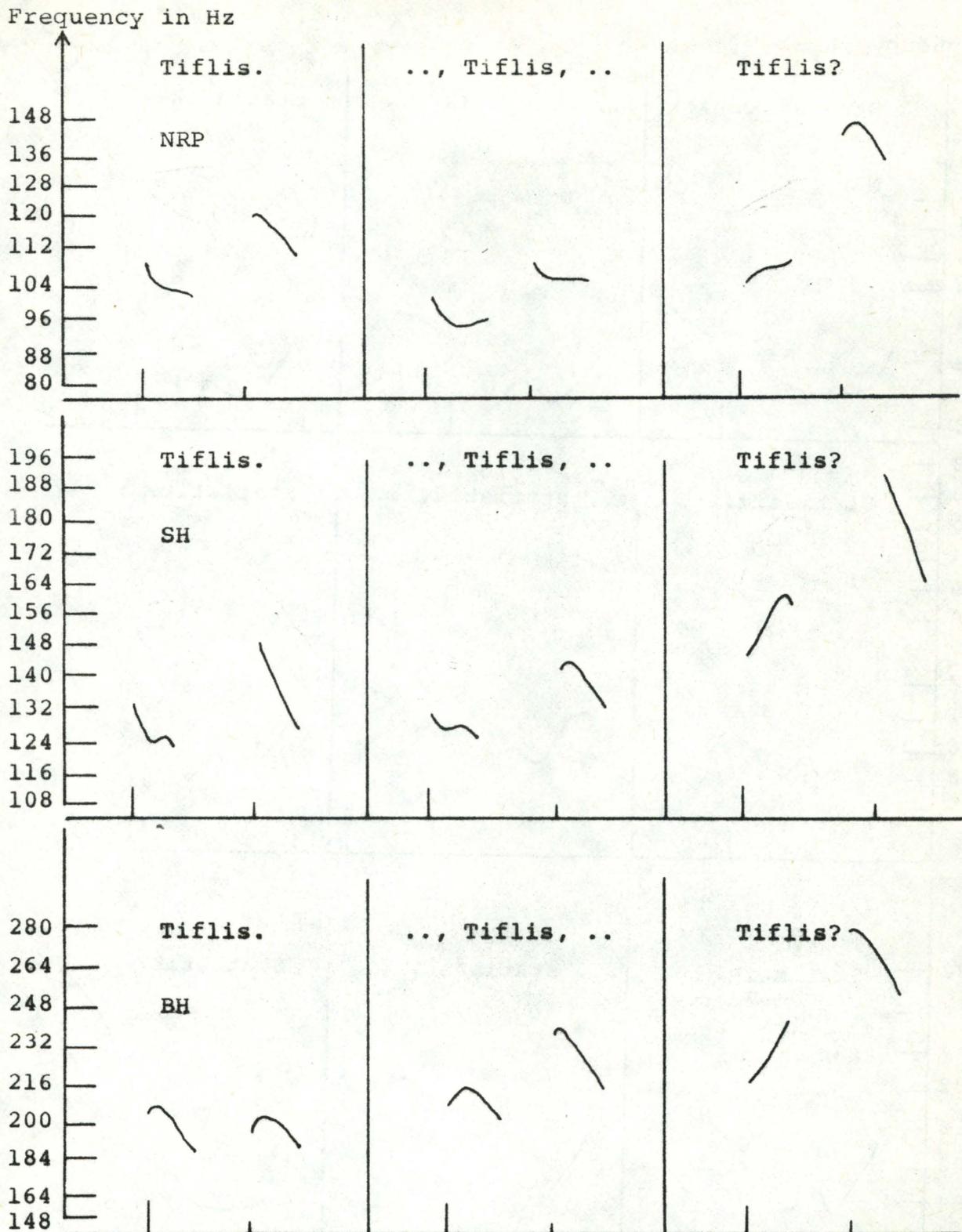
Figure 10

F$_0$ tracings (averages of five recordings) of the word Tíflis by three subjects, from top to bottom: NRP, SH, and BH. The word occurs 1) in isolation, as a declarative utterance (left column), 2) as a member of a reeling-off (middle column), and 3) in isolation, as an interrogative utterance (right column).

Frequency in Hz
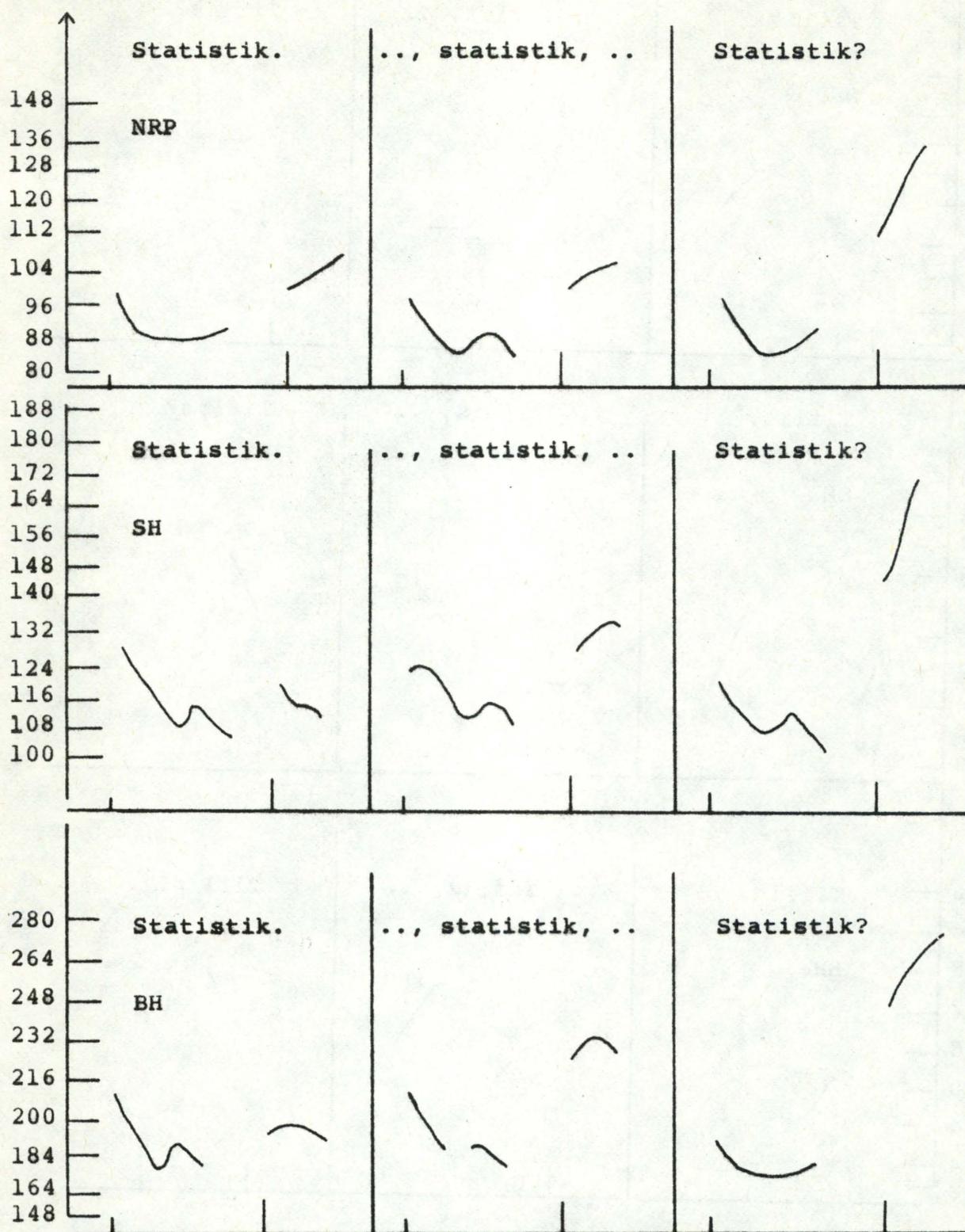
Figure 11

F$_0$ tracings (averages of five recordings) of the word statistík
by three subjects, from top to bottom: NRP, SH, and BH. The
word occurs 1) in isolation, as a declarative utterance (left col-
umn), 2) as a member of a reeling-off (middle column), and 3) in
isolation, as an interrogative utterance (right column).

that relationship may or may not be reversed in sentence medial and final positions (cf. section 3.1.1.2), a direct comparison with pipipí in those positions is not possible.

All three syllables are more or less falling with SH and BH. With NRP the stressed syllable is clearly rising (this is true of all five recordings). This corroborates a subjective impression of NRP's statement being slightly 'non-final' in character, i.e. more or less leading on to the following sentence in the dialogue.

(ii)   Reeling-off members (i.e. non-terminal words)

(ii, a)   Tíflis

With NRP and SH the description would be the same as for the statement, cf. above, except that the two syllables are about 8 Hz lower on the frequency axis with NRP.

With BH there is in this function a clear difference between the stressed and unstressed syllable.

The two syllables are, roughly, falling with all subjects.

(ii, b)   statistík

With NRP the description would be the same as for the statement, cf. above.

With SH and BH there is a clear difference between both of the unstressed syllables on one hand and the stressed syllable on the other, which difference is brought about by the stressed syllable being (considerably) higher than in the statement, whereas the position of the unstressed syllables is the same.

Note that the stressed syllable is rising (or rising-falling) with all subjects.

(iii)   Questions

(iii, a)   Tíflis

There is a large and clear F_o difference between the (low) stressed and (high) unstressed syllable with all subjects.

The stressed syllable is rising, the unstressed one falling with all subjects.

(iii, b)   statistík

The F_O difference between unstressed and stressed syllables is considerable. The unstressed syllables resemble those of the statement and reeling-off member with all subjects, i.e. they are low and falling (or falling and slightly rising).

The stressed syllable is steeply rising with all subjects, and the extent of the rise is greater than in sentence initial pipipí.

(iv)   Conclusion

We may conclude that the stress/F_O relationships observed in the nonsense words recur in the natural words. Further, the general impression is one of falling unstressed syllables all over and of falling stressed syllables in the declarative, rising in the interrogative utterances. In the non-terminal words Tíflis is falling, statistík is rising.

3.2.1.2   Intonation contours in the words Tíflis and statistík

It may not be proper to speak of intonation contours in isolated words, intonation contour being defined as the figure described by a succession of stressed syllables in a (simple) sentence. However, in utterances containing only one stressed syllable the F_O movement in that syllable, together with its relative position on the frequency axis may be an adequate definition of intonation contour.

BH is very straightforward to describe: The stressed syllable is falling in the declarative, rising-falling in the non-terminal words, and steeply rising in the interrogative word. The position on the frequency axis becomes progressively higher as we go from declarative, through non-terminal, to interrogative words. The distance between unstressed and stressed syllables increases progressively in the same fashion.

SH deviates slightly from the simple pattern of BH in that declarative and non-terminal Tíflis look very much alike, and the stressed syllable of non-terminal statistík is rising.

NRP deviates from BH in that declarative and non-terminal statistík look alike, and both have rising stressed syllables. Furthermore, the stressed syllable of non-terminal Tíflis is lower than that of declarative Tíflis, and the stressed syllable of interrogative Tíflis is not higher than that of declarative Tíflis, but it is rising, and the succeeding unstressed syllable is considerably higher than in the other two words.

Conclusively it can be said that declarative isolated words are characterized by low and falling stressed syllables. The distance to any preceding (lower) and succeeding (higher) un-stressed syllables is rather small. Interrogative isolated words are characterized by high and rising stressed syllables. The distance to any preceding (lower) and succeeding (higher) unstressed syllables is rather great. Non-terminal words may resemble the statements or they may be intermediate between de-clarative and interrogative words.

3.2.2 Declarative sentences

There were six statements in the material (cf. section 2.1.2). Two of these have been accounted for already (cf. sec-tion 3.1.2 and fig. 9). The other four, where Tíflis and sta-tistík occur sentence initially and -medially, only confirm the results and the model in fig. 9 with regard to stress/F_o rela-tions as well as to intonation contour: The intonation contour describes a smooth fall from the first stressed to the last stressed syllable in the sentences.

3.2.3 Interrogative sentences

The intonation contours in the questions will be described in relation to the statement contour. Figs. 12-15 are tracings
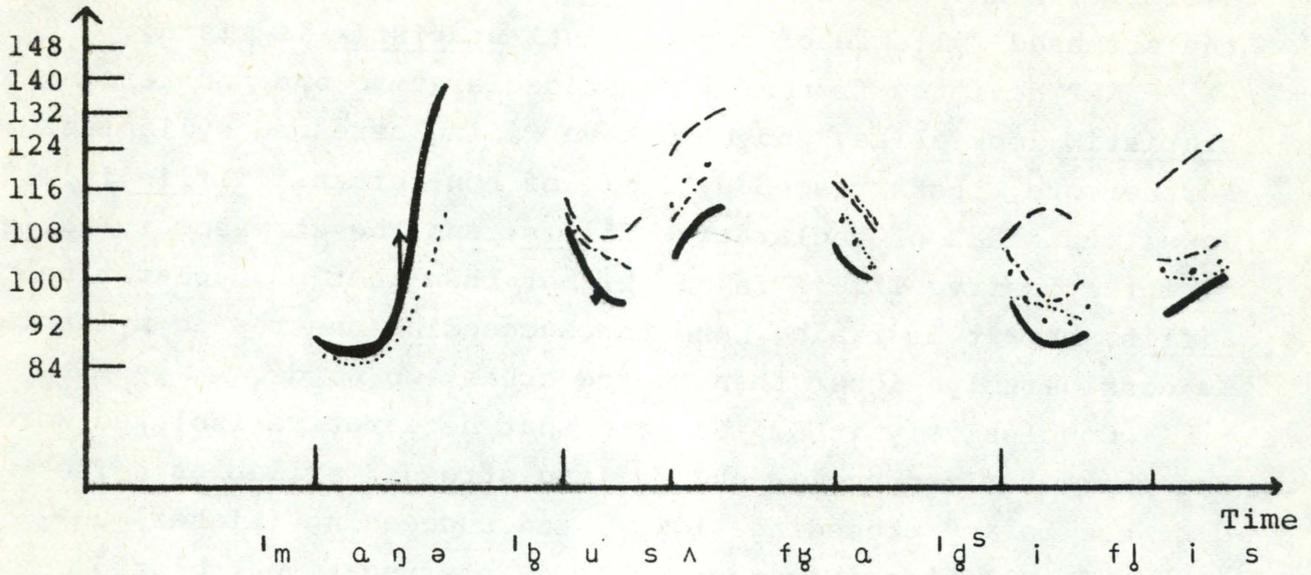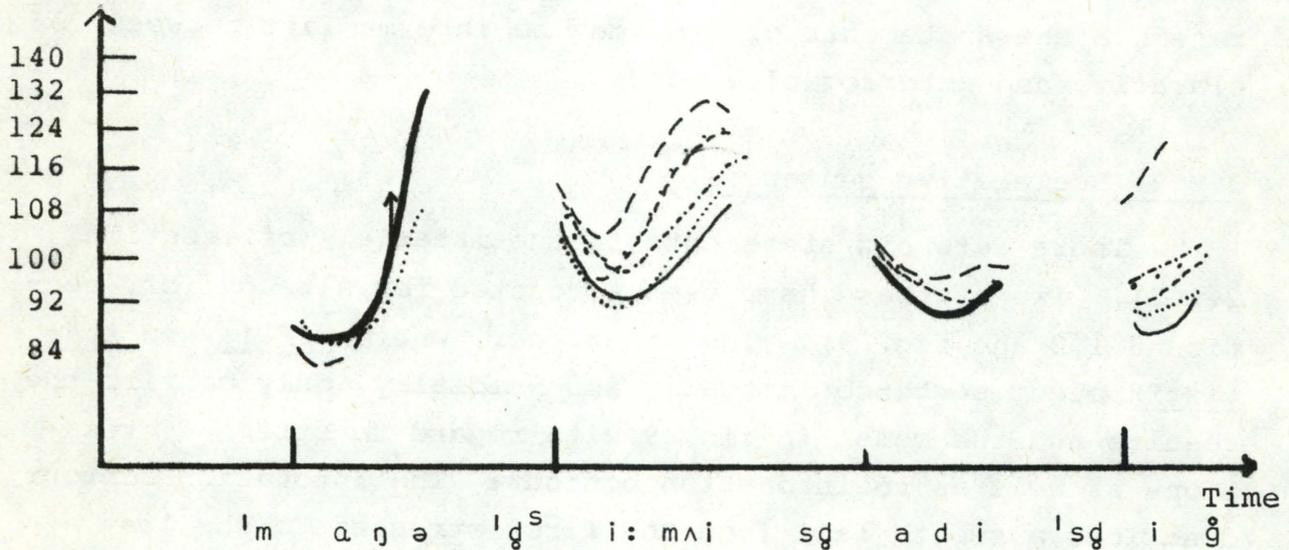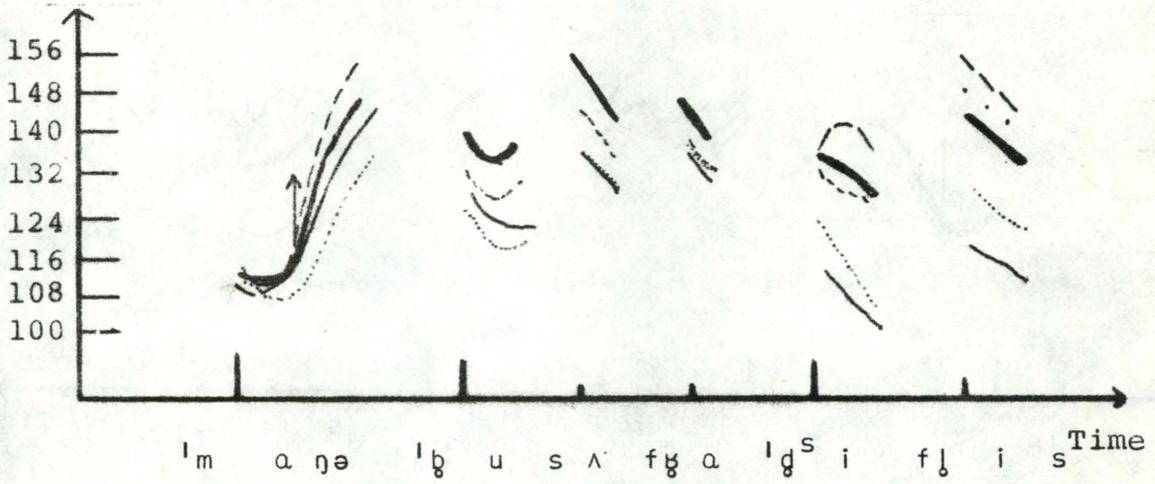
Figure 12

(Subject NRP)

F$_\circ$ tracings (averages of five recordings) of the sequence '.. mánge bússer fra Tíflis' (above) and '.. mánge tímer i statistík' (below) in 6 editions: one statement and five different types of questions. See the text for an account of the lines.
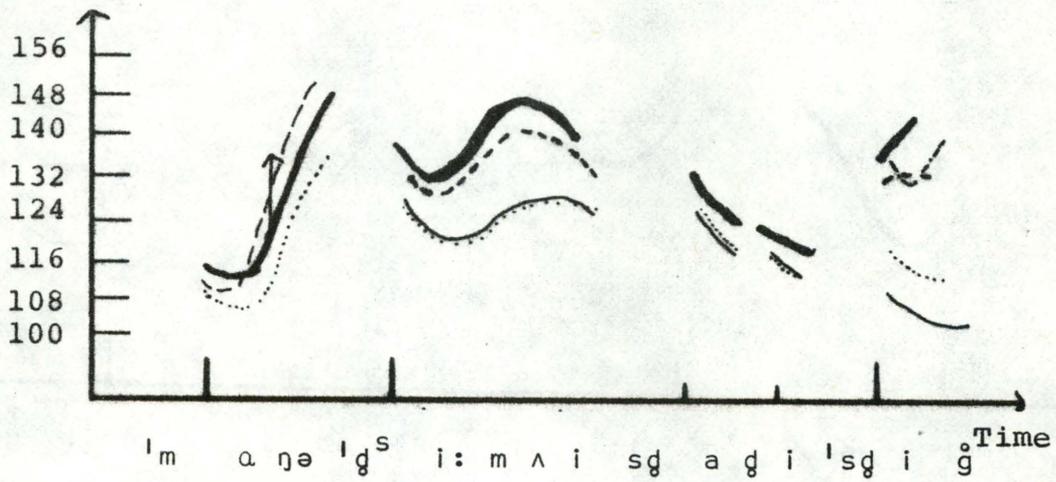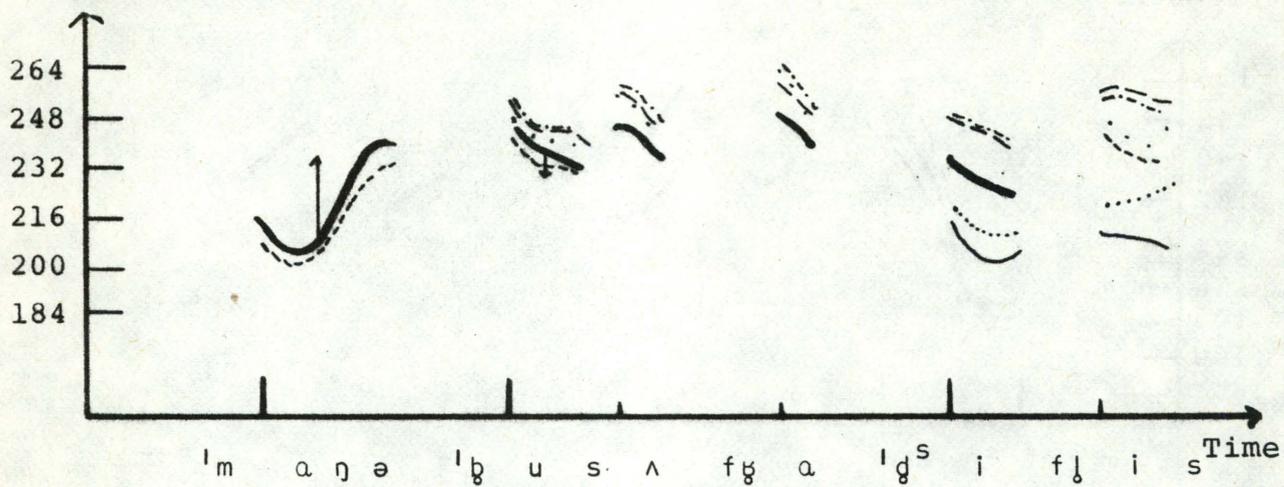
129

Figure 13
(Subject SH)
See the legend to figure 12.

Frequency in Hz

ˈm ɑ ŋ ə ˈb u s ʌ fʏ ɑ ˈdˢ i fʏ i s Time

Frequency in Hz

ˈm ɑ ŋ ə ˈdˢ i: m ʌi sʏ aʏ i ˈsʏ i ʏ Time
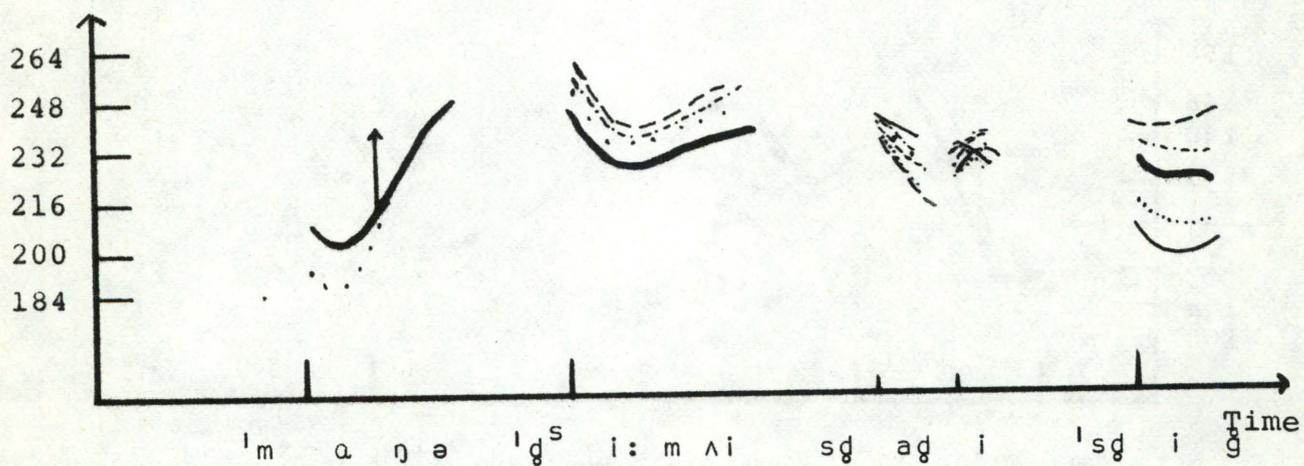
Figure 15

(Subject JG)

See the legend to figure 12.
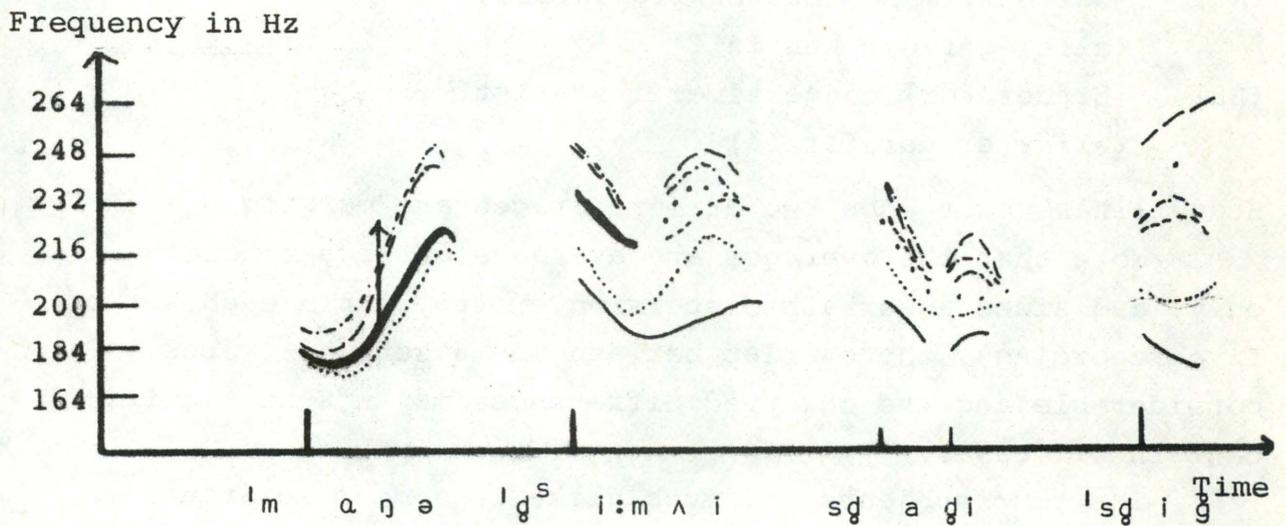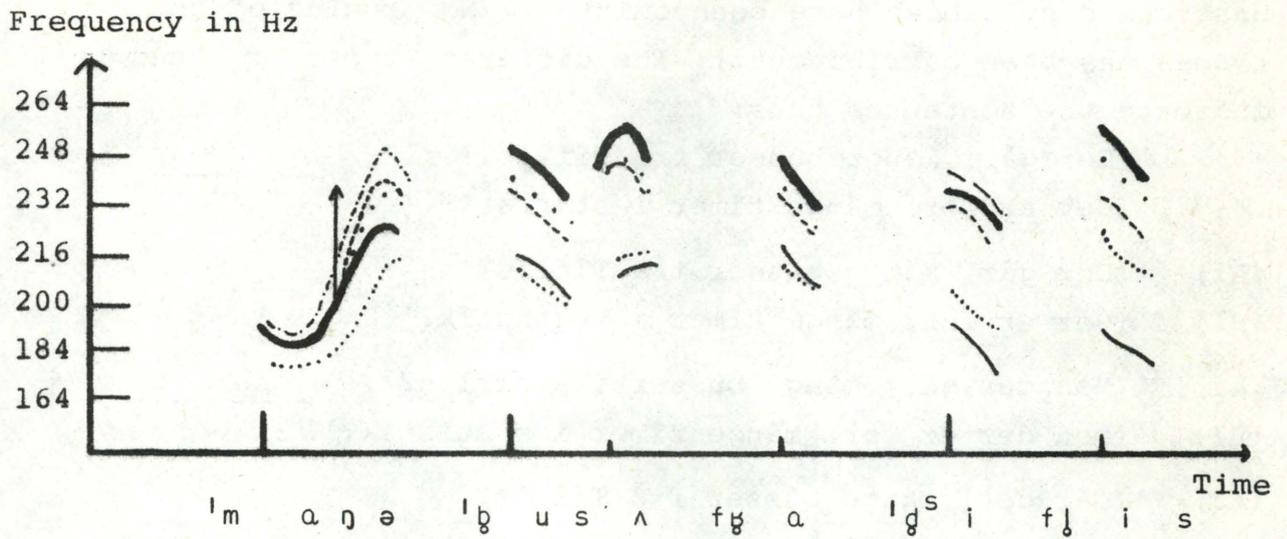
Figure 14

(Subject BH)

See the legend to figure 12.

(averages of five recordings) of the F_o courses in the declara-
tive and five interrogative sentences. (The sentence initial
unstressed syllables have been omitted. No pruning of the
traces has been carried out.) The different types of lines
indicate the sentences thus:

'(Der går) mánge bússer fra Tíflis.'

'(Der er for) mánge tímer i statistík.'

(A1) '(Der går) mánge bússer fra Tíflis?'

(B1) '(Der er for) mánge tímer i statistík?'

(A2) '(Mon der går) mánge bússer fra Tíflis?'

(B2) '(Mon der er for) mánge tímer i statistík?'

(A3) '(Går der) mánge bússer fra Tíflis?'

(B3) '(Er der for) mánge tímer i statistík?'

(A4) '(Hvórfor går der) mánge bússer fra Tíflis?'

(B4) '(Hvórfor er der for) mánge tímer i statistík?'

(A5) '(Går der) mánge bússer fra Tíflis,
      (eller går der kún fá)?'

(B5) '(Er der for) mánge tímer i statistík,
      (eller er der for fá)?'

Heavy lines occur when two or more traces are merging.

Note that the averages are averages of five recordings
only, and since a certain dispersion exists within each set of
five recordings, the overlap between the 6 sets of traces is
considerable and the observed differences may not be signifi-
cant in all cases.

The four subjects are much alike in some (important) re-
spects, but they differ in others:

With NRP and JG the differences in the F_o courses appear
clearly only in the second and third stress groups, and with JG
they are even rather small in the second stress group. (In the
sentences A4 and B4 ('Hvórfor går der mánge .../ Hvórfor er der
for mánge ...' (closely dotted line)) mánge is the second stress

group, which is reflected, by the lower position, in NRP's tracings.)

SH and BH have small differences in the first stress group and rather large differences throughout the rest of the sentences. mánge of sentences A,B4 ('Hvórfor ...') is also lower than in the remaining sentences with these two subjects.

The F_o courses are (remarkably) scattered in the figures, a scattering which must reflect several different interrogative intonation contours (related to the semantic/syntactic function of the sentences, cf. below). (NRP may be an exception, though! His intonation questions (A,B1) are rather clearly separated from the other questions and the statement, which in their turn are relatively low and close together, at least in the last stress group. Perception experiments will reveal whether the contours in these other questions are perceived as being inter-rogative at all. If not, then it seems as if NRP only utilizes an interrogative contour when absolutely necessary, i.e. when it is the only cue distinguishing the question from a statement. However, the four questions and the statement do not merge, and their order of succession on the frequency axis does follow the same pattern, as far as a pattern can be established, as the courses of the other subjects.)

Keeping the abovementioned differences between NRP and JG on one hand and SH and BH on the other in mind, the F_o courses (i.e. F_o in stressed as well as unstressed syllables) in the sentences can be accounted for in the following manner:

The lowest course is always the statement and the highest is always the intonation question (A,B1), provided they do not merge with some other sentence. The lowest but one is the 'hvórfor' question (A,B4). - The remaining three questions do not form a clear pattern which recurs with all subjects. With SH they are all rather high and close together and with NRP they are, as mentioned above, all rather low and close together. There is a tendency, however, for the non-terminal clause (A,B5)

to be the highest of the three. This is so for NRP and JG
(A5 is very slightly above and almost merges with the intona-
tion question with JG). With SH and BH only B3 is higher than
B5 (merging with the intonation question with SH), whereas A3
and A5 merge. There is also a tendency for A,B2 ('Mon' questions)
to be the lowest of the three. This is so with NRP (A3 merges
with the statement) and BH and also with SH and JG, insofar as
it does not merge with one of the other questions.

Looking at the statements and the questions as singular
events, their contours may be described thus: The intonation
contour, i.e. F_o of the stressed syllables, describes a smooth
slanting slope in statements. (The order of magnitude of this
fall is about 30 Hz for NRP and SH, about 40 Hz for JG and
50 Hz for BH, which is seen most easily in figs. 4-7.) In into-
nation questions the contour seems to be almost horizontal.
In the remaining types of questions the contours are more or
less slanting slopes between these two extremes. The steepest
fall is found in the 'hvórfor' questions. Cf. fig. 16. With
some subjects some of the contours in the interrogative sen-
tences may not be identified as interrogative contours at all.

Thus it seems as if there is a tendency for the intonation
contour to be more steeply falling (i.e. more statement-like)
the more syntactic information about the interrogative status
is contained in the sentence. This presupposes that an inter-
rogative particle overrides word order inversion with respect
to 'interrogative information', which is not unreasonable in
light of the fact that word order inversion is also found in
non-interrogative sentences.

3.2.3.1 Stress and F_o in interrogative sentences

The different slopes of the intonation contours may in-
fluence the F_o relations between stressed and unstressed syl-
lables, and any changes in relation to the statements would be
most marked in the intonation questions which are treated in
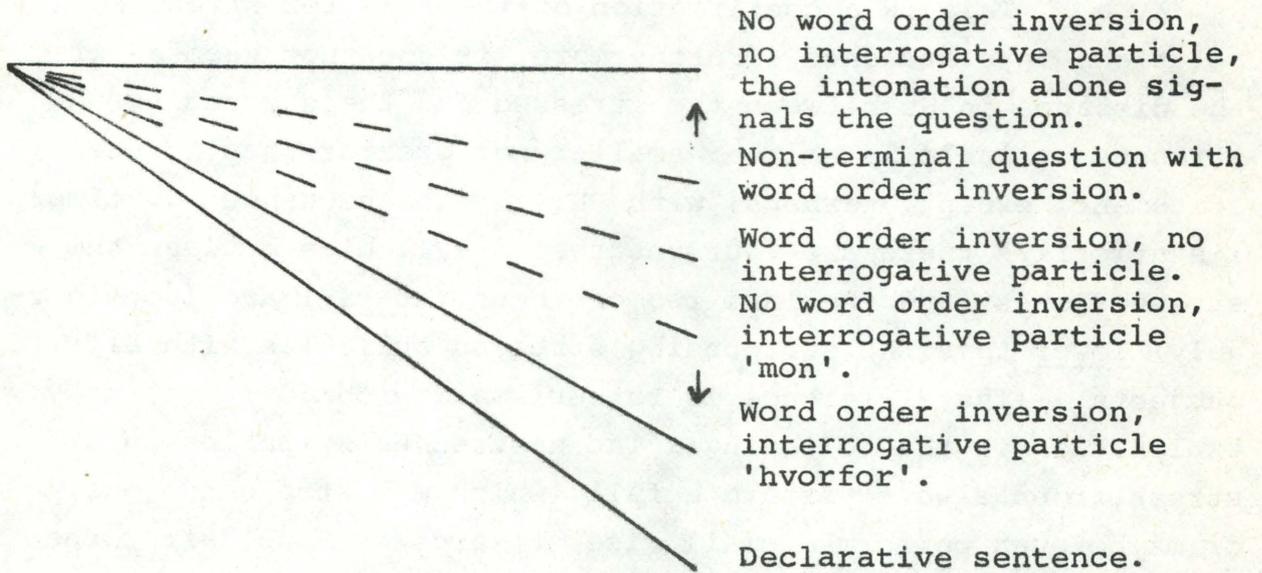the following:

No word order inversion, no interrogative particle, the intonation alone signals the question.

Non-terminal question with word order inversion.

Word order inversion, no interrogative particle.

No word order inversion, interrogative particle 'mon'.

Word order inversion, interrogative particle 'hvorfor'.

Declarative sentence.

Figure 16

Stylized graph of the intonation contours in declarative sentences and in five different types of interrogative sentences.

The last two stress groups in each of the two intonation questions show a difference from the F_o course in the statements with all four subjects: In the sentence '... bússer fra Tíflis?' there are only two unstressed syllables between the stressed ones, and they seem with all subjects to be higher than both surrounding stressed syllables if one introduces a correction (even if slight) for intrinsic F_o differences between [i] and [u, ʌ, a]. This is a confirmation of the rule for stress/F_o relations in statements. Furthermore, it does not seem as if the distance in Hz between the stressed and the succeeding un-stressed syllable is neither smaller nor greater than in the statement, except, perhaps, with BH. In the sequence '.. tímer i statistík?' there are four unstressed syllables between the stressed ones, and the last two of these (stati-) are (consider-ably) lower than the surrounding stressed syllables with all subjects. (The difference is rather small with JG.) A likely explanation is the following: The unstressed syllables in a stress group always perform a fall (which may straighten out, or maybe even perform a small rise, towards the end) after the initial rise from the preceding stressed syllable. The slope of this fall may vary (cf. the difference between SH as against NRP, BH, and JG in fig. 8, and the full versus broken line in fig. 9) and if there are enough unstressed syllables in a stress group the last ones will be below the succeeding stressed syl-lable. The higher this stressed syllable is in relation to the preceding stressed syllable, the 'sooner' the level of the pre-ceding unstressed syllables will be brought below it. This is, in fact, corroborated by the traces in figs. 12-15 (and 17-20).

If the description of the stress/F_o relationship is correct, i.e. the same basic F_o pattern recurs with slight modifications which are predictable from the intonation contour, then the definition of intonation contour as the contour described by the stressed syllables alone seems justified. - This does not mean, of course, that the F_o course of the unstressed syllables is irrelevant for the perception of a given contour, but it is, strictly speaking, redundant.

3.2.4 Non-terminal clauses

The intonation contours in the non-terminal clauses will also be described in relation to the statements, cf. figs. 17-20. The different types of lines indicate the sentences thus:

'(Der går) mánge bússer fra Tíflis.' _____

'(Der er for) mánge tímer i statistík.'

(A6) '(Der går) mánge bússer fra Tíflis,
(så vi kan gódt lade bílen stá).'

(B6) '(Der er for) mánge tímer i statistík, — — — — — —
(så vi bliver nødt til at strýge et
par stýkker).'

(A7) '(Hvis der går) mánge bússer fra Tíflis,
(kan vi gódt lade bílen stá).'

(B7) '(Hvis der er for) mánge tímer i sta- · · · · · ·
tistík, (bliver vi nødt til at strýge
et par stýkker).'

(A8) '(Går der) mánge bússer fra Tíflis,
(kan vi gódt lade bílen stá).'

(B8) '(Er der for) mánge tímer i statistík, —·—·—·—·—
(bliver vi nødt til at strýge et par
stýkker).'

The reservation mentioned in section 3.2.3 about the averages must be kept in mind.

The four F_O courses are dispersed to a greater or lesser degree with each subject in much the same fashion as in interrogative sentences, except that the dispersion in the first stress group is small or non-existent with all subjects. The mutual relations between the four courses are almost exactly the same for all subjects (insofar as traces do not merge completely): The lowest F_O course is the statement, next (from bottom to top) sentences no. 6, 7, and 8 (cf. above).
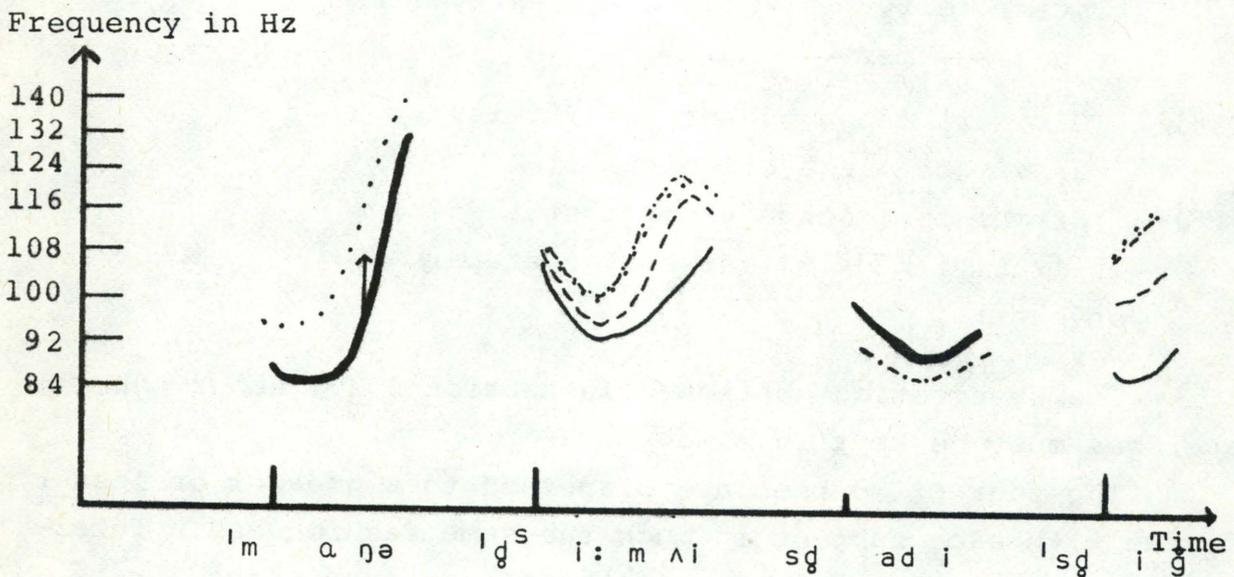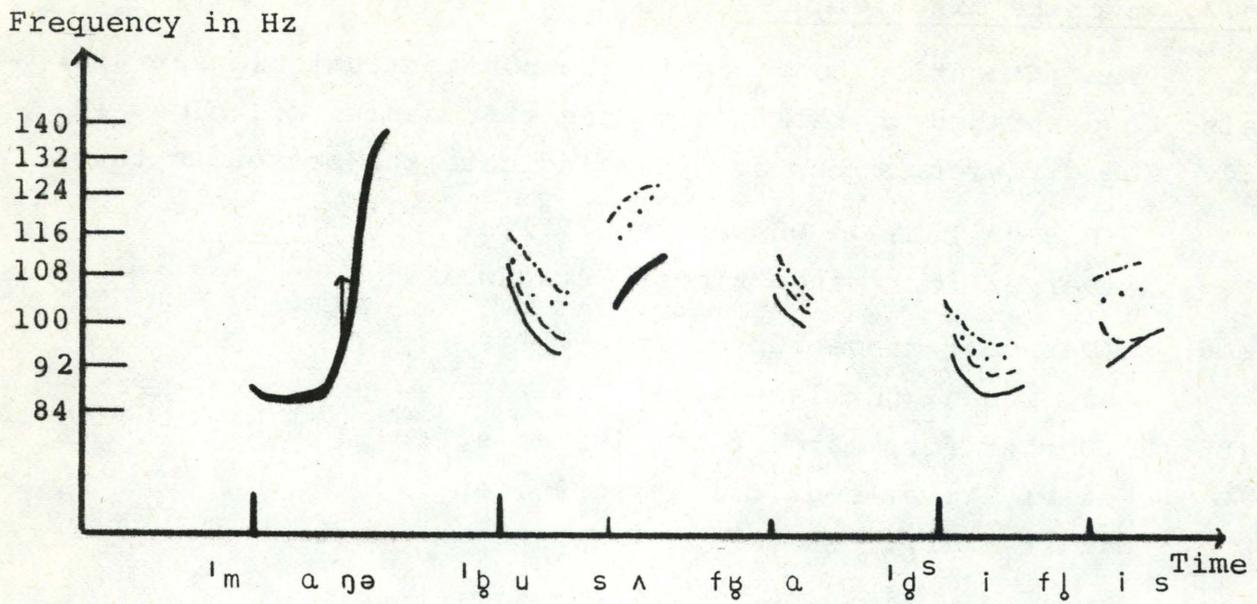
Figure 17

(Subject NRP)

F_o tracings (averages of five recordings) of the sequence '.. mánge bússer fra Tíflis' (above) and '.. mánge tímer i statistík' (below) in 4 editions: one statement and three different types of non-terminal clauses. See the text for an account of the lines.
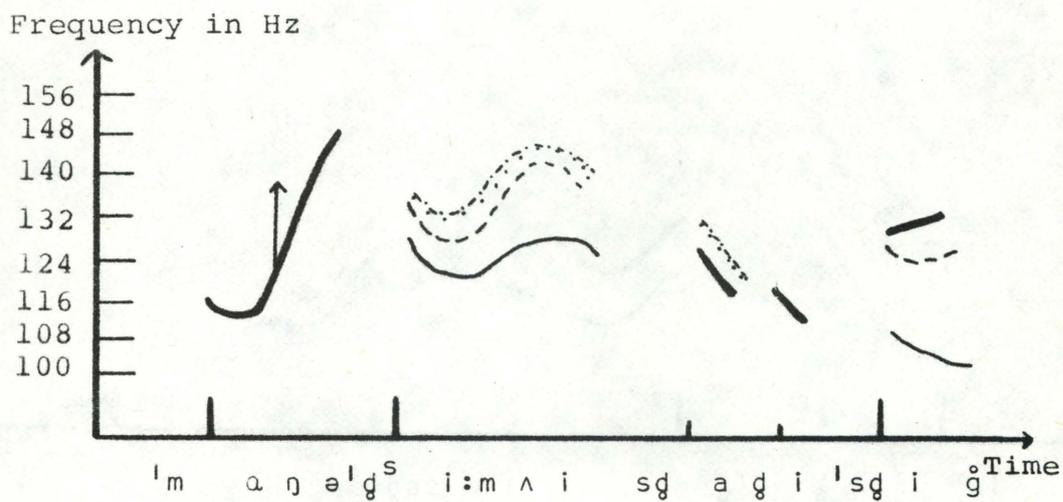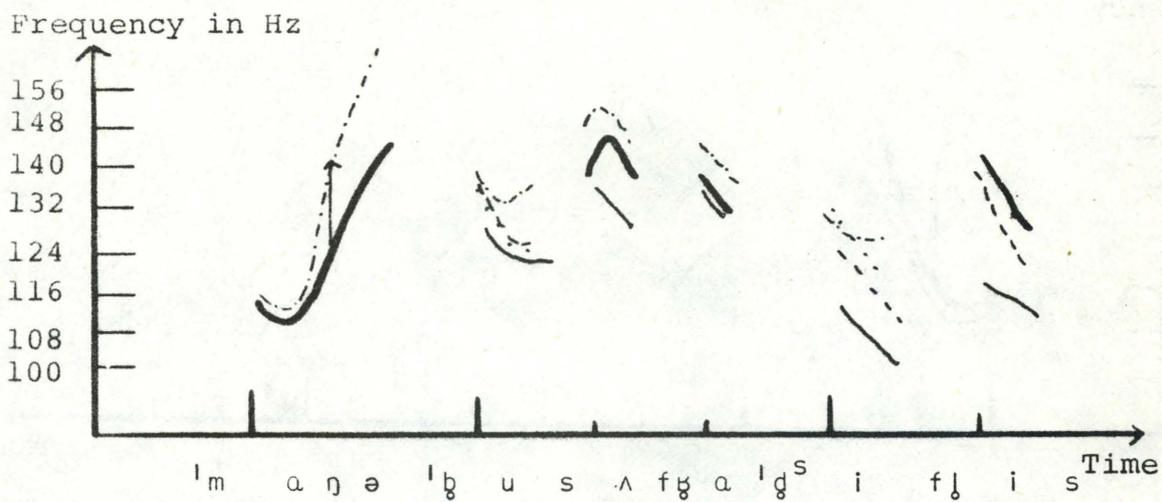
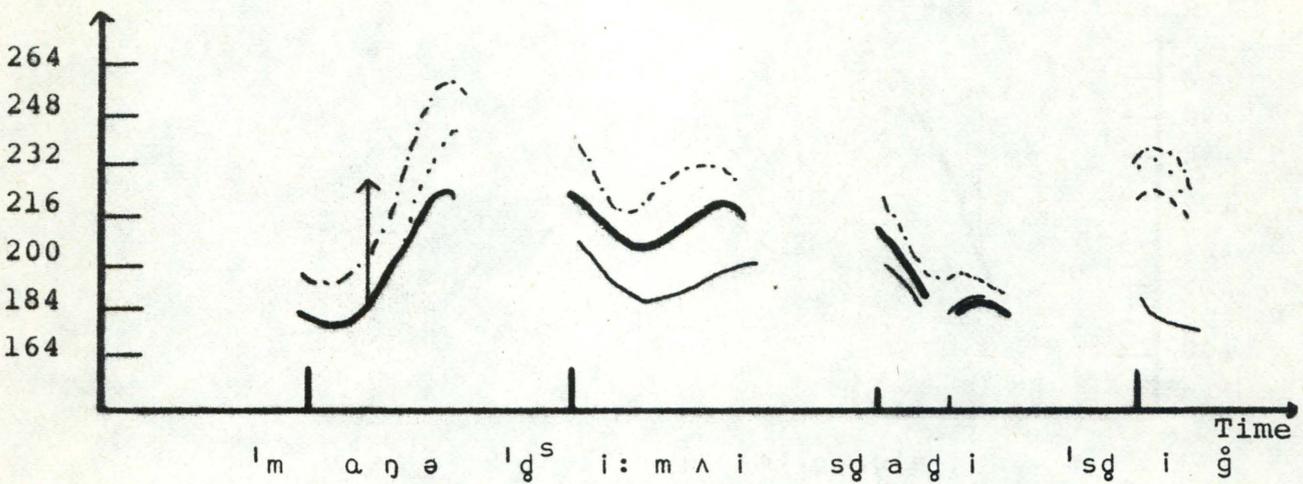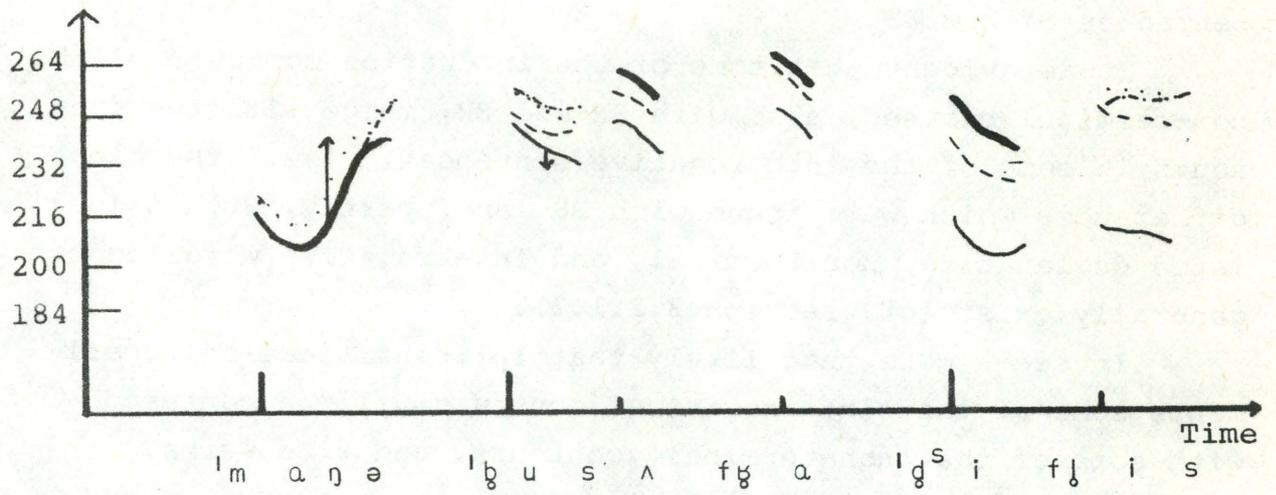Figure 18

(Subject SH)

See the legend to figure 17.

Figure 19
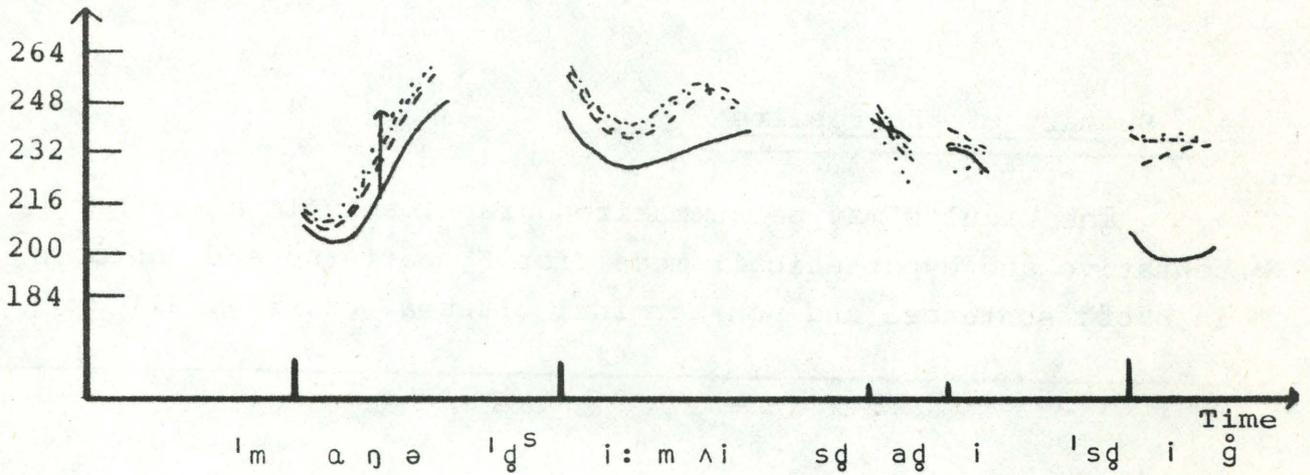
(Subject BH)

See the legend to figure 17.

Figure 20

(Subject JG)

See the legend to figure 17.

With JG the three non-terminal courses are all relatively high and close together, - in fact, they are as high as her intonation questions, cf. fig. 15. The same is true of NRP, sentences B7 and B8.

It is evident that some of the intonation contours of the non-terminal clauses, also with SH and BH, merge with the contours in some of the interrogative sentences. I.e., the clear differences which were found with BH (and, partly, SH) in iso-lated declarative, non-terminal, and interrogative words do not generally exist (cf. section 3.2.1.2).

It seems more than likely that in identification experi-ments some of the 'interrogative' contours will be confused with some of the 'non-terminal' contours, and vice versa. It is possible, also, that the intonation contours will tend to be classified in two, rather than three, groups as terminal versus non-terminal, i.e. subjects may not be able to distinguish interrogative from non-terminal intonation (and such a distinc-tion may not exist with the speaker, either).

4. Summary of the results

The results may be summarized graphically in a (rather tentative and hypothetical) model for F_0 patterns and contours in short sentences and non-terminal clauses, cf. fig. 21.

Figure 21

A model for F_0 in short sentences and non-terminal clauses:
1) Questions where the only cue distinguishing the sentence from a statement is intonation, 2) questions with word order inversion, non-terminal clauses in compound sentences,
3) questions with interrogative particle, 4) declarative sentences.
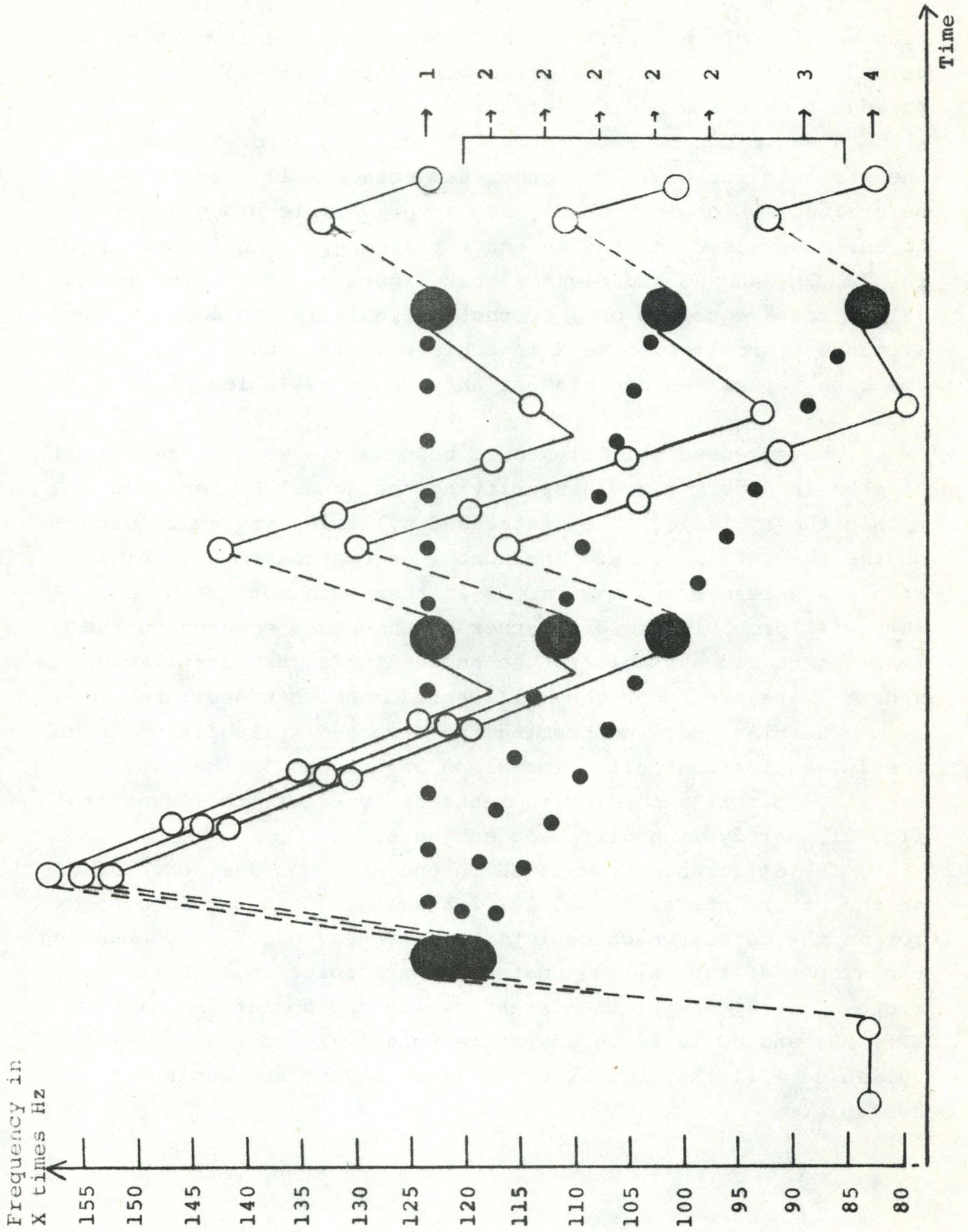
This model is, of course, a greatly stylized and simplified version of the facts as they were found and it must be regarded as provisional. Future experiments will probably lead to adjustments and modifications.

A scale has been tentatively indicated. For X equal to one it would fit NRP. For other subjects X will have to be determined, which is probably not a very simple procedure. If one measures F_o of the second stressed syllable in statements for SH, BH, and JG and divides that figure by 100 (which is the value, for X equal to one, of that syllable in the model), then a coarse approximation to X is achieved, although it predicts too high values for the highest unstressed syllables of BH and JG.

The stressed syllables have been marked with large filled circles in order to avoid specifying the actual F_o movement within the syllable. The unstressed syllables are small circles on the thin, full lines. The number of unstressed syllables within a stress group may vary. If there are more than four they will probably crowd together on the lines connecting the first posttonic syllable to the succeeding stressed syllable, and if there are fewer they will spread out on those lines.

The black dots connecting the stressed syllables describe the intonation contours. The slope of the one in the middle may vary, partly according to sentence type, cf. the legend to fig. 21, partly with different subjects.

The difference between SH on one hand and NRP, BH, and JG on the other (cf. section 3.1.1.2 and fig. 8) is not accounted for in the model, which depicts the strategy of SH. By changing the slopes of the falling unstressed syllables in each stress group, i.e. by making them steeper, the appropriate model for NRP, BH, and JG is achieved, since relatively more unstressed syllables will thus be below the level of the succeeding stressed syllable.

The course of the unstressed syllables rests on three
assumptions (which cannot be rejected on the basis of the pre-
sent material), namely 1) that the difference in Hz between
a given stressed syllable and the immediately succeeding un-
stressed syllable is, for a given subject, the same, irrespective
of intonation contour; this difference decreases from the first
to the last stress group. 2) The slope of the unstressed syl-
lables is the same (for a given subject) in all stress groups,
which is why the last (of several) unstressed syllables are
below the level of the succeeding stressed syllable in non-
statement contours, and most clearly so in the second stress
group, and 3) unstressed syllables where /ə/ elision or -assimi-
lation has taken place are assimilated tonally to the preceding
syllable.

The model covers single word utterances (including reeling-
offs) if the initial low unstressed syllables are connected to
the second or third stress group (depending on the individual).

Presumably, the model would also cover slightly longer
sentences, containing four and five stress groups, which would
just cause a given falling contour to be slightly less steep,
its upper and lower limits remaining the same.

5. Perspectives for the future

Some of the investigations which at present seem called for
are listed below. Some concern universal phenomena, others are
specific to Danish. When these investigations are undertaken,
they will probably in their turn bring new problems to light
and create a demand for still further research.

We need to know more about:

Intrinsic F_0 in unstressed syllables

Intrinsic F_0 in stressed and unstressed syllables in various
positions in the sentence

F_O variations caused by surrounding sounds (there are various
        papers on this subject but it is far from exhausted yet
        (see Jeel 1975 and the references therein))

The F_O course of varying numbers of unstressed syllables within
        a stress group, in different positions in the sentence

The mechanism of tonal assimilation in words with unstressed
        /ə/ syllables

F_O in syllables with secondary stress, as opposed to main- and
        unstressed syllables

Intonation contours in compound sentences, and the dependence
        of stress groups upon clause boundaries

Intonation contours in free speech.

        The perceptual relevance of F_O courses needs to be estab-
lished. There are several, more or less sophisticated, lines
of procedure:

Trained phoneticians could take down (i.e. transcribe) pitch/
        intonation as discrete phenomena (i.e. by means of lines
        and dots) and as continuous phenomena (i.e. in as good
        accordance with the perceived course as possible)

Trained phoneticians could beat rhythm and mimic intonation
        with a succession of identical CV syllables (i.e. they
        would hear, or produce themselves, a given sentence,
        whereafter they mimic it solely with CV syllables)

Listeners should adjust sinus- or triangular waves to the pitch
        of a given vowel (or consonant) in a given syllable

Listeners should identify the various F_O courses in the sequences
        '.. mánge bússer fra Tíflis' and '.. mánge tímer i
        statistík.' These sequences can be mutilated in various
        ways: One can cut out words in the beginning of the
        sequence in order to find out how much information is
        necessary for the identification, and one can cut out
        words from the end to find out how early identifiable
        differences in the sequences occur

The difference limen for F_O in connected speech should be estab-
    lished. How well do listeners localize a given perceived
    difference in relation to the chain of segments? How are
    several simultaneous differences perceived?
Are F_O courses in voiced obstruents and (non-syllabic) sonorants
    of any perceptual relevance?

    Finally, other varieties of Danish, besides Advanced
Standard Copenhagen Danish, should be investigated.

References

Basbøll, Hans   1968:   "The phoneme system of Advanced
                               Standard Copenhagen", ARIPUC 3,
                               p. 33-54

Jeel, Vivi   1975:   "An investigation of the fundamental
                               frequency of vowels after various
                               consonants, in particular stop
                               consonants", ARIPUC 9, p. 191-212

Malmberg, Bertil   1961:   "Haut-bas" ou "montant-descendant"?
                               Language and Society, Essays pre-
                               sented to A.M. Jensen on his 70th
                               birthday (Copenhagen)

Petersen, Niels Reinholt   "Intrinsic fundamental frequency
1976:   of Danish vowels", this issue.

SOME DATA ON NORTH GERMAN STOPS AND AFFRICATES

Eli Fischer-Jørgensen

Abstract: The present paper presents some data on voicing (3.1),
         duration (3.2), intra-oral air pressure (3.3), airflow
         (3.4), subglottal pressure (3.5), and lip pressure
         (3.6) of German stops and affricates. On the basis
         of these data the last section discusses some more
         general problems, viz. the features distinguishing
         ptk and bdg (4.1), the reduction of aspiration after
         s (4.2), the distinction between stops and affricates
         (4.3), differences due to place of articulation (4.4),
         "heightened subglottal pressure" (4.5), and the re-
         lation between air pressure and airflow (4.6).

1. Introduction

        The present paper brings some data on duration, intra-oral
and subglottal pressure, airflow, lip pressure, and (to a certain
extent) voicing of North German stops and affricates. The material
was collected in the course of the years 1955-1968. The main part,
comprising stops and affricates in stressed position (lists I and
II) was meant as a background material for a description of Danish
stop consonants (which has never been completed). A smaller part
(list III) comprising stops in unstressed position is taken from
a material intended for the description of close and open contact
in German (see Fischer-Jørgensen 1969) and not particularly set up
for the purpose of analysing stops.

        Although the material is restricted and, as far as the un-
stressed position is concerned, unsystematic, I have found it
useful to publish it in a separate paper because relatively little
has been published on North German stop consonant duration and,
as far as I know, hardly anything on air pressure and airflow.
Some (rather restricted) instrumental data on duration, particularly

the duration of aspiration, can be found in E.A. Meyer (1901 and
1904), Hentrich (1925), Weitkus (1931), Schmitt (1931 and 1947),
v. Essen (1934), Bennett (1935), and Rositzke (1944 and 1947).
Some scattered data can be found in Zwirner and Zwirner's text
lists (1936-37). A more comprehensive material on duration and
voicing (but only comprising one person's pronunciation of t and
d) is found in Bothorel-Witz and Pétursson (1972). The three
most comprehensive studies (Lotzmann 1958/59 on aspiration,
Meinhold and Stock 1963 and Esawa 1972 on voicing) are not in-
strumental (Meinhold and Stock have taken a few oscillograms
for control), but based on auditory impressions. Lotzmann has
not been accessible to me. Meinhold and Stock have listened to
more than 1000 bdg-sounds in absolute initial position and in
the position after voiceless sounds spoken by 34 professional
speakers and actors (in connected texts). They have made the
tape recordings at a speed of 30 ips and played them back at
very slow speeds, and they have only made decisions on presence
or absence of voicing. This seems to be a very reliable procedure.
Esawa has listened to voicing of bdg and ptk in various positions
in 90 samples of 5 minutes length from a connected text spoken
by 90 subjects from all over Germany.

The emphasis of this paper is on the presentation of data,
and I am not going into any details concerning the general problems
of stop features and stop production. A detailed discussion of
these problems can be found in my paper "Voicing, tenseness and
aspiration in stop consonants" (1968), and a more recent treatment
with a very useful survey of the literature is given in the paper
by Bothorel-Witz and Pétursson (1972). But it is, of course, my
hope that the data presented here may contribute not only to the
description of German (and, ultimately, to the teaching of German)
but also to the description of the distinctive features used to
separate the two stop categories designated as ptk and bdg, and
that it may throw some light on the mechanism of stop production.
In the final section I will return briefly to these problems.

2. Subjects, material, and recordings

2.1 Subjects

The investigation is mainly based on six subjects: CH, WS, KV, HP, and (to a somewhat lesser degree) GJ and GR. For subsidiary information about the duration of unstressed stops recordings made by three further subjects (HT, HWL, and HL) have been used. They all speak various types of Standard Northern German except for GJ, who is from the Western part of Germany (Koblenz), but speaks Standard German.

CH (born 1934) grew up in Lübeck and Stettin and had at the time of recording (1955) been in Stockholm for one year.

WS (born 1939) grew up in Velbert in the Northern Rheinland. He has lived in Copenhagen since 1961. The main recording was made in 1966, two others in 1964 and 1968. He speaks German at home, and his German seems uninfluenced by Danish.

KV (born 1941) grew up in Elmshorn near Hamburg. At the time of recording he had been in Denmark for six months.

HP (born 1907) lived in Hamburg until the age of 28. He has stayed in Copenhagen since 1935, i.e. at the time of recording (1956) for 21 years. He is married to a Dane, and he reported that he spoke Danish at home, but German at his job. In spite of his long stay in Denmark he was included among the main subjects because he was within reach and because his German sounded all right. However, the measurements showed that he differed from the other subjects on several points, which could be due to Danish influence. He should therefore not be taken as a typical German speaker.

GJ (born 1912) grew up in Koblenz. At the time of the recording (1955) he had been several years in Copenhagen. He spoke German at home, and his German seems to be completely uninfluenced by Danish.

GR (born 1912) grew up in Berlin. At the time of recording (1955) he had lived for seven years in Stockholm. He speaks German at home, and his German seems quite uninfluenced by Swedish.

HT and HWL are from Northern Germany; HL from the Ruhr area.

2.2 Texts

The main material (list I) consisted of a series of words with initial stop or affricate in stressed position, preceded by an unstressed vowel and followed by one of the vowels a, i or u

The words used in most of the recordings were:

I (a) b   die Bahre, die Busze, die Biese
        d   die Dame, die Dusel, die Diebe
        g   die Gabe, die Guten, die Giebel

        p   die Paare, die Pute, die Pike
        t   die Tafel, die Tute, die Tiefe
        k   die Kate, die Kugel, die Kiefer

I (b) pf   die Pfanne, der Pfuscher, die Pfiffe
        ts   die Zahlen, die Zuber, die Ziege
        p   die Panne, die Puppe, die Pille
        t   die Tasche, die Tunke, die Tinte

In the case of GJ, GR, and CH's first recording (CH I) an earlier version of list Ib was used containing the words:

        pf   die Pfanne, der Pfuhl, der Pfiff
        ts   die Zahl, die Zunge, die Ziege
        p   die Panne, der Putz, der Pilz
        t   die Tat, die Tunke, die Tiefe.

List Ia was intended for a comparison between ptk and bdg, list Ib for a comparison between pt and the affricates. In the second version all words were made disyllabic, and p and t were followed by short vowels and could be compared to the words with long vowels in list Ia. But in this way the t-words got less comparable to the ts-words. However, as the differences are significant in both cases, this does not matter.

A subsidiary word list (called II) consisted of the words:

I (c) bdg das Band, das Dach, das Gatt
             das Bett, das Deck, das Gift
        ptk das Paar, das Tal, das Kalb
             das Pech, das Tempo, das Kind

This list was only spoken by GJ, GR, and CH.

The words appeared twice in the list in groups of three, the second time in reversed order, to avoid influence from rhythmic factors.

List I was intended for comparison with Danish, and was restricted to initial position. It contained, however, by chance some stops in unstressed medial position, and these have been included in the present investigation. They were supplemented by some words from list III.

List II consisted of a list of isolated words. The words utilized in the present paper were: _piepen_, _Lippe_, _tapern_, _tappen_, _bieten_, _beten_, _bäten_, _baten_, _Betten_, _Latte_, _lieben_, _lebe_, _bibern_, _beben_, _Ebbe_, _bibbern_, _wieder_, _laden_, _Widder_, _Kladde_.

2.3 Recordings and measurements

Tape recordings were made of lists I, II, and III, spoken by GJ and GR, lists I and II by CH, and lists I and III by WS and KV. The subsidiary data from subjects HT, HL, and HWL were also based on tape recordings. The recordings were made on professional tape recorders in sound treated rooms, those of GR and CH at the Technical High School in Stockholm, the others in Copenhagen. The tape recordings were used for mingographic registration comprising oscillograms and intensity curves. These curves, supplemented by spectrograms, were used for measurements of duration.

As start of the consonant closure that point has been chosen where the preceding vowel ends, i.e. where the intensity curve drops abruptly. The delimitation of the open interval (i.e. the distance from release of the closure to the start of the following vowel) was sometimes dubious in the case of voiced _bdg_, and some cases have been omitted from the averages.

Air pressure and lip pressure of HP and CH (lists I and III) were recorded in the Cardiologic Laboratory of the University Clinic (Rigshospitalet) in 1956 in cooperation with Oluf M. Thorsen. The intra-oral air pressure was recorded by means of a small plastic tube, seven cm long, with an outside diameter of 1.5 mm

and a bore of O.8 mm inserted into the mouth. It was connected
to an electrical manometer (for more technical details see
Fischer-Jørgensen and Tybjærg Hansen 1959). Lip pressure was
recorded by means of a rubber bulb placed between the lips and
connected to the manometer. The registration was made on an
Elema oscillograph. The calibration was in mm H_2O. The velars
were left out in these recordings of air pressure.

Airflow curves comprised lists I and III spoken by HP and
KV. The instrument used was the aerometer built by Frøkjær-
Jensen who assisted at the recordings. As the instrument could
not be calibrated at the time, the measurements have been made
in mm. These curves have also been used for measurements of
duration.

A simultaneous recording of intra-oral and subglottal air
pressure and of airflow was undertaken at the Institute of
Phonetics in 1966. In this recording the tube used for picking
up the intra-oral air pressure was inserted through the nose.
The text was list I, and the speaker was WS. The subglottal
pressure was recorded by means of a balloon placed in the oeso-
phagus. This technique had been used by Schuhmacher (WS) for
his investigation of German vowels, and he was kind enough to
swallow the balloon once more for the recording of stops (for
more technical details see Schuhmacher 1972). Unfortunately,
the calibrations of these curves have disappeared, so that they
had to be measured in mm. As both air pressure and airflow
scales are linear, the relations between the values measured in
mm will be correct, and generally only the relations are of
interest, but for a comparison between the three types of regi-
stration the real values would have given more information. The
word list was repeated four times, but only repetitions number
three and four included subglottal pressure. For these two record-
ings the amplitude of the air pressure and airflow curves was reduced,
but since the number of words was the same in all recordings a
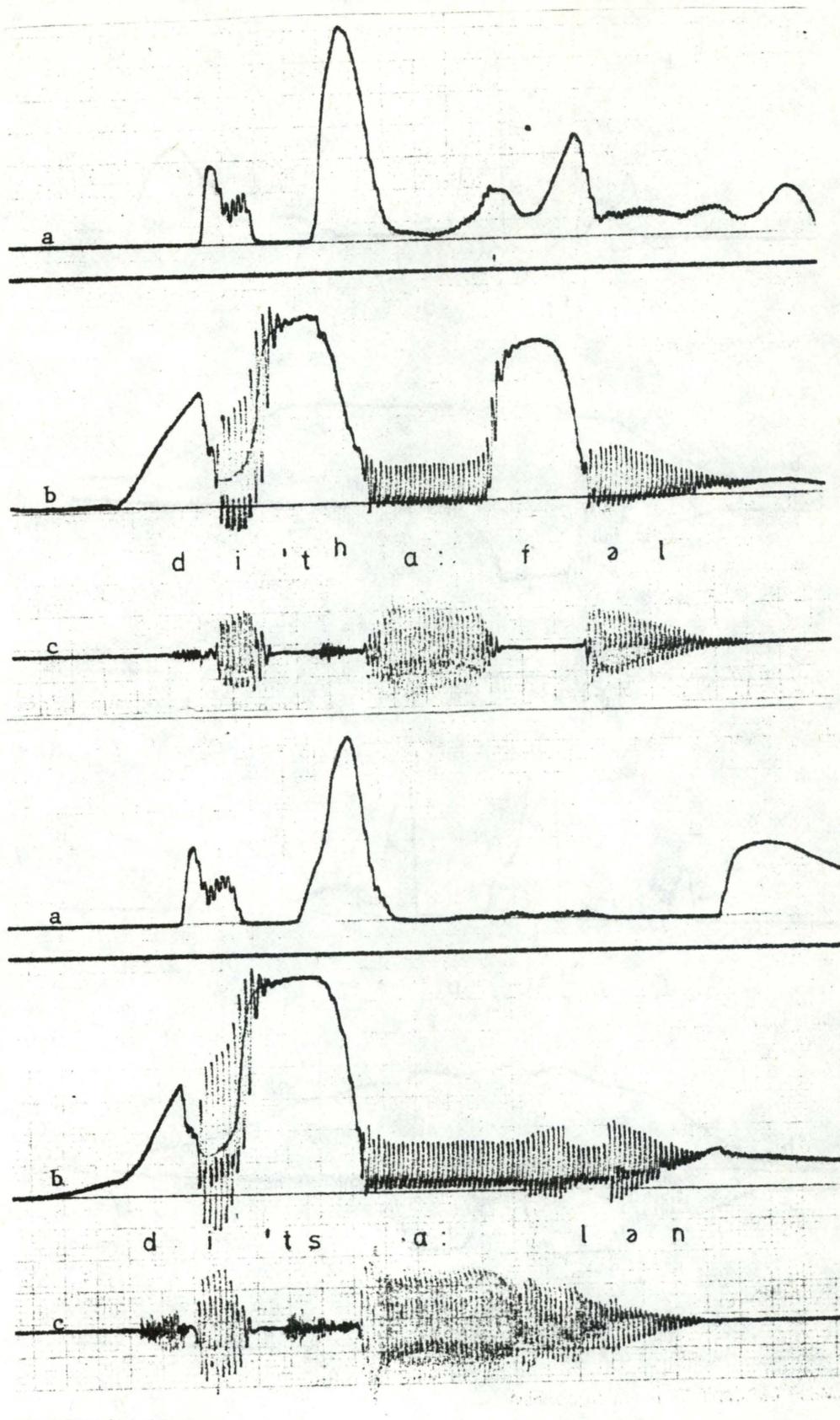common mean has been taken of all four recordings. A specimen of

Figure 1

a) airflow, b) intra-oral air pressure,
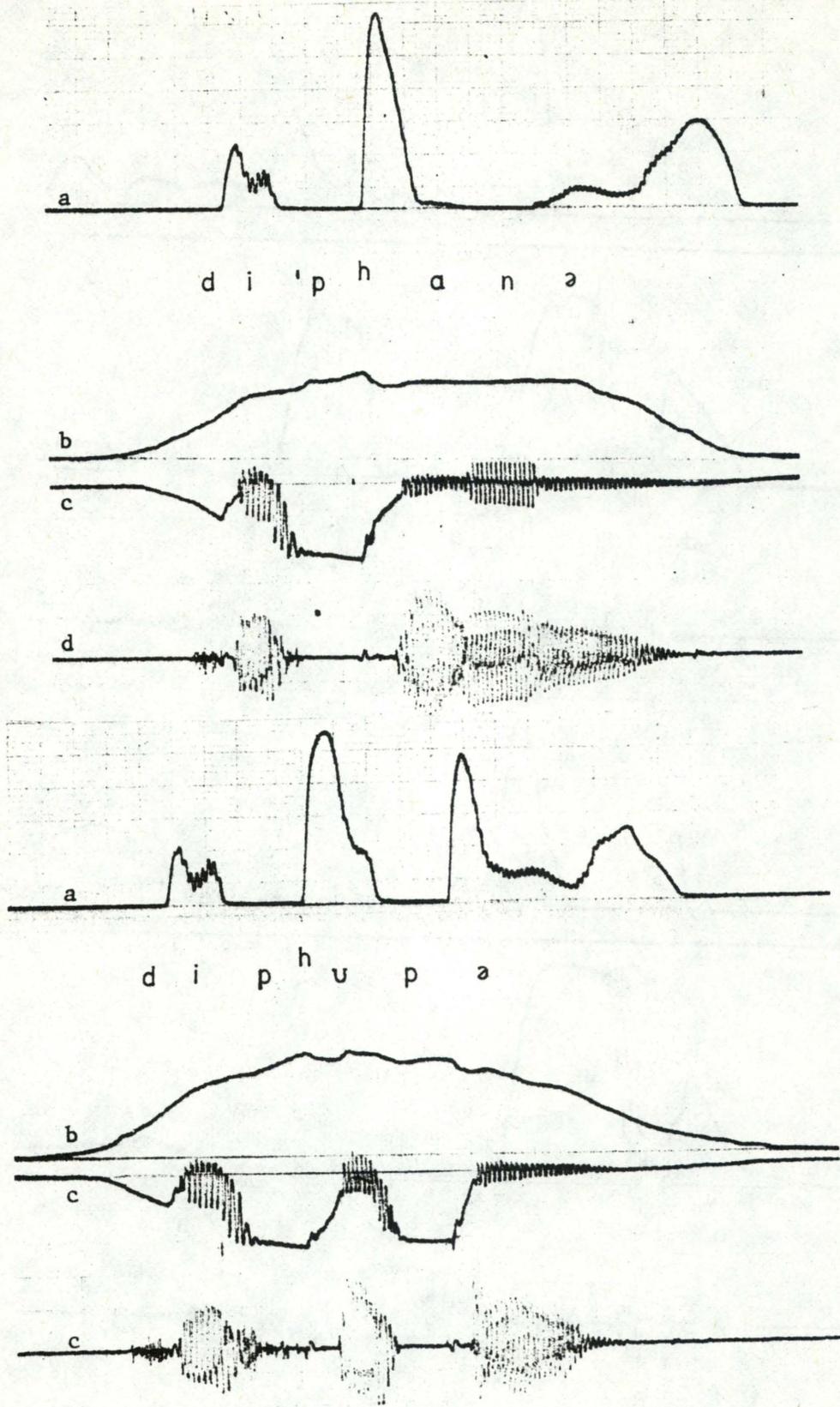c) oscillogram (subject WS).

d i 'pʰ a n ə

d i pʰ u p ə

Figure 2

a) airflow
b) subglottal pressure
c) intra-oral pressure (inverted)
d) oscillogram

these two types of curves is shown in fig. 1. These curves have also been used for measurements of duration.

The duration of voicing has been measured in all the curves of the six principal subjects, and in the case of total voicelessness and complete voicing the measurements are safe, but in the numerous cases of partial voicing with decreasing amplitude the delimitation is rather arbitrary, and depends to some degree on the amplification used in the individual recording. For a material recorded at different times and with different instruments the measurements can be compared only with great caution.

A good deal of the measurements and computations have been made, according to my instructions, by Oluf M. Thorsen, Kirsten Gregersen, Karsten Vogel, Jente Andresen, H.P. Jørgensen, and Mimi Jacobsen.

3. Results

3.1 Voicing

As mentioned in section 2.3 I am somewhat sceptical as to the exact measurement of voicing where it is only partial. This section will therefore not contain large tables of averages, but some information is necessary as background for the judgement of other parameters.

3.1.1 ptk

The stops ptk are always voiceless in this material except for 20-30 ms of weak voicing in the beginning after a voiced sound. But it should be kept in mind that the list consisted of isolated words. In connected speech ptk may be voiced (see Esawa 1972).

3.1.2 bdg

As for bdg their degree of voicing is very variable depending on stress and preceding sounds.

3.1.2.1 bdg in initial position in stressed syllable

(a) **Absolute initial position**

The main material (list I) did not contain any examples of this type (the d̲ of unstressed "die" was always voiceless). But there are a number of examples in list III. CH, HP, and KV always have voiceless bd̲g̲ in this position, GJ and GR normally have voiceless bd̲g̲ with a few random exceptions, whereas WS varies unsystematically between voiced and voiceless b̲ and d̲, but always has voiceless g̲.

(b) **Initial position after word boundary preceded by s**
(s#'CV)

The material only contains 30 examples of bd̲g̲ in this position spoken by CH, GJ, and GR. They are always completely voiceless. GJ has also voiceless bd̲g̲ after other voiceless sounds (see Fischer-Jørgensen 1952, p. 123).

(c) **Initial position in stressed syllable after word boundary preceded by unstressed vowel** (ˏV#'CV)

The main part of the material consists of words of this type. bd̲g̲ in this position are normally partly voiced, but there is a large variation between subjects and also in the recordings of the same subject and the same word. As a rough indication of the variation between subjects the following values may be given (duration of voicing in % of the closure):

GJ 75%, GR 61%, WS 54%, KV 49%, CH II 40%, CH I 28%,
HP II 20%, and HP I 14%.

GJ has a number of fully voiced examples, GR only one. The voiced portion is always found in the first part of the closure, never at the end, and the amplitude of the vibrations is always decreasing, which makes the delimitation dubious in many instances.

HP's percentage of voicing corresponds to 10-30 ms. The averages are often very slightly higher for bd̲g̲ than for pt̲k̲, but there is complete overlapping. His bd̲g̲ must therefore be considered voiceless in this position.

3.1.2 Medial position before unstressed vowel after stressed vowel ('VCə)

In this position bdg are often fully or almost fully voiced, with decreasing amplitude of the vibrations. This is true of all subjects: GJ, GR, CH, WS, and KV.[1] The voicing often stops at the release, but may also continue through the release. This is almost always the case for GJ and GR, and often for WS. The degree of voicing in CH's recording is very variable. She may have fully voiced or almost voiceless bdg. She has sometimes less voicing after short than after long vowel. In the case of HP there are great differences between different recordings. In one recording he has more than 50% voicing, in another the consonants are practically voiceless. KV has voiceless bdg after short vowel (e.g. Widder), fully or partly voiced bdg after long vowel.

The great variation in the voicing of bdg confirms the observations by Esawa (1972), who finds that complete or partial voicelessness is found very often, also in Northern Germany in all positions of the word, and it is also in agreement with the indications of Bothorel-Witz and Pétursson (1972) on the voicing of d. The finding that bdg are completely voiceless after s and, in most cases, voiceless in absolute initial position is in agreement with Meinhold and Stock (1963), who found only 1.3% of voiced bdg after a voiceless sound (and they consider these few examples to be cases of hypercorrection) and 22.6% voicing initially with great individual variation (the individual percentages ranging from 14 to 100).

---

1) Voicing was not measured for the subsidiary subjects.

160

Figure 3

closure of ptk
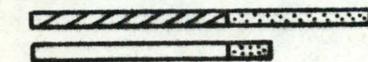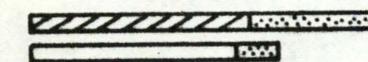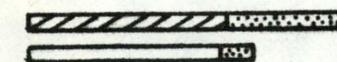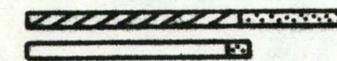closure of bdg
open interval

0   100   200 ms
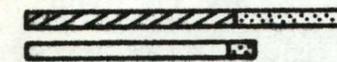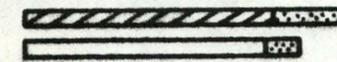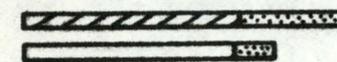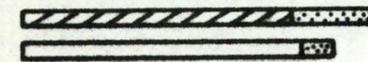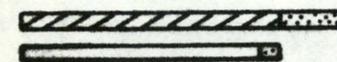
Table 1

Duration (in msec) of ptk and bdg
in the position V'CV:

S = subject, N = number, C = consonant,
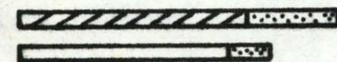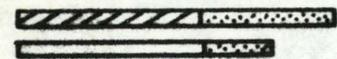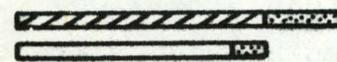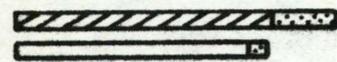cl. = closure, op.i. = open interval,
tot. = total duration, % op.i. = op.i.
in % of tot.

S	N	C	cl.	op.i.	tot.	%op.i.
GJ	(6)	p	131	32	163	20
	(6)	b	118	11	129	9
	(6)	t	129	37	166	22
	(6)	d	111	18	129	14
	(6)	k	94	67	161	42
	(6)	g	96	35	131	27
	(18)	ptk	118	45	163	28
	(18)	bdg	108	21	129	24
GR	(12)	p	134	32	166	19
	(11)	b	124	9	133	7
	(12)	t	139	42	181	23
	(11)	d	143	17	160	11
	(12)	k	113	55	168	33
	(12)	g	108	22	130	17
	(36)	ptk	129	43	172	25
	(34)	bdg	125	16	141	11
CH I	(6)	p	108	62	170	36
	(6)	b	106	11	117	9
	(6)	t	109	58	167	35
	(6)	d	105	10	115	9
	(6)	k	91	71	162	44
	(6)	g	86	30	116	26
	(18)	ptk	102	64	166	39
	(18)	bdg	99	17	116	15
CH II	(21)	p	113	66	179	37
	(21)	b	106	20	126	16
	(23)	t	100	73	173	42
	(18)	d	101	21	122	17

Figure 3
(continued)

0   100   200   ms
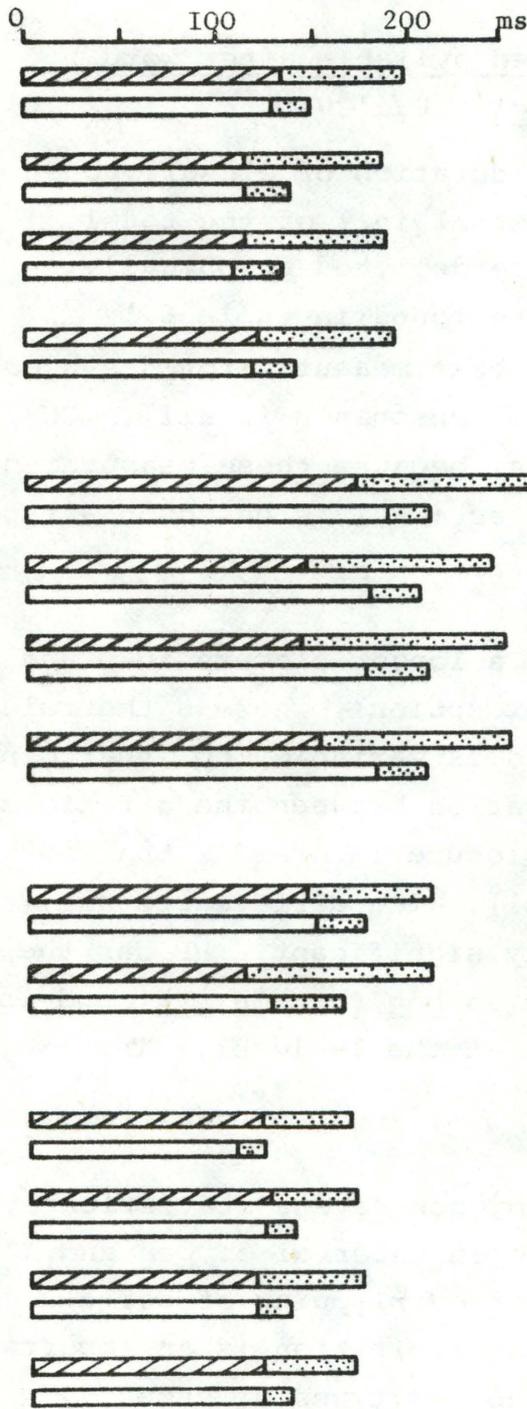
Table 1
(continued)

S	N	C	cl.	op.i.	tot.	%op.i.
WS	(24)	p	133	65	198	33
	(24)	b	129	19	148	13
	(24)	t	114	72	186	39
	(24)	d	114	24	138	17
	(24)	k	115	74	189	39
	(24)	g	108	26	134	19
	(72)	ptk	120	71	191	37
	(72)	bdg	117	23	140	16
HP	(18)	p	170	92	262	35
I	(18)	b	188	22	210	10
	(18)	t	145	96	241	40
	(18)	d	178	24	202	11
	(18)	k	142	108	250	43
	(18)	g	176	33	209	16
	(54)	ptk	152	99	251	39
	(54)	bdg	181	26	207	16
HP	(16)	p	144	66	210	31
II	(15)	b	149	26	175	15
	(18)	t	111	98	209	47
	(18)	d	126	37	163	23
KV	(12)	p	120	48	168	28
	(12)	b	108	15	123	13
	(12)	t	125	45	170	26
	(12)	d	122	16	138	12
	(12)	k	118	56	174	32
	(11)	g	117	18	135	13
	(36)	ptk	121	49	170	29
	(35)	bdg	115	16	132	13

3.2 Duration

3.2.1 Differences between ptk and bdg

3.2.1.1 Initial position in stressed syllable after word boundary preceded by unstressed vowel ($_{\circ}$V#'CV)

Table 1 gives a survey of the duration of closure, open interval, total length and open interval in % of the total length. In fig. 3 the same facts are depicted graphically. (The durations of pt from list Ib are found in table 6.)

Duration in this position has been measured for 6 subjects who have spoken list Ia and Ib (1092 consonants in all). CH II and HP II do not comprise the velars, because these measurements are taken from the air pressure curves (CH I is based on a tape recording, HP I on airflow curves).

(a) closure

Five of the six subjects have a longer closure in p and k than in b and g (GJ k < g forms an exception), whereas the relation between the closures of t and d is variable, cp. that Nina Thorsen (1971) found a variable relation between the durations of t and d in English, but longer closure in p and k than in b and g (in a more restricted material). The differences are, however, small and not statistically significant. HP has the opposite relation: a longer closure in bdg than in ptk, and in HP II the difference is significant at the 1% level. This may be due to Danish influence.

(b) open interval

It is evident that all six subjects have a statistically significant difference between the open interval of ptk and that of bdg. There is not one instance of overlapping of corresponding consonants. This means that aspiration is an important factor in the distinction between the two types of stops. But the duration of the aspiration shows considerable variation between subjects.

The general averages for all three stops in list Ia and b (in ms and as percentage of the total duration) are:

	GJ	GR	CH	WS	HP	KV
ms	45	40	61	68	94	48
%	28	24	33	36	40	28

The differences between the subjects cannot be due to dialectal background only. Of the three subjects with the shortest aspiration one is from Koblenz (GJ), one from Berlin (GR), and one from Hamburg (KV). This is peculiar since the subject who has the longest durations (HP) is also from Hamburg. His very long aspirations may be due to his slow tempo of speech (only the absolute values are deviant), but may also be due to Danish influence.

The indications on the duration of aspiration in ptk found in the literature show a considerable variation. Some have found still shorter aspirations than the minimum found in the present material, e.g. Meyer (1904) p 24, t 26, and k 32 ms (in his own North German pronunciation), Weitkus (1931) 36 ms for t (22%), v. Essen (1934) 22, 22, and 25% for p, t, and k, respectively; others have found durations corresponding to our maximum: Schmitt (1947): t 92 ms, 39%, Bothorel-Witz and Péturs-son (1972): t 80 ms open interval (50%). Of the 80 ms 48 are described as "duration of explosion", which probably means: fricative phase. This distinction has not been made in the present investigation.

(c) total duration

Mainly due to the consistent difference in open interval the total duration of ptk is always greater than that of bdg, and the difference is significant for all consonant pairs and all subjects, in almost all cases at the 1 or 0.1% level.

3.2.1.2 Stressed position after word boundary preceded by s (s#'CV)

Three subjects: GJ, GR, and CH have spoken list Ic containing examples of the type: das Band, das Pech, etc. Table 2 contains the average durations, and in fig. 4 they are given in graphical form.

164

Figure 4

closure of ptk
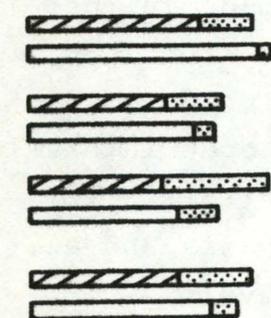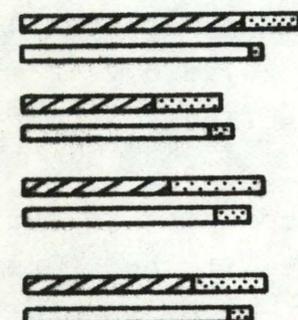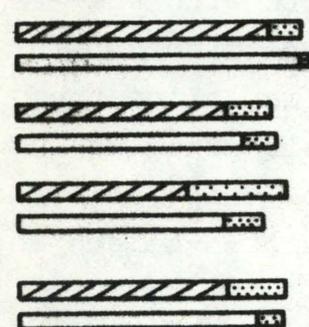closure of bdg
open interval

0   100   200 ms

Table 2

Duration of ptk and bdg in the
position s-'CV

S	N	C	cl.	op.i.	tot.	%op.i.
GJ	(2)	p	132	15	147	10
	(2)	b	148	5	153	3
	(2)	t	108	25	133	19
	(2)	d	117	18	135	13
	(2)	k	88	52	140	37
	(2)	g	108	20	128	16
	(6)	ptk	109	31	140	22
	(6)	bdg	124	14	138	10
GR	(4)	p	115	31	146	21
	(4)	b	120	6	126	4
	(4)	t	69	34	103	33
	(4)	d	96	14	110	13
	(4)	k	76	49	125	39
	(4)	g	100	18	118	15
	(12)	ptk	87	38	125	30
	(12)	bdg	105	13	118	11
CH	(4)	p	90	28	118	23
I	(4)	b	120	6	126	4
	(4)	t	71	30	101	30
	(4)	d	86	11	97	11
	(4)	k	69	54	123	44
	(4)	g	76	20	96	21
	(12)	ptk	77	37	114	32
	(12)	bdg	94	12	106	11

(a) <u>closure</u>

In the position after <u>s</u> the relations between the closures of <u>ptk</u> and <u>bdg</u> are reversed compared to the position after vowel. The material is small, but the relations are stable and statistically significant at the 1% level. Compared to the position after vowel, GJ has a slight shortening of <u>ptk</u> and a lengthening of <u>bdg</u> (-9 and +16 ms, respectively). The two others have shortening in both types, but most for <u>ptk</u> (GR -42, -20 ms, CH -25, -6 ms).

(b) <u>open interval</u>

Compared to the position after vowel the aspiration of <u>ptk</u> has been shortened considerably and the open interval of <u>bdg</u> slightly, but there is still a consistent difference without any overlapping. After vowel the difference between <u>ptk</u> and <u>bdg</u> was: GJ 24 ms, GR 55 ms, and CH 47 ms, whereas after <u>s</u> it is 17, 25, and 25 ms, respectively. GJ's <u>p</u> and <u>t</u> are practically unaspirated. We shall come back to this problem in section 4.

(c) <u>total duration</u>

Since both closure and open interval have been shortened more in <u>ptk</u> than in <u>bdg</u>, the difference in total duration is no longer significant and sometimes reversed.

<u>3.2.1.3 Medial position before unstressed ə</u>
<u>after vowel</u> ('VCə)

The material was not set up for the purpose of investigating this position (in the following called unstressed position) and does not allow of a systematic comparison between the six consonants, but the main characteristics will be mentioned briefly.

(a) <u>comparison with the stressed position</u>

The relations between stressed and unstressed <u>ptk</u> are rather variable and irregular. After a long vowel the closure may be either shorter or longer than in stressed position, the same subject may have different relations for different consonants, and different relations in different recordings. Also the duration of the aspiration is rather irregular in unstressed <u>ptk</u>. GJ, GR, CH and KV have practically unaspirated stops in this position; whereas WS and HP have rather long aspirations (see table 3). KV has, however, a somewhat longer aspiration in a different recording than in the one used here.

Table 3
Open interval (in ms) of ptk in position 'VCə

	GJ	CH I	CH II	GR	KV	WS	HP II
p	15	26	22	11	13	23	31
t	26	27	43	20	14	61	66
k	33	29		24	14	46	–

bdg show much more regularity than ptk. They are shortened in unstressed position for all subjects. As the material contained relatively many examples of b, this consonant is chosen as an example in table 4.

Table 4
Duration (in ms) of the closure of b in stressed and unstressed position (V'bV: and 'V:bə)

S	GJ	GR	CH	HP	WS	KV
N	6/6	11/6	26/30	33/27	24/8	12/16
V'bV:	118	124	106	169	129	108
'V:bə	93	66	80	82	85	82

These differences are statistically significant. CH, HP, and WS have also a number of d's in this position. Their average durations are 60, 54, and 68 ms. Similar relations have been found for other subjects.

(b) ptk versus bdg after long vowel

As ptk are only slightly longer or shorter in unstressed position after a long vowel than in stressed position, and bdg are always considerably shortened, the consequence is a clear difference in the duration of the closure between ptk and bdg in this position. The average difference for the recorded examples is 44 ms, and this difference is evidently significant. Table 5 compares the closures of p and b, t and d (after long vowel) for a number of subjects, including some of the subsidiary subjects (HWL, HT, and HL).

Table 5

Duration (in ms) of the closure
of p vs. b and t vs. d in position 'V:Cə

S	CH	HP	HT	KV	HWL	HL
N	19/15	36/27	15/12	25/30	6/8	6/9
p	107	135	122	121	85	124
b	80	82	65	82	60	72

S	HP	WS	HT	KV
N	72/12	36/12	35/6	27/6
t	123	97	108	93
d	54	68	49	65

As for the open interval of ptk and bdg HP has a clear distinc-
tion and WS has a significant distinction with some overlapping
for p/b. On the other hand, KV has complete overlapping. GJ,
GR, and CH have a distinction between the averages but with some
overlapping (in CH II the difference is significant for t/d). But,
as the open interval of bdg is often very difficult to measure
and the number restricted, the averages are not always quite
reliable.

The result is that there is a tendency to distinguish the
open intervals of ptk and bdg in this position, but this distinc-
tion is rather inconsistent.

The open interval has not been measured for the subsidiary
subjects.

(c) ptk and bdg after short vowel

The preceding tables only included the position after long
vowel. After short vowel there is more irregularity.

German only possesses a very small number of words with
bdg after short vowel. List III, which was spoken by 10 subjects
in all, contained some words of this type (bibbern, Kladde, Widder,
Egge). Most subjects (GJ, WS, HT, HL, HWL) had a slightly longer
stop in these words, others (CH, GR, HP, and partly KV) showed a
considerable lengthening of the stop. The extremes are WS, who

had a difference of 1 ms between bdg after long and short vowel, and HP who had a difference of 90 ms (and this cannot be Danish influence). Those who have lengthened consonant normally pronounce it with less voicing.

As for ptk, most subjects have also a somewhat longer consonant after a short vowel than after a long vowel. This is true of 5 subjects out of 6 for p, and for 8 and 7 out of 10 for t and k, but the differences are often small. Only CH, GR, and HP (who had also long bdg after short vowel) have an appreciable lengthening.

Because of these irregularities the relations between bdg and ptk are not so clear in this position. bdg have longer closures than ptk in some cases, but as the words with bdg are very rare, this is less essential.

The tendency to longer and only partly voiced b after short vowel has been observed by Hentrich (1925). Rositzke (1944) found t and k, but not p, to be slightly longer after short vowel, cf. also E.A. Meyer (1901 a and b).

3.2.2 Differences between stops and affricates

Recordings of affricates have been made by GJ, GR, CH, WS, HP, and KV. The values of the durations of closure and fricative phase are given in table 6 and displayed in graphical form in fig. 5.

All subjects have a shorter closure and, conversely, a longer open interval in pf and ts than in p and t (the only exception is KV's pf). As the difference in open interval is larger than the difference in closure, the total duration of the affricates is in all cases longer than that of the stops, the averages ranging from 111% (HP I) to 142% (CH I).

Table 6 contains 8 pairs of averages for each place of articulation, thus 16 pairs of averages for closure, open interval, and total duration. As for the closure there are 5 cases which are not statistically significant, whereas all differences of aspiration and all but one difference of total duration are significant, normally at the 0.1% level.
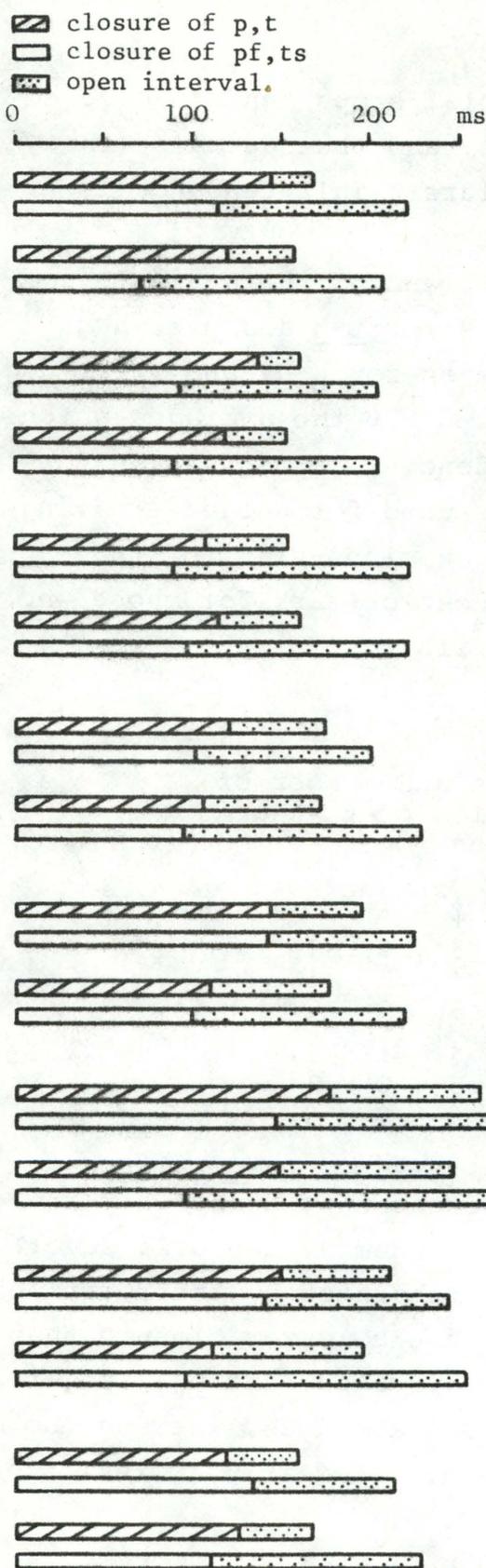
Figure 5

- ▨ closure of p,t
- ☐ closure of pf,ts
- ▩ open interval.

0   100   200   ms

Table 6

Duration of p,t and pf,ts in the position V-'CV

S	N	C	cl.	op.i.	tot.	%op.i.
GJ	(3)	p	142	25	167	15
	(3)	pf	113	107	220	49
	(3)	t	118	38	156	24
	(3)	ts	70	138	203	68
GR	(6)	p	137	23	160	14
	(6)	pf	91	112	203	55
	(6)	t	118	33	151	22
	(6)	ts	88	114	203	56
CH I	(7)	p	105	46	151	30
	(5)	pf	89	131	220	60
	(6)	t	112	47	159	30
	(6)	ts	95	125	220	57
CH II	(18)	p	117	55	172	32
	(20)	pf	101	99	200	50
	(24)	t	104	66	170	39
	(20)	ts	94	134	228	59
WS	(24)	p	142	52	194	27
	(24)	pf	141	83	224	37
	(24)	t	108	68	176	39
	(24)	ts	98	121	219	55
HP I	(18)	p	175	86	261	33
	(18)	pf	145	145	290	50
	(18)	t	147	99	246	40
	(18)	ts	95	193	288	67
HP II	(13)	p	148	62	210	31
	(11)	pf	139	104	243	43
	(17)	t	109	86	195	44
	(18)	ts	94	158	252	63
KV	(12)	p	117	40	157	26
	(12)	pf	131	80	211	38
	(12)	t	124	42	166	25
	(12)	ts	109	118	227	52

3.2.3 Differences due to place of articulation

3.2.3.1 Closure

It has often been found that labial stops tend to have a longer closure than dentals and velars, and that dentals (though less often) tend to be longer than velars. This tendency is also found in the present material.

In stressed position (table 1, 2, and 6) there are 19 comparable pairs of averages for p-t and 9 for t-k and p-k, and there are 11 comparable pairs of averages for b-d, and 9 for d-g and b-g. In table 7 a plus sign is used for the averages which are in agreement with the general tendencies, a minus sign for the others. Similarly, a plus-sign is used for subjects having the mentioned difference consistently, a minus-sign for those having the opposite relation, and a question mark for those who show different relations in different lists.

Table 7

Number of pairs of averages and number of subjects showing the relation $p > t > k$ and $b > d > g$ for closure duration

	averages		subjects		
	+	−	+	?	−
$p > t$	15	4	3	2	1
$t > k$	7	2	4	1	1
$p > k$	9	0	6	0	0
$b > d$	9	2	4	1	1
$d > g$	8	1	5	0	1
$b > g$	8	1	6	0	0

For four of the six subjects the differences p > k and b > g are statistically significant (for three of them at the 0.1% level).

This tendency can be further corroborated by the relations found in unstressed position. Here p has a longer closure than t in 13 out of 15 comparable pairs of averages, and p a longer closure than k in 8 out of 11 comparable averages, whereas the relation between t and k is more random. Similarly, b is longer than d in all 8 comparable pairs.

Finally in the affricates <u>pf</u> is found to have a longer closure than <u>ts</u> in 7 out of 8 averages.

Bennett (1935, p. 16-21) has found similar relations: <u>b</u> > <u>d</u>, <u>b</u> > <u>g</u>, but <u>g</u> > <u>d</u>.

3.2.3.2 Open interval

For the open interval the opposite tendency has often been found, k > t > p and g > d > b.

This tendency is still more obvious in the present material. This is illustrated in table 8 which is set up in the same way as table 7.

Table 8

Number of pairs of averages and number of subjects showing the relation k > t > p and g > d > b for open interval duration

	averages		subjects		
	+	-	+	?	-
k > t	9	0	6	0	0
t > p	17	2	4	1	1
k > p	9	0	6	0	0
g > d	9	0	6	0	0
d > b	11	0	6	0	0
g > b	9	0	6	0	0

The same relation is found for the fricative phases of <u>ts</u> and <u>pf</u>. In all 8 pairs of averages the fricative part is longer in <u>ts</u> than in <u>pf</u>. In figs. 6, 7, and 8 the average values of closures and open interval for the stressed stops of tables 1, 2, and 6, respectively, have been plotted in a diagram with closure vertically and open interval horizontally. A comparison between the three places of articulation (connected by lines) shows that there is in most cases an inverse relationship between the duration of closure and open interval (i.e. the lines follow mostly the direction from upper left to lower right).

Figure 6

Duration of closure and open
interval of the stops in
table 1.

○ b ● p
▲ d ▲ t
□ g ■ k

closure
ms

closure
ms

open interval
(ms)

Figure 7

Duration of closure and open
interval of the stops in
table 2.

○ b ● p
▲ d ▲ t
□ g ■ k

closure
ms

open interval
(ms)

<u>Figure 8</u>
Duration of closure and open interval
in the stops and affricates of table 6.
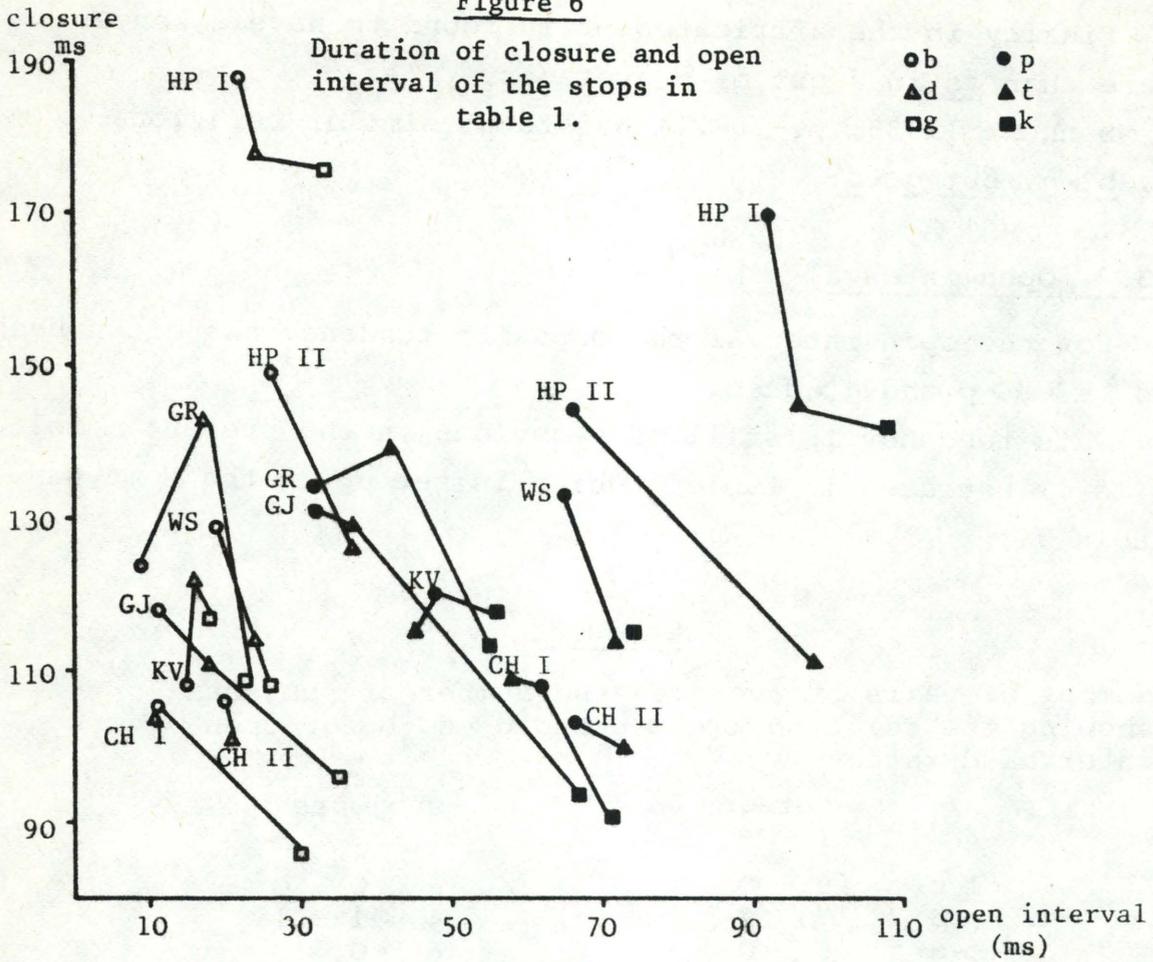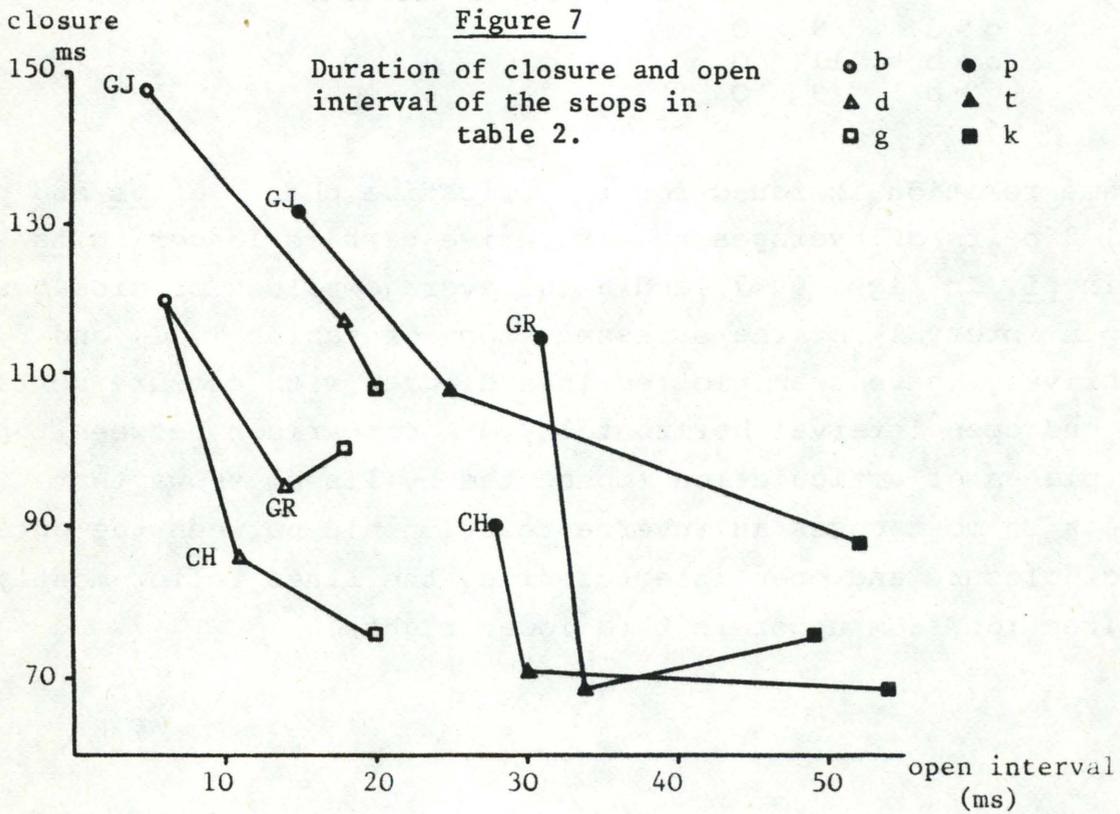
3.2.4 Differences due to the following vowel

It has also often been observed that the open interval tends to be longer before high vowels than before low vowels. This tendency can also be seen in the present material, though less consistently. In list Ia four of the six subjects have a longer aspiration in pu than in pa and pi, and a shorter aspiration in ka than in ki and ku, and all have a shorter aspiration in ta than in ti.

There is also an almost consistent difference between the durations of the aspiration in the stops of list Ia (table 1) and those of list Ib (table 6). The former, which are followed by a long vowel, have a longer aspiration than the latter, which are followed by a short vowel. However, a closer inspection of the measures shows that the difference is mainly (though not exclusively) due to differences before i: and u: versus I and U, less to differences between a: and a, which means that the crucial factor also in this case is vowel height, not vowel length. This is corroborated by the observation that in Danish, where short and long vowels have approximately the same quality, no difference in the aspiration could be observed.

There is a concomitant shortening of the closure of p in list Ia compared to list Ib, which might be interpreted as a compensation, if it were not for the fact that the t-lists show the opposite difference (t before long vowel has both longer closure and aspiration), and no such compensation is seen in the cases of stop closures before low and high vowels.

3.3 Intra-oral air pressure

Intra-oral air pressure has been recorded for subjects CH, HP, and WS.

3.3.1 Stressed position ($_{\circ}$V# 'CV)

3.3.1.1 Peak pressure

Peak pressure averages in stressed position are given in table 9.

Table 9

Intra-oral air pressure (peak pressure) of stops and affricates in stressed position ($_{\circ}$V'CV)

S	L	N	C	mm H_2O	N	C	mm H_2O			
CH		(14)	b	43	(17)	d	50			
	(a)	(16)	p	53	(15)	t	59			
	(b)	(13)	p	58	(19)	t	59			
		(13)	pf	56	(16)	ts	62			
HP		(23)	b	86	(17)	d	110			
	(a)	(22)	p	90	(17)	t	104			
	(b)	(14)	p	108	(17)	t	106			
		(10)	pf	90	(18)	ts	116			
				mm[1]			mm			
WS		(24)	b	18.9	(24)	d	20.6	(24)	g	20.7
	(a)	(24)	p	19.8	(24)	t	20.1	(24)	k	20.8
	(b)	(24)	p	20.3	(24)	t	20.8			
		(24)	pf	18.3	(24)	ts	20.3			

The two sets of _pt_ values are averages of list Ia and of list Ib in which the consonants stand before long and short vowel, respectively. The averages are slightly higher before short vowel. There is a tendency for _p_ to have higher air pressure than _b_, and a tendency for dentals (and velars) to have higher air pressure than labials (but the latter tendency is reversed in unstressed position). However, none of the differences are significant except for that between _p_ and _b_ of subject CH, which is signifi-

---

1) As mentioned in section 2.3, WS's curves had to be measured in mm since the calibration curve had disappeared.

cant at the 5% level. CH and HP differ as to the voicing of b
and d, HP having voiceless bd, CH 40% voicing of the closure;
and as there is normally an inverse relation between air pressure
and voicing, it is understandable that CH has a more consistent
difference between pt and bd in air pressure. But it is aston-
ishing that WS, who has 68% voicing in bd, has no difference in
peak air pressure. He has, however, a slower rise of the curve
in bdg. The peak value of the affricates do not differ from
that of the stops.

3.3.1.2 Decay of pressure

The three types of consonants are, however, clearly di-
stinguished by the decay of the pressure curve, bdg having an
abrupt decay, ptk a somewhat slower decay, and the affricates
a very slow decay. Moreover, it often happens that the pressure
of the affricates continues to rise after the release, particular-
ly in ts, so that the maximum is found in the fricative part of
the consonant (cf. fig. 1). In HP's ts the distance from release
to maximum is 142% of the duration of the closure. On the other
hand, KV's pf has a shorter distance from implosion to maximum
than his p, and this is therefore not a reliable measure of
affrication. It has been attempted to quantify the differences
in decay by two measures: (a) the distance (in ms) from the re-
lease to the point where the decay has reached 50% of the peak
value, and (b) the fall in pressure value 20 and 40 ms after the
release, indicated in % of the peak value. Table 10 shows the
average values of these measures for p and t (list Ia and Ib com-
bined) and for pf, ts. No measurements have been made for bdg,
since they always fall abruptly. In fig. 9 the same differences
are depicted graphically. The differences between p and pf and
t and ts are evidently significant.

Table 10

Decay of intra-oral air pressure in stops and affricates
(a) distance from release to 50% pressure value (in ms)
(b) fall in pressure value 20 and 40 ms after the release
in % of the peak value.

S	N	C	dist. ms	pressure fall +20ms %	pressure fall +40 ms %	N	C	dist.	pressure fall +20ms %	pressure fall +40 ms %
HP	(30)	p	17	56	82	(34)	t	56	20	41
	(10)	pf	57	22	48	(18)	ts	141	4	4
WS	(48)	p	41	30	53	(48)	t	46	27	44
	(24)	pf	70	3	18	(24)	ts	105	1.5	2
CH	(29)	p	11	74	90	(34)	t	21	53	78
	(12)	pf	65	3	25	(16)	ts	107	-10	16

Apart from the differences between stops and affricates
it also appears from the table that stops tend to have a slower
decay in t and k than in p (except for pu/tu). They also have a
slower decay before u and i than before a, particularly when the
vowel is long (list Ia). These differences correlate with ten-
dencies observed in the duration of the aspiration, and similar
tendencies are found in Danish stops. The differences according
to following vowel (list Ia) have been quantified in the same
way as the differences between stops and affricates. They are
depicted in graphical form in figs. 10 A and B. The number of
examples included in the averages of different vowels are 4-7
for CH, 6 for HP, and 8 for WS, HP's distance values for pV:
(list Ia) have disappeared. HP has a very slow decay in ti, i.e.
he has affrication. This may be Danish influence.

3.3.2 Unstressed position ('VCə)

In unstressed positions all consonants have a lowered peak
pressure. The average values for the examples found in list Ia
and b are given in table 11.
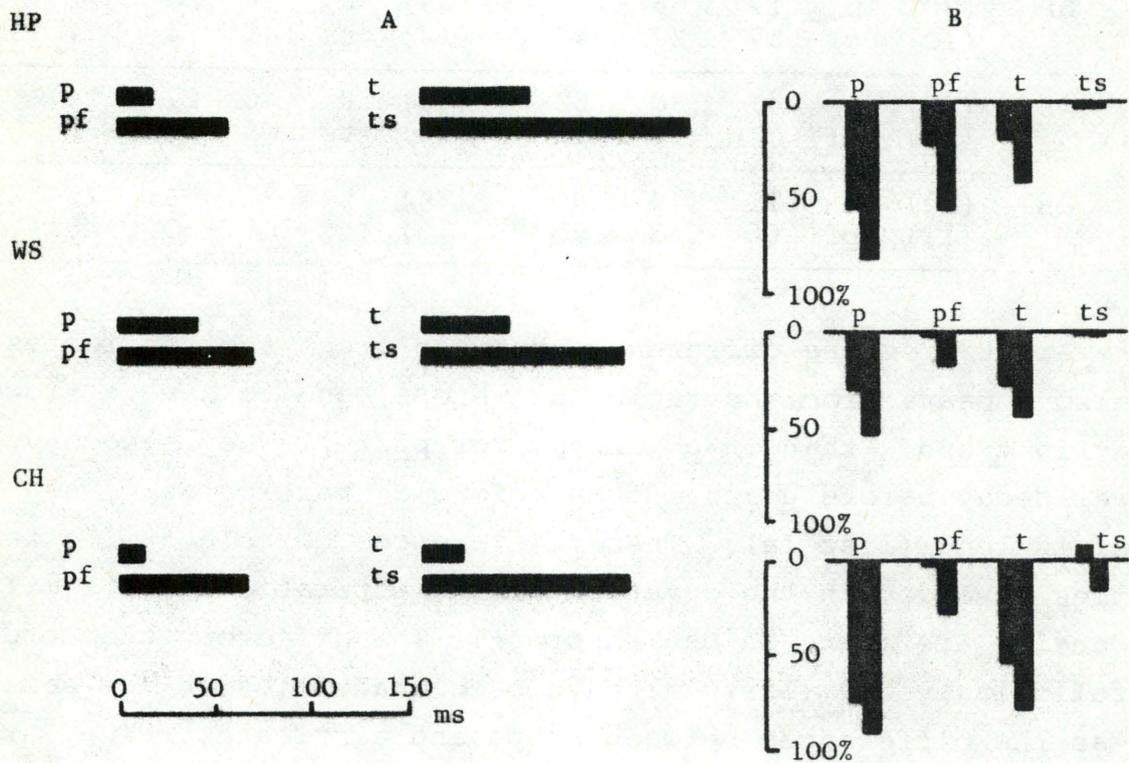
Figure 9

Decay of intra-oral air pressure

A. Distance from release to 50%
   of peak value

B. Decay 20 and 40 ms after release
   in % of peak value.
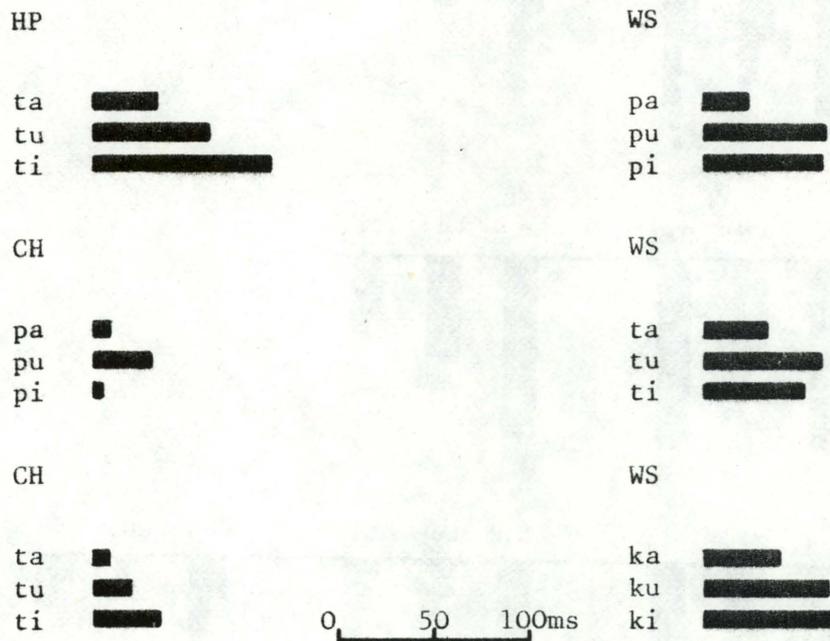
Figure 10A

Decay of intra-oral air pressure before
different vowles. Distance from release
to 50% of peak value.
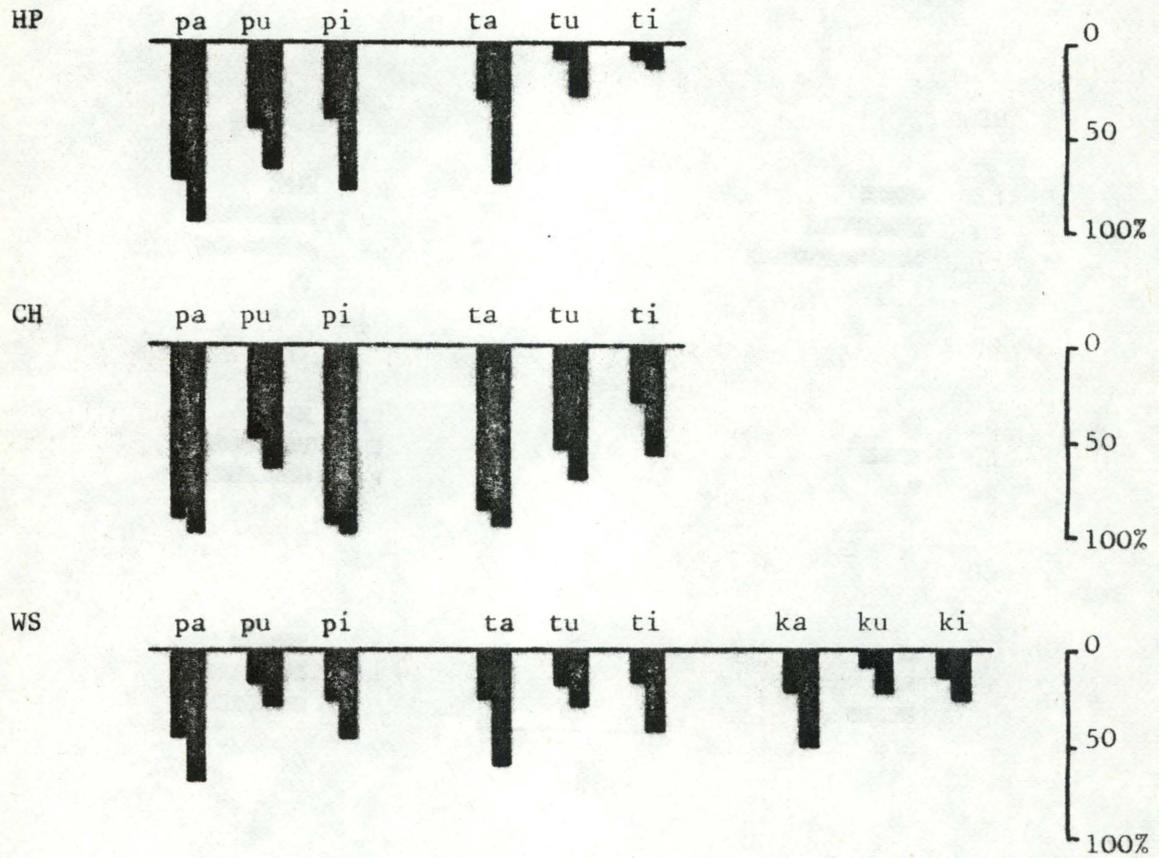
Figure 10B

Decay of intra-oral air pressure before
different vowels. Decay 20 and 40 ms
after release in % of peak value.

Table 11

Intra-oral air pressure (peak pressure averages)
in stops in medial position before unstressed
vowel after stressed vowel ('VCə)

S	N	C	mm H_2O	N	C	mm H_2O			
CH	(16)	p	49	(20)	t	46			
	(22)	b	27	(7)	d	19			
HP	(6)	p	98	(18)	t	78			
	(12)	b	86	(7)	d	54			
			mm			mm			mm
WS	(8)	p	18.1	(32)	t	16.9	(8)	k	17.6
	(32)	b	11.6		d	–	(16)	g	12.0

The peak pressure of ptk is only slightly lowered in un-
stressed position (cf. table 9), and the same is true of HP's
voiceless b and d, but the fully voiced bdg spoken by CH and WS
have a considerably weakened pressure in unstressed position
compared to initial bdg, and there is therefore a significant
difference both between their bdg in stressed and unstressed
position and between their ptk and bdg in unstressed position.
The pressure curves of unstressed bdg also have a slower rise.
The difference between stressed and unstressed position also
appears clearly from an inspection of the individual words con-
taining two stops. WS has, for instance, seven words of this
type (Puppe, Pute, Tute, guten, Pike, Diebe, Gabe) with 8 examples
of each, and the peak pressure is always lower in the second stop
except for one example of Pike and two of Puppe.

3.4 Airflow

Airflow recordings have been made of the subjects HP, WS,
and KV.

3.4.1 Stressed position (ͺV# 'CV)

3.4.1.1 Peak values

The average peak values for ptk, bdg, and the affricates in
stressed position after vowel are given in table 12, and a graphical
display is found in fig. 11.

Table 12

Airflow (peak values) of stops and affricates
in stressed position ($_{\circ}$V#'CV) ((a) before
long vowel, (b) before short vowel). Number of
examples: HP and KV 12, WS 17.

S		C	mm	C	mm	C	mm
HP		b	14.2	d	10.8	g	6.5
	(a)	p	18.6	t	18.1	k	18.3
	(b)	p	23.3	t	22.7		
		pf	17.1	ts	12.5		
WS		b	19.0	d	16.0	g	13.5
	(a)	p	28.0	t	26.4	k	22.8
	(b)	p	29.7	t	33.7		
		pf	21.6	ts	21.5		
KV		b	34.4	d	27.0	g	15.6
	(a)	p	40.8	t	36.3	k	30.5
	(b)	p	53.2	t	46.2		
		pf	37.7	ts	23.0		

All three subjects have a clear difference between ptk and
bdg. Furthermore, they all have higher peak values in p and t
than in pf and ts. These differences are all significant at the
0.1% level (KV's p/b at the 1% level). There is further a ten-
dency for the peak value to be higher before short than before
long vowels (6 pairs), and to be higher in labials than in den-
tals (12 pairs with one exception), and in dentals than in velars
(6 pairs with one exception).
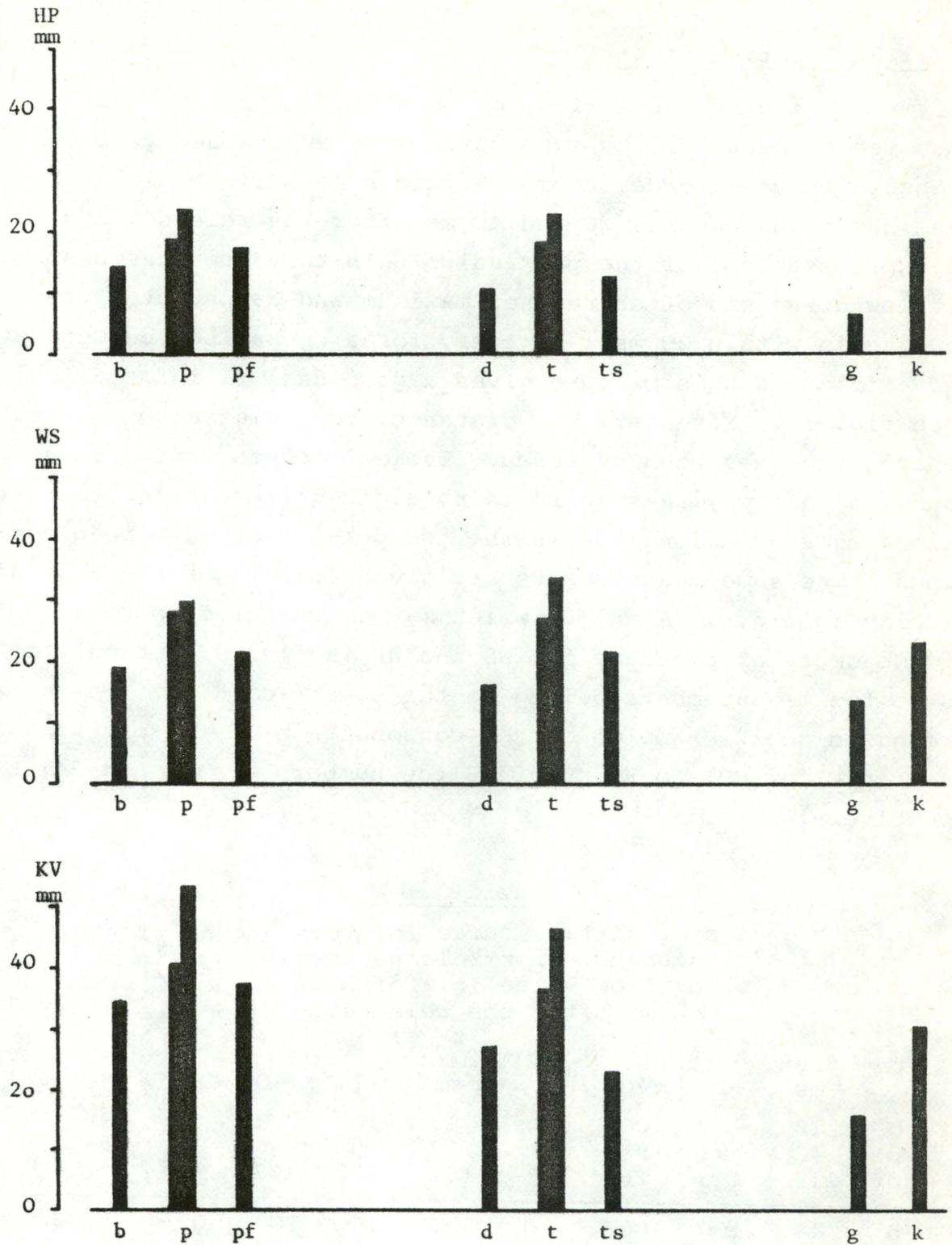
<u>Figure 11</u>

Airflow, peak values. For <u>ptk</u> the left column indicates the position before long vowel, the right column before short vowel.

3.4.1.2 The rise of airflow

The rise of the airflow curve after the release has been measured by means of the same parameters as the decay of the pressure curve in order to facilitate a comparison, i.e. the rise has been measured 20 and 40 ms after the release, and it is expressed in % of the peak value. In the stop consonants the airflow curve has often reached maximum and has started going down again within 20 ms. However, for a comparison between stops and affricates this measure gives a good deal of information (see fig. 1). Moreover, the distance from release to the peak of the curve has been measured. Table 13 contains the values found in this way. As there is no clear difference between stops before short and long vowels the two ptk-series have been combined. The same measurements are given in form of a graphical display in fig. 12 A and B, which may be compared to the display of the decay of pressure for WS and HP in fig. 9; but only for WS do the measurements belong to the same recording. The averages are based on 12 examples of the consonants b, d, g, pf, ts, and 24 of ptk for HP and KV. For WS the numbers are 17 and 34, respectively.

Table 13

Rise of airflow curve in stops and affricates
(a) distance from release to maximum (in ms)
(b) airflow value in % of peak value 20 and
     40 ms after the release.

	C	dist. ms	flow +20	+40	C	dist. ms	flow +20	+40	C	dist. ms	flow +20	+40
HP	b	13	100	100	d	18	100	100	g	24	*	100
	p	42	92	98	t	62	59	86	k	65	53	84
	pf	119	13	26	ts	151	22	46				
WS	b	19	100	100	d	21	100	100	g	25	100	100
	p	26	90	99	t	39	71	99	k	45	54	95
	pf	65	31	66	ts	104	16	35				
KV	b	13	100	100	d	12	100	100	g	14	100	100
	p	16	100	100	t	23	94	100	k	43	67	98
	pf	69	21	34	ts	108	19	38				

*) HP's g has not quite reached 100% at 20 ms distance (the distance to the peak is 24 ms), but the measurement is missing.

HP

b
p
pf

d
t
ts

g
k

WS

b
p
pf

d
t
ts

g
k

KV

b
p
pf

d
t
ts

g
k

0   50   100ms
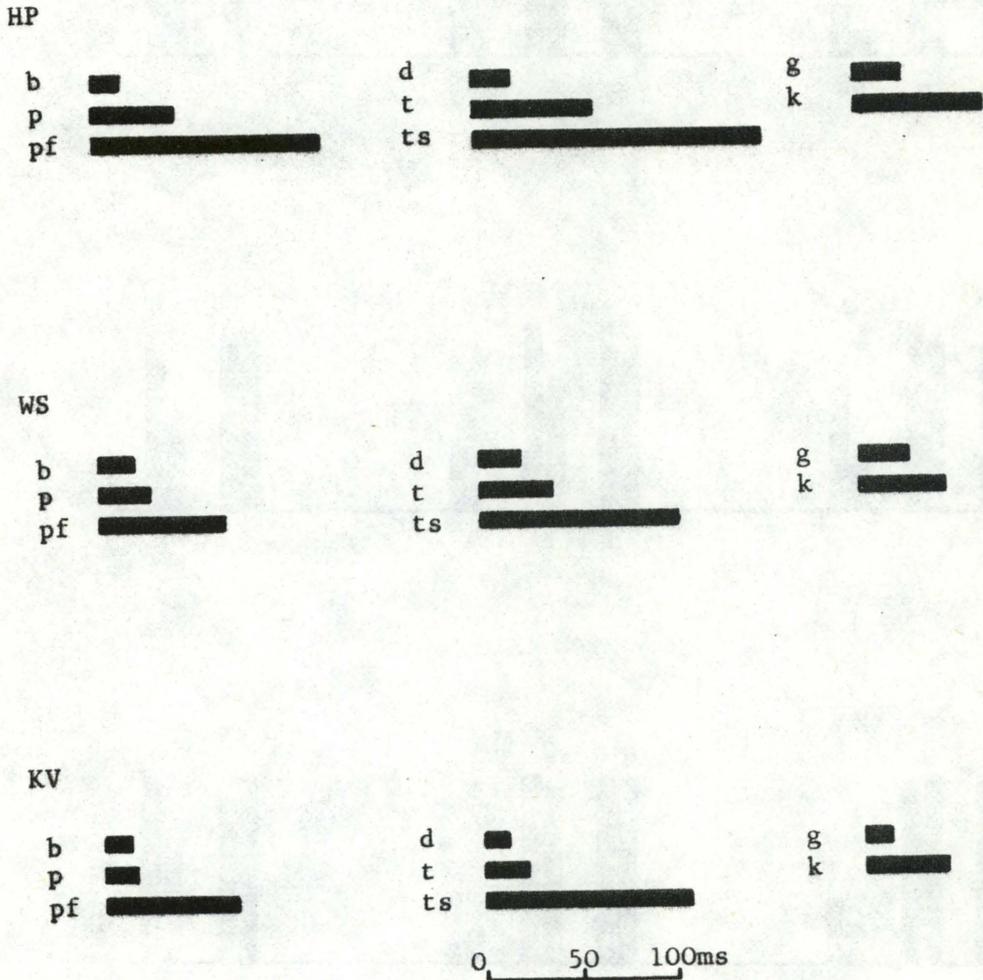
Figure 12A

Rise of airflow. Distance from
release to peak (in ms).
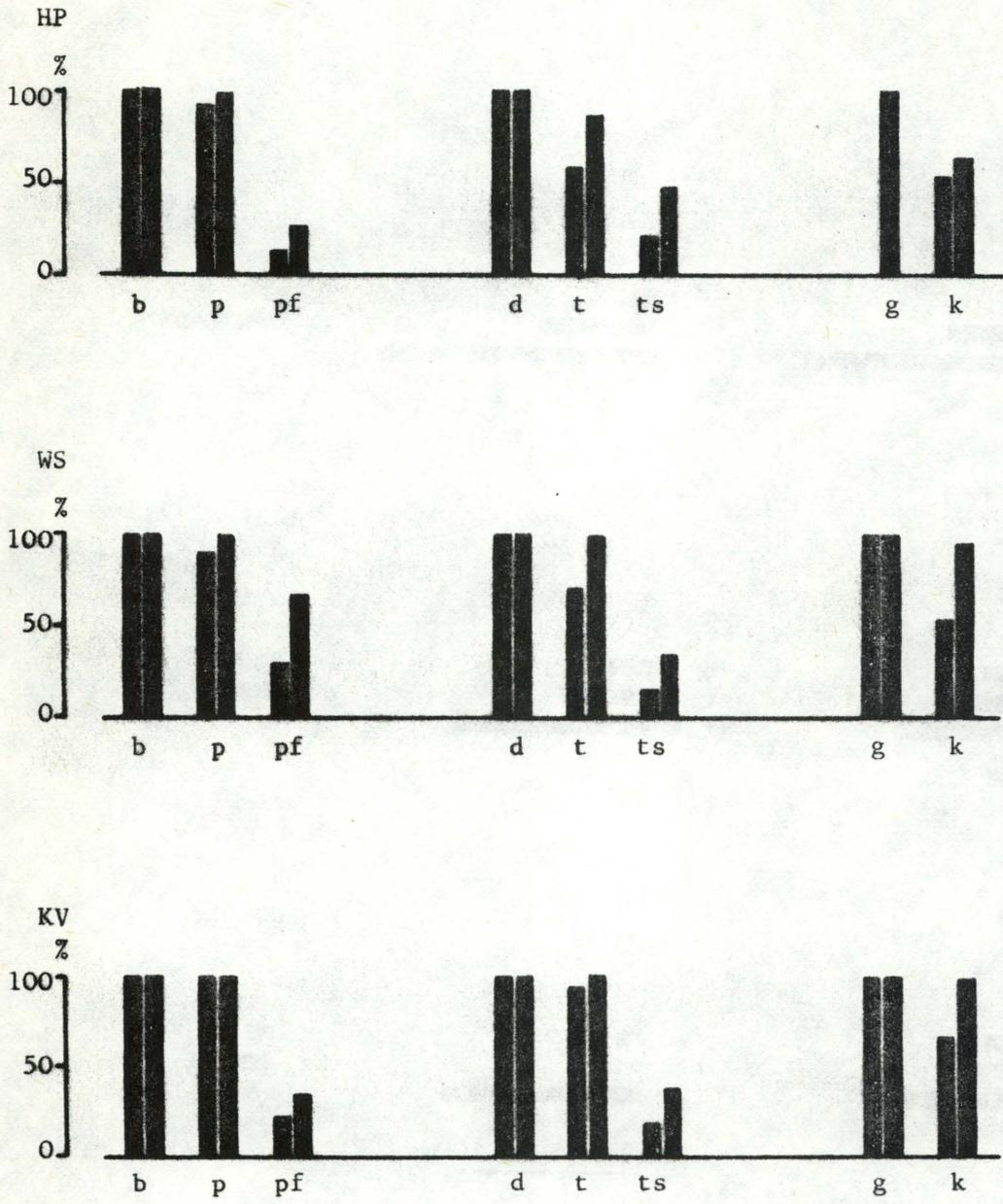
Figure 12B

Rise of airflow 20 and 40 ms after
release in % of peak value.

It appears from table 13 that the airflow rises more quick-
ly in bdg than in ptk and more quickly in ptk than in pf and ts.
There are no exceptions to these relations in the averages, ex-
cept that p and t may reach 100% like bdg. Moreover, it appears
that labials rise more quickly than dentals and these more quick-
ly than velars. The only exception is the relation b-d in KV's
curves, but these were rather dubious to delimit. These relations
are the same as those found in the decay of pressure. There are
also some differences according to vowels, but they are rather
inconsistent.

3.4.2 Unstressed position ('VCə)

In unstressed position the peak of the airflow is normally
lower than in stressed position. The values for WS and KV are
given in table 14.

Table 14
Airflow (peak value) in unstressed position ('VCə)

S	N	C	mm	N	C	mm	N	Ç	mm
WS	(6)	p	24.6	(24)	t	23.9	(6)	k	25.0
	(24)	b	11.9		d	-	(12)	g	11.5
KV	(4)	p	36.3	(16)	t	24.1	(4)	k	27.1
	(16)	b	26.0		d	-	(8)	g	17.9

These values are somewhat lower than those for the stressed posi-
tion except for WS k and KV g.

Generally the second of two stops in a sequence 'CV˳CV has
a lower peak airflow than a preceding stressed stop, but there
are many exceptions. Not only has t in guten a much stronger air-
flow than g, but also in the cases of same type of consonant can
this relation be found, e.g. WS Diebe, Giebel, Pute. There is
a significant difference between p-b and k-g.

3.5 Subglottal pressure

Subglottal pressure has only been recorded for WS. The pressure has been measured at the point of release. This is at the same time normally the maximum of the whole word. The averages for the stressed position are given in table 15.

Table 15

Subglottal pressure at the point of release of stops and affricates ($_{o}$V #'CV)

S	N	C	mm	N	C	mm	N	C	mm
WS	(24)	p	13.9	(24)	t	15.9	(12)	k	15.8
	(11)	b	15.5	(12)	d	15.3	(12)	g	15.6
	(9)	pf	13.4	(12)	ts	15.0			

There is no consistent difference between the peak values of ptk and bdg. But the pressure curves of ptk differ from those of bdg and the affricates by having a small dip starting at the release and ending approximately 10 ms after the start of the vowel. This dip is deeper before a than before i and u (see fig. 2), corresponding to the quicker escape of air at the release before a and the quicker drop in intra-oral air pressure. The dip is slightly deeper before i and u in list Ib, which has short (and thus lower) i and u. The differences are very small, but hardly quite accidental. The measurements are given in table 16.

Table 16

Size of dip after the release of ptk (in mm)

	pa	pu	pi	ta	tu	ti	ka	ku	ki
(Ia)	2.8	0.8	1.1	2.5	1.4	1.2	2.8	0.9	0.9
(Ib)	2.3	1.0	1.6	2.9	1.9	1.6			

Such dips are not found after bdg, and only rarely in the affricates. In some cases, however, a dip is found in affricates but not at the release, but later, when the airflow starts to rise more quickly and the intra-oral pressure goes down.

After the dip the curve rises somewhat again, but it does not reach the value from before the dip. This is particularly true of the dip after a̲ (see fig. 2). The result is that at the start of the vowel p̲t̲k̲ have normally a lower subglottal pressure than b̲d̲g̲ (see also Ohala 1974). These values (measured 10 ms after the start of the vowel) are given in table 17.

Table 17

Subglottal pressure of the stops at
the start of the vowel (in mm)

WS	p	12.9	t	14.6	k	14.3
	b	15.3	d	15.3	g	15.6

After affricates and b̲d̲g̲ the curve sometimes rises very slightly after the release, particularly before u̲.

Unstressed syllables rarely have a separate peak. In most cases the second stop in words of the type T̲u̲t̲e̲ has a slightly lower subglottal pressure than the first, but it may have the same pressure or (in two cases) a somewhat higher pressure.

In voiceless sounds the sub- and supraglottal pressure should be approximately the same. Unfortunately it could only be measured in mm. But there should at least be a close correlation between the measurements. This is also found in so far as p̲ has a lower pressure than t̲, p̲f̲ a lower pressure than t̲s̲, and p̲ a slightly higher pressure than p̲f̲ in both cases. But the relation does not hold for t̲-t̲s̲, and it does not hold in single cases (see the values of intra-oral peak pressure in table 9; the relations remain the same if only those recordings are included which comprise both intra-oral and subglottal pressure).

3.6 Lip pressure

Lip pressure has been recorded for CH and HP. The averages
are given in table 18. Before stressed vowel CH has a slightly
higher pressure in p than in b, but HP has the opposite relation.
Before unstressed vowel both have higher lip pressure in p than
in b, but in both cases the variation is very large and the dif-
ferences non-significant. CH has often assimilated a following
syllabic n to the labial stop and this seems to result in a
higher lip pressure. If these examples are discarded the average
of her unstressed b will be 6.3 mm H_2O and the difference from
p quite clear, but in that case the average is only based on 4
examples.

Table 18

Lip pressure of p and b in the positions
₀V#'CV and 'VCə

S	N	C	mmH₂O	N	C	mmH₂O
CH	(24)	'p	11.7	(12)	'b	10.6
	(16)	₀p	13.3	(16)	₀b	12.3
HP	(36)	'p	25.2	(18)	'b	33.5
	(12)	₀p	22.0	(16)	₀b	17.8

3.7 Intensity of the explosion

The intensity of the explosions has been measured in dB
on the mingograms for CH, GJ, and GR. However, the measurement
of intensity of such brief bursts of sounds on mingograms with a
certain integration time and even some high pass filtering
(these were the first mingograms, taken in 1955) is rather un-
reliable, and there is no sense in giving tables with exact values.
The only thing that can be said is that there was a certain
tendency for ptk to have higher intensity than bdg, but with ex-
tensive overlapping and great variation, and that the affricates
had evidently weaker explosions than the stops.

4. Summary and discussion

4.1 The distinction between North German ptk and bdg

It seems evident that a common description of North German and South German stops is not possible. In Southern Germany ptk are normally unaspirated and bdg are voiceless even in rather educated speech. The problem is whether all North German speakers can be said to have the same distinctive features.

4.1.1 Aspiration and voicing

Aspiration is obviously an important factor in the distinction between ptk and bdg in stressed position. In unstressed position ptk may be almost unaspirated, but there is still a difference in airflow and often a small difference in open interval.

In unstressed position there is, however, also, in practically all cases, a difference in the degree of voicing (we must here leave HP aside since he may be influenced from Danish). Such a difference is not consistently found in stressed position after a vowel, it is very rarely found in absolute initial position, and never after s.

In order to produce a voiced stop, two conditions must be fulfilled: The glottis must be relatively narrow, and there must be a pressure drop across the glottis. After a vowel the glottis will be narrow and the pressure drop may be sufficient for some time, but after a voiceless sound it is probably not sufficient that the vocal cords are approached. Because of the open position of the glottis during the preceding voiceless sound, the intra-oral air pressure will be rather high already at the implosion, and a command for an enlargement of the cavities will probably be necessary. After a pause it may be somewhat easier to produce a voiced stop since nothing prevents the glottis from narrowing down, and the subglottal pressure is rising. In our small material all stops were voiceless after an s in the preceding word, but Esawa (1972) quotes various examples of more or less

voiced stops in this position.

Those Germans who have, or may have, voiced stops after
voiceless sounds (and perhaps also those who have initial clearly
audible voicing) may be assumed to have a separate command for
an enlargement of the cavities, leading to a lower pressure, and
thus to have voicing as a particular factor. For others the
voiced stops after voiced sounds may be a simple case of assimi-
lation (see also Schmitt 1947, and Meinhold and Stock 1963), so
that voicing cannot be considered a separate factor. The dif-
ference from Danish, which normally has voiceless bdg even after
voiced sounds in stressed position (but normally weak voicing in
unstressed position), is only one of degree, and might perhaps
be due to a slight difference in glottal opening. This is pure
speculation, since nothing is known about the opening of the
glottis in German stops. But it is not unplausible.

One may also speculate further and set up the hypothesis
that North German is now developing in the same direction as
Danish. It is conspicuous that the data on aspiration in older
descriptions (Meyer 1904, Weitkus 1931) indicate much shorter
aspirations than modern descriptions. The objections to the
traditional requirement of voiced bdg in all positions have also
been increasing. This may be accidental, but it may also reflect
a development, i.e. North German may be on its way to a new sound
shift, and in such a situation there will be much vacillation.

4.1.2 Fortis-lenis

Some phoneticians think that the common distinction between
German ptk and bdg in stressed and unstressed position is one of
tenseness (Schmitt 1947, Bothorel-Witz and Pétursson 1972). I
have not found much evidence for this assumption in the present
material. In languages where a fortis-lenis difference is more
evident, e.g. in French or in Swiss German (see Fischer-Jørgensen
1972 and Dieth and Brunner 1943), this difference is consistently
realized by a longer closure duration and higher organic pressure
in the tense consonant. But the difference between the closure

duration of ptk and bdg in stressed position in the present
material is small and not significant (whereas there is a dif-
ference in unstressed position). The same is true of lip pres-
sure. Some think that differences in intra-oral air pressure
may be suggestive of a fortis-lenis opposition, but this dif-
ference is not consistent either in the present material. The
intensity of explosion was also very variable (see section 3.7).
It is possible that the assumption of a fortis-lenis opposition
can be corroborated by other facts, e.g. higher speed of the
movement of the articulators and consequently quicker formant
transitions, higher EMG activity of the relevant muscles, etc.,
but this remains to be demonstrated.

4.2   ptk after s

The examples of ptk after an s in the preceding word showed
a remarkable shortening, not only of the closure but also of the
aspiration. This should perhaps be seen in the light of the well
known fact that after an s of the same word ptk in Germanic lan-
guages have not undergone either the Germanic or the High German
sound shift. Kim (1970) has suggested that this should be due to
the fact that the opening-closing movement of the glottis neces-
sary for p may start in the s and therefore come to an end earlier.
Frøkjær-Jensen, Ludvigsen, and Rischel (1971) have found such
common one-peaked glottis movements in glottograms of combinations
of voiceless consonants. Pétursson (1976, p. 188) has found the
same in Icelandic.

4.3   The distinction between stops and affricates

The measurements undertaken have shown the affricates to
have shorter closure, longer open interval, slower decay of intra-
oral air pressure, slower rise of airflow and less intensity of
the explosion compared to stops. This suggests a weaker and
slower articulation.

4.4 Differences due to place of articulation and following vowel

The measurements of German stops have corroborated other findings pointing to some more universal tendencies (see e.g. Ilse Lehiste 1970, p. 27-30). The following tendencies have been found: labial stops have longer closure, shorter open interval, quicker rise of airflow, higher airflow peak, and (as far as ptk are concerned) quicker decay of intra-oral air pressure than dental stops. The same relations hold, although with less consistency, for dental stops compared to velar stops. Moreover, the degree of openness of the following vowel is of importance, the open interval being shorter and the decay of both intra-oral and subglottal air pressure quicker before a than before i and u.

It seems very plausible to assume that the differences due to following vowel must depend on the degree to which the flow of air is impeded by constrictions in the oral cavity: The low vowels permit a quicker escape of air and consequently a shorter open interval.

A similar reasoning may then be applied to the place of articulation of the stops: the labials have no constriction in the oral cavity and permit a quicker escape of air, resulting in a quicker decay of air pressure, a quicker flow of air, a higher peak airflow, and a shorter open interval than dentals and velars. Moreover, it may be assumed that the tongue tip moves more quickly than the bulk of the tongue, and perhaps the lips more quickly than the tongue tip, and this would give the same results. The relative speed of the movements might be concluded from the duration of formant transitions measured from the point of release. This distance was found to be shorter in the order labial < dental < velar in French stops (see Fischer-

Jørgensen 1972)[1]. As for the relation between closure durations, which are normally in inverse relation to the durations of open interval, the order being labial > dental > velar, it might perhaps be interpreted as a compensation phenomenon. In this connection it would have been of interest to measure the duration of the following vowel, but the consonants following the vowel were too different in the present material to allow of such measurements. (In Danish there is not only an inverse relation between open interval and closure duration but also between open interval and following vowel for ptk compared to bdg, but the reduction of the closure is considerably more pronounced than that of the vowel. As for the difference between the different places of articulation t, which in Danish is strongly affricated, has also a clearly shorter duration of the closure, whereas the shortening of the vowel is about the same for the three consonants.) Instead of interpreting the closure durations as caused by compensation phenomena, one might also think of direct physiological mechanisms. It might, for instance, be assumed that as the closure of p can take place quite independently of preceding and following vowels, it may start earlier and last longer (to decide this it would be necessary to measure preceding vowels, too). Moreover, the closure may be more firm at the lips than e.g. at the place of articulation of k and g (in Danish it is quite evident that t has a rather loose and thus a short closure as is also the case with German affricates).

More investigations are needed on this point.

---

1) A confirmation for German stops might be sought in Vieregge (1969), but the mathematics of this book exceeds my capacity of understanding.

4.5 The assumption of heightened subglottal pressure in aspirated sounds

It is well known that Chomsky and Halle have assumed a higher subglottal pressure in aspirated sounds. In the present material no such difference was found between ptk and bdg. On the contrary, there was a small dip in the pressure during the aspiration (also found in other languages, cf. Ohala 1972 and 1974), and it is remarkable that even in such small details as the larger dip before a subglottal pressure seems to depend on the conditions in the supraglottal cavities. One might, perhaps, modify the hypothesis of Chomsky and Halle by speaking of increased pulmonary effort instead, i.e. by assuming that in order to produce a rather constant subglottal pressure during different sound types, a higher effort of the expiratory muscles would be needed for sounds with a wide open glottis and with escape of air through the mouth, i.e. for aspirations, h, and - to a certain extent - fricative sounds. If the pressure went up again after the dip, this might be an argument in favour of such an assumption, but, as shown in section 3.5, the pressure only rises very slightly and is lower than the pressure in bdg at the start of the following vowel. It is, however, astonishing that during affricates it very often happens that subglottal and supraglottal pressure remain at the same level, even if there is a rather considerable escape of air. But in order to prove that this requires more energy on the part of the expiratory muscles more investigations are needed, and the curves will have to be calibrated.

4.6 Air pressure and airflow

It has often been assumed that the degree of air pressure during the closure determines the amount of airflow at the explosion. But in the present material there is no close correlation between these two phenomena. For instance, in a word like

guten the g has a higher intra-oral pressure than the t, but the latter a much higher peak airflow. And the consistent difference in airflow between ptk and bdg is not paralleled by a corresponding difference in air pressure. What is probably more important for the airflow is the degree of opening of the glottis at the moment of release. (See also Pétursson 1976, p. 180.)

4.7 Final remarks

It appears clearly from this investigation that much more research needs to be done before we can give a sufficiently well documented description of stop production in general and of German stop production in particular.

References

Bennett, S.G. 1935: Vergleichende experimentalphonetische Untersuchung der ungespannten Plosive im Englischen, Deutschen und Französischen (Bonn)

Bothorel-Witz, A. and M. Pétursson 1972: "La nature des traits de tension, de sonorité et d'aspiration dans le système des occlusives de l'allemand et de l'islandais", Travaux de l'Institut de Phonétique de Strasbourg 4, p. 277-356

Dieth, E. and R. Brunner 1943: "Die Konsonanten und Geminaten des Schweizerdeutschen experimentell untersucht", Festschrift Jud. Romanica Helvetica 20, p. 737-762

Esawa, K. 1972: Die Opposition stimmhafter und stimmloser Verschlusslaute im Deutschen (Tübingen)

v. Essen, O. 1934: Das Kompensationsprinzip beim Sprechvorgang (Hamburg)

198

Fischer-Jørgensen, E. 1952:   "Om stemthedsassimilation", Fest-
                                 skrift til L.L. Hammerich, p. 116-129

Fischer-Jørgensen, E. 1964:   "Sound duration and place of arti-
                                 culation", Zs.f.Ph. 17, p. 175-208

Fischer-Jørgensen, E. 1968:   "Voicing, tenseness and aspiration
                                 in stop consonants", ARIPUC 3,
                                 p. 63-114

Fischer-Jørgensen, E. 1969:   "Untersuchung zum sogenannten
                                 festen und losen Anschluss", Kopen-
                                 hagener germanistische Studien 1,
                                 p. 138-164

Fischer-Jørgensen, E. 1972:   "ptk et bdg français en position
                                 intervocalique accentuée", Papers
                                 in Linguistics and Phonetics to
                                 the Memory of Pierre Delattre,
                                 p. 143-200

Fischer-Jørgensen, E. and   "An electrical manometer and its
A. Tybjærg Hansen 1959:   use in phonetic research", Phon-
                                 etica 4, p. 43-53

Frøkjær-Jensen, B.,   "A glottographic study of some
C. Ludvigsen and J. Rischel   Danish consonants", F&S, p. 123-140
1971:   (also in ARIPUC 7, p. 269-295)

Hentrich, K. 1925:   "Ueber den Einfluss der Dauer auf
                                 die Stimmhaftigkeit von Verschluss-
                                 lauten", Vox, p. 21-24

Kim, Ch.W. 1970:   "A theory of aspiration", Phonetica
                                 21, p. 107-116

Lehiste, Ilse 1970:   Suprasegmentals (Cambridge, Mass.)

Lotzmann, G. 1958/59:   "Zur Aspiration der Explosivae im
                                 Deutschen", Wissenschaftliche Zeit-
                                 schrift der Humboldt-Universität
                                 Berlin 8, p. 150-185 (has not been
                                 accessible)

Meinhold, G. and
E. Stock   1963:

"Stimmlosigkeit und Stimmhaftigkeit
der Verschlussphase(plosion) bei
deutschen Medien im absoluten An-
laut und nach stimmlosen Lauten",
Zs.f.Ph. 16, p. 137-148

Meyer, E.A.   1901a:

"bitte - biete", Le Maître phonéti-
que, p. 39-41 and 59-60

Meyer, E.A.   1901b:

"Konsonantendauer im Deutschen",
Le Maître phonétique, p. 111-115

Meyer, E.A.   1904:

"Zur Vokaldauer im Deutschen",
Nordiska Studier tillägnade A.
Noreen, p. 347-356

Ohala, M. and J. Ohala
1972:

"The problem of aspiration in Hindi
phonetics", Annual Bulletin, Re-
search Institute of Logopedics and
Phoniatrics, University of Tokyo 6,
p. 39-46

Ohala, J.   1974:

"A mathematical model of speech
aerodynamics", ARIPUC 8, p. 11-22

Pétursson, M.   1976:

"Aspiration et activité glottale.
Examen expérimental à partir de
consonnes islandaises", Phonetica
33, p. 169-198

Rositzke, H.A.   1944:

"Short and long stops in High German",
Journal of English and German Philo-
logy 43, p. 88-93

Rositzke, H.A.   1947:

"A note on the length of German
stops", Le Maître phonétique, p.
22-24

Schmitt, A.   1931:

"Media, Tenuis und Aspirata",
Teuthonista VII, p. 287-297

Schmitt, A. 1947:   "Die neuhochdeutschen Verschluss-
                            laute", Zs.f.Ph. 1, p. 149-177

Schuhmacher, W.W. 1972:   "Beitrag zur Bestimmung des physio-
                            logischen Korrelats des deutschen
                            Vokalgegensatzes", Linguistics 90,
                            p. 35-78

Thorsen, Nina G. 1971:   "Voicing in British English t and
                            d in contact with other consonants",
                            ARIPUC 5, p. 1-39

Vieregge, W.H. 1969:   Untersuchungen zur akustischen
                            Struktur der Plosivlaute (Bonn)

Weitkus, M. 1931:   Experimentelle Untersuchung der
                            Laut- und Silbendauer im deutschen
                            Satz (Bonn)

Zwirner, E. and K.   Textliste neuhochdeutscher Vorlese-
                            sprache schlesischer Färbung,
                            Phonometrische Forschungen B 1,
                            1936

A SYSTEM FOR COMPUTER AIDED PROCESSING OF PHONETIC MEASUREMENTS[1]

Peter Holtse and Jens H. Stellinger

Abstract: A set of computer programs is described which will
sample and average e.g. intonation, air pressure, or
EMG activity. Data processing is essentially user
controlled, but trivial decisions may be left to the
computer if this is desired. The programs are acces-
sible to users with little experience with computers.

1. Introduction

During the last two years a system of computer programs has
been developed which aims at providing the phonetician with a
processing tool for the kind of DC-signals normally recorded on
e.g. Mingograms. The programs (called the Multi-Channel Pro-
cessing system, or MCP) should enable the researcher to process
statistically whole patterns of simultaneous events as well as
single, selected measuring points.

It has always been a problem, especially when dealing with
physiological recordings, that for practical reasons the investi-
gation is limited to selected points in time which the researcher
hopes will clarify his theory. It has been found, most notably
in the field of EMG, that single points are in most cases only
poor descriptions of the complex behaviour of the speech organs
during the articulation of a particular speech utterance. Intona-
tion is another good example of the importance of the pattern
rather than a point.

The programs described here is an attempt to provide a means of describing such complex measurements combined with the usual description of points. And it is hoped that they will be found useful in dealing with a variety of different measurements, such as for instance air pressure, air flow, EMG activity, intonation, and combinations between these and the audio signal.

Although the original reason for building the system was a project in EMG (see Fischer-Jørgensen 1974), the system is a general one, not written with only one kind of signal in mind. This means that in many ways it is incomplete, since obviously it cannot have every conceivable function ready made. However, the system is modular, so that many functions not thought of during the first planning stages can be implemented by changing only small sections of the complex of programs.

It was intended that the programs should be fairly easy to operate for the user with little or no knowledge of computer programming. This requirement is of course in direct conflict with the design features mentioned above. However, we believe that the original goal has been partially reached, at least for some of the more obvious processing functions.

2. System survey

The complete MCP system consists of four main parts:
1. The analogue recording process, in which the material to be investigated is recorded on a frequency modulating tape recorder.
2. Data editing and preparation. During this part of the process the FM-recordings are transferred to Mingograms and inspected visually. All information describing contents, the quality of the recorded material, where it is to be found, etc. is noted and transferred to punch tape.

3. The actual sampling process, in which the recorded material is transferred from analogue tape to digital representation and stored in digital form with its appropriate description of contents as it was specified during stage two.

4. The processing stage, during which the experimenter causes the digitized, stored data to be averaged and displayed in various ways.

3. Hardware set up

The MCP system is run on the lab.'s PDP-8/E computer with 32k of memory. It relies heavily on a 1.6M disk and uses a dual DEC-tape drive as additional storage. Data is sampled through an eight channel mulitplexed analog-to-digital converter controlled by a real-time-clock. Data, results, etc. are manipulated and displayed via a Tektronix 4014-1 graphic terminal and a DEC-writer LA30.

4. Description of the system components

4.1 Analogue recording process

The MCP system assumes that the material to be processed is recorded on an eight channel FM-tape recorder. Of these eight channels one is used by the system for a special time code which uniquely identifies any position on each FM-tape. The remaining seven channels are available to the user. The time code is generated by a computer program and may be recorded on channel eight during the actual experimental recording, or it may be inserted some other time. Fig. 1 shows a system flow chart of the analogue recording process.

204
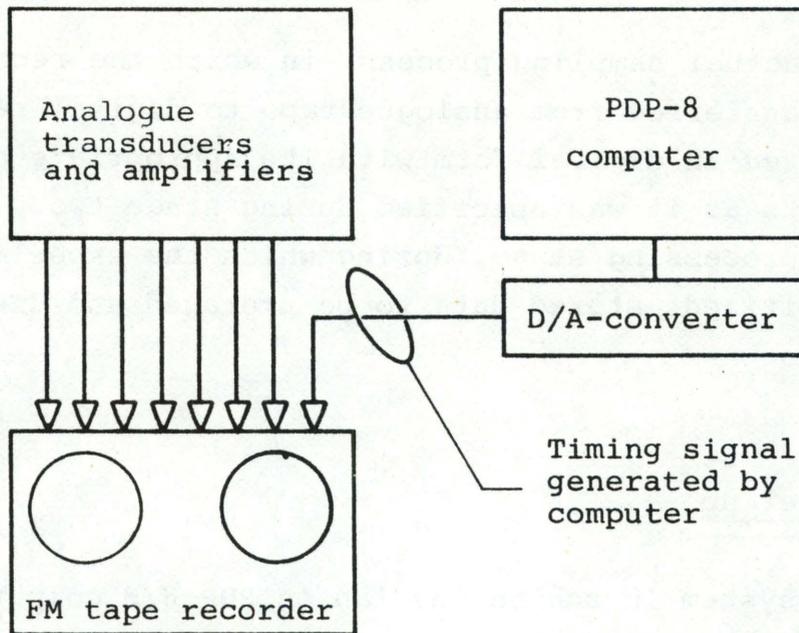
Figure 1

System flow chart of the analogue recording
process. Seven data channels are recorded on
the analogue tape recorder together with a
computer generated timing signal.

Interpretation
of octal code   6 0 4 7
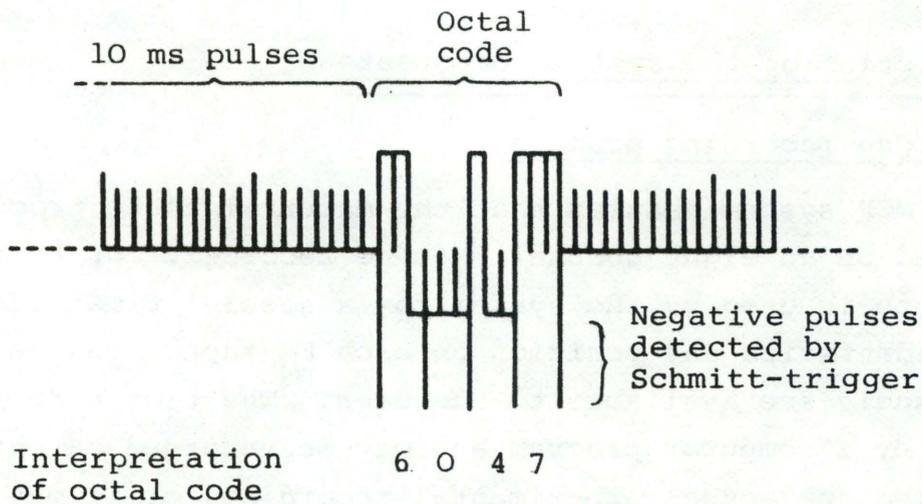
Figure 2

Time signal and twelve bit binary code displayed
on Mingograms. Upward shift of zero line is inter-
preted as a binary one. Downward shift as a binary
zero.

4.1.1   The time code

The time code is generated by a special computer program
(MCTIME) and recorded on the analogue tape via the computer's
analog-to-digital converter. The code consists of a short
positive pulse every 1/100 second. Every two seconds the zero
line is shifted up or down during twelve cycles to form a twelve
bit binary pattern as exemplified in fig. 2. As may be seen
from fig. 2, the basic 10 ms marking is recognizable even during
the shifting of the zero line. The twelve bits are combined by
threes and interpreted as a four digit number to the base of
eight. Hence the code is spoken of as the Octal Code.

In order to facilitate recognition of the code signal
during replay into the computer and to ease visual reading of
the code, the beginning of every octal code and the end of every
three-bit octal digit is signalled by a short, but large, nega-
tive pulse. During replay the time signal is led to both the
analog-to-digital converter of the computer and to a schmitt-
trigger entrance. The schmitt-trigger detects the negative pulses
at the beginning of the code and starts the A/D-converter samp-
ling the base line during the following twelve centiseconds.
These samples are interpreted as explained above.

4.2   Off-line preparation of data

The contents of the FM-tapes, complete with time codes,
is printed on Mingograms and inspected visually. This phase is
of paramount importance, since the system is not an automatic
data evaluator. Determining the quality of the recordings and
where the interesting periods are located is left entirely to
the experience of the researcher, who must later pass on his
judgements to the MCP system. The results of the researcher's
evaluation are collected in special files, known as Sampling
Control files (or S-files). For each combination of speaker
and utterance type one separate file must be collected. These
files will later control the sampling and further data processing.

A Sampling Control file must contain the following infor-
mation: The first record of the file identifies the speaker
and utterance type. The second record describes the contents
of the seven data channels to be sampled.[1] One record describes
where every repetition of this utterance is located, i.e. the
number of the analogue tape on which the individual item is
recorded, its nearest preceding octal code, and the distance in
centiseconds from the octal code to the beginning of the sampling
window. The sampling window is a 1.25 second wide window placed
round the item to be sampled. Only information within the samp-
ling window will be stored in digital form by the computer.

Another record describes a number of time references within
each sampling window. As many as seven time references may be
associated with each item. One or more time references are used
as line-up-points during the later averaging procedure, but the
researcher is free to use the time references for his own pur-
poses as well, such as describing the time distance between
various events within the sampling window.

Finally, the S-file contains a record describing the
quality of each channel of each sampled item. Thus each item/
channel combination is assigned a number in the range zero to
seven, where zero is normally assumed to describe an empty
channel. Otherwise there is no inherent hierarchy of the numbers
of the quality description, and the researcher may use them at
will to describe the quality of his recordings.

Special coding forms are available for preparing the
Sampling Control information. This information is later punched
on paper tape. The punch tape files are read into a computer
program, known as the MCP compiler, which checks the tapes for
spelling mistakes, missing information, etc. From each input

---

1) If an experiment needs less than seven channels the remaining
    channels are marked as empty. They are sampled and stored
but will not appear on any display.

tape the compiler produces a proofread S-file containing a complete description of all the repetitions of the utterance in question. This file is stored on magnetic tape (DEC tape) for easy reference later on. Furthermore, the compiler produces a printout of the complete S-file. The prints of the S-files provide the researcher with a useful survey of his sampled material.

The MCP compiler is equipped with an extensive default system, which allows the user to enter only really necessary information from punch tape, while information which is in any way predictable is supplied automatically by the default system. For instance, describing the contents of a channel will make the compiler assume that all repetitions of this utterance as spoken by that particular speaker are of acceptable quality. Therefore, the quality of the recordings need only be described explicitly in the (hopefully rare) cases where the recordings for some reason or other are not up to expectations.

Fig. 3 shows a listing of the Sampling Control information for one utterance.

4.3 Digital sampling process

The sampling process is controlled from the program MCSAMP. First the user enters a list defining the speakers and utterance types he wants sampled during the particular run of the program. MCSAMP then reads the appropriate Sampling Control information from the S-files created during stage two and merges the control information into one file. This file is sorted so as to minimize tape movement on the analogue tape recorder.

When sorting is completed the program makes its first request to the operator for a specific analogue tape to be mounted and starts reading the time code. Every time the program encounters a code combination defined in the control file it will sample a 1.25 second wide window from all seven channels with a sampling frequency of 200 Hz and store this record on the disk.

```
####################################################################
                              208                    IPUC DATALAB.
S S I - D A T A
                                                     DATE: 76/ 9/ 8

####################################################################
                            #
MCP PROJECT NAME:           #  EMG-74
                            #
                            #
FILENAME:                   #  DEMONZ. JS
                            #
SUBJECT:                    #  JS
                            #
                            #
RESEARCH ITEM:              #  DEMONSTRATION
                            #
                            #
CONTENTS OF CHANNEL-NR. :   #  1      BICP
                            #  2      TEMP
                            #  3      FRON
                            #  4      CT
                            #  5      VOC
                            #  6      F0
                            #  7      I
                            #
                            #
NUMBER OF RECORDS           #
IN DATAFILE:                #  6
                            #
                            #
ANALOGUE TAPES              #
AND SAMPLING-POINTS:        # REP. NR.      AT-NR.      OCT-CODE      DEC-CODE
                            #  1            67           316            10
                            #  2            67           325            32
                            #  3            67           404            41
                            #  4            67           413           126
                            #  5            67           477            39
                            #  6            68           051           142
                            #
                            #
TIME-REFERENCES:            # REP   TA    TB    TC    TD    TE    TF    TG
                            #  1    40    79    24    53    70    93   112
                            #  2    40    80    23    51    70    93   112
                            #  3    40    80    24    54    72    93   110
                            #  4    40    82    15    52    73    97   114
                            #  5    40    83    22    53    75    98   118
                            #  6    40    85    21    57    77    98   118
                            #
                            #
ACCEPTABILITY LEVELS        #
IN CHANNEL NR.              #        1     2     3     4     5     6     7
                            #"""""""""""""""""""""""""""""""""""""""""""""
                            #  1     4     6     2     3     5     4     1
                            #  2     1     2     3     4     5     6     0
                            #  3     0     1     2     3     4     5     6
                            #  4     5     5     5     5     5     5     5
                            #  5     7     7     7     7     7     7     7
                            #  6     7     7     7     7     7     7     7

####################################################################
```

Figure 3

Listing of Sampling Control information

When necessary the program asks the operator to rewind the ana-
logue tape or mount a new tape.

The relatively low sampling frequency implies that most
signals need to be hardware preprocessed before they can be
sampled into the computer. Thus EMG activity must, for instance,
be hardware filtered and integrated. And intonation is investi-
gated by sampling the output of a pitch meter. However, due to
the modular design of the system this limitation can, if desired,
be circumvented by altering only the sampling program (MCSAMP).
These alterations need not affect the controlling and sorting
programs at all.

When all requested items have been sampled a special sorting
program (MCSORT) is called. This program reads the sampled records
and sorts them into separate files (called Z-files) for each
speaker/utterance type. At the head of each Z-file is inserted
the Sampling Control information from that speaker/utterance
type combination. Thus all sampled items for a given utterance
and descriptions of contents and quality for that utterance are
stored in one file. Finally, the Z-files are transferred to
DEC-tape for long term storage.

Fig. 4 shows a system flow chart of the complete data-
entry process as described in sections 4.2 - 4.3.

4.4 Processing of data

During the design phase of the MCP system it was realized
that it is impossible to foresee in all detail what the pro-
cessing requirements of the future users of the system will be.
It was therefore decided to let the user control the procedures
directly in some detail rather than making a complicated auto-
matic system. This means, of course, that users must spend some
time typing out orders to the system. But the interactive nature
of the system will, on the other hand, allow the user to see the
results of his orders almost immediately, if he wishes.

Analogue data and
time code printed on
Mingograms

Visual inspection
of analogue data

Analogue
tape
recorder

Sampling Control
information on
punch tape

Analogue
preprocessing
of signals

Compilation and
correction of
Sampling Control
information

MCCOMP

Sampling
from
analogue
tapes

Sampling Control
information stored
on DEC-tape (S-files)

Listing of
Sampling Control
information

MCSAMP

Sampled data
stored on disk

MCSORT

Sampled data
stored on
DEC-tape

Sorting of sampled data to
appropriate (Z-) files

Figure 4

System flow chart of the MCP data entry process.

At the moment the processing procedures consist of the following five programs:

MCMODF - A data modification program which works directly on the Z-files. The contents of one specified record of the file is displayed on the terminal screen, and the user may enter various modification into this record and its corresponding description in the file head. Erroneous data values may, for instance, be deleted or time references may be altered. It is possible to delete single measurements in the middle of a specified channel without discarding the whole channel.

MCAVEX - Calculates mean values, standard deviations and number of tokens from one or more Z-files. The results are stored on the disk in a special X-file. For each input file (i.e. Z-file) to the program the user must define which of the time references should be used as line-up-point for the averaging process. Averaging is carried out with the input records aligned to these points, and the results are a 250 points wide mean window.

This possibility of defining a separate line-up-point for each channel is an important feature of the system. It is possible, for instance, to average the contents of one channel round a line-up-point at, say the explosion of a stop consonant, while another signal in the same recording is averaged round a line-up-point located at the beginning of a neighbouring /s/.

Channels to be averaged are defined by their contents - not by number. Thus it is possible to ask for, for instance, the EMG activity of the Vocalis muscle to be averaged without worrying about whether this was stored in channel two or three. This feature is particularly useful if somebody wants to average a given type of recording from two different recording sessions where the contents of the channels for some reason have not been ordered in the same way.

MCAVEM - Is an averaging program similar to MCAVEX, but calculates median values and quartile distances instead of means and standard deviations. The output is known as an M-file.
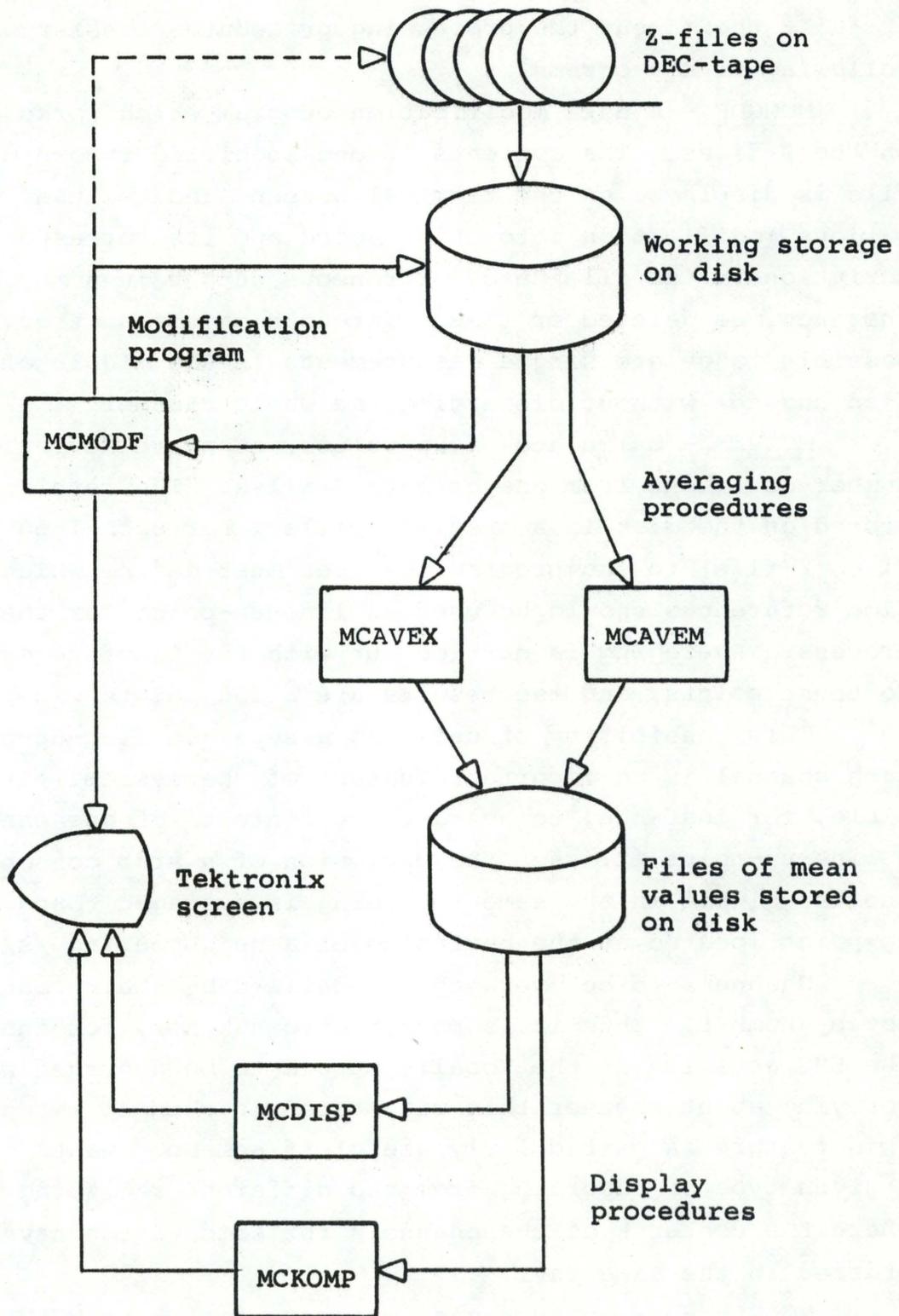
Figure 5

System flow chart of the processing part of the MCP system.

MCDISP - Displays the contents of all the channels of one average file (X- or M-) on the graphic screen.

MCKOMP - Displays several specified channels from specified average files for comparison on the graphic screen.

The logical interaction of the processing programs is shown in fig. 5.

Great pains has been taken to make communication with the processing procedures as simple as possible to the user not familiar with the working of computers. This problem has been approached by providing a set of supervising programs which allow the user to control the processing with approximations to normal words. Furthermore, a default system has been incorporated so that the programs will assume certain standard values for the input parameters if they are not explicitly defined by the user. This feature will in most cases cut down appreciably the amount of typing to be done by the user.

Hard copies of the displayed data and results are at the moment obtained by photographing the display screen. Figs. 6, 7,8,9 show examples of displays.

5. System capacity

The MCP system will handle almost any number of different utterance types and speakers, although the number of DEC-tapes available will usually put a practical limit to the amount of material that can be stored at one time.

The maximal length of one Z-file (containing raw, sampled data) is 17 repetitions. Since, however, the averaging programs will combine a maximum of 49 different input files into one mean file, utterances repeated more than 17 times per speaker are simply stored in two or more files and combined during the averaging process.

During one run the sampling and sorting programs will sample from a maximum of ten different analogue tapes, sort the sampled records into a maximum of 63 Z-files (one for each speaker and utterance type), or sample and sort a maximum of 450 items - whichever of these three limits is reached first. Experiments exceeding the limits must run the sampling and sorting programs several times.

References

Beauchamp, K.G. 1973:   Signal processing: Using analog
                             and digital techniques (London)

Fischer-Jørgensen, E. 1974: "Electromyographic investigation
                             of Danish consonants, stress, and
                             stød", ARIPUC 8, p. 203-206

Knuth, D.E. 1973:   The art of computer programming
                             Vol. 3: Sorting and searching
                             (London)

Lorin, H. 1975:   Sorting and sort systems (London)

Madnick, S.E. and   Operating systems (New York)
J.J. Donovan 1974:

Port, D.K. 1971:   "The EMG data system", Haskins SR
                             25/26, p. 67-72

Legend to figs. 6,7,8,9 (pp. 216-219):

Figure 6

Display of sampled data from utterance: 'Det er wiener-
brødet de siger' [de:'viː:?nʌbrœð?ddi'si(ː:ʌ)] as spoken
once by speaker 'BM'.
Time references associated with this item are 'TA' at
the boundary between [v] and [iː] and 'TB' at the ex-
plosion of [b].
The contents of each data channel is printed to the left
of the display. The two top channels are EMG recordings.
The two bottom channels are fundamental frequency and
intensity. (The figures to the right of the display
are internal scaling factors used by the display rou-
tines.)
The display was obtained by photographing the Tektronix-
screen, which is slightly convex. This reproduction
technique unfortunately introduces a slight distortion
to the scale, especially towards the edges of the screen.

Figure 7

Results of averaging all repetitions of the utterance
by speaker 'BM' which is exemplified in fig. 6.
Channels 'CT', 'VOC' and 'I' have been averaged with
time reference 'TA' (fig. 6) as line-up-point. Channel
'FØ' was averaged with 'TB' as line-up-point. The un-
broken lines represent mean values. The broken lines
indicate ±1 standard deviation from the mean.

Figure 8

Sampled data from utterance: [de:bibi'biː:bibi(...)]
spoken once by 'BM'. Seven time references (TA...G) are
used to indicate various points of the utterance.

Figure 9

Average of all repetitions of the utterance exemplified
in fig. 8. Time reference 'TF' was used as line-up-
point for all channels.

Figure 6

Figure 7

218

Figure 8

219
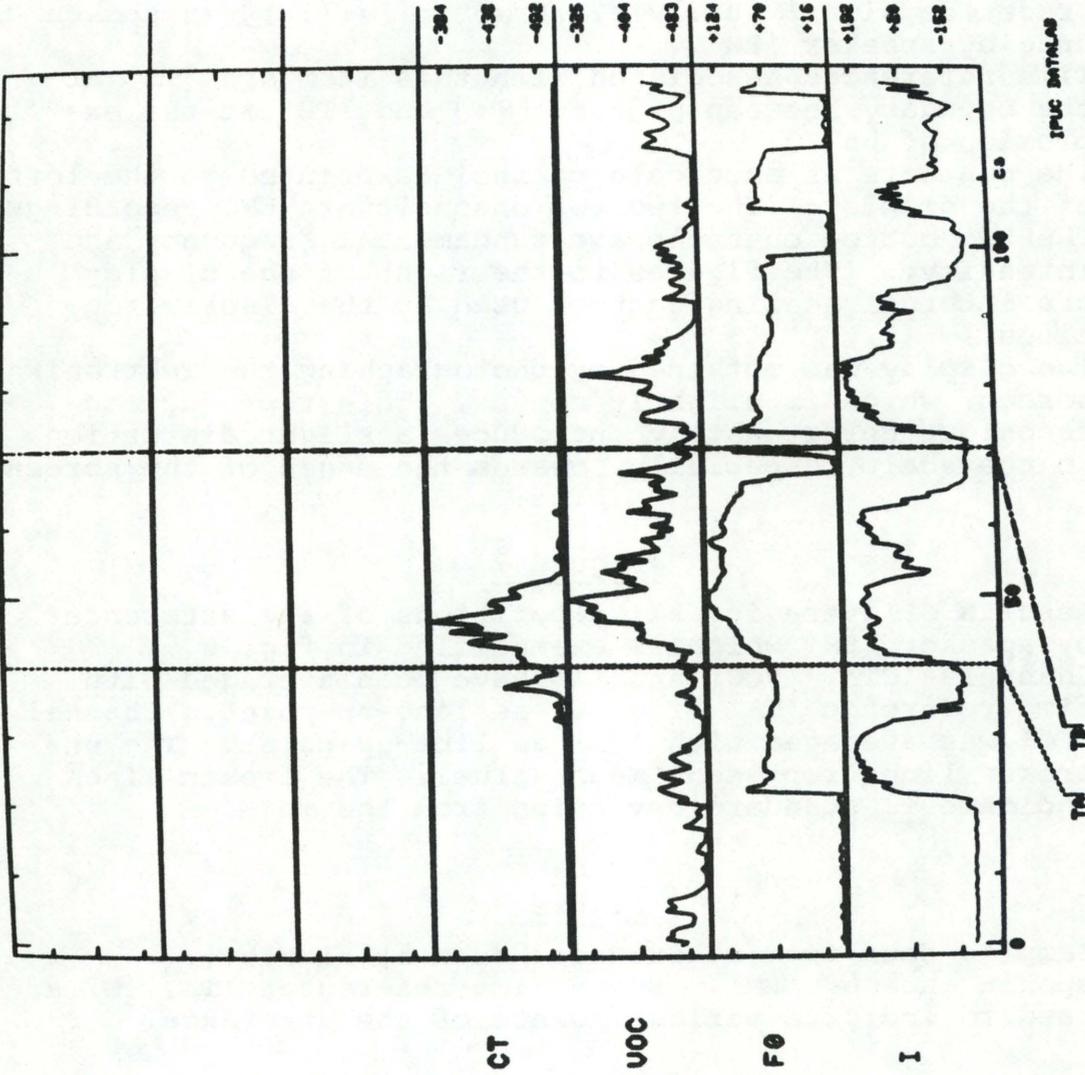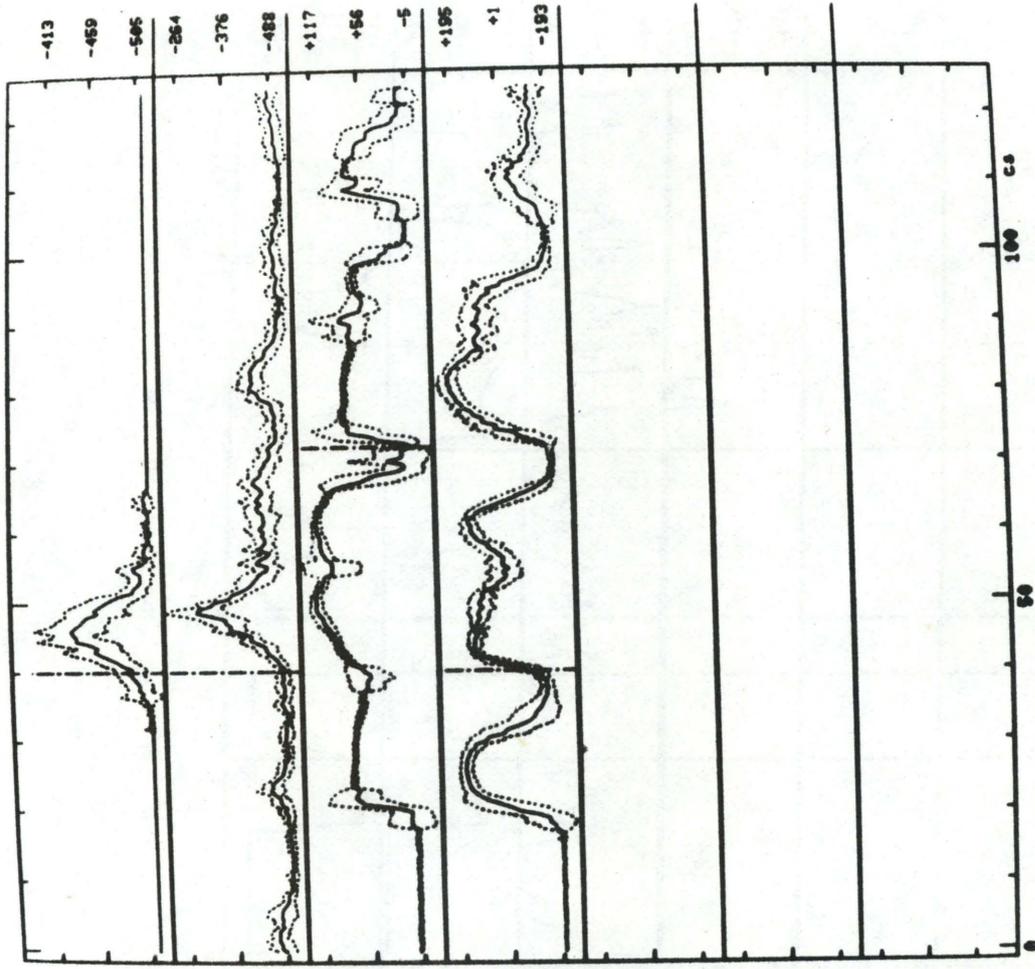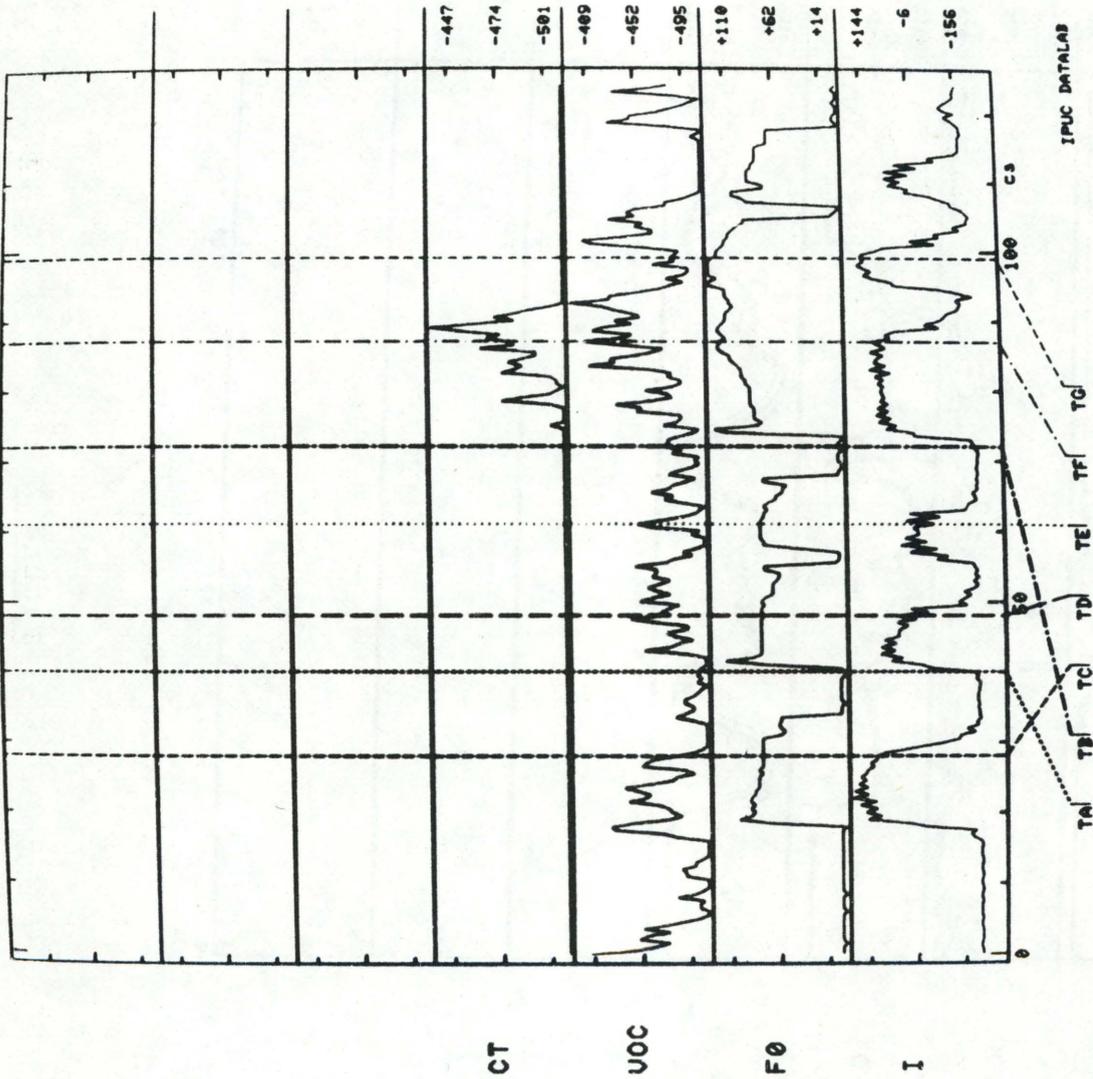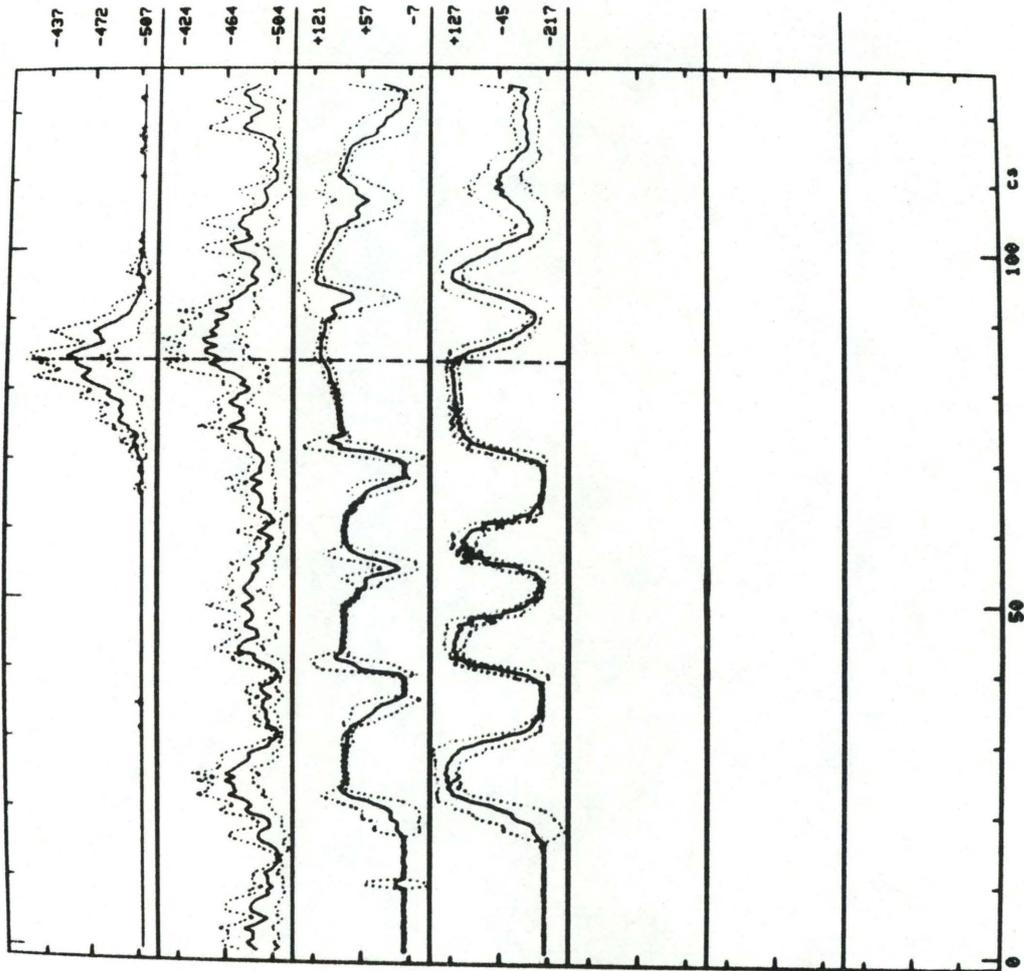
Figure 9

MUSICAL ACOUSTICS
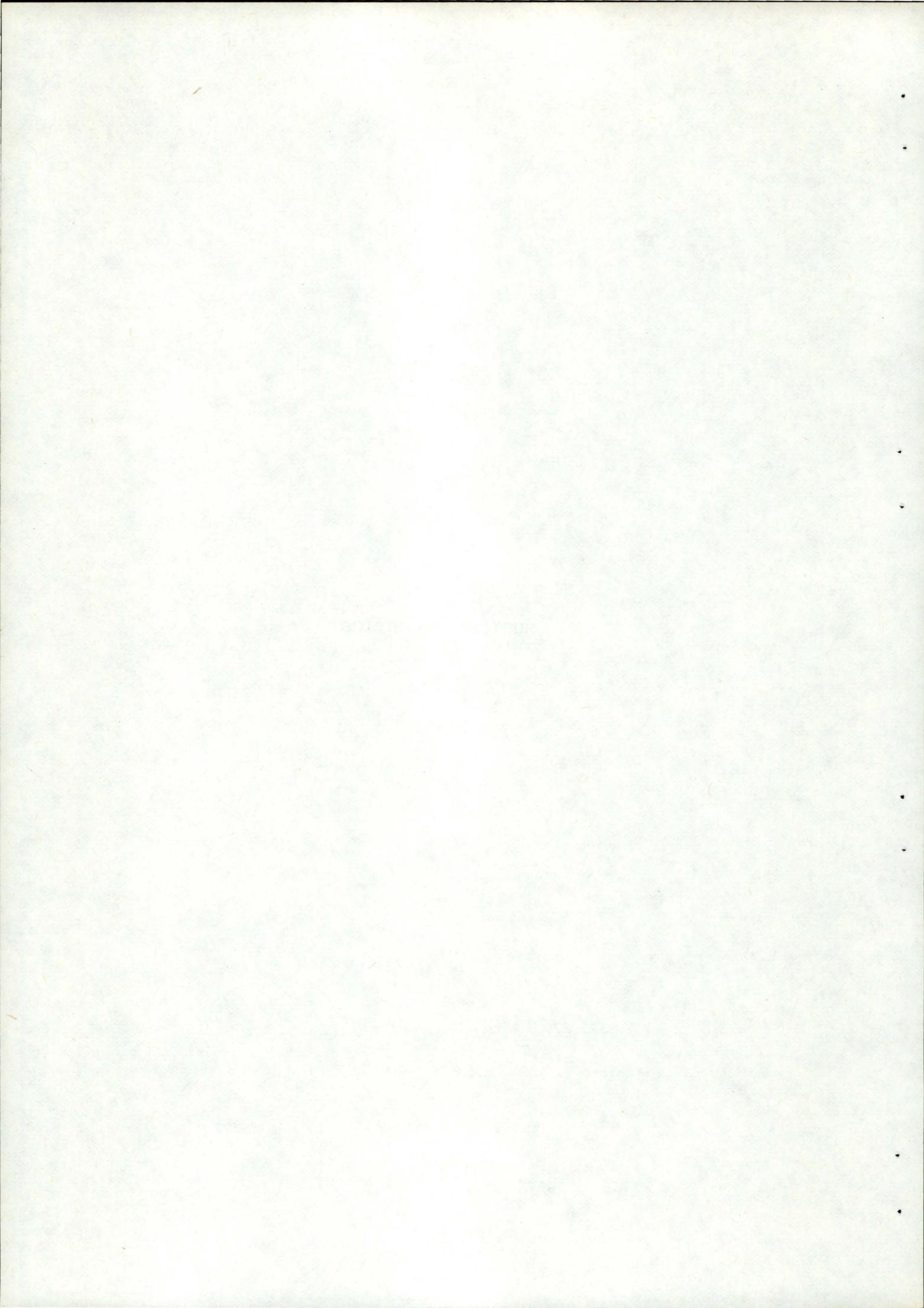
A PHYSICAL AND PHYSIOLOGICAL STUDY OF BLOWING TECHNIQUE
IN RECORDER PLAYING

Niels Bak

Abstract: The research summarized in this paper has revealed
            some of the specific physiological conditions of
            recorder playing. In order to understand the nature
            of the instrument it is essential to be aware of
            these physiological conditions, although they seem
            to have been largely neglected by pedagogues.
            It appeared during the investigation that registra-
            tion of blowing pressure raises a number of problems.
            The author succeeded in identifying some of the spe-
            cific causes of measurement error, and the paper de-
            scribes an approach which makes it possible to detect
            such errors.
            The measurement errors encountered were found to
            provide interesting information on the technique used
            by the individual recorder player.

1. Introduction[1]

    The changes in musical styles which took place in the eigh-
teenth century led to the replacement of the recorder by the
transverse flute in orchestral and chamber music, but in the
twentieth century the recorder has experienced a revival due to
such factors as growing interest in the music of earlier eras,
recognition of the value of the recorder in school music programs,
and widespread use of the recorder as a folk instrument.

Rising interest in this historical instrument has led in-
evitably to a pressing need for proper guidance, a need which
leading artists and experienced instrumentalists idealistically
attempted to provide by numerous reports of personal experience,
including detailed instructions concerning blowing technique, in
the pedagogical literature. This initiative has resulted in an
interesting documentation of the characteristics of different
methodological schools, but has hardly contributed to clarifica-
tion in the area of methodology, a fact which is reflected in the
following quotation from the literature: "Is there anywhere else
in the whole field of educational publishing a literature so
remarkable for its inadequacies and inconsistencies?" (Rowland-
Jones 1973, p. 41).

It is the uncertainty with respect to pedagogical questions
such as that reflected in the above mentioned literature which
has motivated the research reported here, the purpose of which
is to contribute to the required clarification of the subject by
providing objective information about the physiological and phy-
sical aspects of recorder playing.

The aspect of sound production in wind instruments which
so far has been particularly investigated is <u>blowing pressure</u>.
The results of the measurements reported by various investigators
are, however, very dissimilar, which is probably due to differ-
ences in the methods used (see, e.g., Anastasio and Bussard 1971, p.
69, and Fletcher 1975, p. 233). Therefore, in this paper, con-
siderable emphasis is placed on the measurement of blowing pressure,

including critical evaluation of the methods employed. The paper
should be seen in this context as an introduction to the problems
involved and as a plea for standardization of measurement prac-
tices with a view towards facilitating the exchange of ideas
among all interested parties.

The present paper deals exclusively with the recorder it-
self. It may be appropriate to add that, because of its relatively
simple blowing mechanism,[1] which reduces the number of parameters
that determine the instrument's character, the recorder presents
itself to the researcher as a particularly convenient object for
such investigations, and that the results of this project indi-
cate that research into the problems of recorder playing will
contribute to a better understanding of the playing techniques
of related instruments as well.

2. Previous studies and theoretical perspective

The studies reported here are a continuation of the author's
previous research which is summarized in this section by way of
introduction.

The first series of experiments, in which F-alto recorders
were blown by means of a mechanical apparatus, was intended to
clarify the relationship between blowing pressure and fundamental
frequency (Bak 1970). The experimental data obtained confirmed
that, as shown by earlier investigators (Lüpke 1940; Mühle 1966),
the main principle in playing the recorder is that the blowing
pressure[2] must be increased note by note when playing an ascending
chromatic scale. The author's studies showed, furthermore, that

---

1) See fig. 1 for a schematic drawing of the sound generating
   system.

2) "Blowing pressure" is, strictly speaking, not an unambiguous
   term, for, as the word "blowing" implies, besides the (excess)
   static pressure which acts equally in all directions, flow, and
   hence aerodynamic forces, are also present in the air channels. The
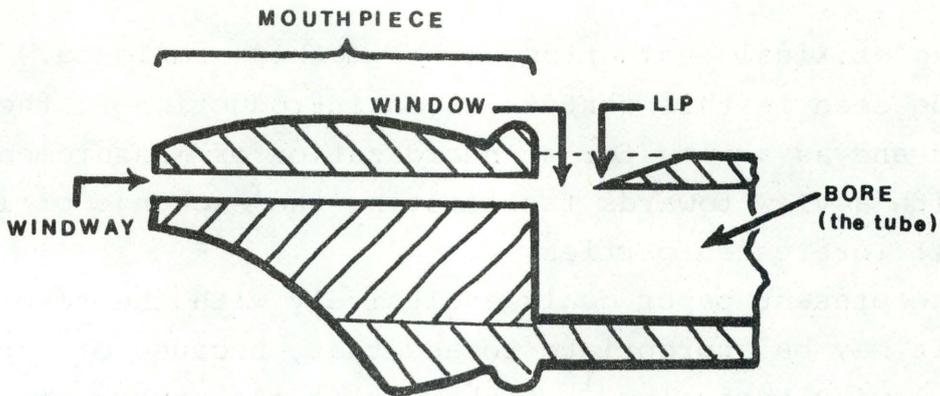
Figure 1

Parts of the recorder which are referred to in the paper.
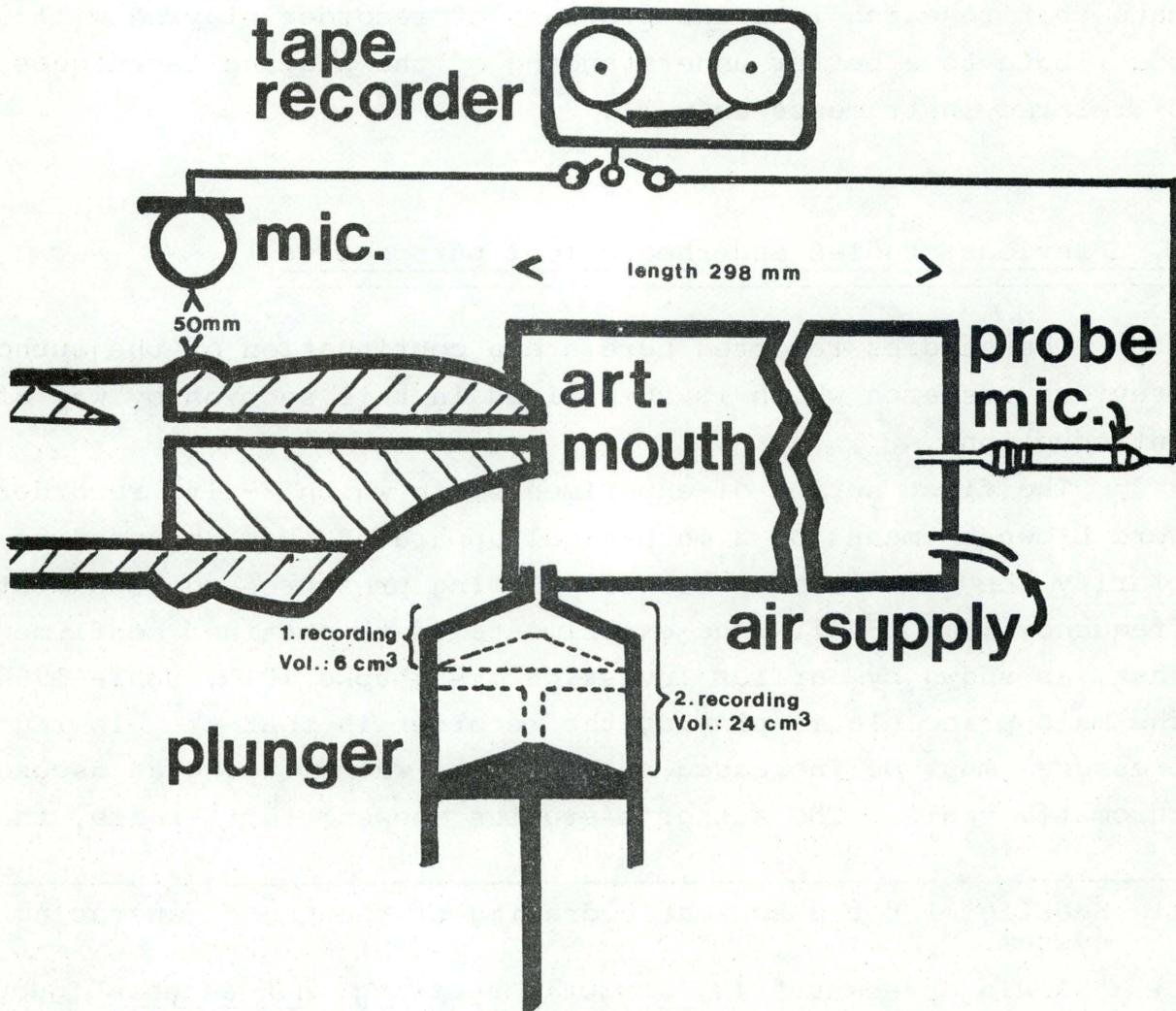
Figure 2

Block diagram of the experimental setup used for evaluating the sound spectrum of the recorder tone as a function of the resonant properties of the artificial mouth during mechanical blowing.

there is a clear positive correlation between blowing pressure and the fundamental frequency with unchanged fingering. Finally, in pilot studies made in connection with the present project, investigations were made to find out whether changes in the volume of that section of the air channel which corresponds to the instrumentalist's mouth and throat can influence the acoustic output of the recorder, as contended by schools based on the theory referred to in this paper as the "vowel-" or "resonance-" theory.[1] The final and most important studies in the latter series were based on the assumption that if, as believed by some teachers, the musician may influence tone and pitch appreciably by changing the resonances of the mouth cavity, then it should be possible to register at least a tendency toward such a relationship when using mechanical blowing technique via an artificial "mouth".

(footnote, continued from the preceding page)

transfer, during play, of muscle power to the sound generating system of the instrument (the functioning of which influences the pitch of the tone) is due to just this combination of the two forms of energy.

Because of the simultaneous presence of the two forms of energy, it is only in special cases that an unambiguous connection may be found in flute playing between oral air pressure alone and frequency. Such a case exists when one is dealing with a large reservoir with negligible air velocity inside the mouth and a well defined geometry of the outlet from the mouth, where all pressure energy ultimately is converted into kinetic energy, cf. Bouhuys (1964, p. 974). By employing mechanical blowing, it was possible to fulfil this condition and, under such stable conditions, the author succeeded in demonstrating that the frequency (f) of the tone emitted by the recorder is related to the blowing pressure (p) (within certain limits), according to the equation

$$f = k \times p^{\alpha}$$

where k and α are constants (Bak 1969; 1970).

The changes in measurement conditions during natural blowing by human subjects are discussed later in this paper.

1) The terms "vowel theory" and "resonance theory" were invented for this paper for the sake of clarity of exposition. The terms are synonymous, the former being used particularly in the context of pedagogical issues, and the latter in dealing with acoustic aspects.

After various futile attempts to affect tone and pitch by
varying the resonance conditions of the artificial mouth, the
series was terminated with experiments using the apparatus
sketched in fig. 2.

The tone blown in the experiment described here was D_5
(frequency 587 Hz).

The artificial mouth itself consisted of a cylindrical tube
closed by flat plates at either end. Normally, the sound field
within the "mouth cavity" is not audible, but in this experiment
it was registered by means of a probe microphone as shown in
fig. 2, whereas the tone produced by the recorder was recorded
by a microphone placed 50 mm above the mouthpiece. Both micro-
phone signals were recorded on tape for use in subsequent analyses
in which the spectral composition of the two signals could be
compared.

With the probe microphone pilot studies of the acoustics
of the artificial mouth were carried out before attachment of the
variable resonator. Standing waves with maximum sound pressure
variation at the end plates of the tube and a sound spectrum com-
prising a set of partials corresponding to those of the recorder
signal were registered. The first harmonic was of considerable
intensity (with a sound pressure level of more than 100 dB, re
2×10^{-5} N/m^2) which was expected, since the length of the "mouth
cavity" was approximately half the wave length for a tone equal
to 587 Hz.[1]

Finally, the extra resonator, whose resonance frequency
could be adjusted by moving a plunger, was coupled to the arti-
ficial mouth. In the experiments which followed, the plunger was

---

1) As wave length (λ) depends upon frequency (f) and the velo-
   city of sound (c) according to the formula $\lambda = \frac{c}{f}$, then, when
c is set equal to ca. 34,400 cm/sec and f is 587 Hz, $\lambda \approx$
(34,400/587) cm = ca. 58.6 cm. An effective half-wave resonator
must therefore have the length $\ell = 1/2 \lambda = 1/2 \times 58.6$ cm = 29.3 cm.

adjusted in such a way as to produce a considerable variation in sound pressure level for each of the first three harmonics within the artificial mouth in turn. The object of these experiments was to investigate the degree to which such internal acoustical changes could influence the tone produced by the recorder as registered by the external microphone.

The following example illustrates a typical result: The plunger was first set for 6 cm^3 volume, and then for 24 cm^3. At 6 cm^3 the probe microphone in the artificial mouth registered a total sound pressure level of 133 dB (re 2 x 10^{-5} N/m^2), and the sound pressure level of the first harmonic was also 133 dB. At 24 cm^3 (maximum resonance), the total sound pressure level fell to 109 dB and the level of the first harmonic to 94 dB.[1] The tone emitted by the recorder, however, underwent no measurable changes.

All additional tests in this series confirmed what is indicated by the results mentioned above: no changes in the acoustic conditions in the cavities behind the "windway" can in any way influence the tone produced by the recorder as heard in the auditorium. The possibility that the musician's sensation of the tone can be influenced by the sound field in the mouth and throat through transmission of the sound by tissue and bone conduction cannot, of course, be excluded. An examination of this question has not, however, been included in the present studies.

Since the results mentioned above, together with the studies reported below, suggested that the "vowel theory" must be abandoned, interest centered upon the pedagogical works on the recorder written in English, where traditionally no resonance effect is ascribed to the mouth and throat cavities (Rowland-Jones 1973 p. 107;

---

1) The acoustic conditions under which a resonator has a damping effect are discussed, among other places, in Rayleigh 1945 p. 2, and in Beranek 1954 p. 69.

Hunt 1963 p. 113). But even in English methodology, viewpoints are encountered which are irreconcilable with the conclusions of the research reported here. For, although the English school advocates the principle that control of variations in blowing pressure is the decisive factor in playing the recorder, the respiratory muscles are apparently considered the only legitimate source of variations in pressure. Thus, aside from special activity during attack, the articulatory organs are assigned the passive role of adjusting themselves so as "to allow an unimpeded flow of breath"[1] (Rowland-Jones 1973 p. 44).

However, as will be documented below, the X-ray recordings made during the present investigations clearly show that the articulatory organs <u>are</u> active, and that this activity displays such significant common characteristics among the different experimental subjects that it must be regarded as an important, integral part of the blowing technique for the recorder.

Since an interpretation of the experimental results presented here thus cannot be based upon existing theories, emphasis will be placed in this paper upon continual evaluation of the degree to which the material collected provides a basis for an alternative theory. This possibility will be examined in the concluding discussion after presentation of the graphs and radiographic material.

3. The experiments

The remaining part of this paper is concerned with a research program in which the activities in mouth and throat during play were investigated with the use of X-ray equipment in combination with recordings of tone and blowing pressure.

---

1) "Impeded/unimpeded flow of breath" will, for the sake of simplicity, be used to indicate that a narrowing - especially at the lips - creates a resistance to the airflow which is either greater ("impeded") or less ("unimpeded") than the normal resistance produced during passage of the exhaled air through the "windway".

3.1   Procedure

The experiments were carried out at the X-ray department of the Copenhagen Dental Hospital in mid-January, 1972, all using the same F-alto recorder.

While the subjects played a previously determined program, a continuous X-ray recording of the movements of the lips, jaw, tongue, and throat was made on video tape, in synchrony with a tape recording of the sound produced.

The blowing pressure was recorded through a hole bored in the tip of the mouth piece, as shown in fig. 3a. The aperture of this bore was connected by a polyethylene tube to a pressure transducer (an electromanometer), the output voltage of which was recorded by a Mingograph (ink jet recorder). Thus, graphs of pressure as a function of time were recorded as the subject played. - The instruments used for this part of the experiment were moved to the dental hospital from the Institute of Phonetics (University of Copenhagen).

In later experiments the static pressure in the windway was also registered, by means of a small probe tube, probe II, as shown in fig. 3b. The reason for introducing this extra measurement of pressure will be discussed below.

Graphs indicating variations in frequency were drawn up later at the Experimental Phonetics Laboratory of the Institute of Phonetics during follow-up analyses of the microphone signals recorded on tape. The same Mingograph was used for tracing of both pressure curves and frequency curves, making it possible to assign approximately equal graphic areas to equal time periods and, later, to identify temporally related features of the two curves and thereby identify those pressure and pitch curves which belonged together.

The registration technique using the Mingograph was later employed in the same manner in order to compare variations in pressure and frequency with the resulting changes in the harmonic structure of the tone. The results of this investigation will be discussed in a subsequent paper.

3.2   Subjects

The seven recorder players who offered their services re-
presented all levels of proficiency from the complete beginner
to artistic players. They have been assigned the letters A to G,
arranged roughly according to descending level of technical
ability, so that the first letters of the alphabet represent
subjects who are fully developed professionals, while subject G,
who had no prior experience in playing wind instruments, played
the recorder for the first time during the experimental recording
session.

A valuable supplement to the experimental data was obtained
by repeating the experiments with Professor Ferdinand Conrad of
Hannover as a subject. Professor Conrad was visiting Copenhagen
on a concert tour and was kind enough to place himself at the
author's disposal. A different instrument was used during these
experiments. This instrument was provided with an extra probe
as shown in fig. 3b. While the subject played, pressure curves
were thus registered simultaneously both at the tip of the mouth-
piece and in the windway. Therefore, two pressure curves were
available for each recording of Professor Conrad in the examples
cited below.

3.3   Measurement of blowing pressure during natural blowing

In earlier experiments reported in the literature, the intra-
oral pressure during blowing was registered by means of a probe
placed inside the mouth (Roos 1936 p. 4; Roos 1940 p. 139; Bouhuys
1964 p. 968; Bouhuys 1965 p. 453; Anastasio and Bussard 1971, p. 64,
cf. Fletcher 1975 p. 233). Since, in the present investigation,
it was considered particularly important not to interfere with
the activity of the articulatory organs while the X-ray recordings
were being made, it was decided to measure the blowing pressure
by means of a hole bored in the extreme tip of the flat end of
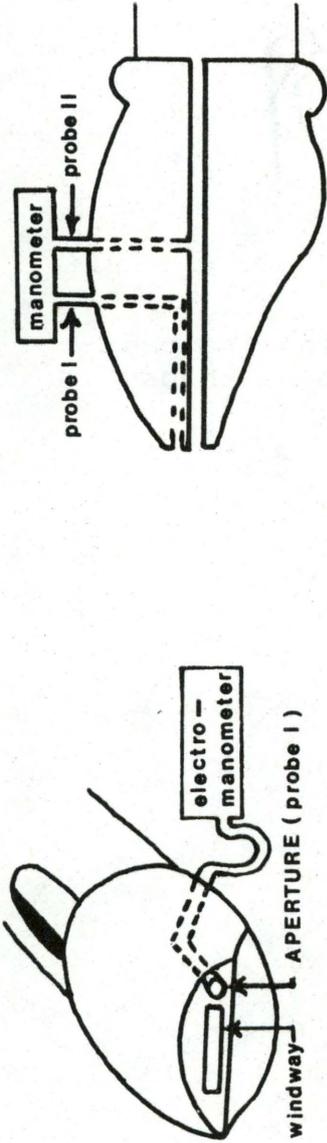
Figure 3a

Figure 3b

Fig. 3a shows the placement of the aperture of probe I used for recording the (total) blowing pressure. - Fig. 3b shows how an additional probe, probe II, was placed in order to measure a local, static pressure in the windway.
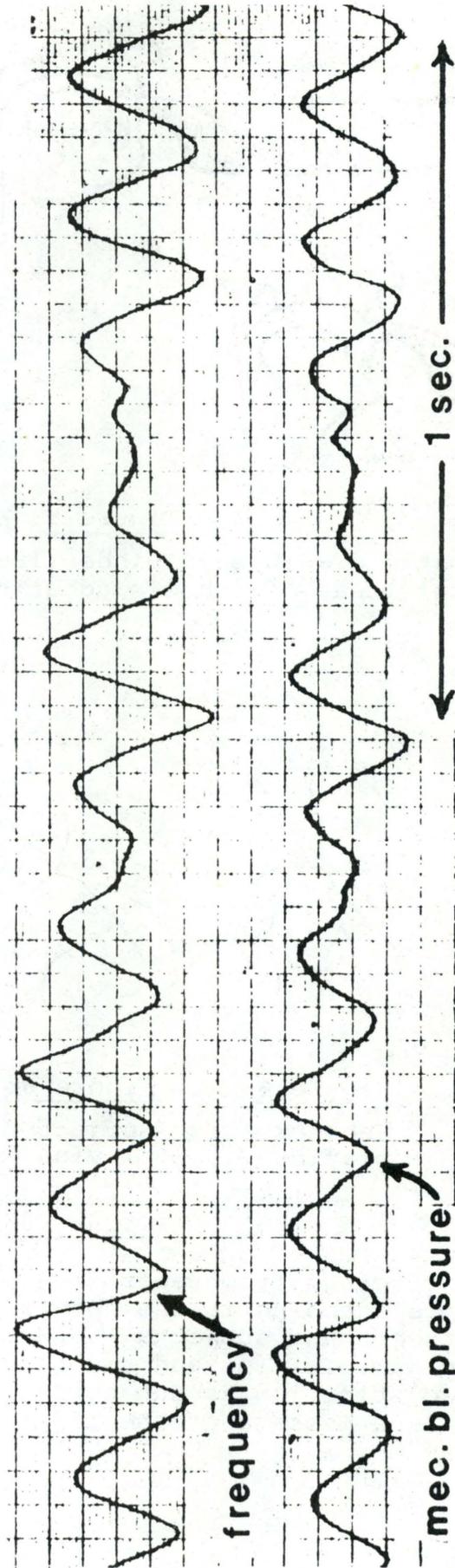
Figure 4

Sample of frequency and pressure curves showing how the fundamental frequency of the recorder depends on the blowing pressure with mechanical blowing. (There is a constant offset in time between the two traces.)
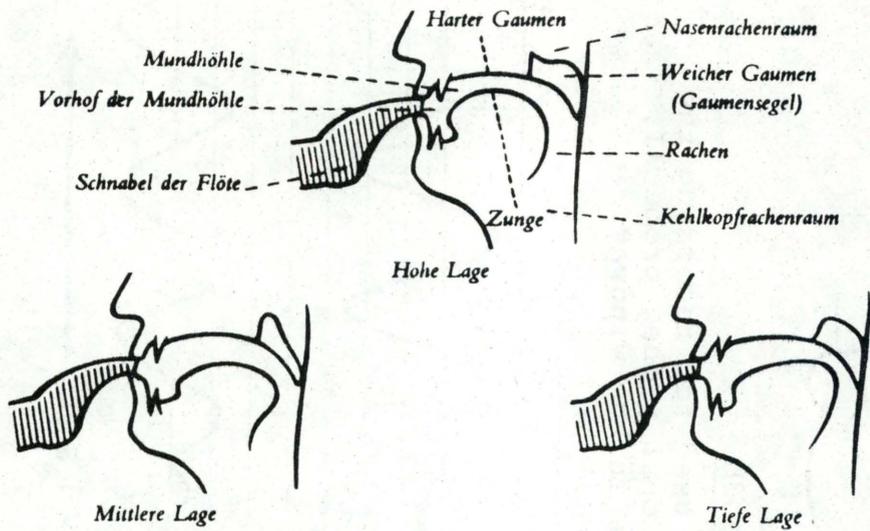
Figure 5

Schematic drawings in Linde (1962), reflecting the peda-
gogical ideas of schools adhering to the "vowel theory".
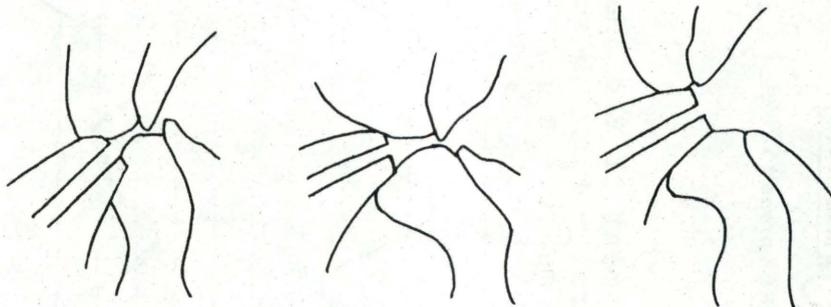
Figure 6a	Figure 6b	Figure 6c
Subject A playing G_4	Subject B playing A_4	Subject C playing G_4

Figure 6

X-ray tracings from recordings of the three professional
subjects intoning in the deep register of the recorder.
It is seen that subjects A and B tend to have a relatively
closed jaw position and a constriction at the lips, in
contradistinction to subject C, whose articulation provides
for a free passage of the airstream at the entrance to the
windway.

the mouthpiece, as shown in fig. 3a. The pressure measured at
that point, i.e. just at the level - perpendicular to the direc-
tion of flow - where the exhaled air begins to flow into the
windway, is here taken to be the blowing pressure.[1]

For comparison purposes, the bottom curve in fig. 4 shows
the pressure registered through the aperture during mechanical
blowing, where it was ensured that the airflow passed unhindered
from the artificial mouth into the windway. It will be noted
that the pressure curve and the simultaneously recorded frequency
curve follow each other in a monotonous up-and-down motion.

The uniformity of the pressure at probe I on one hand, and at
the entrance to the windway on the other, would undoubtedly be
preserved if the human subjects fulfilled the traditional peda-
gogical demand for "free passage for the exhaled air", as exempli-
fied by the schematic drawing in fig. 5, taken from a recorder
textbook in German inspired by the above mentioned "vowel theory"
(Linde 1962 p. 25), or by the English schools' instructions to
maintain "an unimpeded flow of breath" (Rowland-Jones 1959
p. 44). Several of the subjects actually did fulfil this require-
ment, for example subject C, as seen in fig. 6c.

But if the flow of breath is "impeded" by a constriction
of the lips, as indicated by some of the X-rays (cf. figs. 6a and
6b), it is quite possible that the method employed for measure-
ment of pressure will produce misleading results, since the correct

---

1) Note the similarity of the chosen method of measurement to
   measurements of the total pressure by means of a pitot tube,
an open-ended tube facing the airstream (Ower 1966 p. 9).

measurement requires that the aerodynamic conditions at probe I are not disturbed.[1] As long as the narrowing of the lips is moderate, this will presumably still be the case, but with increased reduction of the lip aperture a number of aerodynamic effects may occur which prevent a uniform distribution. In

---

1) It is doubtful whether traditional methods of measuring pressure inside the oral cavity would be of more use under these conditions, for according to theory the oral pressure may only be regarded as identical to the blowing pressure as long as the exhaled air passes unhindered from the mouth into the mouthpiece of the instrument (cf. footnote on p. 239). When, as in the case in question here, resistance at the lips is introduced, the potential energy consisting in the higher pressure within the mouth is expended in overcoming the variable resistance to the flow of air at the lip opening as well as on sound production in the instrument. Consequently, under these conditions, no unambiguous relationship exists between the oral pressure and the energy consumed in blowing the recorder. In order to throw light on these relationships a supplementary experiment was made, whose result is given in the diagram of fig. 7. In this diagram the blowing pressure is given as a function of the constriction between two artificial lips, the system being supplied with air by mechanical blowing. The two notes blown in this manner, viz. D_5 and A_4, were kept at constant frequencies of 587 Hz and 440 Hz, respectively. The experiment was performed in an ambient temperature of 19° C.

As might be expected, the results indicate that variations in the distance (h) between the artificial lips have the greatest effect in the range where the constriction is narrowest. Thus, the note D_5 comes out with practically the same fundamental frequency regardless whether one chooses the combination

h = 0.8 mm; pressure = ca. 29 mm H_2O

or the combination

h = 1.6 mm; pressure = just below 18 mm H_2O,

whereas an increase of h in excess of 1.6 mm does not necessitate any considerable lowering of the pressure in order for the frequency to remain constant. It is conceivable that relationships of this kind may explain why Bouhuys (1964 p. 972) found such a strikingly small difference in pressure when the notes F_4 and F_6 were intoned on a recorder: his subject may have regulated the air supply to the instrument by using his lips as a reducing valve instead of varying the blowing pressure by means of his respiratory muscles.
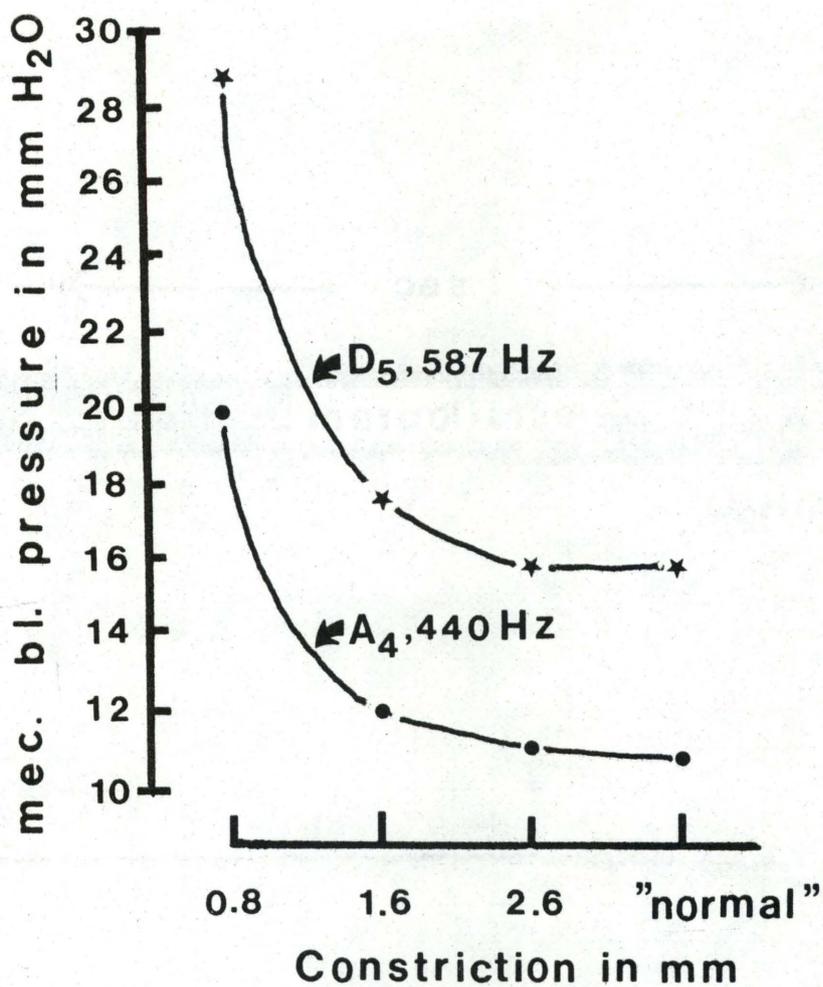
Figure 7

Mechanical blowing pressure required to give a constant funda-
mental frequency with different degrees of constriction be-
tween the upper and the lower lip of the artificial mouth.
The parameter fundamental frequency is shown for fingering of
D_5 and A_4, respectively.
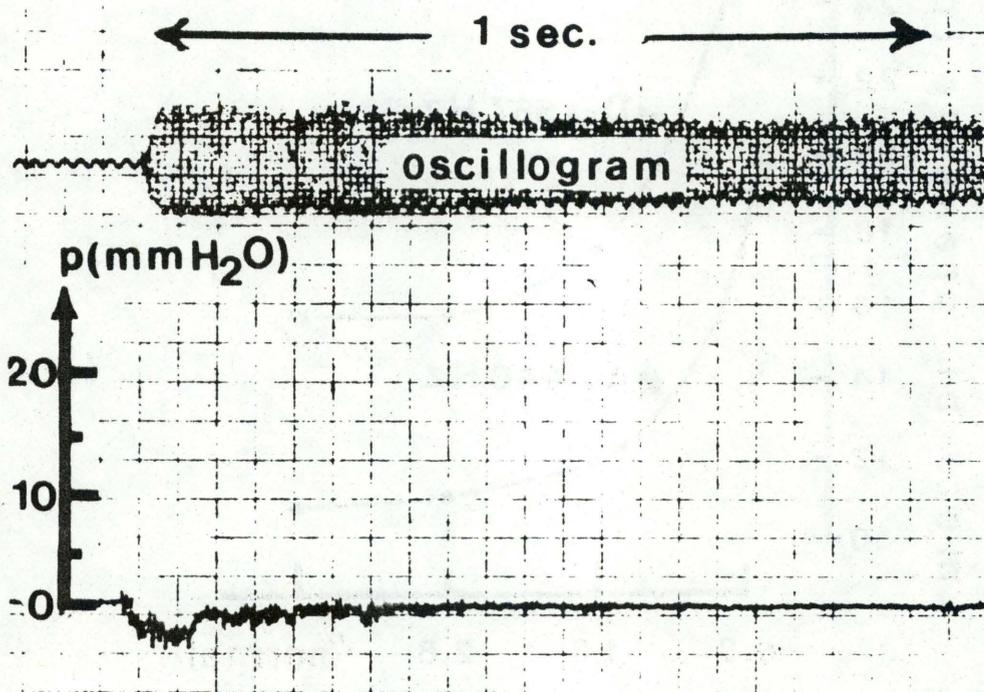
Figure 8

Tracings of pressure curve and oscillogram showing that erroneous measurements of blowing pressure may occur at probe I. Subject A intoned F_5 using "vowel position $\underline{i}$".

extreme cases, the thin, now highly accelerated airstream origi-
nating from the lips may create quite uncontrolable aerodynamic
conditions at probe I which completely invalidate the measure-
ment of blowing pressure at this point.[1]

That such disturbances actually occur is clearly illu-
strated by the pressure curve in fig. 8. The note played was
F_5, produced by subject A. The oscillogram at the top of the
figure shows that the tone was played with "customary intensity".
Nevertheless, the pressure curve indicates "sucking" instead of
blowing during the first part of the tone, since the curve falls
below the base line representing atmospheric pressure. Obviously,
this is a case where the measured value is erroneous, since air-
flow in the windway in the direction from window to the player's
mouth would necessarily prevent the blowing mechanism of the
recorder from functioning.

---

1) Space does not permit a detailed discussion of the many factors
which influence the measurement of pressure in air which is
in motion in a specific direction. For general orientation, it
may be mentioned that - in the relatively low range of blowing
pressures employed in the lower registers of the recorder - the
(mean) jet velocity, v, issuing from a lenticular orifice some-
what like that of the windway, can be derived from the blowing
pressure, p, by the formula

$$v = K \times \sqrt{p}$$

where the constant, K, under the conditions in question can be
set to ca. 3.78, velocity, v, being expressed in m/sec, and blowing
pressure in mm water. This value of K is calculated on the basis
of data taken from Coltman 1968 p. 988, and 1966 p. 104-105.
From this formula it is easily seen that at a blowing pressure
of only 4 mm H_2O (being about the lowest blowing pressure en-
countered in recorder playing), the velocity exceeds 7.5 m/sec,
and at a pressure of 9 mm H_2O it exceeds 11 m/sec. Because of
turbulence (cf. Reynold's number), the velocity will increase at
a lower rate than indicated by this formula if the pressure is
increased still more (Bouhuys 1964 p. 974).
For detailed information on the effect of a fluid's viscosity
on measurement, the boundary layer, the effect of a velocity
gradient on measurement, turbulence, and other aerodynamic con-
ditions, the reader is referred to the literature dealing with
the measurement of airflow, for instance Ower (1966) and Eck
(1944).

The discovery of this measurement error necessitated a critical re-examination of the data. The object here was to ascertain not only how often or seldom errors might occur, but also the degree to which it was possible to assign such errors to specific playing conditions.

For this purpose a minor part of the recorded material was compared on a statistical basis with the pressure-frequency relationships established with mechanical blowing. The material in question consisted of sections from a program of eighteen minutes' duration performed by six different players. The program, which was largely the same for all subjects (except that it had to be modified for subjects E and F, see below), comprised various specific exercises as well as selected passages from compositions for the recorder.

The material was used to furnish data on three different notes, viz. A_4, F_5, and C_6. For each of these notes four tokens were selected from different parts of the material, representing the subject's responses to different instructions with regard to deliberate use of oral articulation, viz. 1) blowing with "vowel position i", 2) blowing with "vowel position a", 3) blowing with "vowel position u",[1] and 4) "free" performance, i.e. a tone cut out of a melodic sequence played according to the subject's own habits.

The subjects E and F were too inexperienced to perform the program in its entirety and had to be given less demanding tasks.

---

1) Concerning the playing with vowel positions the following instructions were given: "Intone A_4 as a long, sustained, and completely steady tone (no vibrato), using the three "vowel positions" in turn: the first time the vowel position a , the second time i , and the third time u ". The same instructions were also given prior to the intonation of the notes F_5 and C_6.

The material for the investigation dealt with here is, therefore, somewhat deficient with regard to them, there being 5 tokens missing for subject E, and 1 for subject F. - Subject G (representing the beginner with absolutely no experience in recorder playing) intoned only the note D_5 and is not included in the statistical treatment of the material under consideration.

The tokens were analysed with regard to the relationship between blowing pressure and frequency. Since the values do not remain constant during the production of a tone, a series of different pressures with corresponding frequency values can be measured within each token and, moreover, a given pressure value may recur so that there are several - possibly different - frequency measures corresponding to that specific pressure within one single token.

The data used in the statistical treatment consisted of pairs of pressure and frequency values, each pair being defined by its pressure component. A finite number of pairs was obtained by choosing the pressure levels coinciding with divisions on the pressure scale employed (i.e. whole multiples of 2 mm H_2O). For each pressure level, all occurrences of it within the token in question were identified and the corresponding frequency values read off. The pressure-frequency pair was then obtained by taking the pressure value in question together with the mean value of the frequency measures.[1] (Standard deviations for the frequency measures were found to be less than 0.5.) This gave a total of 267 pairs of numerical values, which in turn were confronted with the numerical values for mechanical blowing,

---

1) Typical curves presented schematically:
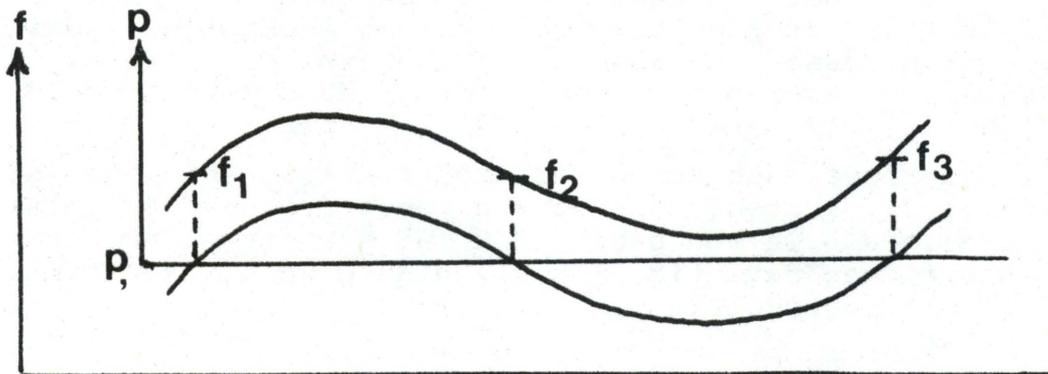
each (mean) frequency measure in the recorded material (natural blowing) being compared with the frequency measure that would occur with mechanical blowing under condition of the same pressure and a temperature of 27° C. The difference between these two measures was calculated and expressed in percentage of the latter. The resulting (non-dimensional) quantity is here termed "deviation".[1]

For six tokens the calculation of deviation turned out to be impossible because the pressures measured were too low to produce the notes in question with mechanical blowing (cf. later in this section). For each of the remaining tokens all deviations were pooled, and a "mean deviation" was calculated for each token. These mean deviations - 60 in total - are shown in the histogram, fig. 9.

It appears from fig. 9 that there is good agreement between natural and mechanical blowing. The arithmetic mean of the deviations is 0.11% (standard deviation 0.7),[2] which shows that the temperature of the recorder has been in the vicinity of 27° C. This is consistent with the fact that the recorder was almost

---

1) For a given subject (S) and a given blowing pressure (p_i) the deviation is given by the expression

$$\frac{f(p_i(S)) - f(p_i(M))}{f(p_i(M))} \times 100 \text{ (per cent)}$$

where $f(p_i(S))$ is the mean value of the frequencies occurring with blowing pressure p_i when the note in question is blown by subject S, and $f(p_i(M))$ is the frequency occurring with mechanical blowing of the same recorder, with the same blowing pressure p_i, and with temperature correction to 27° C (see Bak 1970 concerning temperature correction).

The definition of $f(p_i(M))$ requires an addition with respect to the note C_6: this note is understood to be intoned with a rather small aperture at the thumb hole during mechanical blowing. This addition is necessary because C_6 might have been intoned with covering of a lesser part of this finger hole, which would result in higher $f(p_i(M))$-values.

2) As mentioned above, subject G was excluded from this statistical treatment. Instead, one of the tones produced by this subject was examined. On the basis of 9 pairs of frequency and pressure values a mean deviation was calculated which was found to be -0.0202% (standard deviation: 0.1929).
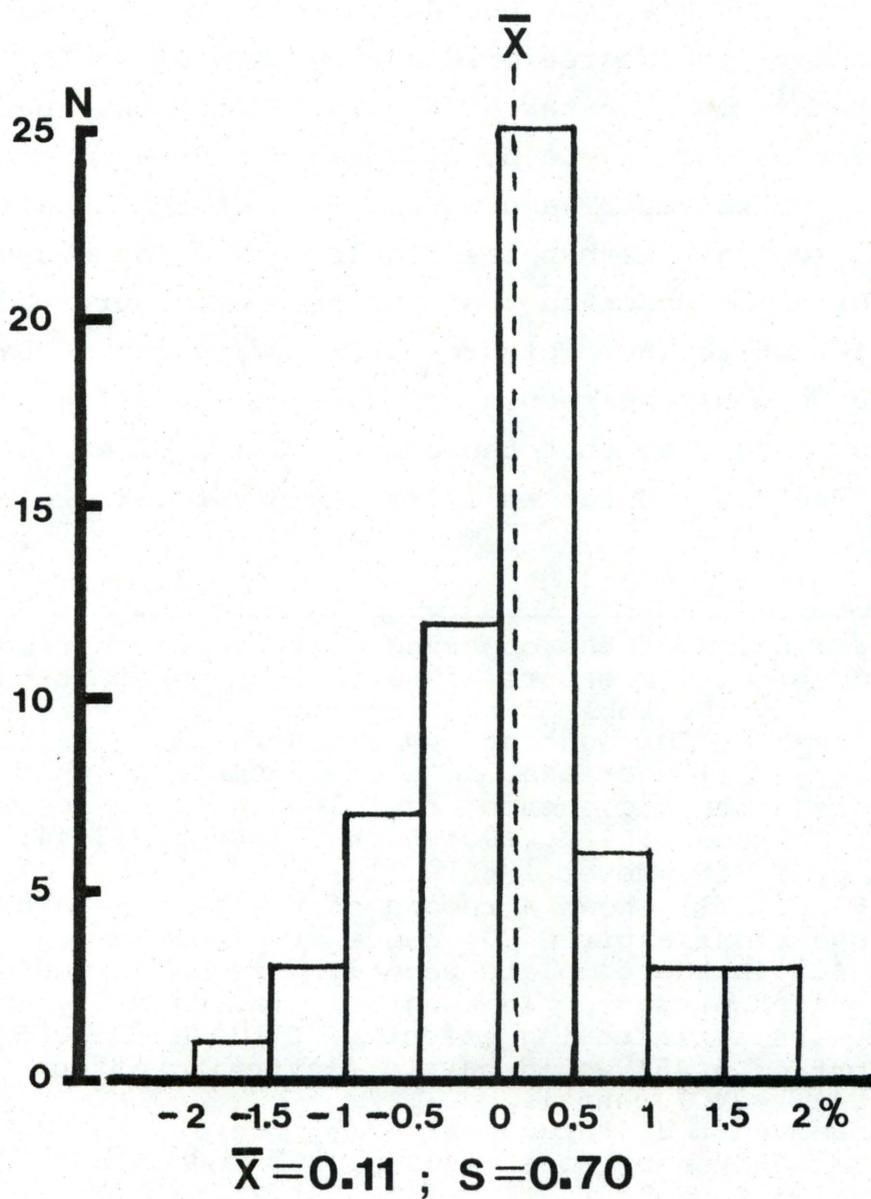
Figure 9

Histogram showing how tones blown by human subjects deviate
from the frequencies calculated on the basis of mechanical
blowing. (Deviations are expressed in per cent.)

constantly in contact with the subject's fingers during the ex-
periments. (Note that heating of the recorder via the fingers
plays a considerably greater role than heating via the airstream
entering the mouthpiece during blowing.)

It is understandable that the deviations have values other
than zero. Firstly, the degree of accuracy as well as the fluctu-
ation of temperature must be taken into consideration. But,
secondly, the deviations may be due in part to variations in the
concentration of water vapour and carbon dioxide in the airstream
during natural blowing. With notes requiring a great consumption
of air it can often be observed that the frequency curve declines
smoothly in relation to the pressure curve, which can be explained
like this: the bore of the recorder is gradually filled with air
exhaled from the lungs, so that the concentration of carbon dioxide
is increased. See fig. 10 for an illustration of this relation-
ship.[1]

---

1) This interpretation of the observed differences in frequency
    for the same blowing pressure is based upon the fact that the
exhaled air entering the tube of the recorder will have a higher
density with increased CO_2 content. With increasing density
the sound velocity will decrease, as a consequence of which
the resonance frequency and, hence, the blowing frequency will
be lowered (cf. Coltman 1966 p. 103; Backus 1969 p. 43-44;
Nederveen 1969 p. 17-18; Meyer 1966 p. 13).
    Bouhuys (1964 p. 971) shows a record of the breath-to-breath
variations in end-tidal expired CO_2 concentrations, which range
from 1.94 to 5.48% in the complete recording from the subject in
question (a clarinet player). In recorder playing a CO_2 content
of 1.94% would have a flattening effect of 0.51% (-8.8 cents),
and a CO_2 content of 5.48% would have a flattening effect
amounting to 1.4% (-24.7 cents).
    (It may be added that different values of the CO_2 concentration
in exhaled air are given by Roos (1936 p. 23) and Coltman (1966).
Roos measured a CO_2 content ranging from 5.772% to 6.726% in the
rest of the supplemental air collected in a sample tube after
playing a tone for 19 and 43 seconds, respectively. Coltman, who
probably collected air from the very beginning of the exhalation,
filled a tube with breath and found that the resonance frequency
of this closed tube was 12 cents (some 0.7%) lower than that of
the same tube filled with saturated moist air at the same tempera-
ture. Coltman concludes that this change in frequency corresponds
to a calculated CO_2 content in the exhaled air of 2.5%.)

Fig. 9 above showed how the (mean) deviations for all tokens cluster around a mean value. According to this histogram the curve material is not subject to any considerable uncertainty. Exception must, however, be made for the six tokens referred to earlier, which were left aside because the pressures measured were inadequate with mechanical blowing. The reliability of the remaining bulk of data is suggestive of a basic difference between the six tokens and the rest, and it is worth while looking for a specific explanation for them.

Out of the six tokens four were played by subject A, and the remaining ones by subject B, i.e. the two most professional players in the whole group of subjects. As for the oral articulation of the player, four out of the six tones were played with intended vowel position $\underline{u}$, and two with intended vowel position $\underline{i}$, i.e. the two positions which - unlike vowel position $\underline{a}$ - may serve to impede the airstream at the lips. It is, moreover, noteworthy that the lower notes are predominant among these six special cases, cf. the following survey:

	tone F_5,	intended	vowel	position	$\underline{i}$
subject A	"	"	"	"	$\underline{u}$
	tone A_4,	intended	vowel	position	$\underline{i}$
	"	"	"	"	$\underline{u}$
subject B	tone C_6,	intended	vowel	position	$\underline{u}$
	tone A_4,	intended	vowel	position	$\underline{u}$

On the basis of these instances it seems reasonable to venture a hypothesis according to which the aberrant measures have a physiological cause: professional players - like subjects A and B - use a constriction of their lips to reduce the airstream from the mouth to the windway, if such a use of the lips as a controlled valve is necessary, e.g. in the production of tones of low frequency.
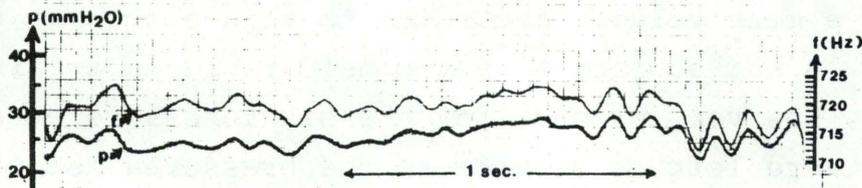
Figure 10

Variations in blowing pressure (lower curve) and funda-
mental frequency (upper curve) during subject D's intoning
of note F_5. The deviations of instantaneous frequencies
from calculated frequency are found to vary from ca.
-0.1% at the beginning of the tone to ca. -1.2% towards
the end of the tone. The pressure curve shows that subject
D used a blowing pressure of ca. 25 mm H_2O on an average.
This is a high value considering that the test instrument
requires only 12.5 mm H_2O for the intonation of F_5 in
accordance with the equal-tempered scale based on A_4 =
440 Hz (the frequency of F_5 then being 698 Hz), cf. fig.
20 at the end of this paper. - In this particular case,
the decrease in "deviation" during the sustained tone,
which has a duration of only some 2.5 sec, is unusually
large, probably because the consumption of air is extremely
high as a consequence of the relatively high blowing pres-
sure employed.
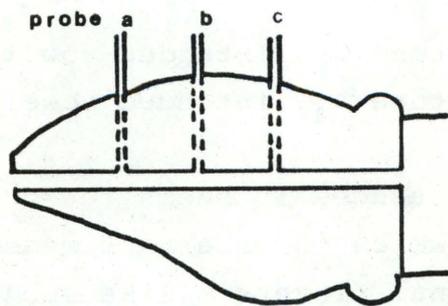
Figure 11

Schematic drawing showing the placement of probes a, b, and c
for measuring the static pressure in the windway during tone
production. Cf. fig. 12.

In fact, the X-ray recordings do show an activity of this very kind, although it is only occasionally that the constrictions seem strong enough to produce the effect on measurements which is found in the six aberrant tokens.

In order to estimate more concretely the risk of erroneous measurement and to develop an effective control on the reliability of the pressure measurements for use in supplementary studies, experiments were made with another recorder, which was provided with three extra probes (apart from that corresponding to probe I in fig. 3a) for measurement of the static pressure at three places in the windway.[1] The placement of these extra probes is shown in fig. 11.

Using this arrangement, several tests were made to determine, both with mechanical and natural blowing, the static pressure in the windway at probes a, b, and c as a function of the blowing pressure.

Some of the values registered with the additional probes and using mechanical blowing under conditions of "unimpeded flow" (cf. fig. 6c) are depicted in fig. 12. A linear relationship is observed between oral pressure and the static pressure at the various measurement points in the windway, which means that, under these conditions, the oral pressure can be calculated from a pressure measurement taken in the windway. That, for a given blowing pressure, successively falling values occur for the pressures read at probe a, probe b, and probe, is hardly surprising, since a similar decline in pressure always occurs with a conducting tube of constant dimensions. However, these drops in pressure have a special importance in the subjects under discussion.

From figs. 6a and 6b it is easily observed that the protrusion of the lips can extend the windway and, in those cases

---

1) Static pressure may be defined as the pressure acting in all directions at a point in a fluid (Ower 1966 p. 6).
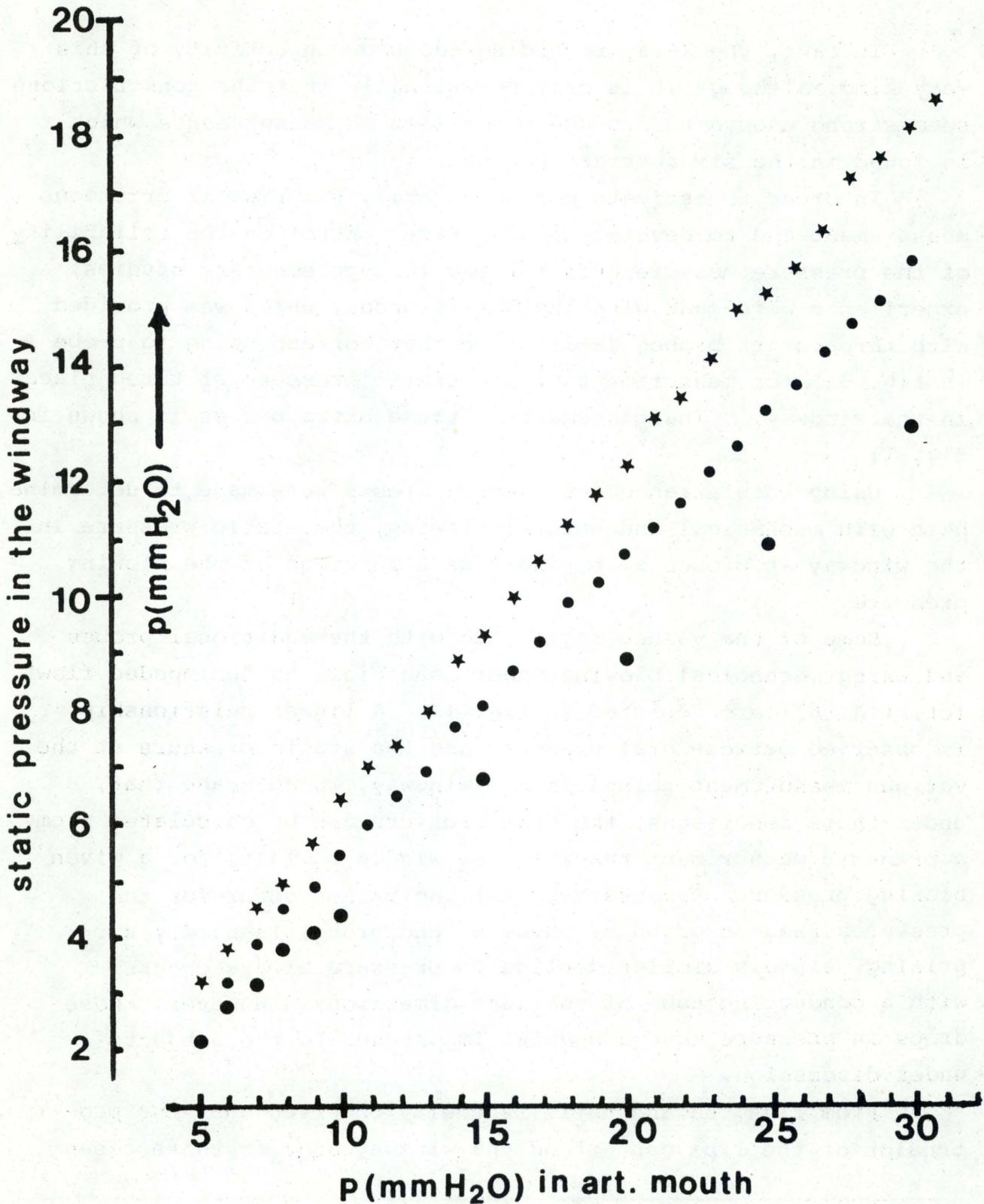
Figure 12

Static pressure measured in the windway of a test recorder during
mechanical blowing. The static pressure is given as a function
of the blowing pressure in the artificial mouth. Stars = measure-
ments made with probe a; small filled circles = measurements
made with probe b; large filled circles = measurements made with
probe c, cf. fig. 11.

where the transition between the lip opening and the windway is so smooth that the airflow does not strike the aperture of bore I in the tip of the mouthpiece, probe I will assume the same status as probes a, b, and c, which register successively falling values for the static pressure as the air flows downstream. However, the strictest requirements for the procedure to be followed when measuring "wall static pressure" (Ower 1966 p. 139-140) will hardly be fulfilled at probe I, so the pressure registered at this probe may be expected to exhibit random variations slightly below the true oral pressure.

A different situation is produced if the opening of the lips is made especially narrow compared to the size of the windway. In this case, the conditions for measuring static pressure at probe I cease to be sufficiently stable (cf. Ower 1966 p. 78 concerning "fully-developed pipe flow", and p. 139 ff), and a drastic reduction of the pressure at probe I may easily result.

After the results of these tests using mechanical blowing were known, a series of supplementary investigations were carried out at the Institute of Phonetics with the cooperation of subject A. The recorder used was the same one that was employed earlier in the main series of recordings with subjects A - G, but it had now been provided with an additional probe (probe II, fig. 3b) for registration of the static pressure halfway down the windway, i.e. at a point corresponding to that at which probe b (fig. 11) was inserted in the other test instrument.

The subject was first requested to improvise passages of as different character as possible on the test instrument, so that the possibility of measurement errors at probe I during free play could be investigated. This was done by comparing the pressure curve from probe I with the frequency curve and the pressure curve from probe II.

Although the test was repeated many times, no irregularities were observed. Only when the subject was asked to form a very marked narrowing of the lips while playing a sustained tone

did an aberration from the normal relationship between the curves
occur, viz. the effect shown in fig. 13. It will be seen that
the frequency curve and the two pressure curves first follow each
other monotonously (in accordance with Bak 1972), but that farther
along, the pressure curve recorded at probe I begins to fall off,
while the pressure curve recorded at probe II and the frequency
curve continue unchanged. This effect of the narrowing of the
lips has been shown to be reproducible.

Earlier in this paper it was hypothesized that the lips
function as a reducing valve in the case of subjects A and B.
This hypothesis can be substantiated by consideration of the
cineradiographic material recorded as part of the present investi-
gation. If one observes the moving X-ray pictures it is possible
to see directly that the jaw, tongue and lips move to a greater
or lesser extent during recorder playing. This is true of all
subjects. Among the more significant movements it deserves
mentioning that the tongue creates a larger oral cavity for high
tones compared to low tones. Similarly, the jaw is lowered for
the production of high tones and thus contributes to the same
effect on the size of the oral cavity. However, it is more
important that movements of the jaw influence the aperture at
the lips and hence the functioning of the lips as a reducing
valve.[1]

Fig. 14 serves to provide a documentation of the most
essential jaw movements during recorder playing. It depicts the
curves obtained by simultaneous X-ray, pressure, and sound re-
cordings of a melodic sequence played by subject A. The notes
to be performed are shown at top, an oscillogram of the audio
signal at bottom, and the pressure curve is placed just above
the oscillogram. Between the pressure curve and the musical

---

1) Note that the tongue and jaw movements, which are charac-
   teristic of all the subjects, are strikingly at variance
with the contention of the vowel theory.
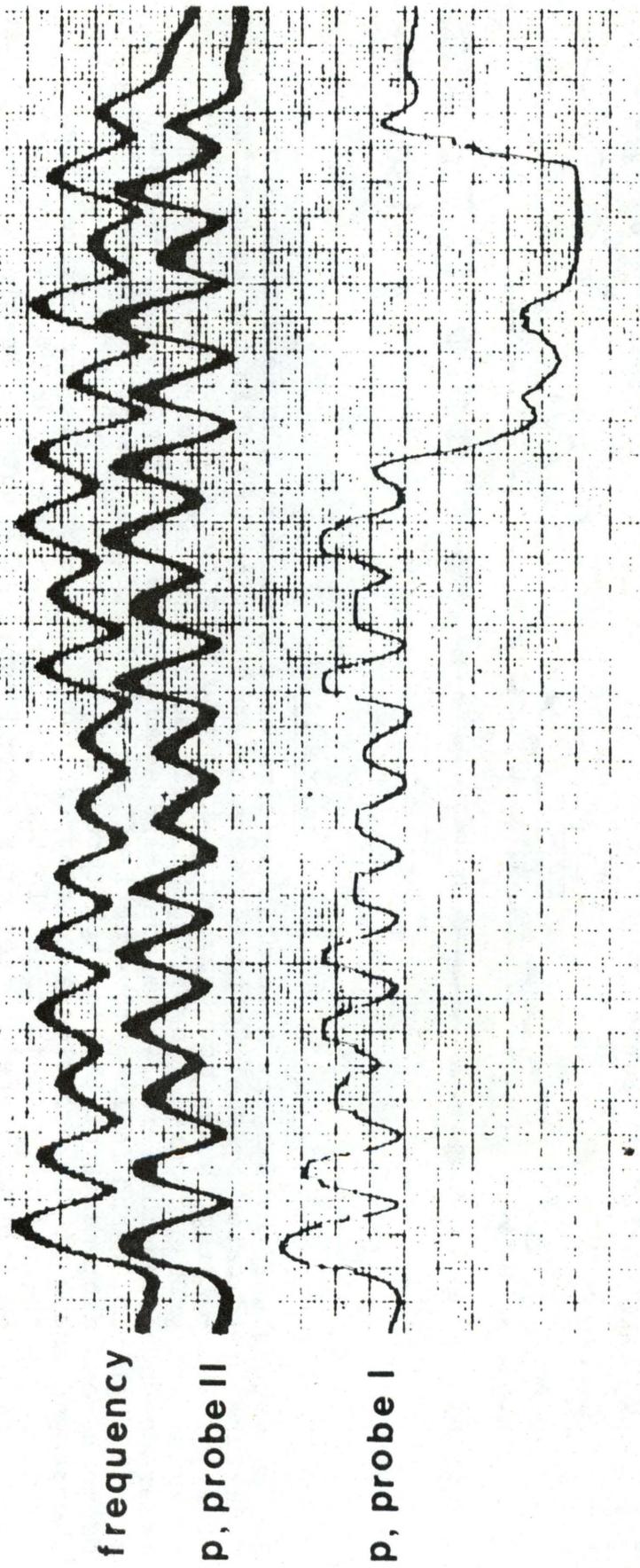
Figure 13

Simultaneous recordings of fundamental frequency and pressure via probes I and II, cf. fig. 3b. The bottom curve shows that erroneous values may occur (because of a narrow constriction at the lips) when measuring the blowing pressure with probe I.
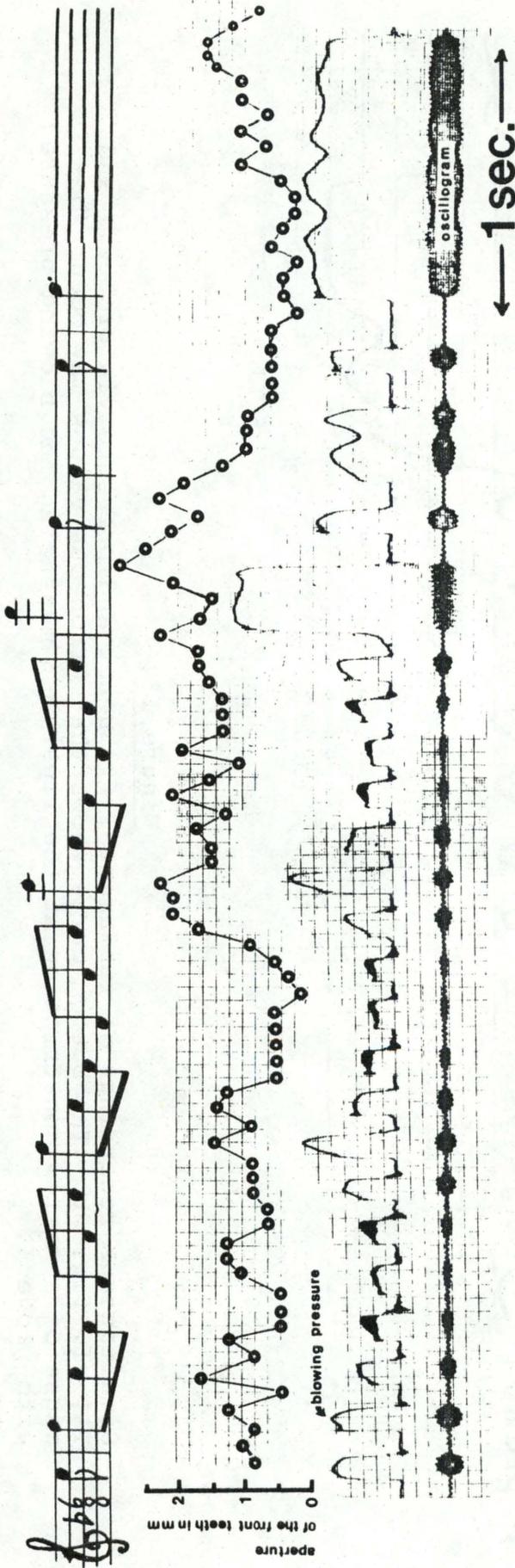
Figure 14

Jaw movements used by subject A in blowing a melodic sequence. The measurements of "aperture" (indicated by small circles) were obtained as explained in the text; also cf. fig. 15. - Although the accuracy with which the temporal relationship between jaw movement and pressure curve + oscillogram could be determined is unfortunately limited (there being a possible error of almost ±1/8 second), the measurements clearly demonstrate that rising and falling movements of the melodic curve are accompanied by increments and decrements of the jaw aperture.
The ragged shape of the oscillogram during the highest note, F_6, is indicative of a failure in the attack. It is possible that the peculiar character of the jaw movement during the following one quarter of a second (until the note C_5 is blown) is explicable as a reaction to the said failure in the attack of F_6.
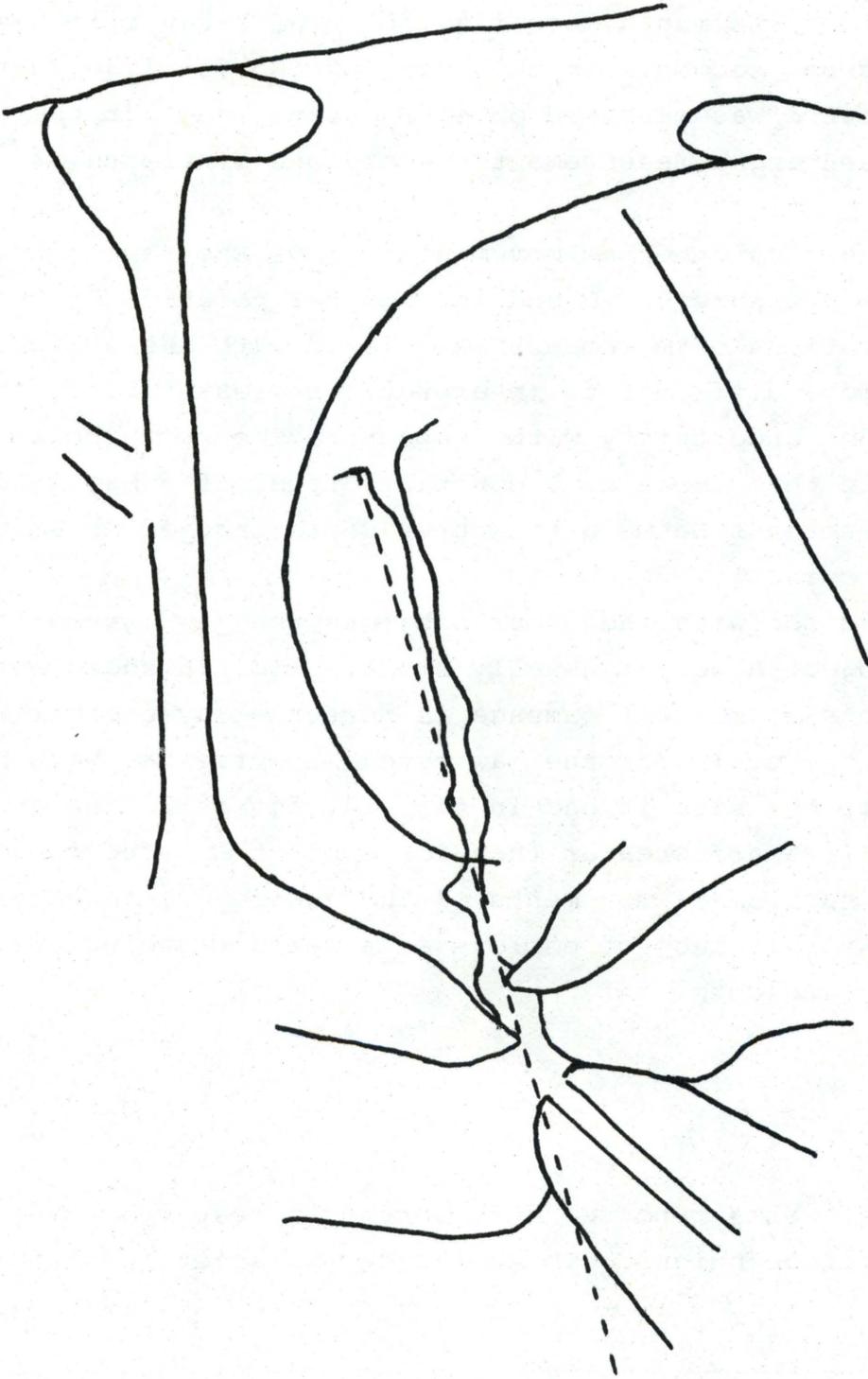
<u>Figure 15</u>

Illustration of X-ray recording with indication of re-
ference line used in measuring jaw aperture.

notation the movements of the jaw are depicted as a broken curve.
Each point on this curve marks the distance from the edge of the
lower front teeth up to a line connecting the edge of the upper
front teeth with a characteristic point on the hindmost molar in
the upper part of the mouth, see fig. 15. The X-ray film has
sixteen frames per second. In this part of the film the just
mentioned distance was measured on every frame, so that there is
a dot - representing a measurement - every one sixteenth of a
second.

It is seen that the jaw movement follows the same general
pattern as the pressure curve and the musical notation.[1] There
are also indications of movements associated with the attacks
but they are more difficult to interpret, one reason being that
there is a minor uncertainty with respect to the synchronization
(it is possible that there is a constant offset of a magnitude
of a few centiseconds between the curve depicting jaw movement
and the other curves).

In comparison with the other subjects the jaw movements
found with subject A were unusually small. Fig. 16 shows how a
section of the same musical passage is performed by another sub-
ject, D; here the scale for the jaw movement curve has been com-
pressed to half the size it had in fig. 14, but still the excur-
sions of the curve are greater than for subject A. The fact that
subject A has such small movements of the jaw may be taken as an
indication that this subject employs a narrower reducing valve
than the other subjects.

4. Discussion

The experiments reported here served to test two predominant
views of blowing technique. These can be characterized briefly
as follows:

---

1) Note the peculiar jaw movement at the note F_6 (in fig. 14
   above), which can be explained as a reflex reaction to a
failure in the attack.
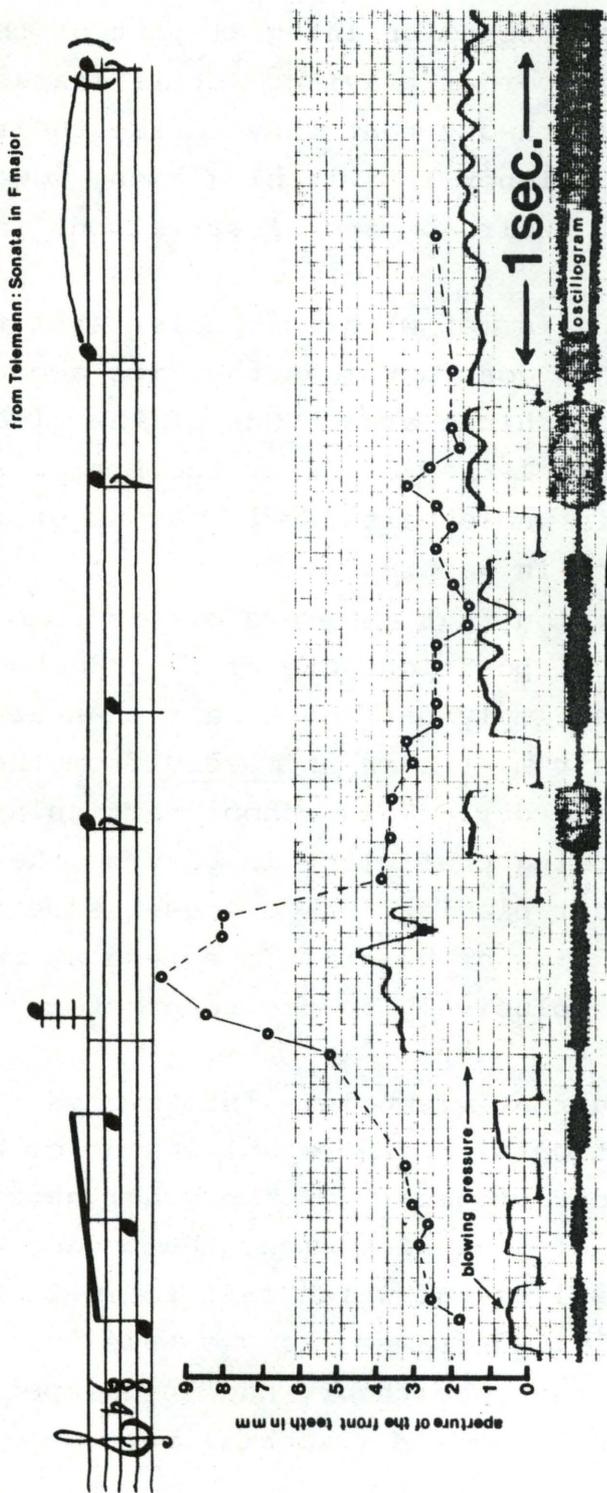
Figure 16

Open circles depict a series of measurements of jaw aperture during subject D's per-
formance of the melodic sequence shown in fig. 14. Only a section of this sequence
is included in fig. 16. (Broken lines indicate that some frames were missing because
the film had been spliced together.)
It is seen that subject D on the whole used a considerably greater jaw aperture than
subject A. However, there is also in this case a correspondence between melodic
movement and jaw aperture.

1) The "vowel theory" or "resonance theory" contends that a good command of alternating vowel positions is a necessary component of the technique of recorder playing. This emphasis on vowel positions has been motivated in terms of an assumption about the mouth-throat resonance being an integral part of the acoustical characteristics of recorder playing. It is characteristic of many adherents of this theory that they advise their students to develop their proficiency in switching among vowel positions rather than training their command of variations in blowing pressure.

The just mentioned theory is met with in specialist literature written in German, but this does not mean that its acceptance is limited to a definite, geographical area. One of the clearest presentations of the theory that I know of is by the French acoustician E. Leipp (1962), who is wholeheartedly in favour of it with regard to wind instruments in general.

2) According to the opinion that is prevalent in school recorder tutors in English vowel positions are of no importance, since the mouth and throat are simply part of an air channel through which the expiratory air must pass unimpeded from the lungs to the windway of the recorder. This school particularly emphasizes the importance of being proficient in varying the blowing pressure during recorder playing. As a consequence of the requirement of "unimpeded flow of breath", the regulation of blowing pressure is assigned exclusively to the respiratory muscles.

The above division of current viewpoints into two main theories was useful in connection with the establishment of a research program for the present project. On the other hand, there are several differences of viewpoint among the pedagogues which are interesting in themselves but which fail to emerge in a discussion within this simplified framework. To take one example, the division into a resonance theory and an unimpeded-flow-of-breath theory leaves no room for a special version of

the latter which is presented quite briefly by the English musicologist A. Baines (1963 p. 72). Baines adheres, without any reservation, to the traditional conception (shared by the different schools) according to which the recorder player is deprived of control of the airstream by means of the lips and hence must resort to other means. For the purpose of "keeping the notes steady in pitch through the rise and fall of loudness demanded by musical expression" fingering is used, but in addition Baines suggests that "it may be that the throat comes into play, being more relaxed in the _forte_ to allow a full stream of air to pass into the instrument; and more tightened in the _piano_, to send forward a thinner stream of air at the same speed, so that the note keeps its pitch but has less volume". Baines' statement seems to imply that the respiratory muscles serve to establish a relatively _constant_ subglottal pressure, the admission of air to the instrument being regulated by the larynx muscles.

The preceding summary of the issue will now be followed by an attempt to explain the above mentioned, strongly divergent views of the use of oral control mechanisms in recorder playing.

The X-ray material and the Mingograph curves of the present project give ample documentation of the activity of the articulatory organs and the variations in blowing pressure. After having occupied himself with this material for an extensive period of time, the author has acquired a definite impression to the effect that recorder playing of good quality requires the same precise interplay among the muscles of the articulatory organs as does speech or singing. This activity is performed with a precision which suggests that it is controlled - like in speech and singing - via highly developed nerve patterns with intimate connections to the conscious and (particularly) the subconscious brain mechanisms, and probably includes the associated, essential feedback mechanisms as well.

On this background it is interesting that the present experiments indicate that this highly developed command of the

articulatory muscles is utilized here for a basically different purpose than in speech and singing: in recorder playing it seems to serve solely to control the stream of air admitted to the instrument. In practice this is implemented by establishing reduction valves as described earlier. As mentioned above, Baines points to the possibility that the larynx muscles are used to control the airstream. Since the function of the larynx was outside the scope of the present investigation, there is no possibility of testing Baines' interesting hypothesis in this context.[1] However, the present investigation gives evidence for another controlled reducing valve being used by skilled professional players, viz. their lips. Moreover, there is reason to assume that the tongue, which functions as a closing/opening valve, also contributes to the ever important airstream control during sustained notes.[2]

It is pertinent now to ask whether any alternative theory can explain the marked theoretical discrepancies among the schools. This question can be answered in the affirmative.

For the sake of clarity we shall start by hypothesizing two different, extreme cases:

(I) A recorder player who really lives up to the stringent requirement of "unimpeded flow of breath" as prescribed by the "English school" must obviously do anything he can to train his respiratory muscles to perform the task of varying the blowing

---

1) It must be stated, however, that the present investigation does not lend support to Baines' assumption as regards the effect of airstream control (i.e. to vary the volume of tone whilst keeping the pitch constant, according to Baines). Under condition of unimpeded flow above the glottis during the blowing of a note a widening of the glottal slit would inevitably result in increased oral pressure. Such a change in pressure would, in turn, cause the fundamental frequency to rise, as established by the research reported in this paper.

2) The question whether a constriction of the airstream (in terms of "tightened throat" or half-closed lips) produces a change in tone or volume is touched upon by Rowland-Jones (1973 p. 107). Rowland-Jones contends that "there is no analogy with the flute here".

pressure (this is, in fact, emphasized by that school). There is no alternative possibility for him to adapt the blowing pressure to the ever changing conditions in his performance of music.

(II) A recorder player who, instead, regulates the admission of air by means of a lip function while keeping the pulmonic air pressure relatively constant, is blowing under markedly different physiological conditions, since (i) the respiratory muscles now work in a very different, slower rhythm, and (ii) the intraoral air pressure will, on an average, be higher because of the resistance introduced at the lips.

These relationships are mentioned here because they may underlie the cleavage between the different schools. This becomes evident if one considers the situation of recorder players trained within the school represented by pedagogical literature in German ("vowel theory"). These persons do not a priori represent any of the two extremes described above, but since they have not been warned against impeded flow to the same degree as pupils of the "English school", it is conceivable that a greater number of them actually use their lips as a reducing valve (even if they are not conscious about their use of this mechanism), as did the most proficient subjects of the present investigation. This would reduce the necessity of using the respiratory muscles for rapid changes in air pressure (see item (II,i) above), and it is natural that such players come to give a lower priority to the training of their command of overt variations in air pressure (air supply) than is usually done within the "English school". This attitude is in good agreement with the general statements of the vowel theory.

It is more difficult to suggest a plausible explanation of why the vowel theory prescribes the use of "vowel positions" going with different pitches, although the positions in question are in fact not used in recorder playing (as shown by the investigations of the present author).

The vowel theory is often explained with reference to the acoustics of the organ. The organ pipe has two separate cavities, the pipe foot and the pipe body, which are almost separated by a metal plate. There is a possibility of coupling between the two cavities via a slit - the flue - through which the blowing air leaves the pipe foot along the edge of the just mentioned metal plate, producing a jet which is an integral part of the sound generating mechanism that sets the air column in the pipe body in vibrations (note that the tone production is based on the same principle as that of the recorder). It has been assumed that there is a significant acoustic coupling between the cavities below and above the flue when the pipe is blown, and a tone is produced, and it is natural enough that this viewpoint has been transferred to the mechanism of tone production in the recorder. The analogy obviously gives weight to the contention that the oral cavity should vary in size according to the pitch intended in blowing the recorder. This may help to understand the theoretical attitude of the school in question. However, recent experiments with organ pipes indicate that the acoustic coupling in fact loses its importance as soon as there is a perceptible airstream, i.e., it can be concluded that the pipe foot does not influence the resonator under normal conditions of tone production.[1] This is in agreement with the results of the present investigation of the recorder: the resonances of the oral cavity are of no importance in recorder playing.

It is entirely possible that there is also an impression-istic reason for the vowel theory. It should be noted that there is always a tendency for the oral pressure to be highest during the intonation of the most high-pitched notes (cf. fig. 17), even in cases in which the magnitude of the pressure is influenced by the function of the reducing valve. It must be

---

1) Kühn 1940 p. 23; Sundberg 1966 p. 89.

assumed that the person playing the recorder is more likely to feel his oral pressure if there is a resistance at the lips (cf. item (II,ii) above). Now, an investigation of spoken syllables with different vowels (Brown 1973 p. 141-151) has revealed that there are specific levels of intra-oral air pressure associated with different vowels. If, possibly, the subjective impression that one is using definite "vowel positions" is not primarily a matter of muscular feeling but rather due to the person's unconscious sensing of his own intra-oral air pressure, we will be facing the following paradox: with high notes the recorder player feels an /i/-like "vowel position" (because of the higher pressure), and with low notes he feels an /u/-like "vowel position" (because of the lower pressure), even if it is really the other way round as regards the shape of the oral cavity.[1]

5. Conclusion

The physiological conditions of recorder playing were investigated by means of continuous X-ray recordings along with simultaneous registrations of blowing pressure and recordings via a microphone of the tone emitted by the recorder.

The results of analyses of this material disagree on essential points with the principles of blowing technique advocated by existing schools. As a consequence of this, an alternative theory is put forward, a theory which implies that the blowing technique used with the recorder is probably less fundamentally different from that used with the transverse flute than has hitherto been assumed: in both cases, it seems, an important lip function is employed.

---

1) Brown (1973) found that the mean intra-oral air pressure during the production of the syllable "di" was slightly above 35 mm H_2O. During the production of "du" and "da", the air pressure was lower.

This paper has dealt at length with problems having to do with the technique of measurement employed in measuring blowing pressure. Hopefully, the solutions found to some of these problems can be utilized in connection with future measurements of a similar kind. For example, it was found that there is frequently a greatly varying resistance to the airstream at the lips, and that, for this reason, a system resembling a pitot-tube (such as probe I of fig. 3) is preferable to a system measuring the intra-oral air pressure. Under extreme aerodynamic conditions the pitot-tube system may fail, however. But such failure may be detected if the static pressure in the windway is measured at the same time, since this pressure has been shown to be proportional to the total pressure measured with the pitot-tube under conditions of no disturbance.

The occasional disturbance of the pitot-tube measurements indicates that at least the most proficient recorder players among the subjects for the present investigation make use of varying constrictions at the lips. The X-ray recordings show that these constrictions serve to control and stabilize the transfer of energy to the instrument.

It may be mentioned in this context that the X-ray recordings make it possible to observe many different movements of the articulatory organs during recorder playing. Several of these are very clearly associated with the musical expression, and may undoubtedly give interesting differential information on the individual characteristics of the subjects. However, with the exception of the control of lip constriction dealt with above, the gestures are generally too complex to be interpreted with any certainty on the basis of the present material.

Legend to figs. 17-19:

Figure 17 shows the blowing pressure of some of the subjects during repeated performances of the notes shown at top. (In subject D's second recording the end of the final note has been cut off for considerations of space.)

The second recording of subject A was dealt with above in connection with fig. 14. The attack of F_6 in subject B's second recording is likewise found to be rather unsuccessful. It is possible that the marked pressure rise at the beginning of note F_6 may be interpreted as a correcting gesture.

As for subject D, possible risks of failure in the attack of F_6 are, as it were, circumvented by this subject's use of unusually high blowing pressures.

Because of a hysteresis effect it seems generally necessary to perform the attack of high notes with a relatively high blowing pressure. After the starting transient, however, the pressure can be safely lowered to its normal level (according to fig. 20 below) during the sustained tone, cf. subject B's third recording.

Figure 18 shows Professor Conrad's performance of the same musical sequence as in fig. 17.

Figure 19a, b, c: pressure curves serving to illustrate individual styles of recorder playing. All curves refer to performances of the melodic sequence given in musical notation at the top of fig. 19a. Only the sections indicated are included in the illustrations of blowing pressure.

For subject B and Professor Conrad, two different recordings are shown. It is noteworthy how similar the repetitions are to the first recordings, although there was a considerable break - filled by recordings of other parts of the program - between the two performances.
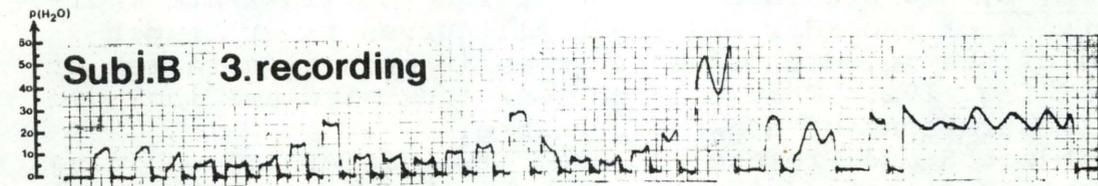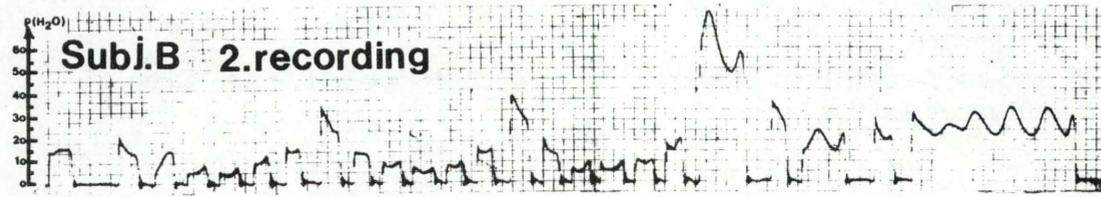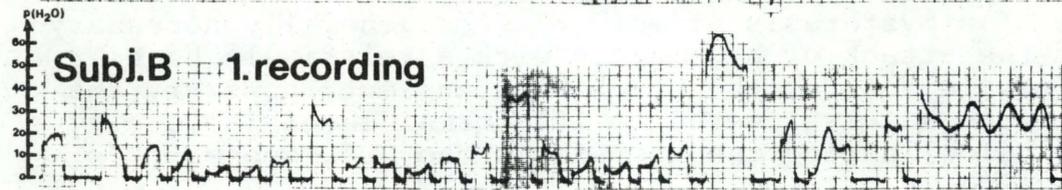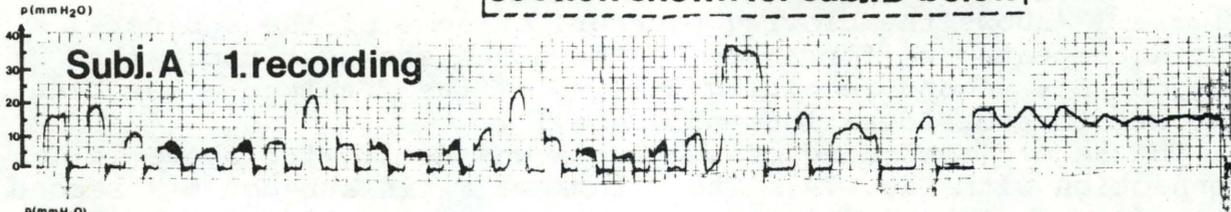
Figure 17

prof. Conrad   1.recording

probe II

probe I

prof. Conrad   2.recording

probe II

probe I
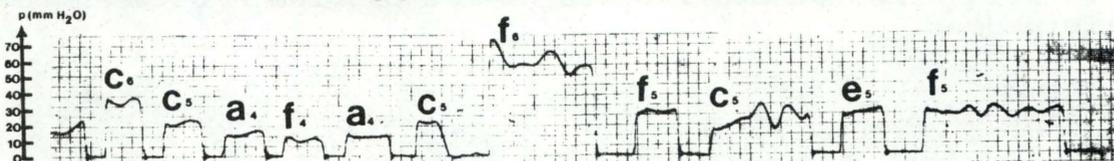
from Telemann: Sonata in F major
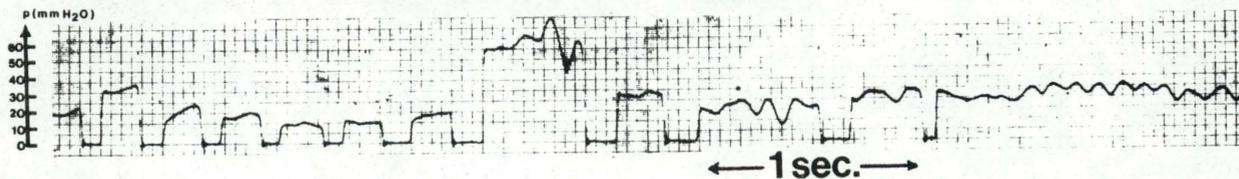
Figure 18

from Telemann: Sonata in C major, 2.movement
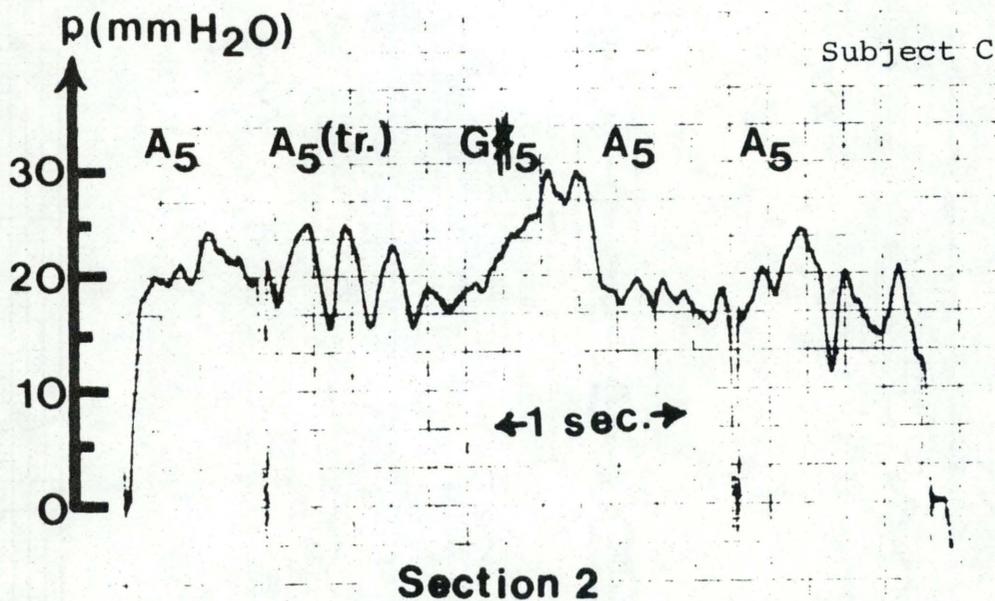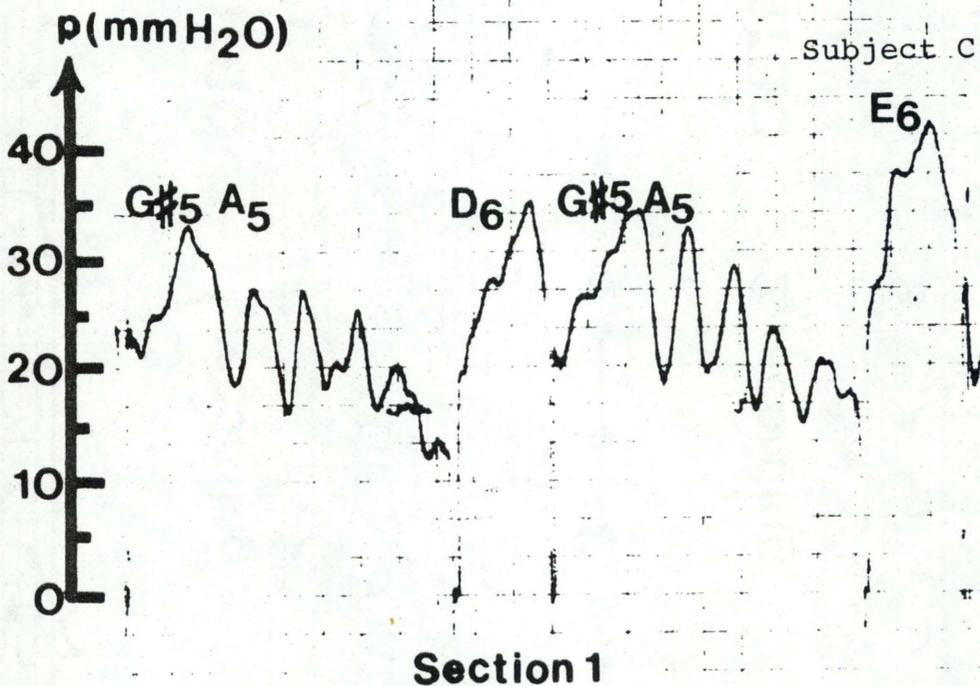
Section 1

Section 2
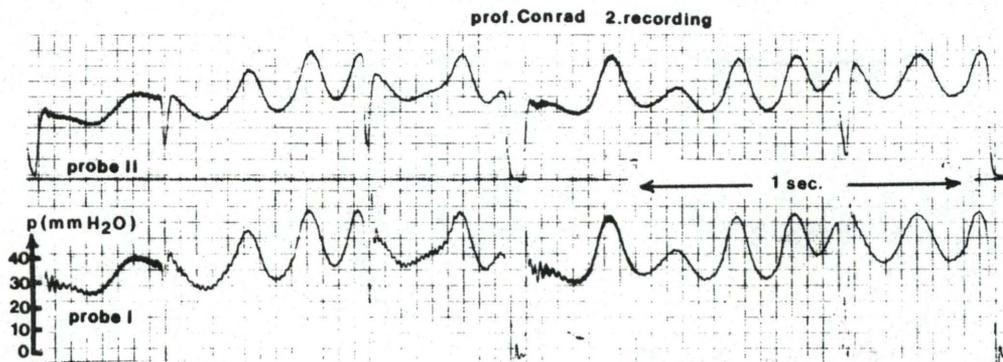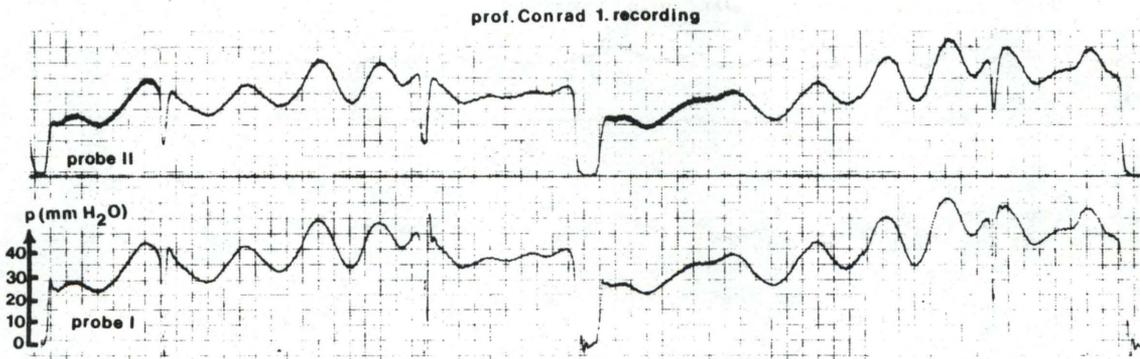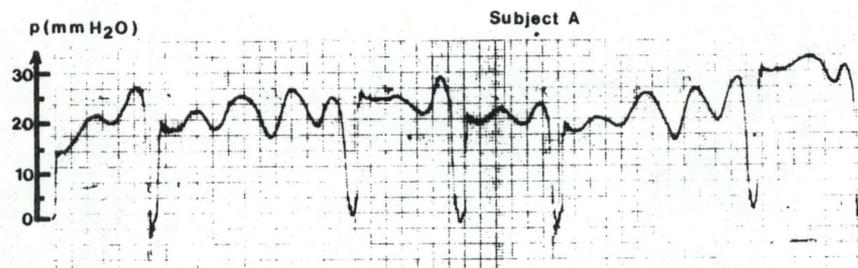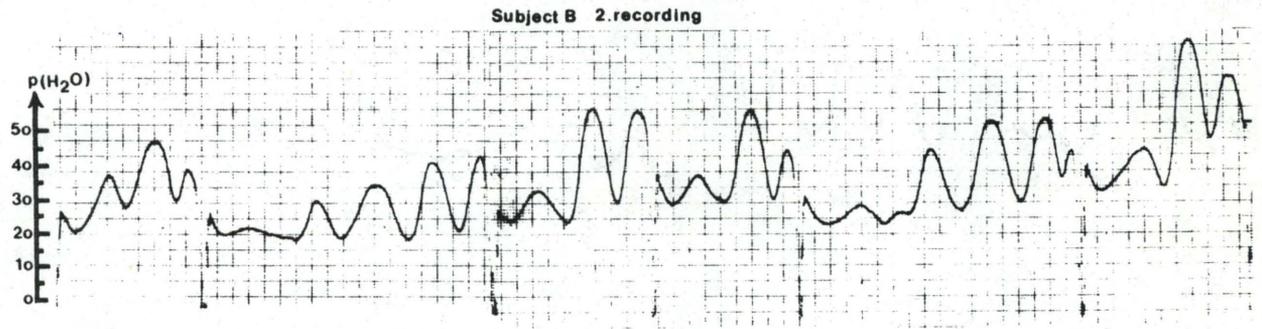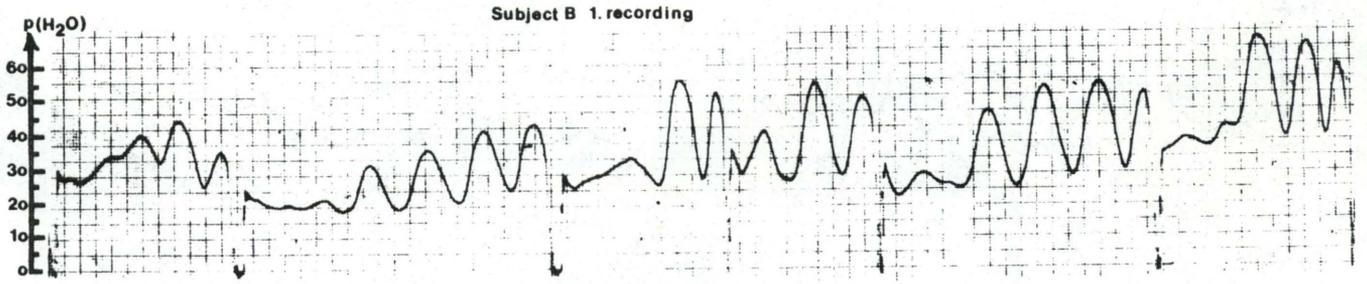
Figure 19a

Section 1

Figure 19b

268
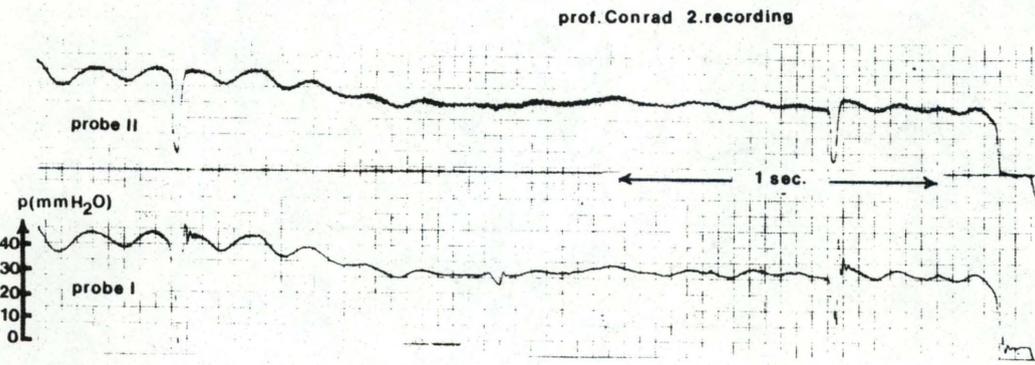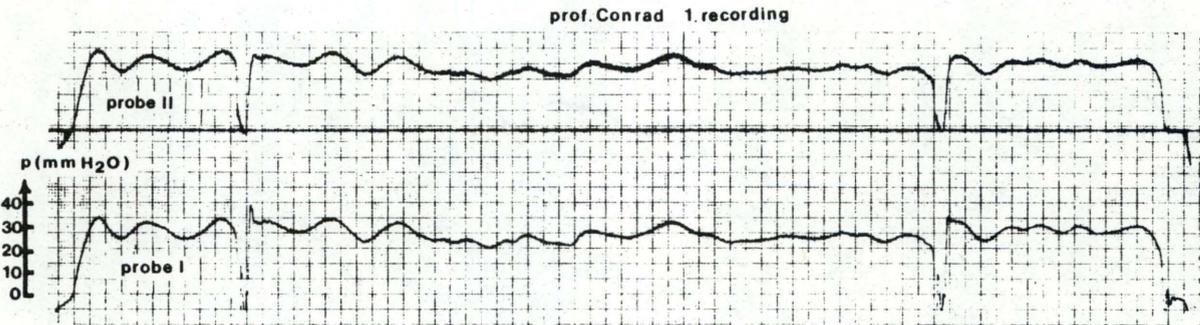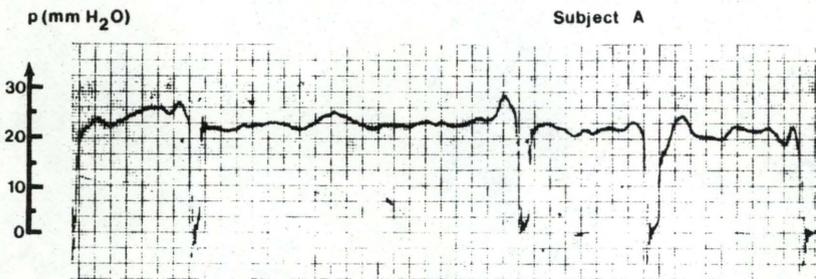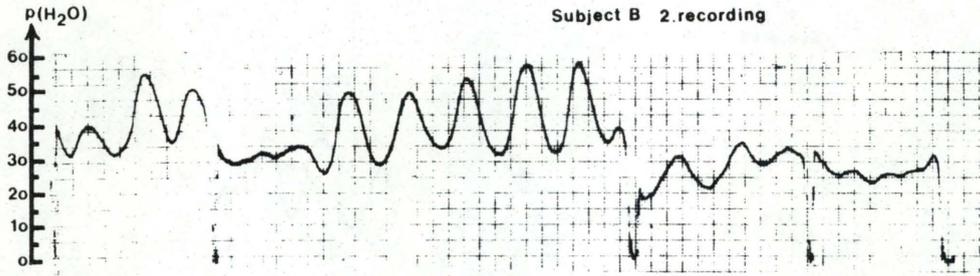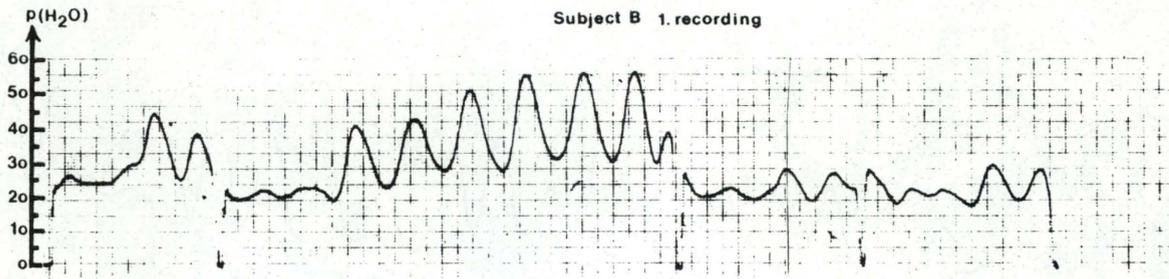
Section 2

Figure 19c
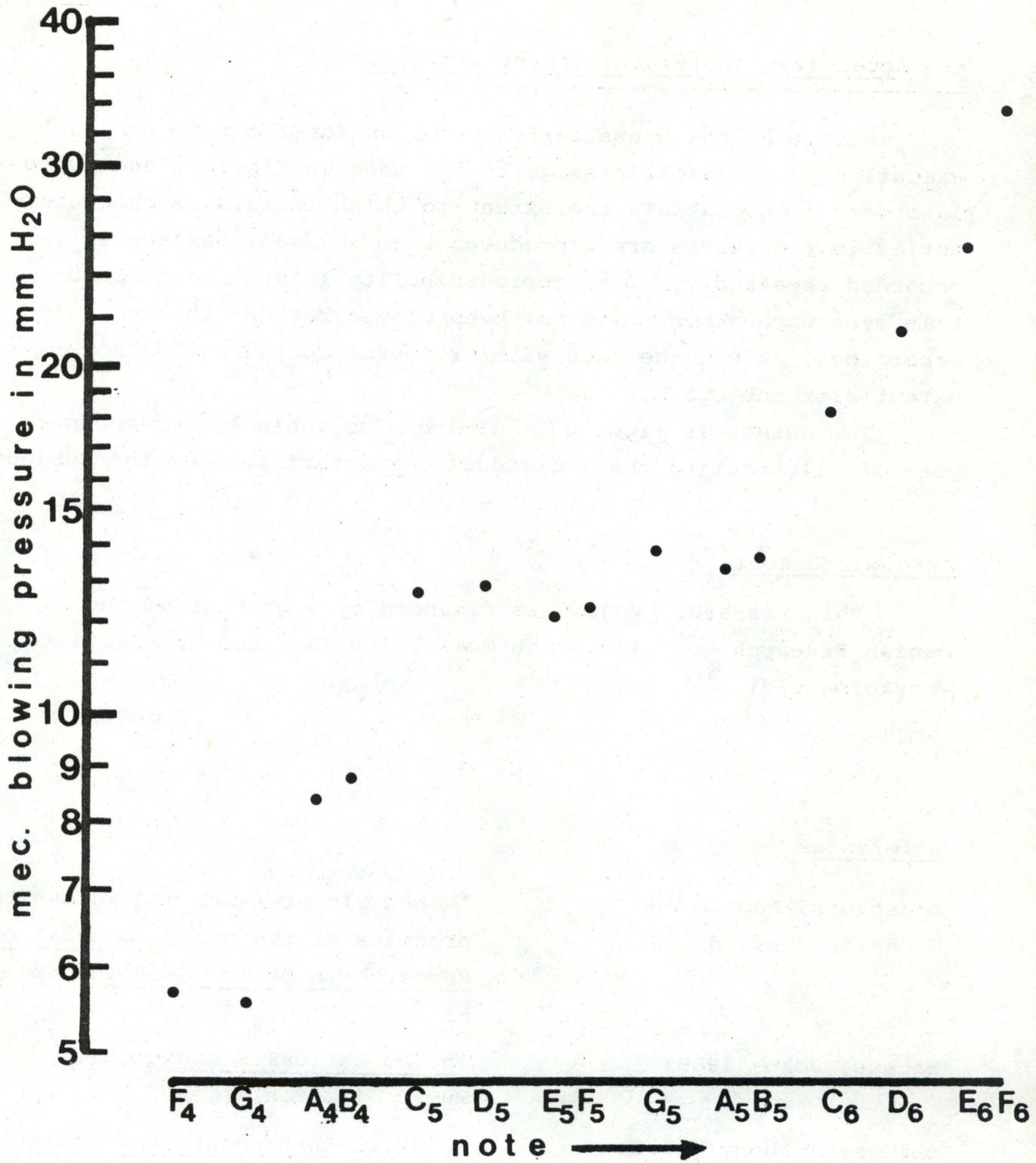
Figure 20

Mechanical blowing pressure adequate for the equal-tempered A 440 scale with the test recorder used in the majority of experiments.

6. Appendix: Individual characteristics

More than one pressure registration for the same subject's execution of a musical passage is included in figs. 18 and 19b,c in order to demonstrate the extent to which individual characteristics in the curves are reproduced when the same passage is recorded repeatedly. This reproducibility is also seen to obtain even when other music has been played between the test recordings, as was the case with Professor Conrad and to some extent with subject B.

The curves in figs. 17 - 19 are also included for the purpose of illustrating the individual characteristics of the subjects.

References

Anastasio, Angelo and Nicholas Bussard 1971: "Mouth air pressure and intensity profiles of the oboe", Journal of Research in Music Education, p. 62-76

Backus, John 1969: The Acoustical Foundation of Music (Norton)

Baines, Anthony 1963: Woodwind Instruments and their History, p. 72

Bak, Niels 1970: "Pitch, temperature and blowing pressure in recorder-playing", Acustica 22, p. 295-299

Bak, Niels 1971:   "Investigating the influence of
                                 blowing technique on pitch and
                                 tone in recorder playing",
                                 ARIPUC 6, p. 307-313

Bouhuys, Arend 1964:   "Lung volumes and breathing pat-
                                 terns in wind-instrument players",
                                 Journal of Applied Physiology 19,
                                 p. 967-976

Bouhuys, Arend 1965:   "Sound-power production in wind
                                 instruments", JASA 37,3, p. 453-456

Brown, W.S. 1973:   "Relationship of lingual and intra-
                                 oral air pressures during syllable
                                 production", JSHR 16, 1, p. 141-
                                 151

Coltman, John W. 1966:   "Resonance and sounding frequencies
                                 of the flute", JASA 40,1, p. 99-107

Coltman, John W. 1968:   "Sounding mechanism of the flute
                                 and organ pipe", JASA 44,4, p.
                                 983-992

Eck, Bruno 1944:   Technische Strömungslehre (Berlin)

Fletcher, N.H. 1975:   "Acoustical correlates of flute
                                 performance technique", JASA 57,
                                 1, p. 233-237

Hunt, Edg. 1963:   The Recorder and its Music (New
                                 York), p. 113

Kühn, Karl Th. 1940:   Klangfarbe und Wirbelform einer
                                 Lippenpfeife in Abhängigkeit von
                                 der Bauweise (Diss.) (Leipzig)

Leipp, E. 1962:   "La cavité buccale, paramètre sen-
                                 sible des spectres rayonnés par
                                 les instruments à vent", 4th
                                 ICA (Copenhagen)

Lüpke, Arn. v. 1940:   "Untersuchung an Blockflöten",
                               Akust. Zeitschrift 5

Meyr, Jürgen 1966:   "Akustik der Holzblasinstrumente
                               in Einzeldarstellungen", Das
                               Musikinstrument (Frankfurt/Main)

Mühle, Christoph 1966:   "Untersuchungen über die Resonanz-
                               eigenschaften der Blockflöte",
                               (Diss.) Das Musikinstrument
                               (Frankfurt/Main)

Nederveen, Cornelis J. 1969: Acoustical aspects of woodwind
                               instruments (Amsterdam), p. 17-18

Ower, E. 1966:   The measurement of air flow
                               (Oxford) (reprinted 1969)

Roos, J. 1936:   "The physiology of playing the
                               flute", Archives Néerlandaises de
                               phonétique expérimentale 12,
                               p. 1-26

Rowland-Jones, A. 1973:   Recorder Technique (Oxford)

Sundberg, Johan 1966:   "Mensurens betydelse i öppna
                               labialpipor", Acta Universitatis
                               Upsaliensis, Studia musicologica
                               Upsaliensia Nova series 3
                               (Uppsala)

REPORTS ON WORK IN PROGRESS AND PILOT STUDIES

PROBLEMS IN THE USE OF THE PHOTO-ELECTRIC GLOTTOGRAPH[1]

Birgit Hutters

Abstract: This preliminary study originates in some of the
            recurrent problems encountered at this institute
            in the course of several years of working with the
            photo-electric glottograph. The relationship be-
            tween the amplitude of the glottogram and the
            glottis aperture is studied by means of synchronous
            glottographic and fiberoptic recordings of the
            glottis during the articulation of long, unvoiced
            consonants in syllables with emphatic stress.
            According to this study, one does not very often
            find a linear relation between the glottis aperture
            and the amplitude of the glottogram. It is pointed
            out that the non-linearity may be caused by one or
            several factors affecting the amplitude of the
            glottogram. This non-linearity is crucial for the
            interpretation of the amplitude of the glottographic
            curve.

1. Introduction

Photo-electric glottography,[2] like most other methods of
registration, entails some important sources of uncertainty,
which must be taken into account in the application of the
method.

---

1) Part of the thesis for the cand.mag. degree, completed in
    May, 1975. I am most grateful to Jørgen Rischel, Svend E.
Lystlund, Mogens Møller, and Hajime Hirose for many valuable
suggestions and an inspiring guidance.

2) The reader is referred to Sonesson 1960, Malécot and Peebles
    1965, Ohala 1966, and Frøkjær-Jensen 1967 for a general intro-
duction to the method of photo-electric glottography.

The present paper summarizes a preliminary study of problems in the interpretation of glottograms. The experiments to be reported originate in some of the recurrent problems encountered at the institute of phonetics in the course of several years of working with the photo-electric glottograph.

2. A view of the relationship between amplitude of the glottogram and glottis aperture

One of the problems associated with the photo-electric glottograph is the relationship between glottis aperture and the amplitude of the glottogram: Is there a linear relation between these two variables, i.e., does a change in the amplitude of the glottogram correspond to a proportional change in glottis aperture?

I know of only one methodological study - in a narrower sense - of that problem. R.F. Coleman and R.W. Wendahl (1968) made synchronous high-speed film and glottographic recordings of the vocal fold vibrations in order to compare the variation of the glottis area and the variation of the glottogram during a vocal fold vibration. The authors (1968, p. 1734) conclude: "... that while photocell monitors, under controlled conditions, may provide an indication of laryngeal periodicity, relating waveforms thus derived to glottal area is not only hazardous but invalid in many cases." The authors may be right, but the glottograms shown in the article look so strange that it is tempting to conclude that there was something wrong with the glottographic set-up.

In contrast to the results just cited R.J. Harden (1975) concludes - likewise on the basis of synchronous high-speed film and glottographic recordings of the vocal fold vibration - that "Although the correspondence between the curves is not exact

in modal and vocal fry register phonations the photoelectric cell does appear to be capable of generating reasonably approximate information." (p. 734).

In the following subsections some of our general experiences about sources of error in the use of the photo-electric glottograph are presented, serving as an introduction to the experiments to be reported later in this paper.

2.1 Factors influencing the amplitude of the glottogram[1]

First of all, it is important to know whether the glottographic set-up itself has a non-linear performance in some part of its registration range, since the input-output characteristics of the set-up are decisive for the relationship between glottis aperture and the amplitude of the glottogram. But even if the voltages occurring during the glottographic recording are within a linear range, many other factors may influence the amplitude of the glottogram.

These other factors are partly associated with the light source and its position in relation to the larynx, and partly with the phototransducer and its position.[2]

2.1.1 The light source

The position of the light source in relation to the glottis, as well as the diameter and direction of the light beam, influence the glottogram to a considerable degree. If, for instance, it is the hindmost part of the glottis that is illuminated, the light will be strongly modulated by small movements of the carti-

---

1) The factors influencing the zero line level and the temporal variation of the glottogram will not be treated in this paper. For a brief survey of sources of error, see Frøkjær-Jensen, Ludvigsen, and Rischel 1971.

2) This study is based on the set-up employed at our institute, in which the light source is placed in a subglottal position and the transducer in a supraglottal position. If the two are placed in the reverse positions, the problems should, in principle, be the same.

laginous glottis. If it is the front part of the glottis that is
illuminated, the light will be strongly modulated by small move-
ments of the vocal folds. Therefore, with the first mentioned
position of the light source, the glottogram will show a great
difference between the maximum amplitudes of the two functions
of the glottis, viz. the vibratory pattern and the gross move-
ments; with the second position, on the other hand, the glotto-
gram will show a smaller difference. This means that it cannot
be taken for granted that there is a linear relationship between
glottis aperture and the amplitude of the glottogram with different
functions of the glottis, and certainly not if a light source is
used that illuminates only a part of the glottis.

If the larynx, and thus the glottis, is moved relatively to
the light source the relationship between glottal configuration
and light entering the pharynx via the glottis slit may be altered.
This will affect the amplitude and thus the relationship between
glottis aperture and the amplitude of the glottogram. Care must
therefore be taken to ensure that, during a recording in which
amplitudes are to be compared, the subject does not move. If such
movements do occur a comparison between successive data is rather
doubtful, even within the same function of the glottis. It is
very difficult, as will be demonstrated later, for the subject to
remain in exactly the same position in relation to the light
source for any long period of time. It is, of course, not possible
to prevent the larynx from altering its position in normal speech.

Light reflected from the structures above the glottis may
also influence the amplitude of the glottogram. This reflected
light may vary both as a function of variations in the position
of the larynx and as a function of changes in the structures
above the glottis.

2.1.2 The phototransducer

The factors mentioned above, which may influence the ampli-
tude of the glottogram, should be seen in relation to variations

in amplitude which are due to the phototransducer and its position in the pharynx.

It may be of great importance that the phototransducer does not alter its position during a recording, especially if the transducer is rather directional, i.e. picks up light only within a rather small angle. Furthermore, the transducer is hardly equally sensitive in all directions within the angle in which it picks up light, which aggravates the problem. The position of the phototransducer may be altered both by movements of the subject's head and neck and by the transducer tube being pushed about by the velum, the back of the tongue, or the pharyngeal wall.

The first problem can be solved by fixing the subject's head. The second problem can be solved to some extent by fixing the polyethylene tube containing the transducer in the oesophagus (Malécot and Peebles 1965, Ohala 1966). Furthermore, sounds with a pharyngeal articulation should probably be avoided. The optimal position of the transducer and whether it is to be fixed or not depends on the light source (its position and the diameter of the light beam), and on the directional selectivity of the transducer.

Structures above the glottis (as, for instance, the root of the tongue and the epiglottis) may with certain articulations and types of phonation enter the light field in the pharynx and reduce the quantity of light picked up by the transducer. This may greatly affect the amplitude of the glottogram.

A potential source of uncertainty is the fact that the phototransducer is sensitive to temperature variations. Malécot and Peebles have eliminated this source of uncertainty "by providing the subject with an air supply heated to body temperature from an electric heater and blower connected to a Bennett face mask." (Malécot and Peebles 1965). But, unfortunately, the authors do not explain how important temperature sensitivity is to the glottographic output.
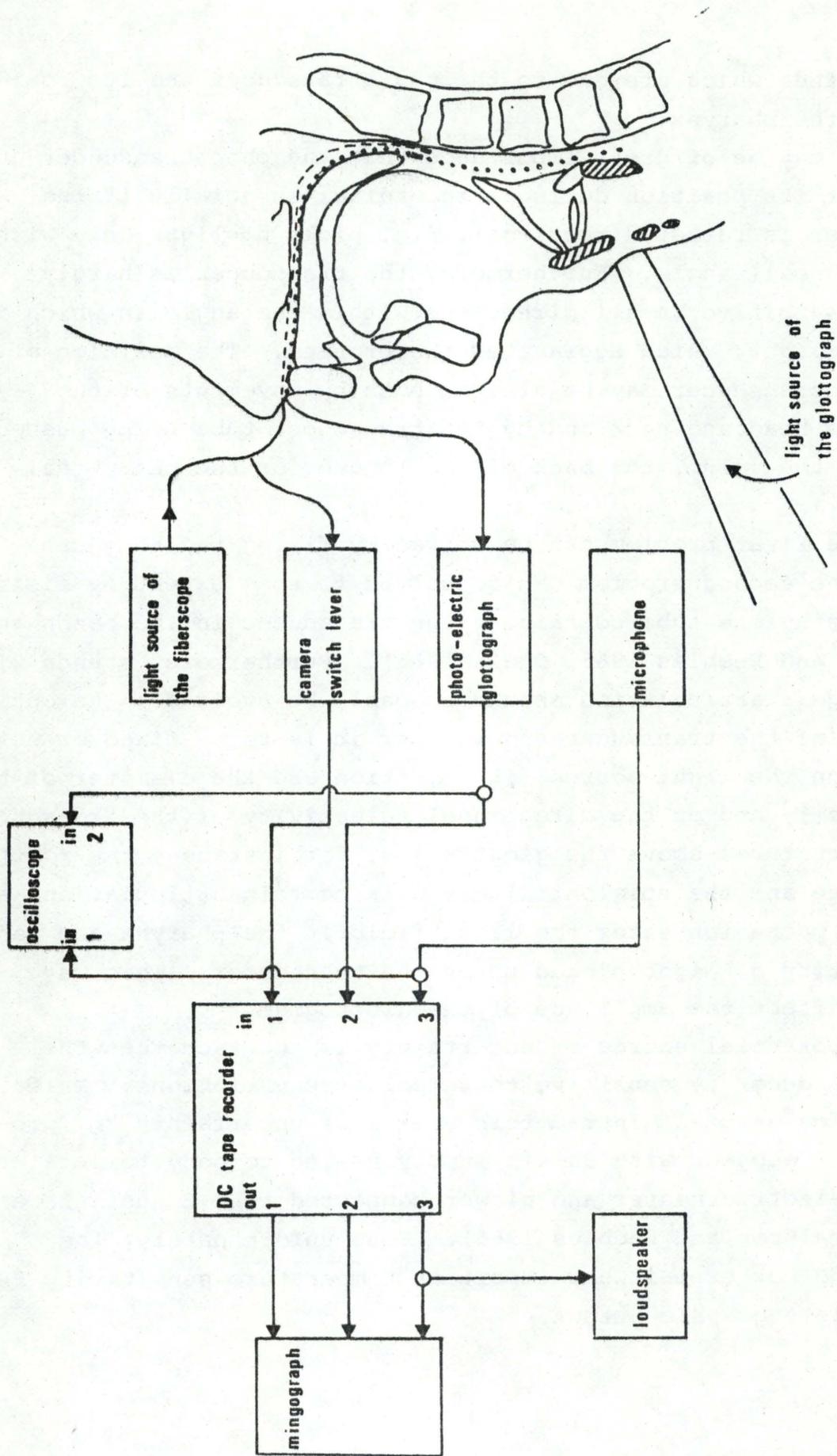
Figure 1

The instrumental set-up.

3. A study of the relationship between the amplitude of the glottogram and glottis aperture

The study I am going to survey here covers only a small fraction of the complex of problems mentioned in the above section.

3.1 The purpose of the study

I chose to approach the problem of the relationship between the amplitude of the glottogram and glottis aperture by studying fairly large glottis apertures, i.e. apertures of such dimensions as may be found in the case of long unvoiced consonants in syllables with emphatic stress. The reason for this choice being that under these conditions it was possible, with our outfit, to take photographs of the glottis which would serve to determine the glottis aperture - viz. by means of a fiberscope connected with a still camera. As there was no possibility of photographing successive variations of the glottis aperture, the study was made on the basis of three fixed glottal positions and synchronous glottographic recordings. At best, this study will show whether it is possible, on the basis of the amplitude of the glottogram, to obtain information about the relative size of rather large glottis apertures.[1]

3.2 The instrumental set-up

The central components of the instrumental set-up, which is shown in fig. 1, are the photo-electric glottograph and the fiberscope with a still camera connected to it.

The glottograph of this study is a modified version of the photo-electric glottograph described by Frøkjær-Jensen (1967, 1968). The light from a light source fed with DC current is led

---

1) One cannot, by this approach, get any absolute information on the glottis aperture, since the photo-electric glottograph cannot be calibrated with living larynges (van den Berg 1968).

to the larynx through a 40 cm long, conic plastic rod. The dia-
meter of the rod is 1 cm at the larynx end.[1] The phototransducer
(Texas Instruments LS 400) is contained in a protecting poly-
ethylene tube with an outside diameter of 3 mm. The transducer
is inserted through the nose into the pharynx. In order to secure
the transducer in a relatively stable position in the pharynx
it is fixed at a distance from the lower end of the transparent
protecting tube which is swallowed into the oesophagus. The tube
containing the transducer is, furthermore, fixed to the subject's
cheek. The glottographic signal is led partly to the oscillo-
scope (Tektronix 5103 N), partly to the DC tape recorder (Lyrec
TR 86). The fiberscope (Olympus VF, type 4A) is described by
Sawashima and Ushima (1971). The diameter of the optic cable is
4.6 mm, and the angle of the lense is 42°. It should be mentioned
here that it is difficult to fix the fiberscope tube in a quite
stable position during the articulation. The M-X synchronizing
switch lever of the still camera attached to the fiberscope was
used to synchronize the glottogram and the picture of the glottis.
Furthermore, a microphone signal (Neuman KM 56) was recorded.
After the recordings mingograms were made of the three signals:
glottogram, synchronization pulse, and microphone signal (paper
speed: 100 mm/sec).

3.3 Measurements on the synchronization and input-output characteristics of the set-up[2]

3.3.1 Synchronization

The pulse from the M-X synchronizing switch lever (set at
M) was used to synchronize the photographs and the glottograms.

---

1) This diameter of 1 cm is an arbitrary choice, since we have
   not yet made a careful study determining the preferable dia-
meter of the light conducting rod.

2) These recordings were made without human subjects.

As the switch lever and the shutter of the camera are not syn-
chronous, the time offset between the movements of these two
parts was measured. In a special set-up the photo-electric glotto-
graph was used to determine the open period of the shutter. The
transducer - placed in a cardboard tube - and the light conducting
rod were placed on either side of the objective and the shutter.
With this set-up the duration of the period during which light
was passing the objective could be registered, as well as its
temporal relation to the pulse from the switch lever. The switch
lever pulse and the "shutter pulse" were led to a storage oscillo-
scope, the screen of which was photographed. The time offset be-
tween the two pulses was determined on the basis of these pictures.
As it could not be taken for granted that the offset was the same,
irrespective of exposure time, it was measured for different ex-
posure times. The difference in time between the start of the
pulses was about 13.0 ms, irrespective of exposure time, the shutter
lagging behind the switch lever pulse. In some cases it was diffi-
cult to determine the moment of cessation of the switch lever
pulse, but it ended about 4.0 ms later than the shutter, regardless
of exposure time.

3.3.2 The input-output characteristics of the glottographic set-up

It is important to know whether a non-linear relationship
between the amplitude of the glottogram and glottis aperture might
be due to the input-output characteristics of the glottographic
set-up.

The sensitivity of the transducer to variations in illumi-
nation[1] was checked. In a darkroom a set-up involving the photo-
electric glottograph, an enlarger, and a voltmeter was assembled.
The transducer was fixed in a stand and placed below the diaphragm
and the bulb of the enlarger. Care was taken to position the

---

1) Or, properly speaking, "luminance".
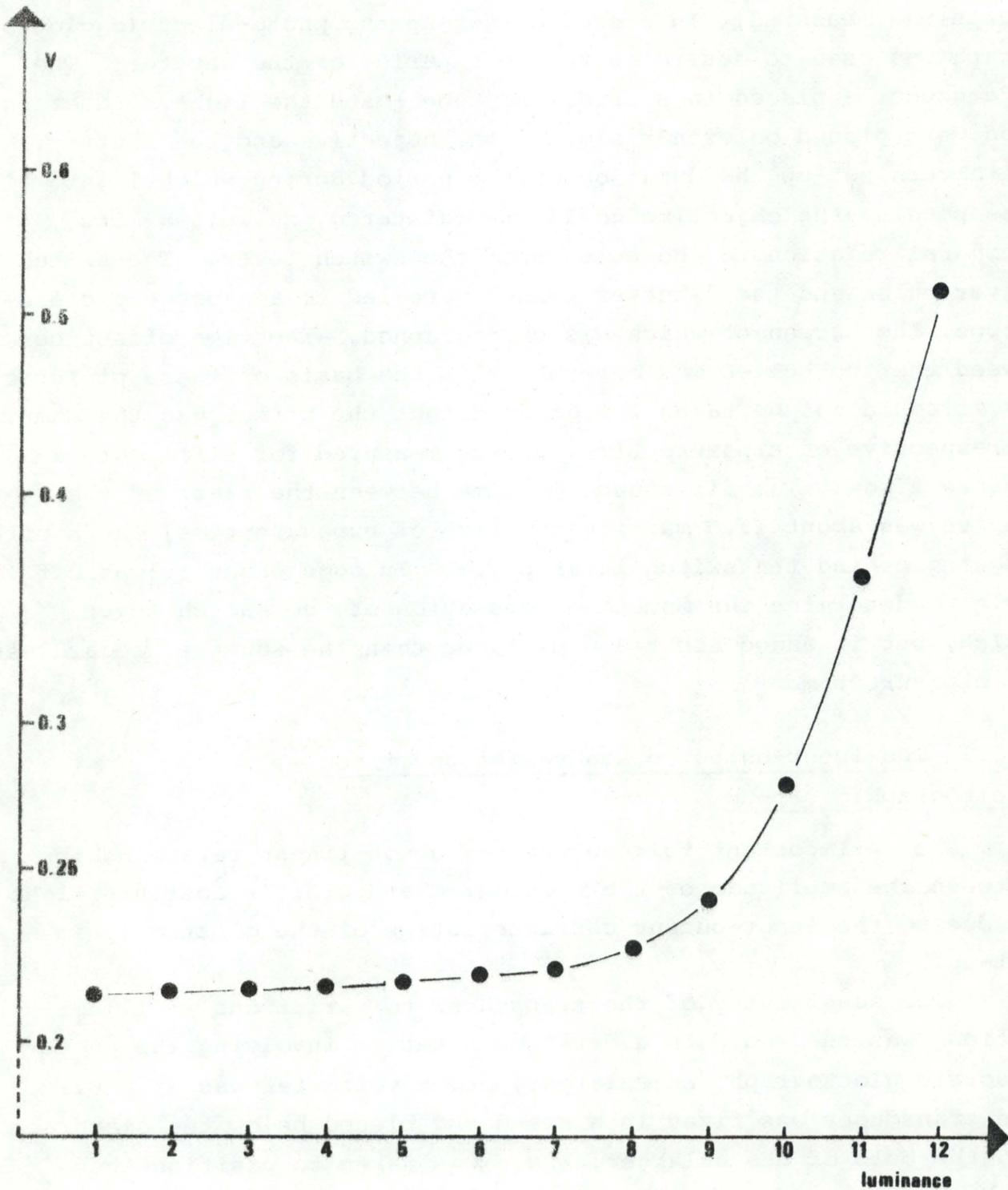
Figure 2

The characteristic curve of the glottographic
set-up - with setting 6 - for variations in
luminance. The division of the abscissa corre-
sponds to steps in the aperture of the diaphragm,
every step corresponding to a doubling of luminance.

transducer so that the whole angle within which the transducer
was sensitive was inside the cone of light, irrespective of the
aperture of the diaphragm. The aperture of the diaphragm was
varied, i.e. the illumination was varied in well defined steps,
and the corresponding voltage for each step of the diaphragm was
read off. In order to obtain enough steps the measurements were
performed with two different distances between transducer and
diaphragm, the greater distance being chosen so that the voltage
for the greatest aperture of the diaphragm with this distance was
equal to the voltage for the smallest aperture of the diaphragm
with the smaller distance. These measurements were performed for
settings 1 to 8 of the glottographic amplifier. The light picked
up by the transducer in the pharynx is diffuse light, and the
light used for these measurements was diffuse light, as well.
In an earlier experiment the transducer was checked for sensi-
tivity to white and red light because the glottographic light in
the pharynx belongs to the red part of the spectrum. There being
no noticeable difference between the characteristic curves for
white and red light, the subsequent measurements were made only
for white light.

In fig. 2 the voltages are shown as a function of illumina-
tion with setting 6 which is the setting used in the recordings
(the division of the abscissa corresponds to steps in the aperture
of the diaphragm; every step corresponding to a doubling of the
luminance).

It seems evident that the glottographic set-up for setting 6
has a non-linear characteristic curve in the lower end of the
voltage range, which means that the relationship between glottis
aperture and the amplitude of the glottogram cannot be linear in
this lower end of the voltage range (unless this non-linearity
happens to be neutralized by other factors). - These findings
imply that, in fact, the voltage ranges occurring in each of the
glottographic recording sessions dealt with in the following
sections should have been measured, in order to reveal whether

the voltage range for a given recording was within a linear range
or not. Unfortunately, I did not realize this at the time when
the recordings were made.

Furthermore, the performance of the transducer at different
angles of incident light with two different degrees of rotation
was measured. The transducer was ground at the free end and the
black paint had been removed in order to make it less directional.
The position of the transducer was varied within an angle of 180°
in relation to an immovable light source, and the output voltages
were read off the glottograph. Two series of measurements were
made, where in the second series the transducer was rotated 180°
around its longitudinal axis relatively to its position in the
first series. In fig. 3 output voltage is shown as a function of
the angle of incident light. It is seen that the transducer picks
up light within an angle of at least 150°, and that its sensitivity
depends on the position of the transducer relatively to the light
source. Furthermore, it is evident that the output voltage varies
when the transducer is rotated around its own longitudinal axis,
but such a rotation is not possible when the lower end of the
protecting tube is swallowed into the oesophagus. On the other
hand, the non-uniform sensitivity of the transducer as dependent
upon the angle of incidence of light may influence the amplitude
of the glottogram. Therefore, whenever light source and trans-
ducer change their position relatively to each other, it cannot
be taken for granted that there is a linear relationship between
the amplitude of the glottogram and glottis aperture, even if
the transducer picks up the same quantity of light. Furthermore,
in the case of two glottis apertures of different sizes, a linear
relationship is not to be expected, even if the glottographic light
is of equal density across the illuminated area of the glottis.

I have not checked whether the sensitivity of the trans-
ducer to variations in heat caused by the inhalation and exhala-
tion of air, seriously affects the performance of the glotto-
graphic set-up used in these experiments.
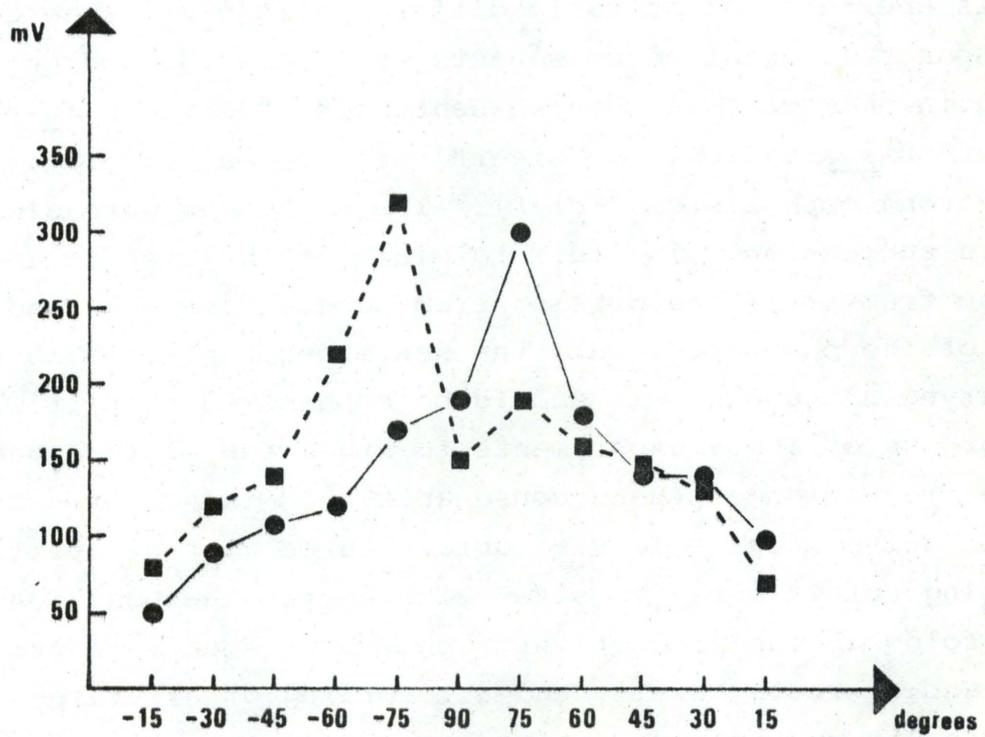
Figure 3

The performance of the phototransducer at
different angles of incident light for two
different degrees of rotation.

3.4 The material

As mentioned above (section 3.1), this study of the relation between the amplitude of the glottogram and the aperture of the glottis is based on recordings of glottis apertures of such dimensions as may be found in long unvoiced consonants in syllables with emphatic stress. In order to obtain information about this relationship it is necessary to record at least three glottis apertures of suitable difference in size. The choice fell upon the sustained consonants [ʂ::], [s::], and [ç::] embedded in the following Danish sentences: "De viste Ib billetten. De viste Lis sibyllen. De higer" (pronounced [ɖi'ç::i:ʌ] with very strong emphasis on "-hi-"). The sentences were always said in this succession. In order to minimize the risk of interference from artifacts arising from tongue movements and movements of the pharyngeal wall the sentences contained no velar or pharyngeal sounds. It should be emphasized that it was not the purpose of these experiments to study the glottal behaviour in the production of these consonants ([ç] is not even found in normal Danish), but solely to obtain three glottal apertures differing sufficiently in size. Therefore, the position of the vocal folds in the production of the three sounds is called a, b, and c, respectively, and a recording of all three glottis apertures in succession is named an abc-series.

3.5 Subjects

Two phoneticians, highly experienced in experimental phonetics, acted as subjects: JR and BFJ. For subject JR 54 abcseries were recorded over three sessions, for subject BFJ 53 abcseries were recorded over two sessions.

3.6 Light and exposure time

Owing to the synchronous recording of the glottograms and the photographs the light source of the fiberscope set-up had to be switched off. The pictures were taken by the light from the

glottograph passing the glottis slit from below. Consequently, these photographs do not show the actual structures of the larynx, but only the contours of some part of the glottis. Therefore, in order to control the position of the fiberscope pictures of the glottis were taken before and after each abc-series by the fiberscope light, i.e. with illumination from above. For these "control pictures" the subject articulated a sustained [s::]. The two [s::]'es are indicated by S and S^+, respectively. If the control pictures of an abc-series were not identical - indicating that the fiberscope had altered its position - it was the intention to reject the series since the pictures of the glottis aperture within one abc-series can only be compared if they are taken under identical conditions.[1] When taking the control pictures the glottographic light was, in most cases, also switched on in order to observe how the light from this source was passing through the glottis slit.

The exposure time of the abc-series was 1/60 sec. The same exposure time could be used for the control pictures with an appropriate adjustment of the fiberscope light.

The film was a Kodak TX 135-36.

3.7 Positioning of fiberscope, transducer, and light source

In the recordings described here the tip of the fiberscope was positioned rather close to the glottis in such a way that the glottis slit took up as much as possible of the field of vision. This was done in order to maximize the accuracy of measurement. But, consequently, it happened rather often that

---

1) It should be mentioned that even if the control pictures are identical this is no guarantee that the fiberscope has not moved back and forth, or up and down, during the abc-series, but this can be checked only by means of a synchronous X-ray recording.

a minor part of the glottis escaped the visual field owing to
movements of the speech organs. Thus, the position of the fiber-
scope had to be readjusted fairly often.

In two recordings (JR 29/5 and JR 22/7) the light conducting
rod of the glottograph was placed in such a way that the light
entered the neck slightly below the cricoid cartilage. In this
position the light passes primarily through the central part of
the glottis slit. As for the position of the transducer of the
glottograph the maximum voltage range was obtained when the
distance between the transducer and the outer nostril was 16 cm.

In all other recordings the rod was placed in such a way
that the light entered the neck between the thyroid cartilage
and the cricoid cartilage. In this position the light passed
primarily through the anterior part of the glottis and the
maximum voltage range was obtained when the distance between the
transducer and the outer nostril was 14 cm.

3.8 Measurements of the glottis aperture and the amplitude of the glottogram

The negatives were converted into contour drawings by
tracing a projection with a magnification of the order of 30:1.
A microfiche viewer was used for this purpose.

As the fiberscope cannot be calibrated (at least no such
calibration has been described), the pictures yield information
only about the relative size of the glottis aperture. The
distance between the vocal processes is generally referred to
as a measure of the glottis aperture. This distance is directly
proportional to the glottis area, provided that the glottis is
a triangle in which the baseline and the line drawn between the
processes are parallel.

The distance between the vocal processes could not be used
in this study, partly because the light source illuminates only
part of the entire glottis slit (in its longitudinal direction)

Control picture [s::]

a [z̧::]
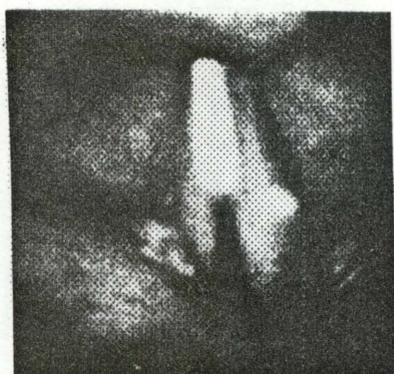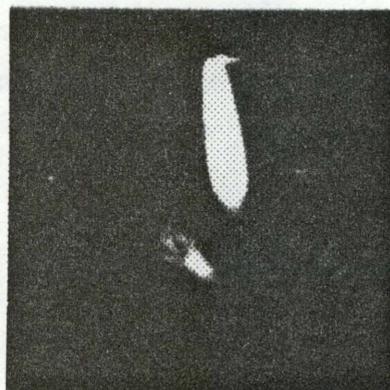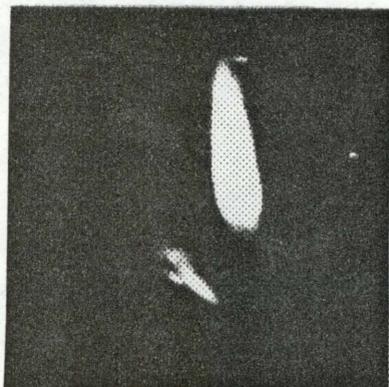
b [s::]
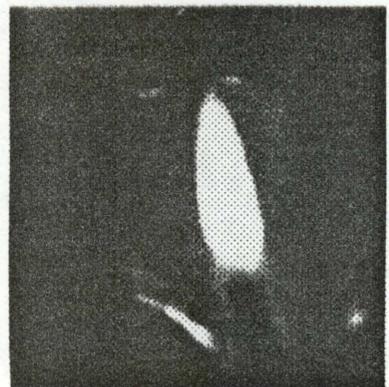
c [ç::]

Figure 4

Photographs of one abc-series including one
of the two control pictures. The glottographic
light passes the anterior part of the glottis.

if the cartilaginous glottis is open, and partly because the abc-pictures reveal only the contour of the glottis. Therefore, another line, parallel to the baseline of the glottis triangle, was used. As this relative measure of glottis aperture is to be used for a comparison, the line has to be drawn at the same distance from the anterior commissure of the glottis in all photographs. In order for the contour of the anterior commissure to be observable the light must pass the anterior part of the glottis; if it does not, the commissure must be determined by extrapolation. In some cases the extrapolation method was applied. In certain other cases I employed a constant distance from the upper limit of the field of vision. This method, which may be even more inaccurate than the extrapolation method, implies that the glottis has a fixed position in relation to the upper limit of the field of vision throughout the recording of an entire abc-series.

It should be borne in mind that the basic prerequisite for a comparison of glottis apertures on the basis of photographs taken through the fiberscope is that the distance between the glottis and the lens of the fiberscope is kept constant during the recording. As mentioned above (footnote to section 3.6), this cannot be taken for granted, even in the case of identical control pictures.

Fig. 4 shows photographs of the glottis contour for one abc-series and for one of the control pictures. The glottographic light is positioned in such a way that the light passes through the anterior part of the glottis.

From the control pictures it is obvious that some trans-illumination of the vocal folds occurs. How much cannot be determined on the basis of the photographs because of the saturation threshold of the film (cf. footnote to section 3.9). However, the error caused by the transillumination is probably the same for the three measurements of the glottis aperture within

one abc-series.[1] Another problem is whether the addition to the amplitude of the glottogram, caused by the transillumination of the vocal folds, is constant - irrespective of the glottis aperture - which is probably not the case. So, here we may have another source of uncertainty to the glottographic method.

The amplitude of the glottogram was measured 10 ms after the start of the switch lever pulse. As mentioned before, the exact time lag between the start of this pulse and the shutter was 13 ms. But truly precise determination of the start of exposure seems to be rather meaningless, considering the inaccuracy introduced by other parts of the adopted procedure. When the amplitude of the glottogram varied no more than 1 mm during the exposure, the average was measured. The glottogram was rejected if the variation exceeded 1 mm. The zero-line level was arbitrarily determined, but it is, of course, fixed within each abc-series.

3.9 Results

Out of 107 abc-series 64 series were rejected for one or several of the following reasons:

1) The picture had been taken too soon or too late in relation to the consonant
2) The glottographic signal was overloaded
3) The fiberscope was not in a stable position
4) Both the amplitude and the relative glottis area differed very little, i.e. 1 mm or less, from a to b or from b to c (i.e. the series is not sufficiently informative)
5) The variation of the glottographic curve exceeded 1 mm during the exposure.

It is obvious that one rejected recording within an abc-series entails a rejection of the whole series.

---

1) Since the difference in size between the observed light spot and the illuminated part of the glottis slit seems to be considerable, it is obviously desirable to develop a different procedure in which the pictures are taken with light from above, the glottographic light being intermittently switched off, triggered by the camera.

Actually, the requirements were satisfied only in a few
cases, inasmuch as the position of the fiberscope (or of the
larynx) changed in most cases during the recording of an abc-
series. Still, if there was only a slight disturbance, the
series was included among the usable material. There is reason
to suppose that the inaccuracy of measurement caused by minor
displacements of the fiberscope does not exceed the inaccuracy
of other parts of the procedure.

It must be emphasized that measurements of amplitude and
aperture can only be compared within one series, as it is only
during this short sequence that attempts have been made to keep
the recording conditions constant. The fiberscope may have been
readjusted between series. Furthermore, the films sometimes re-
veal unintentional disturbances of the glottographic light during
a recording session. It cannot be taken for granted that the
light (from the kind of light source employed in this study) will
remain in a stable position during a recording.

In the table and the figures identical control pictures are
indicated by $S=S^+$, almost identical control pictures are indicated
by $S(=)S^+$. The measures of aperture and amplitude (both in mm)
are called A and GL, respectively.

The results from the serviceable series are seen in table 1.
In figs. 5-9 GL is shown as a function of A. It is seen that the
three points in most cases deviate more or less from a straight
line, and in one third of the cases there is even a negative cor-
relation between the glottis aperture and the amplitude of the
glottogram.

Minor deviations from a straight line may be explained by
the inherent characteristics of the glottographic set-up (see
section 3.3.2 above), i.e. by the characteristic curves showing
the output voltage as a function of varying illumination (lumi-
nance) and as a function of variations in the angle of incidence
of light. As previously mentioned, the output voltage range of
the glottographic recordings were not measured for each session,

since I did not realize until later that that was necessary in
order to know whether the voltage range for a given recording
was within a linear range or not. If the voltage range for a
given recording lies in the low, non-linear range the curve would
bend off "to the right", provided that the glottographic output
were influenced by that factor only. But since several factors
may simultaneously influence the amplitude of the glottogram it
is impossible to propose one single explanation for each devia-
tion from the straight line. There is another essential source
of error, viz. that the transducer and/or the glottographic light
may have moved during the recording of a series. To judge by
the films (see footnote to section 3.6), the light does not seem
to have moved during the recording of any of the series which
show only minor deviations from a straight line. In order to
find out whether the transducer may alter its position during the
production of [ɧ::], [s::], and [ç::] X-ray pictures of those
sounds were taken for subject BFJ, the transducer (and the fiber-
scope) being positioned as during the actual recordings. The
X-ray pictures show that the transducer was pushed slightly down-
wards during the articulation of [s::]. As only one picture of
each sound was taken it cannot be proven whether this varying
location of the transducer is random or systematic. But according
to my experience, minor displacements of the transducer will not
influence the glottographic output very much. Also, the trans-
illumination of the vocal folds could explain part of the devia-
tion from a linear correlation. Of course, the applied procedure
for measuring the glottis aperture causes a certain inaccuracy of
the measurements.

For a large part of the series the results deviate very much
from a linear correlation, and, as mentioned above, one third of
all series show a negative correlation (all but one instances of
negative correlations are found with subject BFJ). This cannot
just be due to the characteristics of the glottographic set-up.

Table 1

Results of the measurement of glottis
aperture and amplitude of glottogram.

Recording JR 29/5

	A (mm)	GL (mm)
1.S(=)S⁺		
a	15.0	9.5
b	22.0	14.0
c	27.0	19.5
2.S(=)S⁺		
a	14.0	8.0
b	21.5	16.0
c	25.5	23.5
3.S=S⁺		
a	17.0	14.0
b	26.0	27.0
c	21.5	22.5
4.S(=)S⁺		
a	17.0	12.5
b	22.0	22.0
c	21.5	20.0
5.S(=)S⁺		
a	16.5	12.0
b	21.5	17.0
c	23.5	24.0
6.S(=)S⁺		
a	17.0	12.5
b	21.0	18.5
c	24.0	25.5
7.S=S⁺		
a	14.0	10.0
b	19.0	13.5
c	22.5	21.5

	A (mm)	GL (mm)
8.S(=)S⁺		
a	15.0	12.0
b	21.0	17.5
c	21.0	23.0
9.S(=)S⁺		
a	15.0	11.0
b	18.5	17.5
c	22.0	26.0
10.S=S⁺		
a	15.0	13.0
b	20.0	22.5
c	21.0	24.5
11.S=S⁺		
a	10.0	4.0
b	20.5	8.0
c	22.0	6.5
12.S=S⁺		
a	16.0	7.0
b	20.0	8.5
c	22.0	12.0

Recording JR 22/7

	A (mm)	GL (mm)
1.S=S⁺		
a	11.0	5.0
b	16.5	9.0
c	26.5	16.0
2.S=S⁺		
a	12.5	22.0
b	15.0	26.0
c	20.5	34.0
3.S=S⁺		
a	13.0	17.5
b	16.0	26.0
c	22.5	33.0
4.S=S⁺		
a	11.5	19.0
b	16.0	32.0
c	20.0	33.0
5.S=S⁺		
a	10.5	15.5
b	15.0	26.0
c	19.5	33.5

Recording JR 28/8

	A (mm)	GL (mm)
1.S(=)S⁺		
a	10.5	5.0
b	11.5	7.5
c	14.5	18.0

Table 1

(continued)

Recording BFJ 18/9

	A (mm)	GL (mm)		A (mm)	GL (mm)
1.S(=)S$^+$			9.S=S		
a	6.0	13.0	a	12.5	14.5
b	8.0	15.0	b	15.5	15.0
c	9.5	29.5	c	19.5	20.0
2.S(=)S$^+$			10.S(=)S$^+$		
a	5.5	5.5	a	12.5	17.5
b	9.5	10.0	b	16.0	18.0
c	16.0	16.0	c	17.0	23.0
3.S(=)S$^+$			11.S=S		
a	7.0	8.5	a	9.5	14.5
b	11.5	12.5	b	16.5	17.5
c	16.0	21.0	c	18.5	11.5
4.S(=)S$^+$			12.S(=)S$^+$		
a	6.0	6.5	a	11.5	11.5
b	11.0	9.5	b	16.5	14.5
c	14.0	15.5	c	19.5	12.0
5.S(=)S$^+$			13.S=S$^+$		
a	6.0	7.5	a	12.5	13.0
b	11.5	9.0	b	17.5	18.0
c	18.5	19.5	c	24.5	18.0
6.S(=)S$^+$			14.S(=)S$^+$		
a	6.5	5.0	a	10.5	11.0
b	11.0	8.0	b	17.0	16.0
c	14.0	10.5	c	21.5	14.0
7.S(=)S$^+$			15.S(=)S$^+$		
a	7.0	6.5	a	11.5	12.0
b	10.5	7.0	b	17.5	19.5
c	18.5	15.5	c	26.0	20.0
8.S(=)S$^+$			16.S=S$^+$		
a	12.0	18.0	a	10.5	14.5
b	13.0	12.0	b	19.5	22.0
c	21.5	21.0	c	24.5	21.5

Recording BFJ 14/10

	A (mm)	GL (mm)
1.S(=)S$^+$		
a	8.5	10.5
b	12.0	15.0
c	12.5	13.0
2.S(=)S$^+$		
a	9.5	10.0
b	12.0	13.0
c	14.0	12.5
3.S(=)S$^+$		
a	8.0	9.0
b	12.5	16.0
c	13.5	13.5
4.S(=)S$^+$		
a	8.0	7.0
b	9.5	10.5
c	16.0	13.0
5.S(=)S$^+$		
a	8.5	8.0
b	11.5	10.5
c	16.5	13.0
6.S(=)S$^+$		
a	8.0	7.0
b	12.0	12.0
c	14.5	11.0
7.S=S$^+$		
a	11.5	8.5
b	13.5	13.0
c	17.5	13.5
8.S(=)S$^+$		
a	8.5	8.0
b	13.5	11.0
c	13.5	6.0
9.S(=)S$^+$		
a	8.0	9.5
b	12.5	14.0
c	15.5	16.0

<u>Figure 5</u>

Recording JR 29/5

Amplitude of the glottogram, GL, is shown as
a function of the glottis aperture, A.
Identical control pictures are indicated by
$S=S^+$, almost identical control pictures are
indicated by $S(=)S^+$.
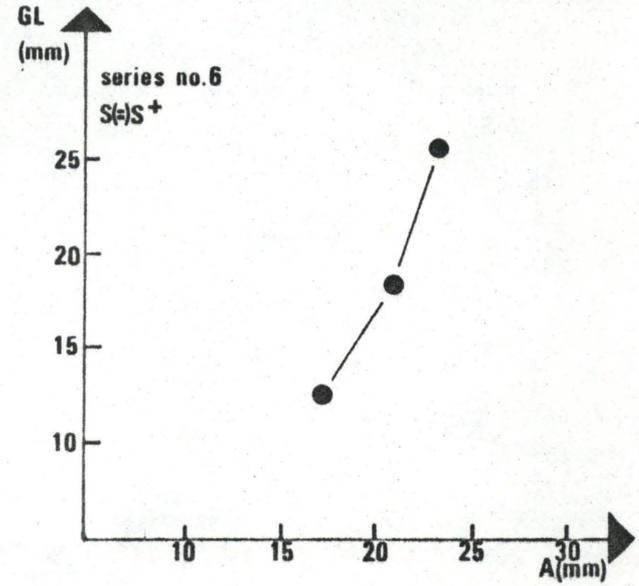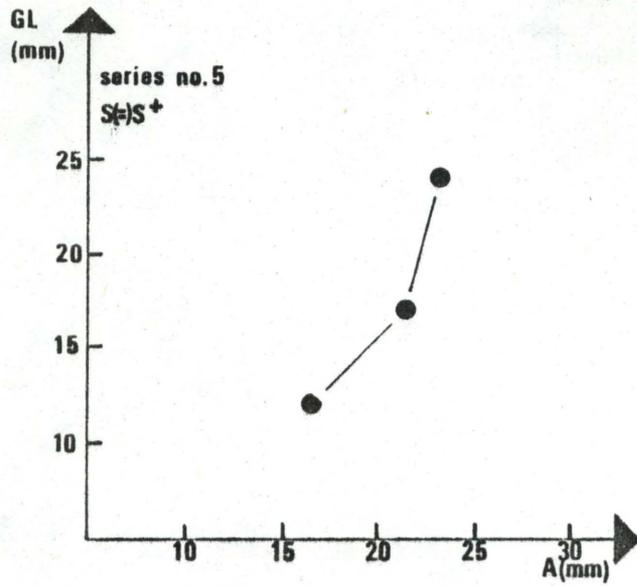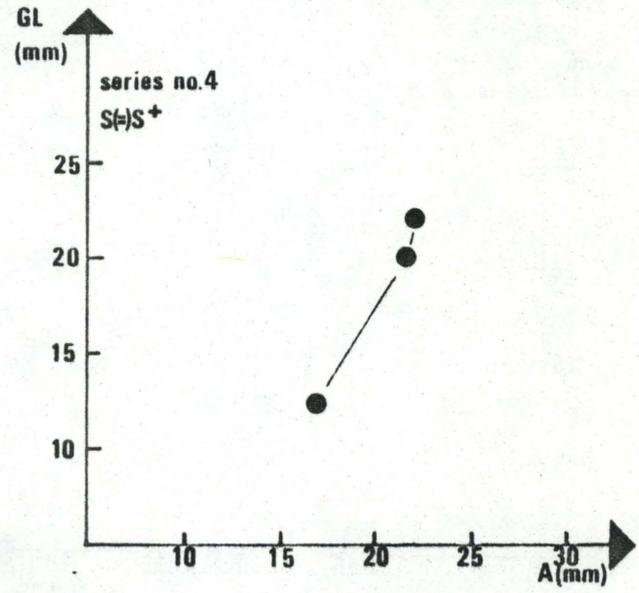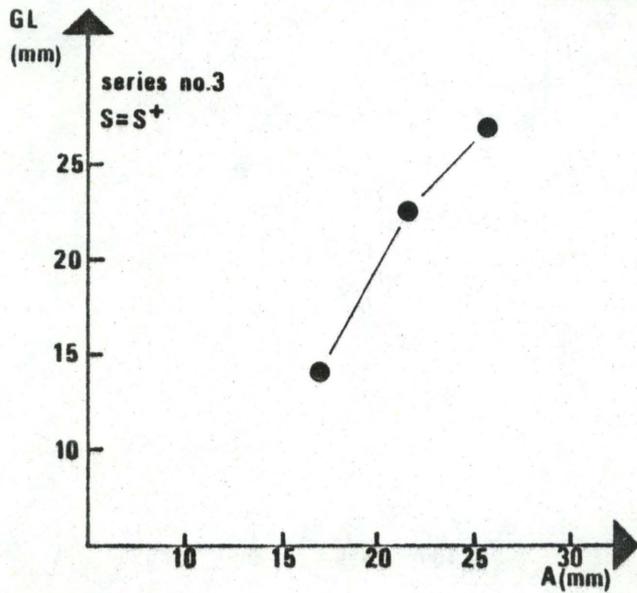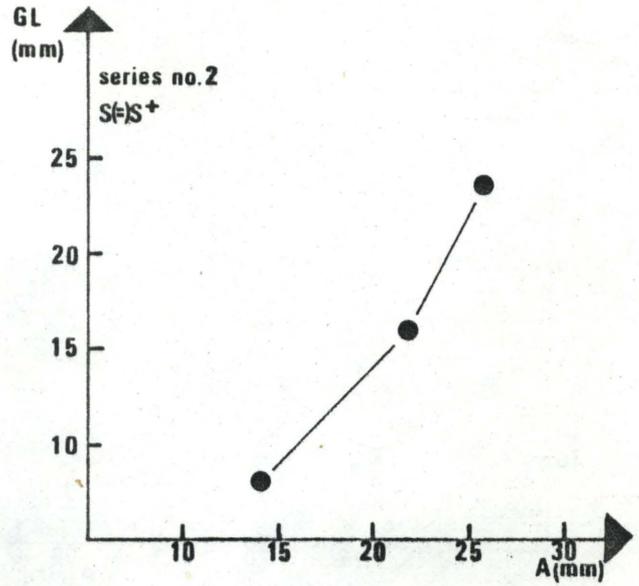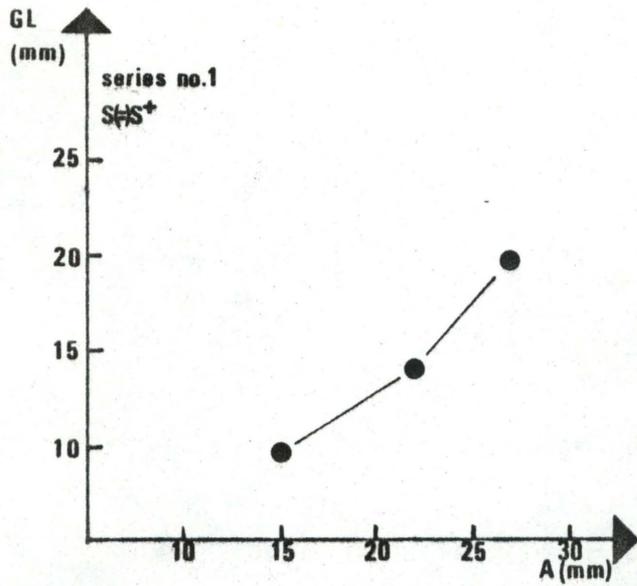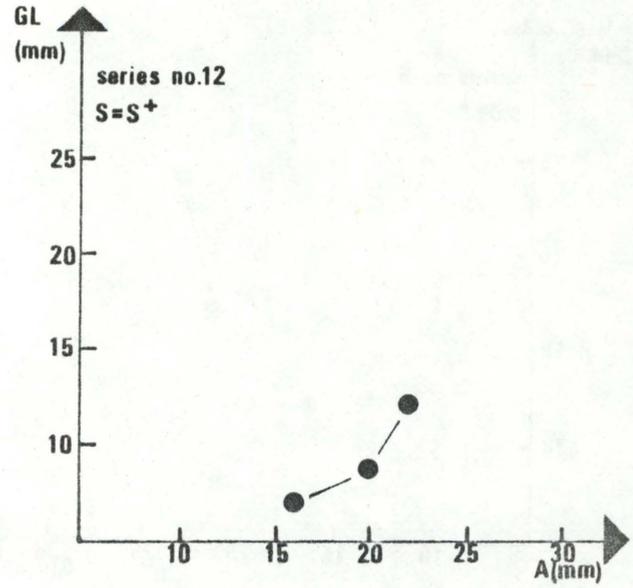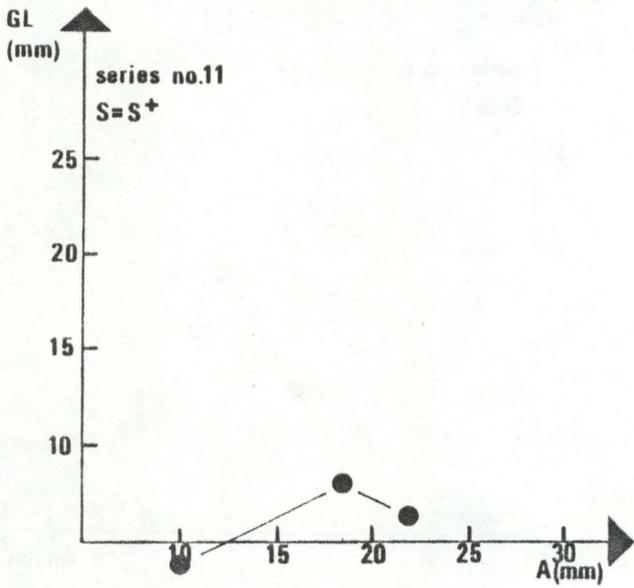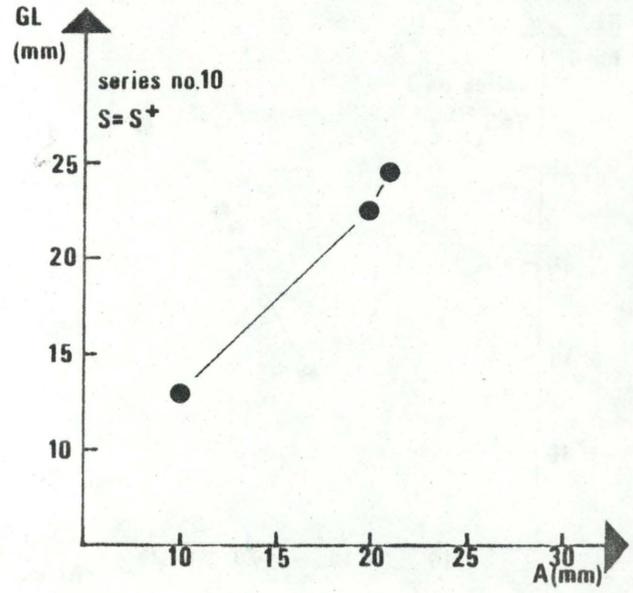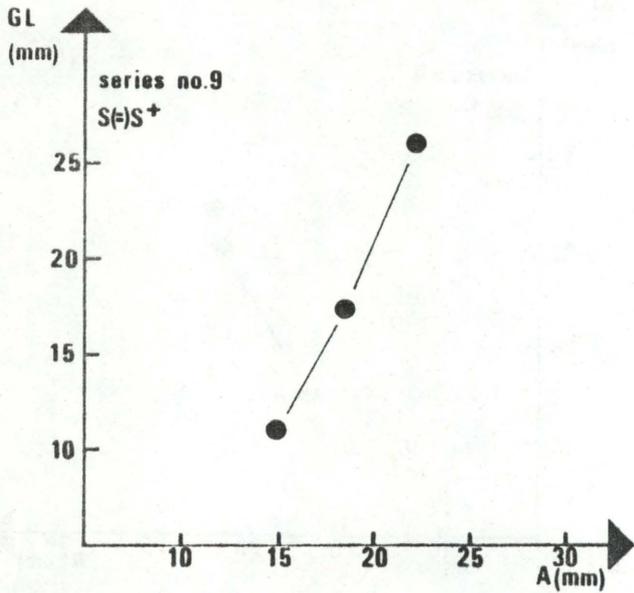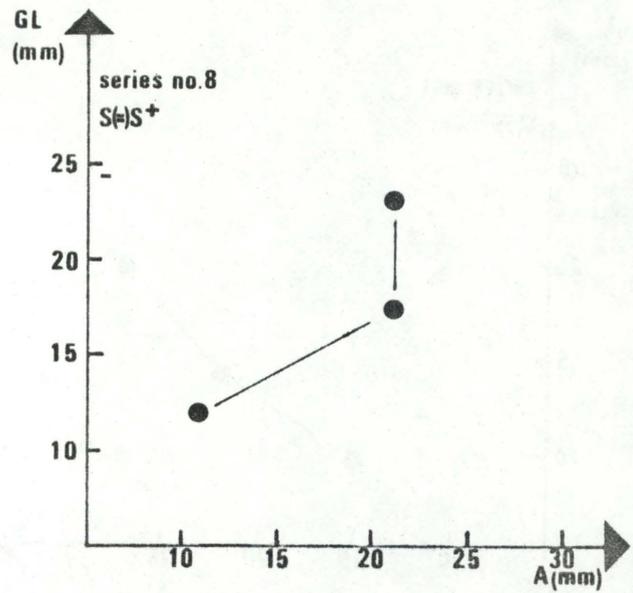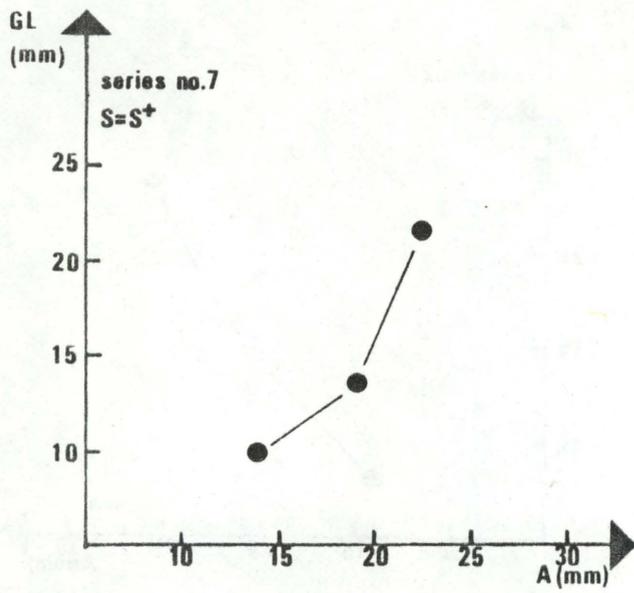
Figure 5 (continued)

Recording JR 29/5
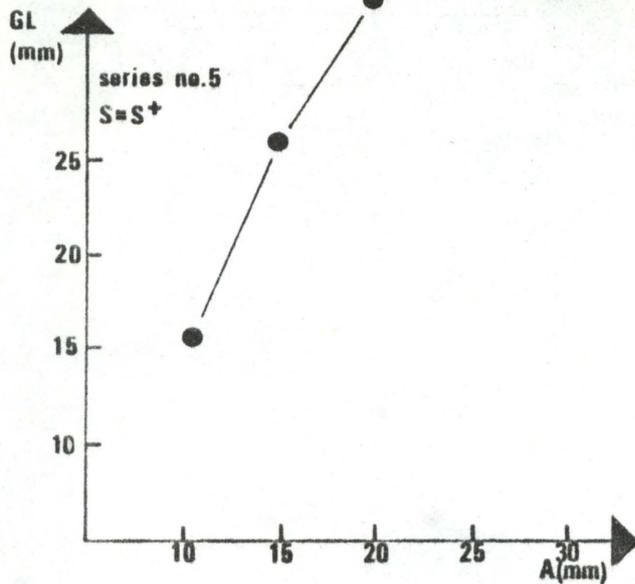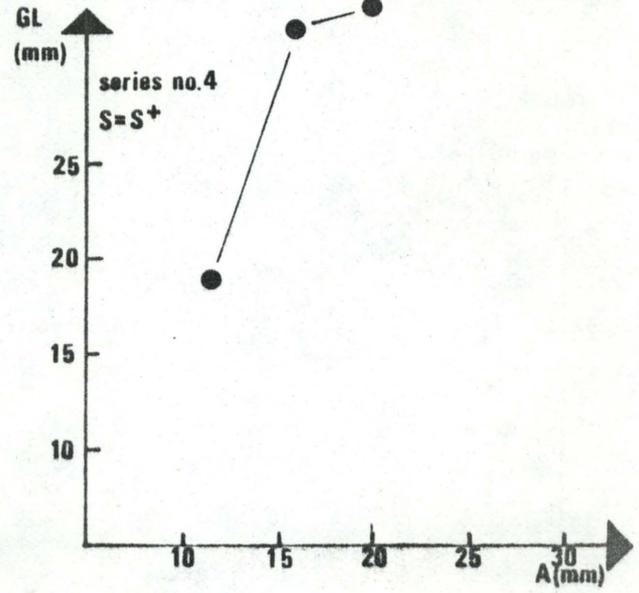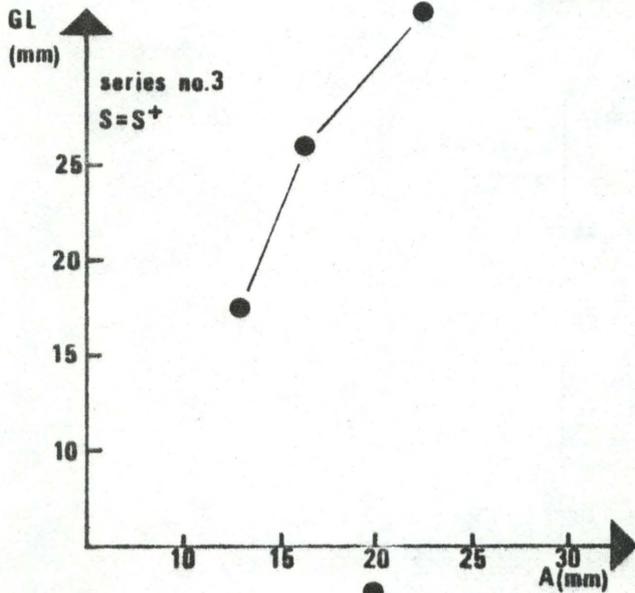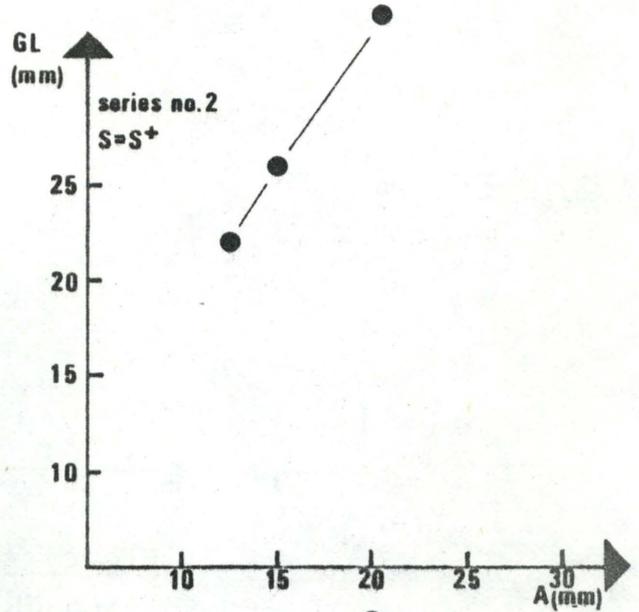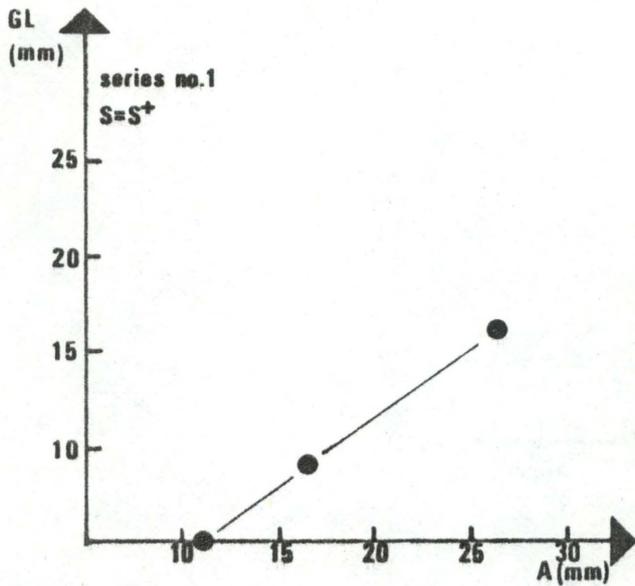
Figure 6
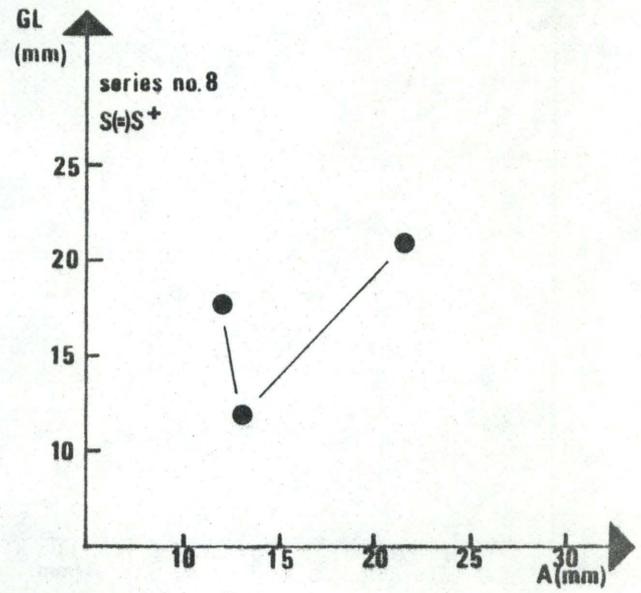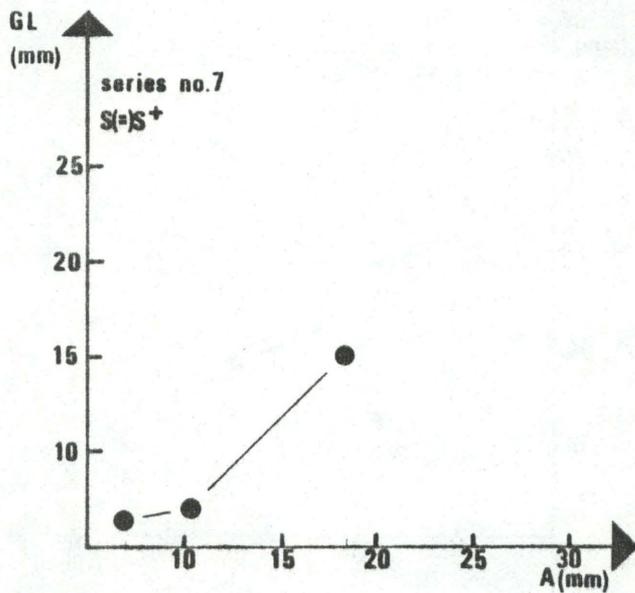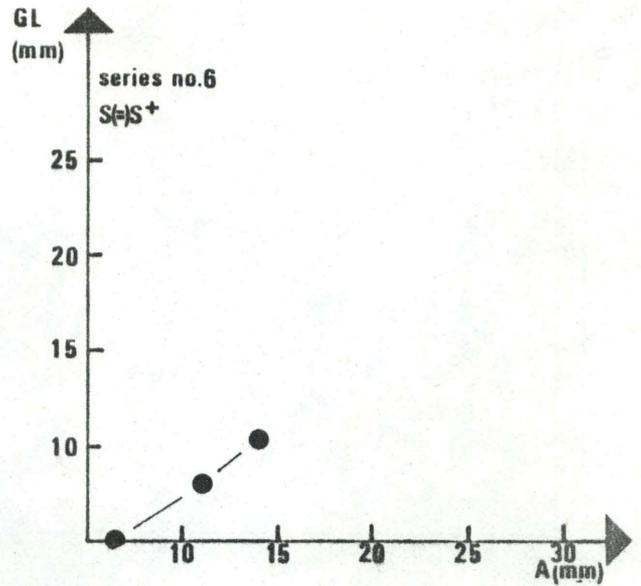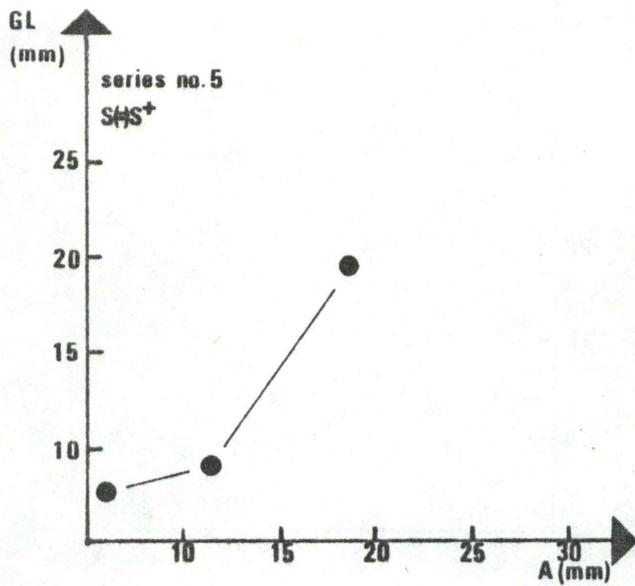
Recording JR 22/7

302
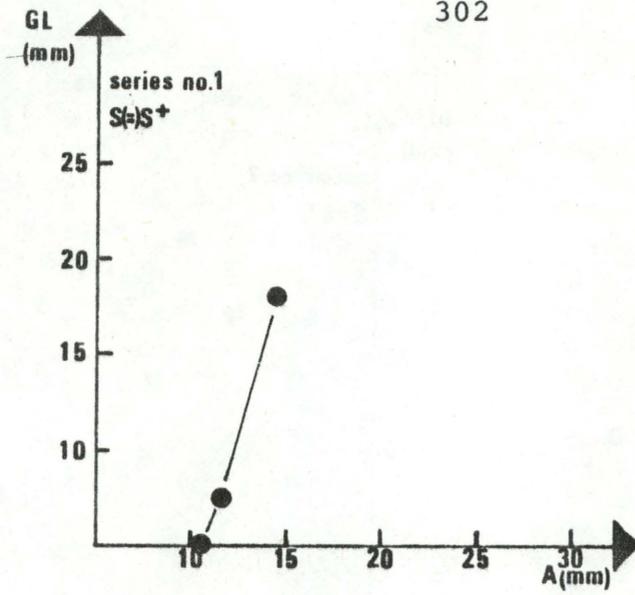
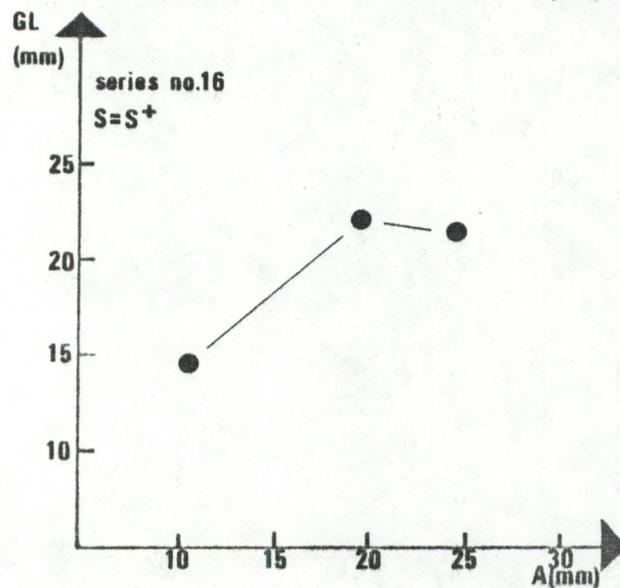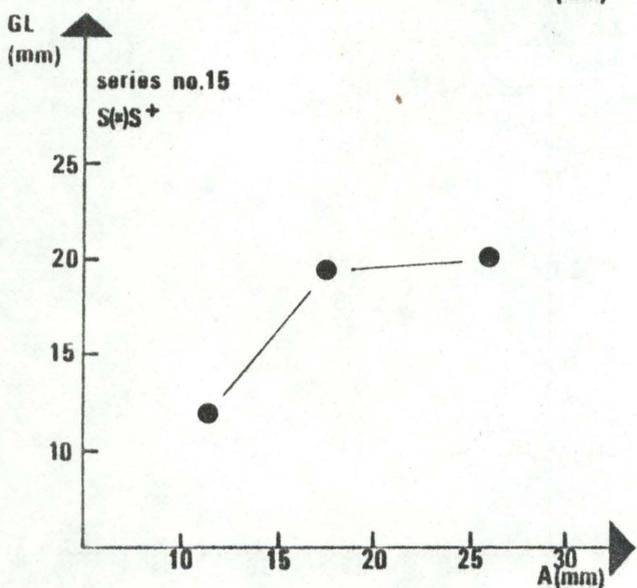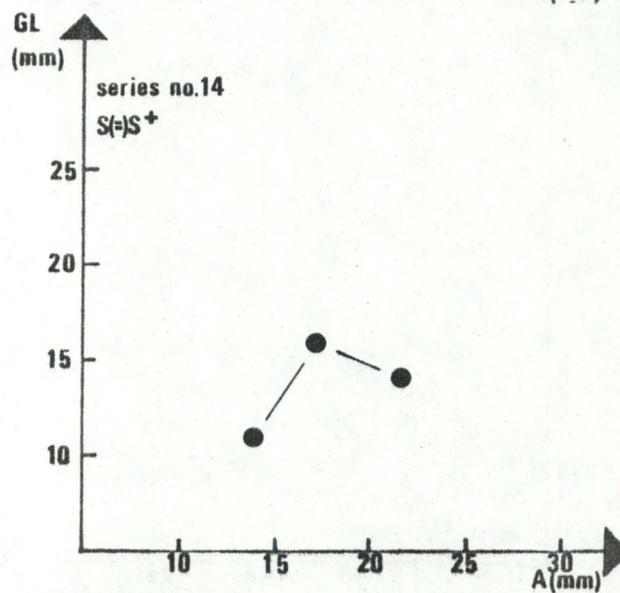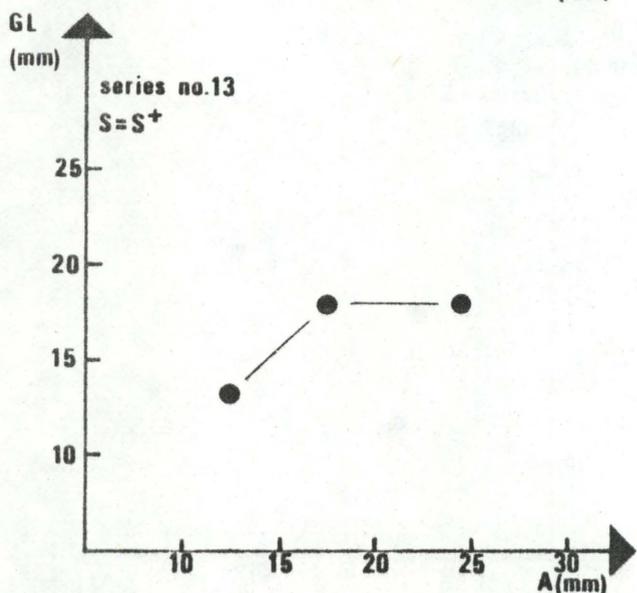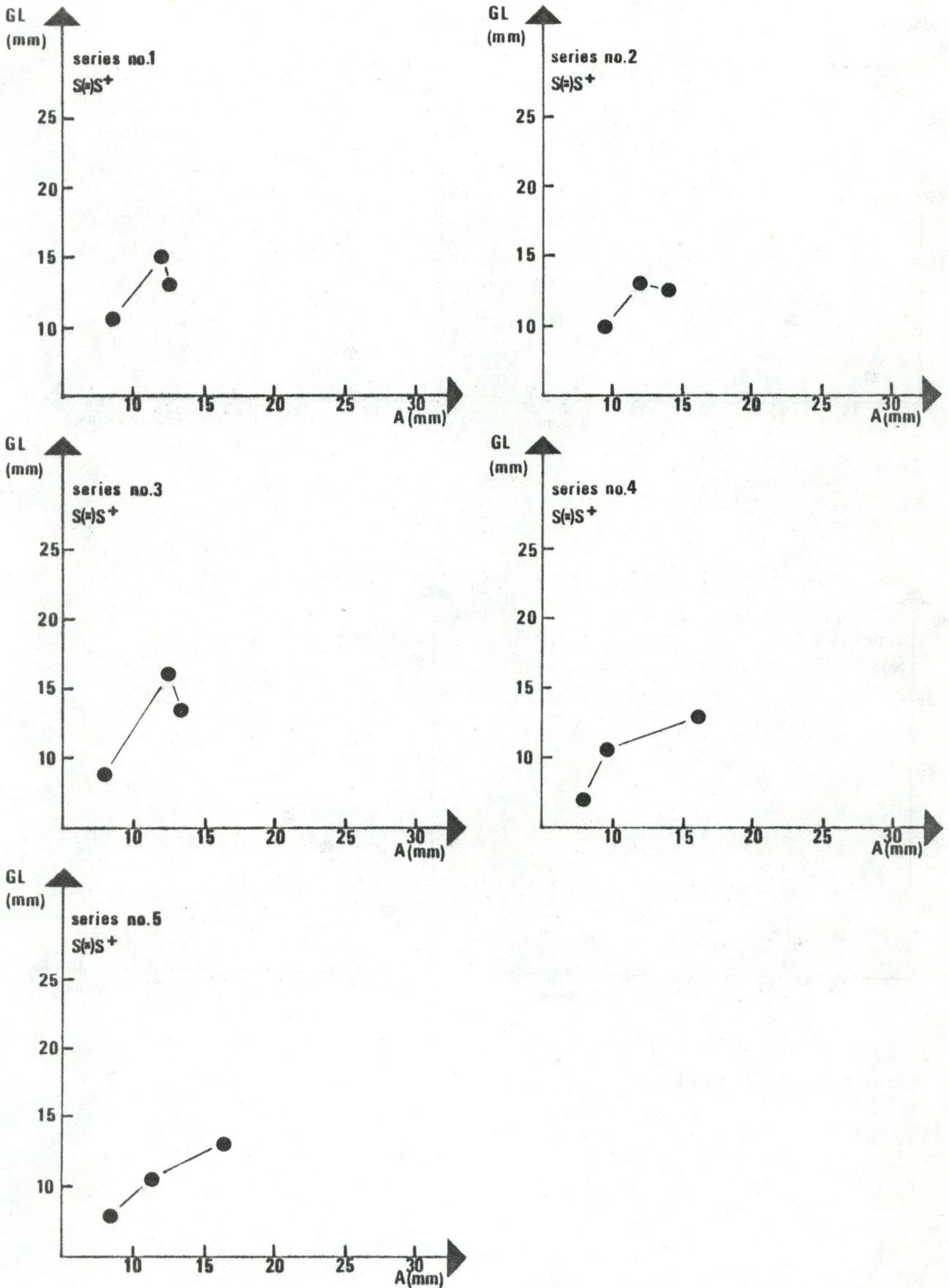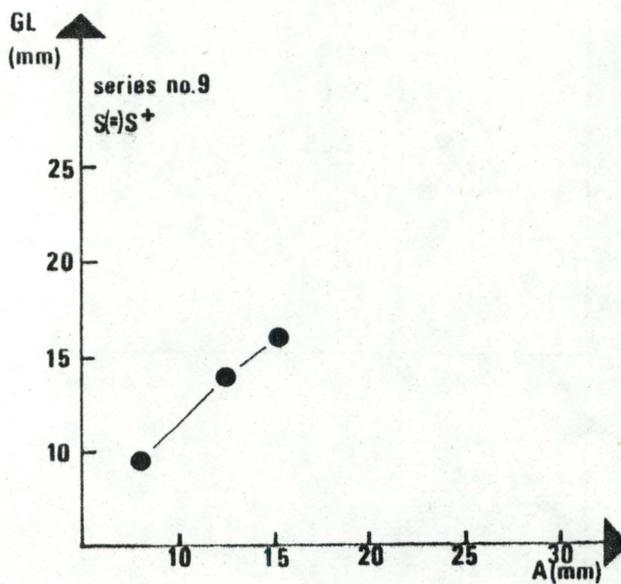Figure 7

Recording JR 28/8
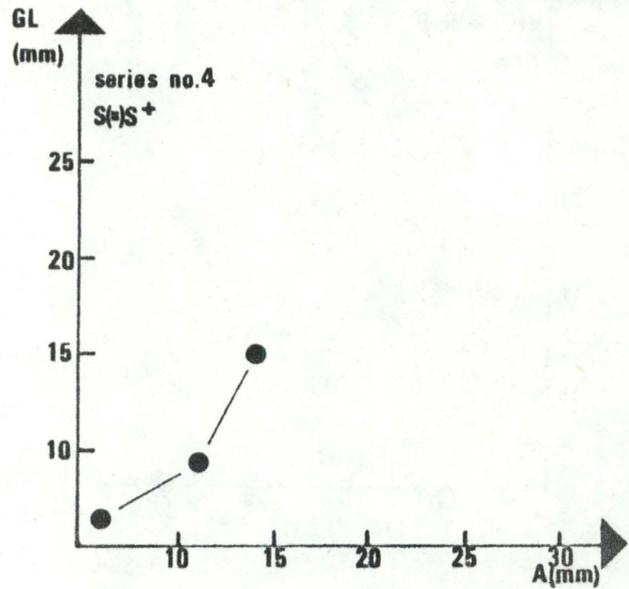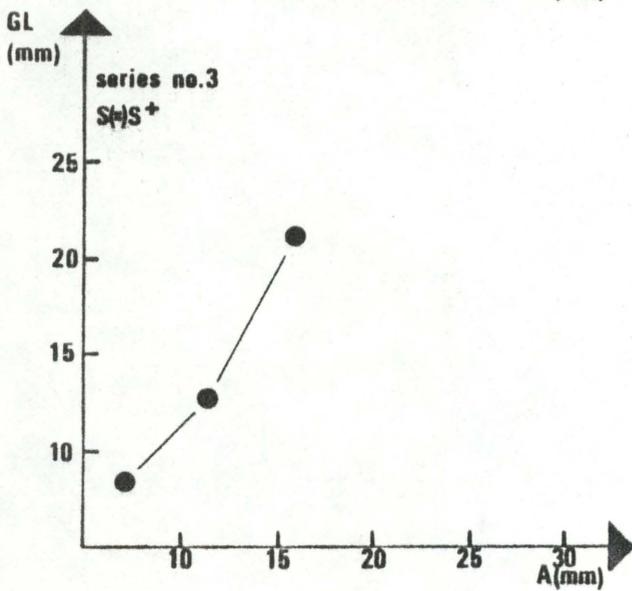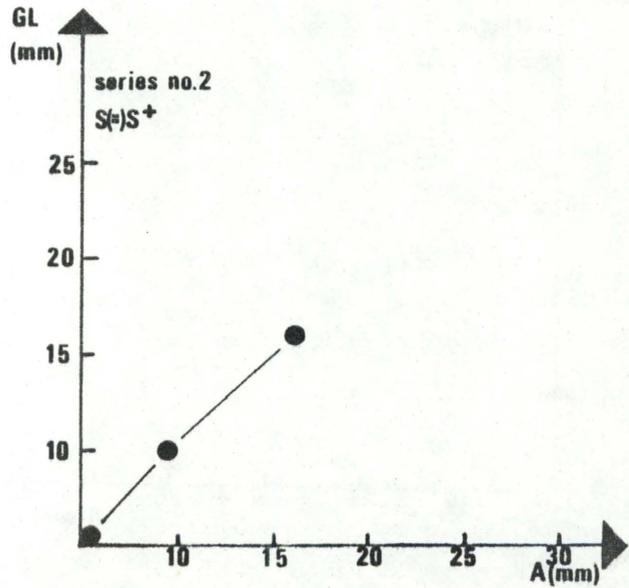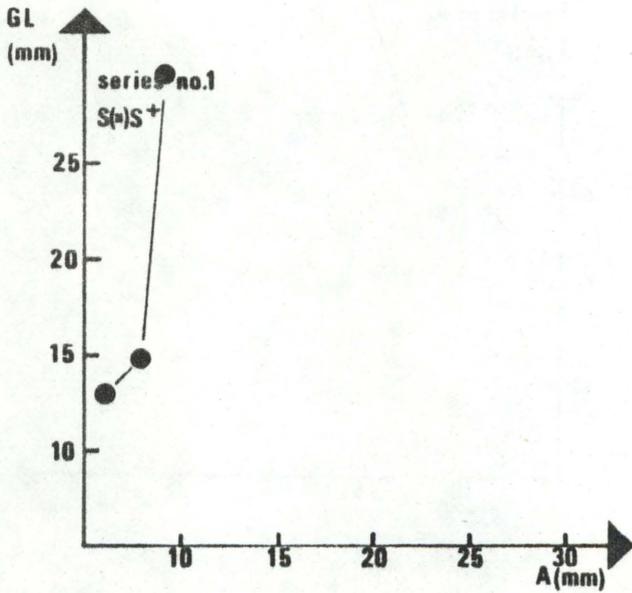
Figure 8

Recording BFJ 18/9
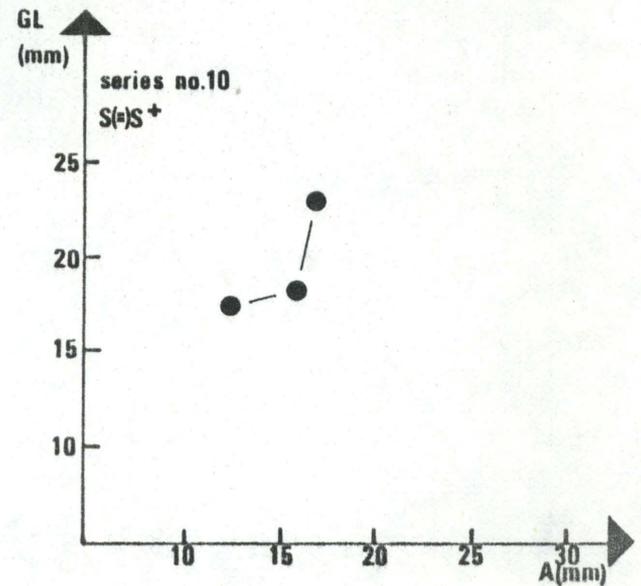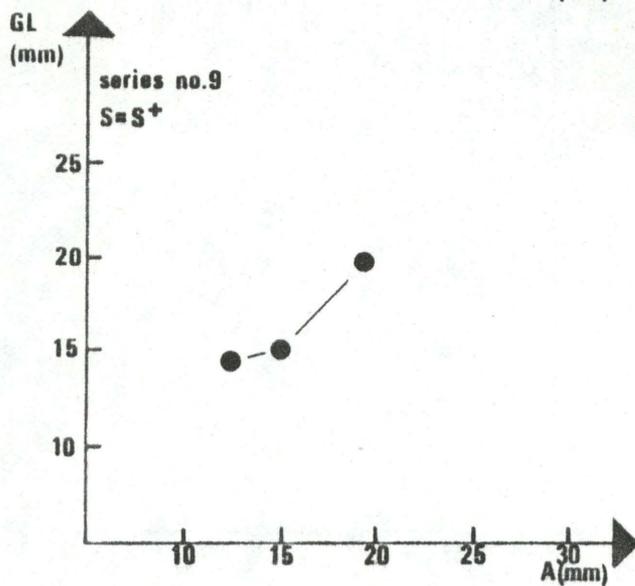
Figure 8 (continued)
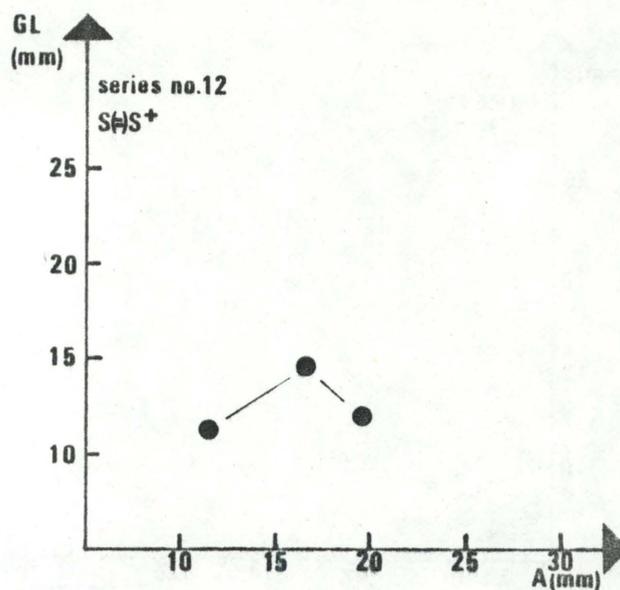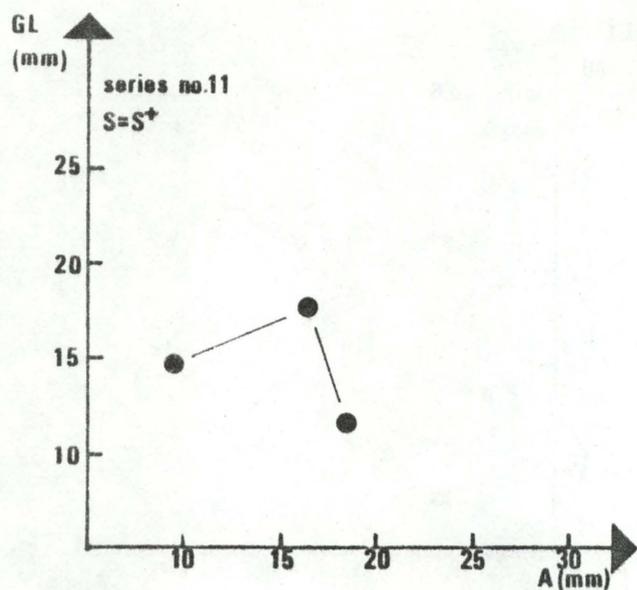
Recording BFJ 18/9

Figure 8 (continued)
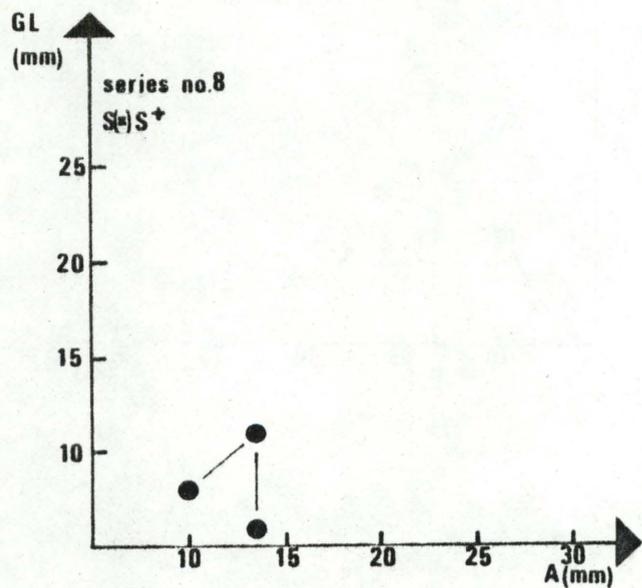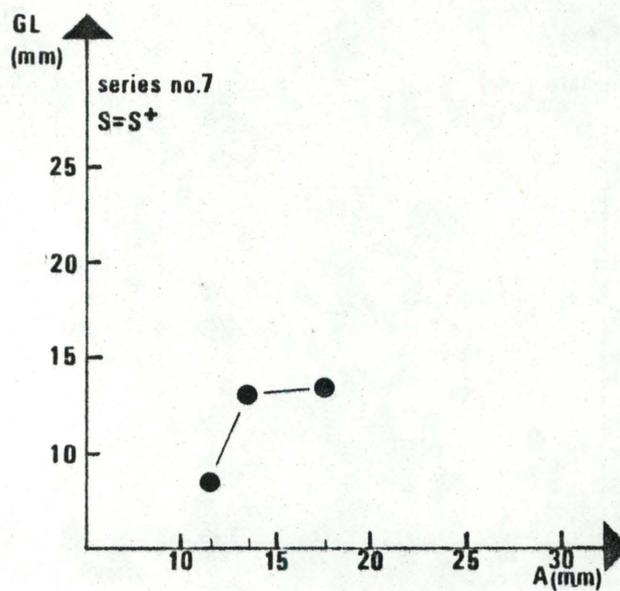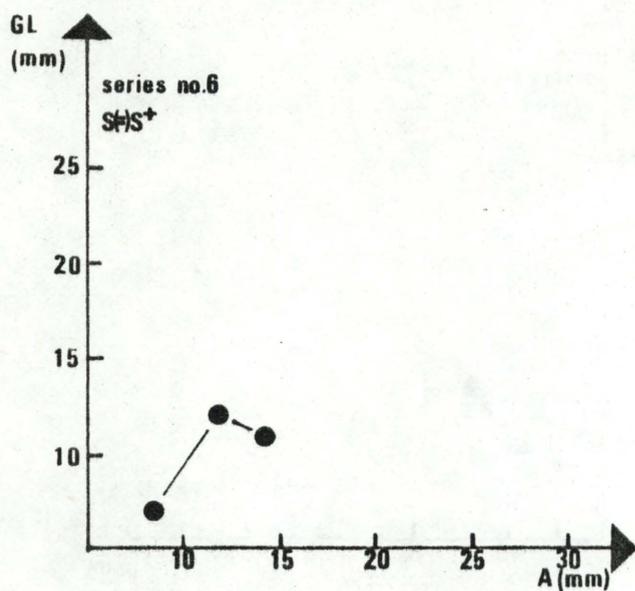
Recording BFJ 18/9

Figure 9

Recording BFJ 14/10

Figure 9 (continued)
Recording BFJ 14/10

It must be brought about by one or several of the following factors:

1) The glottographic light has been disturbed during the
    recording of the series
2) The glottograph transducer has moved considerably during
    the recording of the series
3) Something has shaded the glottograph transducer.

The first mentioned factor may explain the negative correlation
in JR 29/5 series no. 11. The film shows that the part of the
glottis in which the film is saturated is larger for b than for c.
The same factor may explain the negative correlation in BFJ 18/9
series no. 8, because the density of the film in the anterior
part of the glottis is less for b than for a. In all other cases
where the deviation from a straight line is considerable, this
explanation does not apply, judging by the films.[1]

It is not possible to say whether the transducer has been
considerably moved in some cases since this would have required
synchronous X-ray pictures.

The most obvious explanation of the fact that the correla-
tion between a, b, and c is in many cases so far from being linear,
and very often even negative, is that something has shaded the
transducer. It can hardly be the epiglottis or the tongue root
as all the sounds included in the sentences are articulated with
an advanced tongue root. It is rather the fiberoptic tube that
has shaded the transducer. The negative correlation is most fre-
quently seen for subject BFJ, and it is most often the correlation
between b and c that is negative. Therefore, BFJ made a supple-
mentary glottographic recording of 28 abc-series, the light and
the transducer being positioned as in the actual recordings, but
without the tube of the fiberscope.

_____

1) It may be problematic to estimate the relative quantity of
    light passing the glottis on the basis of the films. The film
    has a saturation threshold, and will thus be saturated at and above
    a certain light intensity; thus the film shows maximum density
    irrespective of the light intensity beyond this threshold. Further-
    more, the absolute size of the saturated area seen on the film
    depends partly on the distance between the lens of the fiberscope
    and the glottis, partly on the angle between the plane of glottis
    and the plane of the lens.

If it is correct that the fiberoptic tube under certain conditions shades the transducer, this recording made without the tube should show a greater amplitude for c than for b in the vast majority of cases, as the films normally showed a larger glottis aperture for c than for b. This is exactly what the 28 series show: out of 28 series only 2 have a smaller amplitude for c than for b. This result supports the hypothesis that the fiberoptic tube may have shaded the transducer in certain cases.

It seemed methodologically preferable to confine this study to the relationship between the amplitude of the glottogram and the glottis area under the simplest conditions, viz. with articulatory gestures involving the same type of glottis function. Still, I tried to get a hint of the relationship for series involving two different glottis functions. But as the procedure was fairly inaccurate, and the data very limited, this part of the study will not be reported here, except for one observation pertaining to the fiberscope. The recordings comprised, in addition to the three unvoiced consonants, synchronous fiberoptic pictures and glottograms of sustained vowels [ɪ::] and [æ::] (for the vowels it was the maximum glottis aperture during the vibratory cycle that was registered). It appeared from the film that the whole glottis was positioned considerably closer to the upper limit of the field of vision during phonation than during the articulation of unvoiced consonants. This may be due to movements of the velum, pushing the fiberoptic tube about during the recording. The X-ray recording mentioned above showed that the distal end of the tube was bent forward to the greatest extent in the articulation of [ʂ::], [s::], and [ç::], less in [ɪ::], and least in [æ::]. This displacement, and the subsequent variation in the relative location of the glottis within the field of vision, may be explained by variations in the velopharyngeal closure depending on the type of sound (Lubker 1968, 1970): the higher the position of the velum, the more the distal end of the fiberoptic tube is bent forward because both the velum and the pharyngeal wall press against the tube.

4. Concluding remarks

In this preliminary study the relationship between the glottis aperture and the amplitude of the glottogram has been examined under rather specific conditions of articulation and with one specific glottographic set-up. I did not have any possibility of studying continuous variations in glottis aperture, which of course would be of interest, too. To mention another limitation, only one type of light source was used. Furthermore, the applied procedure for measurement of the glottis aperture causes a certain inaccuracy of the results. Nevertheless, it is justified to make some concluding remarks.

The results demonstrate that the relationship between glottis aperture and the amplitude of the glottogram is sometimes approximately linear, and at other times far from linear; in many cases even a negative correlation is seen.

The negative correlation found in this study is probably caused in most cases by the fiberoptic tube shading the photo-transducer, i.e. it is caused by an artifact stemming from the complex procedure followed in this study and not by an artifact associated with the glottographic method itself. But other factors may give rise to a negative correlation: the tongue root and the epiglottis may shade the transducer and the glottographic light may alter its position during the recording.

It is true of all types of photo-electric glottographs that the linearity of the system has to be established in order to know whether the signal voltages delivered by the glottograph in a given recording are within a linear range or not. In interpreting the data of this study a means to divide the results into two groups would have been desirable, one for which the voltage range is within a linear range, and one for which it is within a non-linear range, in order to preclude that greater or lesser deviations from a linear relationship between amplitude and aperture may be caused by the system's inherent non-linearity for

varying illumination. Since I had not realized that such a
problem exists at the time when the recordings were made, and
hence did not measure the voltage range for each recording
session, this is not possible with my data. It can only be
stated that one does not very often find a linear correlation
between the glottis aperture and the amplitude of the glotto-
gram. Deviations from a linear relationship may be caused by
one or several of the factors influencing the amplitude of the
glottogram: the characteristics of the glottographic set-up,
instability of the light source, shading effects, and so on.
The combined effect of these different factors acting simul-
taneously may sometimes increase and sometimes decrease the
deviation from a linear relationship, depending on the factors
involved.

The non-linearity of the system is a problem in compari-
sons between glottographic curves of different sounds, if and
when the amplitude is taken to be some kind of measure of glottis
aperture. It may, however, also be crucial for the interpreta-
tion of the amplitude variations observed in a glottogram of a
continuous glottis movement. For one thing, the pronounced non-
linearity of the system in the lower voltage range makes it
highly dubious to take the "bottom" of the glottogram of a single
vocal fold vibration as a realistic representation of the true
physiological conditions.

I wish to add that according to my experience it is very
difficult to keep the fiberoptic tube stable during articulation.
This may be a relevant problem when the fiberscope is used for
some sort of quantitative analysis of the glottis area.

As for the glottograph, I venture the general conclusion
that the amplitude of the glottogram may be influenced by so
many factors that it is hazardous to take for granted - without
further proof - that there is a linear relationship between
glottis aperture and the amplitude of the glottogram. Therefore,
it may be "not only hazardous but invalid in many cases" to draw
conclusions about the relative size of glottis aperture on the
basis of glottograms.

References

Berg, Jw. van den 1968:   "Mechanism of the larynx and the
                             laryngeal vibrations", Manual of
                             Phonetics, ed. Bertil Malmberg,
                             p. 278-308

Coleman, R.F. and   "On the validity of laryngeal photo-
R.W. Wendahl 1968:   sensor monitoring", JASA 44, p.
                             1733-1735

Frøkjær-Jensen, B. 1967:   "A photo-electric glottograph",
                             ARIPUC 2, p. 5-19

Frøkjær-Jensen, B. 1968:   "Comparison between a Fabre glotto-
                             graph and a photo-electric glotto-
                             graph", ARIPUC 3, p. 9-16

Frøkjær-Jensen, B., C. Ludvig- "A glottographic study of some
sen, and J. Rischel 1971:   Danish consonants", F&S, Akademisk
                             forlag, p. 123-140 (ARIPUC 7, 1973,
                             p. 269-295

Harden, R.J. 1975:   "Comparison of glottal area changes
                             as measured from ultrahigh-speed
                             photographs and photoelectric glot-
                             tographs", JSHR 18, p. 728-738

Lubker, J. 1968:   "An electromyographic-cinefluoro-
                             graphic investigation of velar
                             function during normal speech pro-
                             duction", Cleft Palate Journal 5,
                             p. 1-18

Malécot, A. and K. Peebles   "An optical device for recording
1965:   glottal adduction-abduction during
                             normal speech", Zs.f.Ph. 18, p.
                             545-550

Ohala, J. 1966:   "A new photo-electric glottograph",
UCLA 4, p. 40-54

Sawashima, M. and   "Use of the fiberscope in speech
T. Ushijima 1971:   research", Ann. Bull. Univ. Tokyo
Logopedics-Phoniatrics, 5, p. 25-34

Sonesson, B. 1960:   On the anatomy and vibratory pattern of the human vocal folds,
Acta Oto-Laryng. Suppl. 156,
Stockholm.

A SURVEY OF THE CONTENTS OF PREVIOUS ISSUES OF ARIPUC

ARIPUC 1, 1967:

Preliminary experiments with the Fabre glottograph
    (Eli Fischer-Jørgensen, Børge Frøkjær-Jensen, and
    Jørgen Rischel)

Phonetic analysis of Danish stop consonants
    (Eli Fischer-Jørgensen)

The Danish long vowels
    (Børge Frøkjær-Jensen)

Changes in formant frequencies and formant levels at high
    voice effort
    (Børge Frøkjær-Jensen)

Studies of the prosodic properties of Danish words
    (Jørgen Rischel)

Studies of diphthongs in Faroese
    (Jørgen Rischel)

Murmured (breathy) vowels in Gujarati
    (Eli Fischer-Jørgensen)

Analyses of French accent
    (Børge Spang-Thomsen)

Features of the acoustical and physiological structure of
    the French nasal vowels
    (Ole Kongsdal Jensen)

Voice assimilation of stop consonants and fricatives in
    French and its relation to sound duration and intra-
    oral air-pressure
    (Oluf Thorsen)

Acoustic analysis of tense and lax vowels in German
    (Hans Peter Jørgensen)

Research on the physiological correlate to the two vowel
    types in German
    (W.W. Schuhmacher)

Perceptual dimensions of speech sounds
    (Eli Fischer-Jørgensen)

Experiments with sharp filtering of Danish and German vowels
   (Eli Fischer-Jørgensen)
Auditory investigation of German tense and lax vowels
   (Eli Fischer-Jørgensen)

ARIPUC 2, 1968:
A photo-electric glottograph
   (Børge Frøkjær-Jensen)
The heterodyne filter of the Institute of Phonetics
   (Jørgen Rischel and Svend-Erik Lystlund)
Speech synthesizer
   (Jørgen Rischel and Svend-Erik Lystlund)
Phonetic analysis of breathy (murmured) vowels in Gujarati
   (Eli Fischer-Jørgensen)
Is Danish 'sj-' one or two sounds?
   (Hans Basbøll)
The word tones of Serbo-Croatian, an instrumental study
   (Per Jacobsen)
Duration of French vowels before fricatives
   (Karen Landschultz)
A formant analysis of the clear, nasalized, and murmured
   vowels in Gujarati
   (Radhekant Dave)
Word tones in Dogri
   (Ved Kumari Ghai)
Statistic calculations of formant data
   (Børge Frøkjær-Jensen)

ARIPUC 3, 1969:
Construction of a Fabre glottograph
   (Børge Frøkjær-Jensen and Poul Thorvaldsen)
Comparison between a Fabre glottograph and a photo-electric
   glottograph
   (Børge Frøkjær-Jensen)

Constructional work on a function generator for speech
   synthesis
   (Jørgen Rischel)
The phoneme system of Advanced Standard Copenhagen
   (Hans Basbøll)
English stops after initial /s/
   (Niels Davidsen-Nielsen)
Voicing, tenseness, and aspiration in stop consonants, with
   special reference to French and Danish
   (Eli Fischer-Jørgensen)
Comparisons between vowel formant frequencies in speech and song
   (Børge Frøkjær-Jensen)
Fourier analyses of photo-electric glottograms
   (Børge Frøkjær-Jensen and Jørgen Rischel)
Vowel quantity in Czech
   (Per Jacobsen)
Acoustic cues for the perception af accent in Japanese when the
   accented vowel is devoiced
   (Hideo Mase)
Notes on the Danish vowel pattern
   (Jørgen Rischel)

ARIPUC 4, 1970:
An attempt to register lip pressure and lip rounding by means
   of the electro-glottograph
   (Carl Ludvigsen)
Construction of a segmentator
   (Poul Thorvaldsen)
Notes on the phonology of Danish imperatives with a digression
   on vowel quantity
   (Hans Basbøll)
Close and loose contact ("Anschluss") with special reference
   to North German
   (Eli Fischer-Jørgensen and Hans Peter Jørgensen)

Falling word tones in Serbo-Croatian
  (Per Jacobsen)
Measurements of the length of some Japanese vowels with special
  reference to their devoicing
  (Hideo Mase)
Morpheme stress in Danish
  (Jørgen Rischel)
Comparative phonetic-acoustic analysis before and after speech
  therapy of voices suffering from recurrens paresis
  (Kai Lauritzen and Børge Frøkjær-Jensen)

ARIPUC 5, 1971:
Voicing in British English t and d in contact with other con-
  sonants
  (Nina Thorsen)
A study of internal open juncture in British English
  (Kirsten Gregersen)
A comment on lexical insertion
  (Jørgen Rischel)
A study of peak amplitude measures
  (P. Mansell)
Syllable and accent in Japanese
  (Hideo Mase)
Energy measurements of speech sounds
  (Carl Ludvigsen)
Comparison of sound pressure level and loudness (GF)
  measurements on speech sounds
  (Carl Ludvigsen and Nina Thorsen)
A study of consonant quantity in West Greenlandic
  (Hideo Mase and Jørgen Rischel)

ARIPUC 6, 1972:
A formant-coded speech synthesizer
  (Jørgen Rischel and Svend-Erik Lystlund)

Spectrographic analysis of English vowels
    (Peter Holtse)
Formant frequencies of long and short Danish vowels
    (Eli Fischer-Jørgensen)
Kinesthetic judgement of effort in the production of stop
    consonants
    (Eli Fischer-Jørgensen)
Perceptual studies of Danish stop consonants
    (Eli Fischer-Jørgensen)
On universals in vowel perception
    (Birgit Hutters and Peter Holtse)
Some conditioning phonological factors for the pronunciation
    of short vowels in Danish with special reference to
    syllabification
    (Hans Basbøll)
Compound stress in Danish without a cycle
    (Jørgen Rischel)
Comparison between auditive and audio-visual perception of
    PB-words masked with white noise
    (Carl Ludvigsen)
Some remarks on acoustic parameters in speech disorders
    (Nils H. Buch and Børge Frøkjær-Jensen)
Construction of an automatic data collection system
    (Poul Thorvaldsen)
On British English p-, b-, sp- and -s + b-
    (Henry Petersen)
A reply to Henry Petersen
    (Niels Davidsen-Nielsen)
Investigating the influence of blowing technique on pitch and
    tone in recorder playing
    (Niels Bak)

ARIPUC 7, 1973:
A study of the role of syllable and mora for the tonal mani-
    festation in Greenlandic
    (Hideo Mase)

Electromyographic investigation of Danish consonants, stress,
    and stød
    (Eli Fischer-Jørgensen)
Acoustical and perceptual properties of the Danish stød
    (Nina Thorsen)
Preliminary work on computer testing of a generative phonology
    of Danish
    (Hans Basbøll and Kjeld Kristensen)
Evaluation of speech disorders by means of long-time-average-
    spectra
    (Børge Frøkjær-Jensen and Svend Prytz)
Preliminary experiments with syntehsis by rule of Standard
    Danish
    (Peter Holtse)

ARIPUC 9, 1975:
Asymmetric vowel harmony in Greenlandic fringe dialects
    (Jørgen Rischel)
On the phonological interpretation of the falling diphthongs
    in Danish
    (Hans Basbøll)
Grammatical boundaries in phonology
    (Hans Basbøll)
Danish r and adjacent short stressed vowels
    (Steffen Heger)
The influence of acoustic and physiological factors on some
    vowel changes in Danish
    (Peter Molbæk Hansen and Bent Møller)
An investigation of the fundamental frequency of vowels after
    various consonants, in particular stop consonants
    (Vivi Jeel)
Perspectives in phonology
    (Eli Fischer-Jørgensen)
Registration of voice quality
    (Børge Frøkjær-Jensen and Svend Prytz)

Topics in West Greenlandic phonology
    (Jørgen Rischel)
Further work on computer testing of a generative phonology of
    Danish
        (Hans Basbøll and Kjeld Kristensen)
Statistic research on changes in speech due to pedagogic
    treatment (the accent method)
        (Svend Smith and Kirsten Thyme)